UNIVERSITA' DEGLI STUDI DEL SANNIO

DIPARTIMENTO DI SCIENZE E TECNOLOGIE

Corso di Laurea Triennale in Biotecnologie

Tesi di Laurea

in

BIOINFORMATICA

L'INTELLIGENZA ARTIFICIALE GENERATIVA IN ONCOLOGIA: IL RUOLO DEI LARGE LANGUAGE MODELS E DEI FOUNDATION MODELS MULTIMODALI

Relatore

Ch.mo Prof.
Francesco Napolitano

Candidata

Giorgia Miano
Matr.171005106

Anno accademico 2025/2026

INDICE

INTRODUZIONE……4

CAPITOLO I - L'INTELLIGENZA ARTIFICIALE IN ONCOLOGIA

1.1 Che cos'è un tumore……...……...…...6

1.1.2 Sottotipizzazione tumorale nel cancro……..…7

1.2 L'intelligenza artificiale e il Machine Learning……...……9

1.2.1 Definizione di AI…….....9

1.2.2 Machine Learning e Deep Learning…….……..12

1.2.3 Principali differenze tra gli approcci……..13

1.3 Applicazione dell'AI in oncologia…….....14

1.3.1 Prognosi e diagnosi……..……..16

1.3.2 Predizione della risposta ai trattamenti……..16

1.4 La complessità dei dati oncologici……..…..17

1.4.1 Dati clinici……..18

1.4.2 Dati genomici e molecolari……...…18

1.4.3 Imaging medico e patologia digitale……..……19

1.5 Limiti degli approcci tradizionali…….20

1.5.1 Modelli progettati per singoli compiti……..…..21

1.5.2 Difficoltà nell'integrazione di fonti eterogenee…….…21

CAPITOLO II - FOUNDATION MODELS E LARGE LANGUAGE MODELS

2.1 Definizione di Foundation Models…….……...23

2.1.1 Pre-training su grandi quantità di dati……....24

2.1.2 Adattamento a compiti differenti…….…..25

2.2 Large Language Models……25

2.2.1 Pre-training e fine-tuning…….……..26

2.2.2 Architettura Transformer……...….28

2.3 Le capacità del Large Language Models……...……29
2.3.1 Comprensione e generazione del linguaggio…….….30
2.3.2 Estrazione e sintesi delle informazioni…….……..31
2.3.3 Question answering……...…….31
2.3.4 Limiti dei modelli basati esclusivamente sul testo ……....31
2.4 Verso i modelli multimodali…….……32
2.4.1 Potenziamento dei Foundation Models multimodali …….32
2.4.2 Integrazione di testo, immagini e dati biologici attraverso i Foundation Models……...……..33

CAPITOLO III - FOUNDATION MODELS E LARGE LANGUAGE MODELS IN ONCOLOGIA
3.1 Il problema dell'eterogeneità dei dati oncologici…….…….35
3.1.1 Frammentazione delle informazioni…….…..35
3.1.2 Dati strutturati e non strutturati……..……36
3.2 Cartelle cliniche elettroniche…….……36
3.2.1 Referti e documentazione medica……...……...37
3.2.2 Estrazione automatica delle informazioni ……..…37
3.2.3 Integrazione dei dati clinici……38
3.3 Dati genomici…….……...39
3.3.1 Mutazioni e biomarcatori…….……..40
3.3.2 Caratterizzazione molecolare del tumore ……...…41
3.3.3 Integrazione dei dati genomici e molecolari……..……42
3.4 Imaging radiologico……..42
3.4.1 Immagini istopatologiche……...……43
3.4.2 Integrazione di immagini e patologia digitale……45
3.5 Approcci multimodali in oncologia……...……46
3.5.1 Fusione delle diverse modalità……...…46
3.5.2 Vantaggi dell'analisi integrata……49

3.6 Il paradigma “Leave no Data Behind”……..50
3.6.1 Superamento della frammentazione dei dati……...…...51

CAPITOLO IV - FOUNDATION MODELS E MEDICINA PERSONALIZZATA
4.1 Dalla medicina tradizionale alla medicina di precisione……..….52
4.1.1 Definizione di medicina personalizzata e di precisione……..…52
4.1.2 Eterogeneità dei tumori…….…….53
4.1.3 Importanza delle caratteristiche individuali sul paziente……...53
4.2 Definizione di stratificazione …….……..54
4.2.1 Identificazione di sottogruppi……...……..55
4.2.2 Caratteristiche cliniche e molecolari……..56
4.3 Clinical Decision Support Systems……...……56
4.3.1 Ruolo dell’AI nel percorso decisionale……..……57

CAPITOLO V - LIMITI DELL’INTELLIGENZA ARTIFICIALE IN ONCOLOGIA
5.1 Limiti e affidabilità dei Large Language Models…….….58
5.1.1 Cosa sono le allucinazioni……..58
5.1.2 Privacy, sicurezza e aspetti etici……59
5.1.3 Responsabilità clinica……...…..60

CONCLUSIONI……61

BIBLIOGRAFIA……..…….63

INTRODUZIONE

Negli ultimi anni, l'intelligenza artificiale (Artificial Intelligence, AI) ha assunto un ruolo sempre più importante nel settore della medicina, contribuendo allo sviluppo di nuovi strumenti per l'analisi, l'interpretazione e l'integrazione delle informazioni cliniche e biologiche. Questo sviluppo è particolarmente significativo in oncologia, un ambito caratterizzato da una forte complessità sia dal punto di vista biologico sia da quello informativo. I tumori non costituiscono, infatti ,una realtà uniforme: anche all'interno della stessa tipologia tumorale possono esistere differenze genetiche, molecolari, istologiche e cliniche che influenzano l'evoluzione della malattia e la risposta ai trattamenti. La possibilità di considerare contemporaneamente queste diverse caratteristiche rappresenta quindi una delle principali sfide della ricerca oncologica contemporanea.

In questo contesto, l'intelligenza artificiale e, in particolare, il Machine Learning e il Deep Learning, hanno permesso di ampliare sistemi in grado di rilevare pattern e relazioni all'interno di grandi quantità di dati. In oncologia, questi approcci sono stati studiati per diverse applicazioni, tra cui il supporto alla diagnosi e alla classificazione dei tumori, la stima della prognosi e la predizione della risposta a specifici trattamenti. La maggior parte dei modelli tradizionali di intelligenza artificiale, infatti, sono stati progettati per svolgere compiti specifici e richiedono molto spesso dataset selezionati e annotati. Questo approccio può risultare limitante quando si cerca di utilizzare contemporaneamente informazioni appartenenti a modalità diverse.

Un'evoluzione importante in questo sfondo è rappresentata dai Foundation Models, modelli di intelligenza artificiale addestrati su grandi quantità di dati e successivamente adeguabili a differenti compiti e contesti applicativi; tra questi i Large Language Models (LLM) hanno ottenuto particolare attenzione grazie alla capacità di elaborare e generare linguaggio naturale, permettendo l'estrazione di informazioni da documenti clinici, la sintesi di testi e il supporto alla ricerca di informazioni rilevanti. L'evoluzione dei Foundation Models sta inoltre portando allo sviluppo di sistemi multimodali, generati per elaborare informazioni provenienti da modalità differenti, come testo, immagini e dati biologici.

La presente tesi mostra l'analisi del ruolo dei Foundation Models e dei Large Language Models in oncologia, concentrandosi in particolare sulle loro potenzialità nell'integrazione di dati eterogenei e sulle prospettive per la medicina di precisione. La prima parte verterà sull'intelligenza artificiale e sulle sue principali applicazioni in oncologia, verranno introdotti i Foundation Models e i Large Language Models, descrivendone le caratteristiche fondamentali, il processo di pre-training e il meccanismo di attenzione. Sarà analizzato oltretutto il passaggio dai modelli esclusivamente linguistici verso approcci multimodali, capaci di combinare informazioni provenienti da differenti tipologie di dati.

L'attenzione in seguito si sposterà verso l'applicazione di questi modelli in ambito oncologico, con particolare riferimento all'integrazione dei dati clinici, genomici e molecolari, dell'imaging radiologico e della patologia digitale. In questo ambito sarà approfondito il paradigma "Leave No Data Behind", inteso come prospettiva il cui intento è quello di ridurre la frammentazione delle informazioni e contemporaneamente valorizzare, quando possibile, le diverse fonti di dati disponibili per la caratterizzazione del paziente e della malattia.

Infine, la tesi si concentrerà sul contributo dei Foundation Models alla medicina personalizzata e alla stratificazione dei pazienti, considerando anche il loro potenziale ruolo nei sistemi di supporto alle decisioni cliniche. A tal proposito, verranno trattati i principali limiti e le criticità associate all'utilizzo dell'intelligenza artificiale in oncologia, tra cui : l'affidabilità delle informazioni generate dai modelli linguistici, il rischio di allucinazioni, la privacy e la sicurezza dei dati. L'obiettivo principale non è quindi considerare l'intelligenza artificiale come sostitutiva della figura sanitaria, ma analizzarne le possibilità e i limiti come mezzo di supporto all'interpretazione e all'integrazione di informazioni sempre più numerose e articolate.

Il concreto sviluppo dei Foundation Models ad oggi rappresenta un'evoluzione tecnologica molto interessante per l'oncologia, ma il loro impiego clinico necessita di una valutazione appropriata e veritiera della qualità dei dati, ma soprattutto dell'affidabilità dei risultati e, tentare di comprendere sia le potenzialità sia le criticità di questi strumenti. Quest'ultima caratteristica in particolare risulta essere estremamente fondamentale per valutare il possibile contributo di questi modelli alla medicina di precisione e allo sviluppo della ricerca e della pratica oncologica.

CAPITOLO I – L'INTELLIGENZA ARTIFICIALE IN ONCOLOGIA

1.1 Che cos'è un tumore

Il termine tumore indica una massa anomala di tessuto che si forma quando le cellule dell'organismo proliferano più del normale o quando non vanno incontro alla morte cellulare nei tempi fisiologicamente previsti. Secondo il *National Cancer Institute (NCI)*, i tumori, definiti anche neoplasie, possono essere classificati in benigni o maligni. Questa distinzione è fondamentale, poiché la presenza di una massa tumorale non corrisponde necessariamente alla presenza di un cancro. Per quel che concerne i tumori benigni, non vanno ad invadere i tessuti circostanti e non si diffondono in altre parti dell'organismo; al contrario, i tumori maligni sono caratterizzati da cellule capaci di colpire i tessuti adiacenti e, talvolta, di diffondersi attraverso il sangue o il sistema linfatico.

Il cancro può essere definito come una malattia caratterizzata dalla crescita sregolata di alcune cellule dell'organismo e dalla loro capacità di diffondersi in altre regioni del corpo. Esso, infatti, non rappresenta una singola patologia, ma un insieme eterogeneo di malattie che possono svilupparsi in qualsiasi tessuto o organo. Le diverse forme di cancro vengono di solito denominate in base all'organo o al tessuto nel quale si sono sviluppate oppure in base al tipo cellulare da cui derivano.

All'origine dello sviluppo tumorale vi è un'alterazione dei normali meccanismi che regolano la crescita, la divisione e la morte delle cellule. In condizioni fisiologiche, infatti, le cellule si moltiplicano quando l'organismo ne ha bisogno e quelle vecchie o danneggiate vengono eliminate attraverso processi di morte cellulare programmata, come l'apoptosi. Nel processo di cancerogenesi, questo equilibrio può essere intaccato, determinando la proliferazione di cellule che non dovrebbero continuare a dividersi o la sopravvivenza di queste che normalmente dovrebbero essere eliminate. Ciò comporta che l'accumulo progressivo di queste cellule può causare la formazione di una massa tumorale.

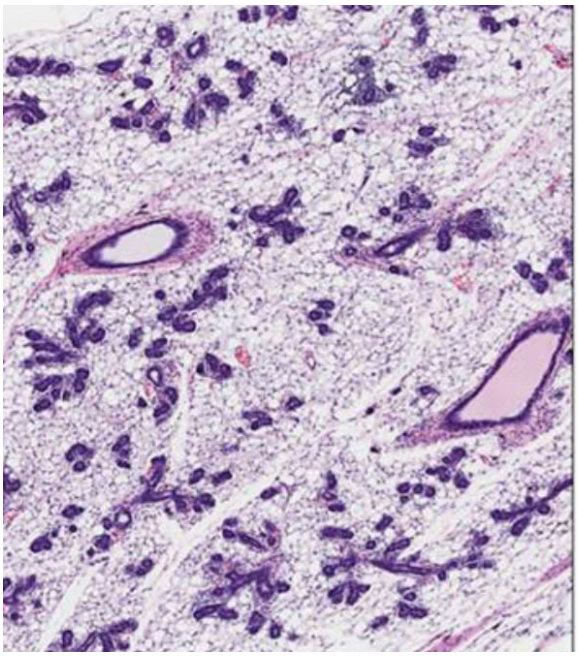
Il *NCI* definisce il cancro una malattia genetica, in quanto la sua comparsa è associata ad alterazioni dei geni che regolano il funzionamento cellulare, in particolare quelli coinvolti nella crescita e nella divisione cellulare. Queste alterazioni, comunemente definite mutazioni o cambiamenti genetici, possono presentarsi in maniera del tutto spontanea durante la divisione cellulare, essere determinate dall'esposizione a sostanze o agenti cancerogeni oppure, in alcuni casi, essere ereditate dai genitori *(Istituto Nazionale dei Tumori).*

È importante sottolineare che lo sviluppo di un tumore rappresenta un processo progressivo: una singola alterazione genetica non determina necessariamente la manifestazione di un cancro. Nel corso del tempo possono accumularsi molteplici cambiamenti genetici, generando cellule con caratteristiche sempre più diverse rispetto alle cellule normali, infatti, quelle presenti all'interno dello stesso tumore possono mostrare combinazioni genetiche differenti, contribuendo all'eterogeneità biologica della neoplasia.

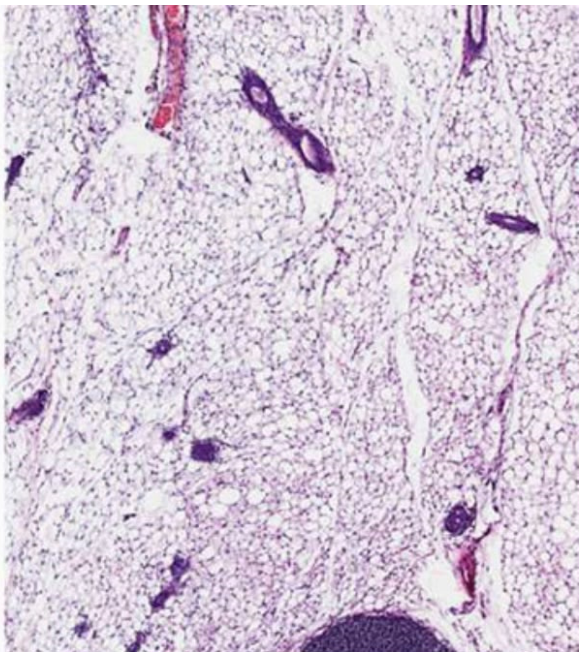
Le cellule tumorali presentano caratteristiche che le distinguono dalle cellule normali, tra queste ritroviamo la capacità di continuare a proliferare anche in assenza dei normali segnali che stimolano la crescita e di ignorare i segnali che dovrebbero arrestare la divisione cellulare o indurre la morte programmata.

La differenza tra tumore benigno e maligno si basa principalmente sul comportamento biologico della neoplasia: i tumori benigni portano alla formazione di masse di dimensioni imponenti, ma solitamente non colpiscono i tessuti circostanti e non sviluppano metastasi. La loro presenza, inoltre, non deve essere sempre considerata priva di conseguenze cliniche, in quanto la localizzazione e le dimensioni di un tumore benigno possono determinare, ad esempio, alterazione della funzione degli organi vicini. Il *NCI*, mostra che anche alcuni tumori benigni possono causare sintomi piuttosto gravi e risultare pericolosi in relazione alla loro sede *(Istituto Nazionale dei Tumori ▪).*

I tumori maligni, invece, sono caratterizzati dalla capacità di invadere i tessuti circostanti e possono diffondersi in altre parti dell'organismo. Questa capacità di invasione e diffusione rappresenta una delle principali peculiarità che distinguono le neoplasie maligne da quelle benigne; infatti ,queste , possono essere designate anche come tumori cancerosi.



Cellula tumorale



Cellula sana

1.1.2 Sottotipizzazione tumorale nel cancro

Il cancro non costituisce un'entità patologica omogenea, ma comprende un insieme eterogeneo di malattie caratterizzate da diverse alterazioni genetiche e molecolari. Questa eterogeneità può essere osservata sia tra tumori appartenenti a organi o tessuti differenti, sia all'interno dello stesso tipo tumorale. Pazienti con una diagnosi istopatologica apparentemente analoga possono presentare profili molecolari diversi e, di conseguenza, caratteristiche biologiche, comportamento clinico e risposta ai

trattamenti differenti. Questa complessità ha progressivamente evidenziato i limiti di una classificazione basata sulle caratteristiche anatomiche e istologiche della neoplasia, favorendo lo sviluppo di approcci di classificazione molecolare sempre più approfonditi *(Zhao et al., 2019).*

La sottotipizzazione molecolare del cancro rappresenta un approccio finalizzato a suddividere i tumori in sottogruppi sulla base delle loro caratteristiche molecolari, il cui obiettivo è quello di identificare gruppi di tumori caratterizzati da profili biologici omogenei e associati, quando possibile, a specifiche caratteristiche cliniche. La sottotipizzazione può contribuire a descrivere la complessità biologica di una neoplasia andando oltre la classificazione tradizionale e permettendo di mostrare differenze che possono avere rilevanza prognostica o terapeutica *(Zhao et al., 2019).*

Tra i dati utilizzabili per la sottotipizzazione rientrano, a seconda dello studio e della neoplasia considerata, informazioni relative al DNA, all'RNA e all'espressione genica, nonché altri livelli di informazione molecolare. *Zhao et al.* descrivono in particolare l'impiego di differenti piattaforme per la generazione di dati molecolari, tra cui microarray, sequenziamento dell'RNA (RNA-seq), PCR quantitativa, tecnologia NanoString e microarray tissutale.

Uno degli aspetti principali della sottotipizzazione molecolare consiste nella possibilità di considerare numerose variabili biologiche, ad esempio i dati genomici possono fornire informazioni inerenti ad alterazioni del DNA, mentre i dati trascrittomici permettono di caratterizzare i livelli di espressione dei geni. L'integrazione di queste informazioni permette di ottenere una rappresentazione più concreta delle caratteristiche molecolari del tumore rispetto alla sola osservazione morfologica. La sottotipizzazione molecolare rappresenta infatti uno degli strumenti attraverso i quali la ricerca oncologica ha cercato di tradurre la crescente quantità di dati omici in una classificazione biologicamente e clinicamente rilevante.

Dal punto di vista metodologico, la sottotipizzazione molecolare comprende diverse fasi: una prima fase riguarda l'acquisizione, la gestione e la preelaborazione dei dati, necessarie per rendere confrontabili e interpretabili i profili molecolari ottenuti; successivamente, possono essere utilizzati metodi di analisi non supervisionata, tra cui le tecniche di clustering, con l'intento di identificare gruppi di campioni caratterizzati da profili molecolari simili. I gruppi individuati vengono quindi sottoposti a successive analisi di caratterizzazione per determinarne le principali proprietà biologiche e cliniche. *Zhao et al*. descrivono questo processo come un flusso di lavoro comprendente pre-processing dei dati, analisi dei cluster, classificazione supervisionata e caratterizzazione dei sottotipi.

L'uso di algoritmi statistici e di machine learning assume particolare importanza in questo processo, perchè l'elevata dimensionalità dei dati molecolari rende difficile un'analisi esclusivamente manuale. I metodi computazionali permettono di individuare pattern e relazioni tra numerose variabili biologiche e possono essere utilizzati sia per identificare nuovi sottogruppi sia per sviluppare classificatori in grado di assegnare un nuovo campione a un sottotipo precedentemente definito. In questo modo, la sottotipizzazione può pian piano passare da uno strumento prevalentemente esplorativo a uno applicabile alla caratterizzazione di singoli pazienti.

1.2 L'intelligenza artificiale e il Machine Learning

Recentemente, l'intelligenza artificiale ha assunto un ruolo sempre più rilevante nel panorama tecnologico e scientifico, entrando in numerosi ambiti della vita quotidiana, della ricerca e delle attività economiche. L'evoluzione delle capacità computazionali, la crescente disponibilità di dati e lo sviluppo di nuove tecniche di elaborazione hanno reso possibili applicazioni che, fino a pochi decenni fa, appartenevano principalmente al campo della ricerca teorica. L'intelligenza artificiale si è affermata come uno dei principali settori di interesse dell'innovazione tecnologica contemporanea (*Russell e Norvig, 2021; Stone et al.,2016).*

La diffusione di sistemi basati su di essa ha portato a una grande trasformazione del modo in cui vengono affrontati problemi caratterizzati da elevati livelli di complessità. Questi sistemi possono essere utilizzati per analizzare grandi quantità di informazioni, individuare regolarità all'interno dei dati, effettuare previsioni o supportare processi decisionali. La varietà delle applicazioni e delle tecniche coinvolte rende difatti il campo dell'intelligenza artificiale particolarmente ampio e articolato, comprendendo approcci differenti sviluppatisi nel corso della sua evoluzione *(Russell e Norvig, 2021; OECD, 2019).*

All'interno di questo settore assume particolare importanza il Machine Learning, il quale rappresenta uno dei principali approcci attraverso cui vengono sviluppati sistemi capaci di elaborare informazioni e migliorare le proprie prestazioni sulla base dei dati disponibili. Il rapporto tra intelligenza artificiale e Machine Learning non deve essere solamente interpretato come una semplice equivalenza tra i due concetti: il secondo infatti si inserisce all'interno di un quadro più ampio, caratterizzato dalla presenza di differenti metodologie e prospettive *(Mitchell, 1997; Bishop, 2006).*

L'importante diffusione di queste tecnologie ha reso necessario un maggiore approfondimento dei concetti alla base dell'intelligenza artificiale e delle discipline a essa collegate, ma prima di analizzarne le principali applicazioni, è pertanto opportuno chiarire il significato dei termini utilizzati e delineare i confini concettuali tra le diverse aree che compongono questo ambito di ricerca *(Russell e Norvig, 2021; Floridi et al., 2018).*

1.2.1 Definizione di AI

L'intelligenza artificiale (Artificial Intelligence, AI) può essere definita come l'insieme di metodi, tecniche e sistemi computazionali sviluppati con l'obiettivo di realizzare macchine capaci di svolgere attività che, in alcune circostanze, richiedono capacità associate all'intelligenza umana, quali l'apprendimento, il ragionamento, la percezione, il riconoscimento di schemi, la comprensione del linguaggio e la risoluzione di problemi. La definizione di intelligenza artificiale non risulta univoca, perché nel corso della sua evoluzione la disciplina è stata interpretata secondo prospettive differenti. Alcuni approcci hanno posto l'attenzione sulla capacità delle macchine di riprodurre o simulare comportamenti tipicamente umani, mentre altri si sono focalizzati sulla capacità dei sistemi di elaborare informazioni e prendere decisioni razionali in funzione di determinati obiettivi *(Russell & Norvig, 2021).*

Uno dei primi contributi fondamentali sull'intelligenza delle macchine risale ad *Alan Turing,* che nel 1950 si interrogò sulla possibilità di considerare una macchina come intelligente, e propose di affrontare la questione attraverso il cosiddetto *"imitation game",*

noto come *test di Turing (Turing, 1950).* La proposta di Turing rappresenta uno dei primi tentativi di spostare l'attenzione dall'interrogativo astratto inerente alla possibilità che una macchina *"pensi"* alla valutazione del comportamento che si osserva all'interno di un sistema artificiale. Nei decenni successivi, l'intelligenza artificiale si è sviluppata come un campo interdisciplinare comprendente informatica, matematica, statistica, neuroscienze, linguistica e altre discipline.

Una distinzione particolarmente utilizzata riguarda la cosiddetta intelligenza artificiale ristretta o debole *(narrow AI)*, costituita da sistemi progettati per svolgere compiti specifici e delimitati. Questi sistemi possono raggiungere prestazioni molto elevate in determinati ambiti, come la classificazione delle immagini, il riconoscimento vocale, la traduzione automatica, ecc… Le applicazioni contemporanee dell'intelligenza artificiale appartengono prevalentemente a questa categoria *(Russell & Norvig, 2021).*

Un altro concetto relativo all'AI è quello di superintelligenza. Con questo termine si fa riferimento a sistemi ipotetici le cui capacità cognitive supererebbero quelle degli esseri umani in un numero molto ampio di attività. La superintelligenza è stata analizzata, tra gli altri, da *Nick Bostro*m, che ha esaminato le possibili traiettorie di sviluppo di sistemi artificiali altamente avanzati e le problematiche associate al loro controllo (*Bostrom, 2014).*

Un ambito dove l'intelligenza artificiale ha prodotto risultati molto efficienti è la *computer vision*, la disciplina che consente ai sistemi computazionali di interpretare il contenuto di immagini e video, ad esempio riconoscendo oggetti, volti e scene *(Szeliski, 2022).* Un punto di svolta si è avuto nel 2012, quando una rete neurale convoluzionale profonda, nota come *AlexNet*, ha ottenuto risultati di gran lunga superiori rispetoo ai metodi precedenti, utilizzati nella classificazione di immagini su larga scala *(Krizhevsky et al., 2012).* Questo risultato è stato misurato sul *benchmark ImageNet*, un dataset di circa 1,2 milioni di immagini di addestramento suddivise in 1.000 categorie, utilizzato nella competizione annuale ILSVRC *(Russakovsky et al., 2015).*

Negli anni successivi l'introduzione delle reti residuali *(ResNet),* con architetture fino a 152 livelli, ha permesso di raggiungere un errore del 3,57% sul test set di *ImageNet* e di prevalere sul ILSVRC 2015 *(He et al., 2016).* Di recente, invce, l'applicazione dell'architettura Transformer alle immagini, con il Vision Transformer, ha mostrato che un modello basato esclusivamente sull'attenzione può essere uguale o superare le reti convoluzionali quando viene pre-addestrato su grandi quantità di dati (*Dosovitskiy et al., 2021)*.

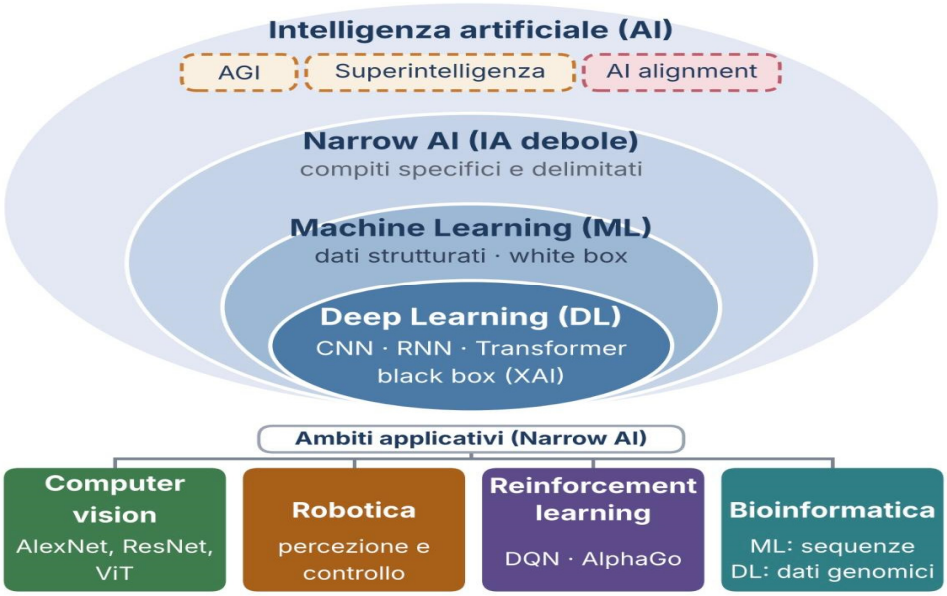
La robotica rappresenta un altro settore di forte integrazione con l'intelligenza artificiale. Un robot può essere inteso come un sistema fisico che percepisce l'ambiente attraverso sensori, elabora le informazioni acquisite e agisce attraverso attuatori, quindi unisce problemi di percezione, pianificazione e controllo *(Siciliano & Khatib, 2016).* Essedo che i sensori sono rumorosi e l'ambiente non molto affidabile, la robotica ha adottato approcci probabilistici per la stima dello stato e la navigazione *(Thrun et al., 2005).*

In realazione alla robotica ritroviamo il *reinforcement learning* (apprendimento per rinforzo), un paradigma in cui un agente impara a prendere decisioni interagendo con un ambiente e ricevendo una ricompensa in funzione delle azioni compiute *(Sutton & Barto, 2018)*. Con l'unione tra reinforcement learning e reti neurali è stato sviluppato un agente in grado di apprendere e giocare a diversi videogiochi *Atari 2600* direttamente dalle immagini

dello schermo, raggiungendo prestazioni comparabili a quelle di un giocatore umano esperto *(Mnih et al., 2015).* Un ulteriore risultato rilevante è stato *AlphaGo*, il primo programma a sconfiggere un giocatore professionista umano nel gioco del Go, battendo il campione europeo per 5 partite a 0 *(Silver et al., 2016).*

Lo sviluppo di sistemi di intelligenza artificiale sempre più sofisticati ha inoltre alimentato un dibattito riguardo non soltanto le opportunità offerte da queste tecnologie, ma anche ai problemi di sicurezza, controllo e allineamento. In particolare, il concetto di *AI alignment*, o allineamento dell'intelligenza artificiale, riguarda in termini generali la necessità di sviluppare sistemi il cui comportamento sia coerente con gli obiettivi e i vincoli stabiliti dagli esseri umani. *Eliezer Yudkowsky* ha dedicato una parte significativa della propria attività alla discussione dei rischi associati allo sviluppo di sistemi di intelligenza artificiale altamente avanzati, sostenendo che la definizione e il mantenimento di obiettivi compatibili con gli interessi umani costituiscano una delle principali difficoltà del problema dell'allineamento *(Yudkowsky, 2022)*. Le posizioni di *Yudkowsky* costituiscono una prospettiva particolarmente orientata all'analisi dei rischi estremi associati all'AGI e devono pertanto essere considerate all'interno del più ampio dibattito scientifico e filosofico sulla sicurezza dei sistemi di intelligenza artificiale.

Un ambito nel quale l'intelligenza artificiale e il machine learning hanno assunto particolare rilevanza è quello della bioinformatica. La bioinformatica costituisce un campo interdisciplinare nel quale informatica, statistica, matematica e scienze biologiche vengono integrate allo scopo di elaborare e interpretare grandi quantità di dati biologici. La crescita delle tecnologie di sequenziamento e delle metodologie di analisi molecolare ha determinato un certo aumento della quantità e della complessità dei dati disponibili, rendendo necessario lo sviluppo di strumenti computazionali in grado di analizzarli in maniera efficiente. Gli algoritmi di Machine Learning possono essere utilizzati per analizzare dati genomici oppure per identificare pattern all'interno delle sequenze biologiche.. L'utilizzo di tecniche di Deep Learning ha inoltre ampliato le possibilità di analisi dei dati genomici, permettendo di affrontare problemi contenenti un'elevata complessità e una grandi quantità di variabili *(Eraslan et al., 2019)*



1.2.2 Machine Learning e Deep Learning

Il *machine learning* (ML) rappresenta un campo interdisciplinare in rapida espansione, nel quale vengono investite ingenti risorse per integrare metodologie informatiche e statistiche nella risoluzione di problemi clinici *(Handelman et al., 2018).* Gli autori definiscono il ML come una tecnologia particolarmente adatta a gestire moli di dati complesse , e contemporaneamente raffigurare una caratteristica intrinseca nel contesto medico, dove informazioni cliniche, laboratoristiche, di imaging e molecolari coesistono con formati e strutture variabili.

I sostenitori di questa tecnologia ne sottolineano il potenziale trasformativo per tre ambiti principali: la ricerca biomedica, la medicina personalizzata, e la diagnostica assistita dal computer, con l'obiettivo ultimo di potenziare in maniera significativa l'assistenza sanitaria su scala globale *(Handelman et al., 2018).* Nonostante queste premesse, gli autori mostrano un limite tuttora rilevante: i concetti teorici alla base dei ML risultano spesso poco conosciuti dalla maggior parte dei professionisti sanitari, un gap culturale e formativo che contribuisce a mantenere in gran parte inespresso il potenziale di questa tecnologia come strumento di ricerca in ambito clinico *(Handelman et al., 2018).*Ragion per cui, gli autori propongono una panoramica dei fondamenti teorici del ML , degli algoritmi più diffusi nel campo medico, analizzandone limiti e criticità e identificandone le prospettive future di applicazione *(Handelman et al., 2018).*

Linearmente con queste premesse, *Swanson et al.* (2023) sottolineano come il ML sia sempre più utilizzato in oncologia clinica lungo l'intero percorso di cura del paziente , dalla diagnosi del tumore alla previsione dei risultati clinici (prognosi), fino al supporto nella pianificazione del trattamento *(Swanson et al., 2023).* Gli autori esaminano nello specifico come queste tecniche vengono applicate a due principali tipologie di dati: da un lato l'imaging medico, dall'altro i dati molecolari ottenuti tramite biopsie tumorali, sia solide che liquide *(Swanson et al., 2023).* Un aspetto rilevante sottolineato dallo studio riguarda le considerazioni chiave da tenere in conto nello sviluppo di modelli di ML, che devono differenziarsi in base alla natura del dato analizzato, proprio perchè l'imaging e i dati molecolari presentano sfide metodologiche distinte in termini di acquisizione, rappresentazione e interpretazione *(Swanson et al., 2023)*.

Infine, gli autori esaminano anche i modelli di ML già approvati dalle agenzie per l'uso clinico in ambito oncologico, offrendo al contempo alcuni approcci volti a migliorarne l'utilità clinica complessiva *(Swanson et al., 2023).*

Il *deep learning* (DL) rappresenta una sottodisciplina dell'intelligenza artificiale che si inserisce come una branca avanzata del machine learning stesso. Esso si fonda sull'utilizzo delle reti neurali artificiali, architetture computazionali ispirate al funzionamento del sistema nervoso biologico, capaci di estrarre pattern complessi ed effettuare previsioni a partire da ampi insiemi di dati *(Tran et al. 2021).*

A differenza degli algoritmi di ML più tradizionali, che richiedono una fase di progettazione manuale delle caratteristiche (ossia la selezione a priori delle variabili rilevanti da parte di un esperto), le reti neurali profonde sono in grado di assimilare in maniera del tutto autonoma rappresentazioni gerarchiche dei dati grezzi, individuando pattern anche molto complessi. Ciò mette in risalto la loro capacità di analizzare dati biologici ad alta dimensionalità , come quelli generati dalle moderne piattaforme "omiche".

La crescente adozione del Deep Learning in ambito sanitario, unita alla progressiva disponibilità di dataset oncologici, ha accelerato in maniera preponderante la ricerca sull'utilità di questa tecnologia nell'analisi della complessità biologica del cancro *(Tran et al., 2021).* Dal punto di vista applicativo, il Deep Learning trova impiego in tre fasi principali: la diagnosi, la prognosi e la gestione terapeutica dei tumori *(Tran et al., 2021)*, configurandosi quindi come una tecnologia trasversale all'intero percorso di cura, dalla caratterizzazione iniziale della neoplasia alla selezione del trattamento più appropriato.

1.2.3 Principali differenze tra gli approcci

A livello metodologico, *Machine Learning* e *Deep Learning* si differenziano innanzitutto per il processo di *feature engineering*. Negli algoritmi di ML tradizionale (ad esempio alberi decisionali, support vector machine o regressione logistica), l'estrazione e la selezione delle caratteristiche ottenute dai dati grezzi richiede intervento di un esperto di dominio, che individua manualmente le variabili più informative per il compito da svolgere *(Hastie, Tibshirani, & Friedman, 2009).* Il Deep Learning supera questa limitazione grazie alla capacità delle reti neurali profonde di apprendere le rappresentazioni delle feature direttamente dai dati grezzi, attraverso strati successivi che catturano pattern via via più astratti *(Krizhevsky, Sutskever, & Hinton, 2012).*

I due paradigmi si differenziano per i requisiti in termini di dati e risorse computazionali. Gli algoritmi di ML classico tendono a fornire prestazioni soddisfacenti anche in presenza di dataset di dimensioni contenute e possono essere addestrati su architetture *hardware standard* (CPU). Al contrario, le reti di Deep Learning necessitano generalmente di grandi volumi di dati per raggiungere buone capacità di generalizzazione e richiedono infrastrutture di calcolo dedicate, quali GPU o TPU, per gestire l'elevato numero di parametri e le operazioni matriciali coinvolte nel processo di addestramento *(Goodfellow, Bengio, & Courville, 2016).*

Un altro aspetto discriminante riguarda l'interpretabilità dei modelli. Numerosi algoritmi di ML tradizionale, come gli alberi decisionali o i modelli lineari, sono considerati modelli *"white box"*, in quanto il loro funzionamento interno è piuttosto trasparente e le decisioni prese possono essere ricostruite e giustificate. Le reti neurali profonde, invece, sono spesso definite modelli *"black box"*, perchè l'elevato numero di parametri e la non linearità delle trasformazioni rendono difficile ricostruire il ragionamento che porta a una determinata predizione; questo limite ha favorito lo sviluppo del filone di ricerca noto come *Explainable Artificial Intelligence* (XAI) *(Murphy, 2022).*

I due approcci si distinguono anche per il tipo di problemi e di dati per cui risultano più efficaci. Il Machine Learning tradizionale è particolarmente indicato per dati strutturati e tabulari, dove le feature sono ben definite e la relazione tra input e output è relativamente semplice da modellare. Il Deep Learning, grazie ad architetture specializzate come le *reti neurali convoluzionali* (CNN), le *reti ricorrenti* (RNN) e i più recenti modelli Transformer, si è invece affermato come l'approccio di riferimento per l'elaborazione di dati non strutturati, quali immagini, audio, testo e video *(Vaswani et al., 2017; LeCun, Bengio, & Hinton, 2015).*

1.3 Applicazione dell'AI in oncologia

L'intelligenza artificiale (IA) applicata all'oncologia sta attraversando una fase di transizione decisiva: da ambito prevalentemente confinato allo sviluppo e alla validazione algoritmica, si sta affermando come componente integrata della pratica clinica quotidiana *(Lotter et al., 2024).* Questo passaggio segna un cambiamento molto rilevante, in quanto sposta il baricentro della ricerca dalla pura dimostrazione di fattibilità tecnica ,ovvero la capacità di un modello di raggiungere elevate performance su compiti circoscritti , alla verifica della sua reale applicazione all'interno dei percorsi di cura oncologici.
Nella loro revisione, *Lotter et al. (2024)* propongono una classificazione delle applicazioni dell'IA articolata secondo due assi principali: il tipo di tumore e l'ambito clinico di applicazione. Gli autori concentrano l'analisi sulle quattro neoplasie a maggiore incidenza, esaminando come l'IA venga impiegata lungo l'intero percorso di cura del paziente oncologico, dalle fasi più precoci a quelle terapeutiche. In particolare, vengono individuati tre momenti del percorso clinico in cui l'IA sta trovando una reale applicazione :
Rilevamento (detection) : identificazione precoce di lesioni sospette o anomalie
Diagnosi : classificazione e caratterizzazione della malattia
Trattamento: supporto alla decisione terapeutica e alla gestione del paziente.

Un elemento centrale sottolineato dagli autori è che queste applicazioni non si basano su un'unica tipologia di dato, ma integrano fonti informative, tra cui diagnostica per immagini, dati genomici e informazioni contenute nelle cartelle cliniche , a conferma di come la natura multimodale dei dati oncologici rappresenti ormai un presupposto strutturale, e non più un'eccezione, per lo sviluppo di soluzioni di IA clinicamente rilevanti *(Lotter et al., 2024).*

La revisione mette in evidenza l'attenzione dedicata al processo di traslazione clinica, ovvero al percorso spesso complesso e non lineare, che porta un algoritmo dalla fase di ricerca al concreto utilizzo pratico. Gli autori mostrano come diverse soluzioni di IA stiano già superando la fase di ricerca e sviluppo per giungere a una diretta integrazione clinica, un dato che testimonia la maturità crescente del settore ma che, al contempo, rende ancora più urgente una riflessione strutturata sulle condizioni necessarie per un'adozione efficace ed equa di questi strumenti (Lotter et al., 2024).

Non mancano sicuramente elementi di criticità: la revisione individua ambiti emergenti ancora da esplorare e soluzioni in evoluzione che il settore sta cercando di sviluppare per affrontarle. Questi ostacoli ,che, come si vedrà nei paragrafi successivi, trovano spazio e approfondimento negli altri contributi analizzati in questa tesi , pongono le basi per comprendere perché il potenziale dell'IA in oncologia, per quanto ampiamente dimostrato a livello algoritmico, richieda ancora un percorso di consolidamento prima di potersi mostrare in una forma sistemica della pratica clinica.

L'oncologia clinica ad oggi è uno dei terreni più complessi per l'applicazione dell'intelligenza artificiale (IA) in medicina. Il percorso diagnostico-terapeutico del paziente oncologico è caratterizzato da una complessità di fonti informative, tra cui: l'istopatologia digitale, i dati genomici e molecolari, nonché le cartelle cliniche elettroniche longitudinali *(Lipkova et al., 2022; Lin et al., 2026).* Ciascuna di

queste modalità racchiude un aspetto fondamentale, ma anche clinicamente molto significativo. La radiologia, ad esempio, fornisce informazioni macroscopiche sulla morfologia e sull'evoluzione della malattia, l'istologia digitale consente di analizzare l'architettura tissutale a livello microscopico, la genomica descrive il profilo molecolare del tumore, mentre le cartelle cliniche restituiscono l'andamento longitudinale della storia clinica del paziente *(Lipkova et al., 2022).*

Come sottolineano *Lipkova et al. (2022),* la maggior parte dei modelli di IA sviluppati finora lavorano in un sistema monomodale , cercando di elaborare un solo dato alla volta. Questa prospettiva, se da un lato ha permesso di raggiungere determinati risultati, quali la classificazione istologica o la segmentazione di lesioni radiologiche, dall'altro comporta un'inevitabile perdita del contesto clinico complessivo, limitando il potenziale interpretativo dei modelli stessi. L'integrazione di più modalità di dati, invece , offre la possibilità di aumentare la l'accuratezza diagnostica e prognostica, avvicinando concretamente l'IA alla pratica clinica reale, utile a spiegare differenze negli esiti clinici o meccanismi di resistenza al trattamento *(Lipkova et al., 2022).*
Questa esigenza di integrazione multimodale è stata ulteriormente confermata e ampliata dai più recenti sviluppi tecnologici.

Lin et al. (2026) ci mostrano come la comparsa di modelli di base (Foundation Models) e di modelli linguistici di grandi dimensioni (LLM) abbia reso possibile una rappresentazione più solida e un ragionamento cross-modale, superando i limiti dei sistemi monomodali tradizionali. Parallelamente, la crescita di modelli computazionali innovativi, come le cosiddette cellule virtuali e i framework di fondazione meccanicistica, introduce un nuovo paradigma per la simulazione delle risposte cellulari e delle interazioni farmaco-tumore, con ricadute dirette sulla progettazione terapeutica e sulla scoperta di nuovi farmaci *(Lin et al., 2026).*

Va però osservato che questi progressi metodologici si collocano in un percorso di applicazione clinica ancora in fase di sviluppo. Già *Elemento et al. (2021),* in un contributo di natura più prospettica, avevano sottolineato come la disponibilità di grandi quantità di dati (big data) non fosse sufficiente a garantire la piena realizzazione del potenziale dell'IA in oncologia, sia in ambito di ricerca sia in ambito clinico: secondo gli autori, restano infatti ostacoli di natura metodologica, regolatoria ed etica che devono essere superati prima di poter tradurre le potenzialità dell'IA in strumenti affidabili per la diagnosi.

Possiamo dire che i tre contributi analizzati seguono un percorso coerente : da una prima fase di riflessione sulle potenzialità e sugli ostacoli dell'IA in oncologia *(Elemento et al., 2021),* si è passati alla definizione di un impianto metodologico per l'integrazione multimodale dei dati e per l'interpretabilità dei modelli (Lipkova et al., 2022), fino ad arrivare agli sviluppi più recenti, che vedono nei modelli di base, negli LLM gli strumenti chiave per superare i limiti dei sistemi monomodali (Lin et al., 2026).

1.3.1 Prognosi e diagnosi

Tra gli ambiti clinici in cui l'intelligenza artificiale (IA) ha raggiunto i risultati più maturi vi sono la diagnosi e la prognosi oncologica. La diagnosi mira a identificare e correttamente la natura della malattia, distinguendo lesioni benigne da maligne o definendo il sottotipo istologico del tumore, mentre la prognosi si occupa di stimare l'evoluzione attesa della malattia e la probabilità di specifici esiti clinici *(Lotter et al., 2024).*

Un esempio dello sviluppo raggiunto nel settore oncologico è offerto da *Bulten et al. (2022),* che in PANDA *(Prostate cANcer graDe Assessment)* hanno dimostrato come algoritmi di Deep Learning applicati a immagini istopatologiche digitalizzate raggiungano una realtà concreta comparabile a quella di patologi esperti nella valutazione del grado di *Gleason* nel carcinoma prostatico. Questo risultato ci mostra come la valutazione istologica del grado tumorale è di base soggetta a un'elevata variabilità interosservatore, e la sua standardizzazione tramite IA ha ricadute dirette sia sull'accuratezza diagnostica sia sulla riproducibilità dei referti a livello internazionale.

L' IA sta assumendo un ruolo crescente nella cosiddetta diagnostica molecolare "indiretta". *Shmatko et al. (2022)* raccontano come i modelli di Deep Learning applicati a immagini istologiche siano in grado di prevedere mutazioni genetiche senza ricorrere a test genomici specifici. Ciò permette di ridurre tempi e costi della caratterizzazione molecolare del tumore, rendendo più accessibile una diagnostica di precisione anche in contesti limitanti.

Come già osservato da *Elemento et al. (2021)*, il potenziale diagnostico e prognostico dell'IA in oncologia, per quanto ampiamente dimostrato a livello di ricerca, rimane condizionato dal superamento di ostacoli legati alla qualità dei dati e alla necessità di convalidazioni precise prima di un determinato utilizzo clinico .

1.3.2 Predizione della risposta ai trattamenti

Accanto alla diagnosi e alla prognosi, un terzo settore in cui l'intelligenza artificiale (IA) sta mostrando un impatto clinico rilevante è quello della predizione della risposta ai trattamenti. Si tratta di un compito caratterizzante in oncologia, in quanto una parte consistente dei pazienti viene sottoposta a terapie che comportano tossicità senza trarne un reale beneficio clinico. Identificare in anticipo, tramite modelli predittivi, quali pazienti risponderanno realmente a uno specifico trattamento rappresenta una delle applicazioni più concrete dell'IA in ambito oncologico.

Un esempio riguarda il carcinoma prostatico localizzato, dove la terapia di *deprivazione androgenica* (ADT) associata alla radioterapia può apportare benefici clinici importanti, ma comporta anche un impatto negativo sulla qualità di vita del paziente. *Spratt et al. (2023)*, in uno studio pubblicato su *NEJM Evidence* e condotto per conto del *NRG Prostate Cancer AI Consortium,* hanno sviluppato un modello di IA integrando immagini di patologia digitale del tessuto prostatico pretrattamento con dati clinici provenienti da *5727* pazienti arruolati in cinque studi randomizzati di fase 3. Il modello, validato in modo indipendente sulla coorte dello studio *NRG/RTOG 9408* con un follow up mediano di quasi quindici anni, si è dimostrato capace di identificare i pazienti destinati a beneficiare della terapia ormonale a breve termine, mostrando come oltre la metà di questi sia a rischio *(Istituto Nazionale dei Tumori ▪).*

Un secondo esempio interessa la chemioterapia neoadiuvante nel carcinoma mammario, dove solo una parte delle pazienti raggiunge una risposta patologica completa, mentre le restanti vengono esposte agli effetti tossici del trattamento. *Guo et al. (2024)*, in uno studio pubblicato su *npj Precision Oncology*, hanno proposto un modello di Deep Learning cross modale in grado di fornire immagini di patologia digitale ottenute da biopsie con immagini ecografiche acquisite in più momenti temporali, al fine di predire precocemente la risposta patologica completa alla chemioterapia neoadiuvante. L'applicazione dell'informazione spaziale, derivata dall'evoluzione ecografica della lesione nel corso del trattamento, ha permesso al modello di raggiungere un'accuratezza predittiva elevata, offrendo alle pazienti la possibilità di un adattamento tempestivo della strategia terapeutica in caso di risposta insufficiente e poco affidabile.

Entrambi gli studi hanno in comune un importante elemento metodologico: la combinazione di più fonti informative, quali immagini istopatologiche, dati di imaging e informazioni cliniche, permette di ottenere una capacità predittiva superiore rispetto a quella raggiungibile con singoli parametri clinico patologici.

Sebbene si tratti di studi solidi dal punto di vista metodologico , la loro effettiva adozione nella pratica clinica quotidiana richiede ancora un consolidamento regolatorio e una diffusione su scala più ampia, aspetti che verranno ripresi nella parte conclusiva della tesi dedicata alle sfide e alle prospettive future dell'IA in oncologia.

1.4 La complessità dei dati oncologici

ll cancro rappresenta oggigiorno uno dei problemi medici più articolati da affrontare dal punto di vista informativo, perchè la sua comprensione richiede l'inclusione di dati provenienti da fonti differenti, come dati clinici, immagini radiologiche, referti, ecc…
Le difficoltà nella ricerca oncologica derivano proprio dalla marcata eterogeneità tumorale, dalla crescita esponenziale dei dati e dalla frammentazione dei sistemi informativi. Questa varietà di fonti rende i dati oncologici non solo voluminosi, ma anche strutturalmente disomogenei, portando a sfide reali in termini di raccolta, standardizzazione e analisi.

Negli ultimi anni, i progressi delle tecnologie di sequenziamento ad alto rendimento hanno stabilito un accumulo di dati omici sul cancro, dando origine al concetto di "big data" applicato all'oncologia, la cui analisi richiede risorse computazionali considerevoli e nuove metodologie di interpretazione .

L'accumulo rapido di grandi quantità di questa tipologia di dati, favorito dalle innovazioni nelle tecnologie ad alto rendimento, ha dato origine a un concetto in evoluzione di *"big data"* applicato all'oncologia, la cui analisi richiede numerose risorse computazionali. A ciò si aggiunge la crescente disponibilità di dati generati automaticamente, come quelli derivanti dalla radiomica e dall'analisi digitale delle immagini, che affiancano i tradizionali dati clinici raccolti nella pratica quotidiana .Gran parte dei dati oncologici non deriva più dai referti clinici della pratica quotidiana, ma da analisi automatiche computerizzate come l'analisi digitale delle immagini.

In molti contesti i dati oncologici restano dispersi tra diverse istituzioni, con cartelle cliniche elettroniche (EHR), immagini, referti istopatologici e dati genomici che mancano di una reale integrazione. In particolare le prime citate rivestono un ruolo fondamentale

nella gestione di tutte le informazioni inerenti al paziente. Se fino a pochi anni fa la gestione dei dati clinici si basava sulle cartelle cliniche elettroniche (EHR), archivi digitali in cui le informazioni venivano inserite e consultate in gran parte manualmente, oggi i modelli linguistici di grandi dimensioni (LLM) permettono di lavorare direttamente sul testo libero: sintetizzare le cartelle, estrarre dati e generare bozze di documentazione. Ne derivano benefici in termini di efficienza, ma anche questioni di affidabilità, privacy e responsabilità. Trattandosi di uno strumento imprescindibile per la continuità delle cure, l'analisi delle EHR sarà ripresa e sviluppata nel capitolo 3.

La complessità dei dati oncologici, non rappresenta soltanto un fenomeno recente legato ai big data: già le prime esperienze di registri tumori ed elettronici avevano evidenziato la difficoltà di raccogliere in modo uniforme e di qualità elevata informazioni cliniche inerenti al percorso oncologico dei pazienti

1.4.1 Dati clinici

I dati clinici rappresentano una delle componenti più notevoli del mondo informativo oncologico. Con l'espressione "dati del mondo reale" *(Real-World Data, RWD)* si fa riferimento alle informazioni relative allo stato di salute e all'erogazione dell'assistenza raccolte al di fuori dei contesti sperimentali controllati. Questi dati generano evidenza del mondo reale *(Real-World Evidence, RWE),* la cui produzione scientifica in ambito oncologico è cresciuta in maniera radicale negli ultimi anni, affiancando sempre più spesso le conoscenze provenienti dai trial clinici randomizzati *(Kokkotou et al., 2024).*

I dati clinici oncologici presentano criticità sul piano della qualità. Gran parte delle informazioni cliniche, come lo stadio della malattia, la risposta al trattamento o i risultati bioptici, non risiede in campi strutturati della cartella clinica, ma in note testuali non strutturate redatte in linguaggio naturale dal personale sanitario: si stima che circa l'80% dei dati clinici real-world sia contenuto in testo libero non strutturato, la cui estrazione richiede tecniche avanzate di elaborazione del linguaggio naturale e apprendimento automatico *(Benedum et al., 2023; Tait et al., 2025).*

Sul piano degli standard, il progetto ICAREdata, promosso dalla *National Cancer Institute's National Clinical Trials Network* in collaborazione con *MITRE Corporation*, ha utilizzato il modello *minimal Common Oncology Data Elements* (mCODE) per raccogliere in formato strutturato dati di esito oncologico direttamente dalla cartella clinica elettronica di dieci centri clinici, , dimostrandone la fattibilità con un impatto organizzativo contenuto *(Casagni et al., 2024; George et al., 2024).* Iniziative di questo tipo rappresentano la standardizzazione dei formati e dei modelli dato, i quali costituiscono una delle strade più promettenti per rendere i dati clinici oncologici realmente confrontabili e riutilizzabili su larga scala.

1.4.2 Dati genomici e molecolari

Accanto ai dati clinici, i dati genomici e molecolari costituiscono una delle fonti più ricche e al tempo stesso più complesse dell'oncologia contemporanea. L'avvento delle tecnologie di sequenziamento di nuova generazione *(Next-Generation Sequencing, NGS)* ha consentito lo studio di un'analisi genomica dei tumori rapida, accurata e a costi progressivamente più contenuti, permettendo di sequenziare simultaneamente centinaia di

geni e di ottenere informazioni dettagliate su DNA, RNA, DNA libero circolante (cell-free DNA) e DNA tumorale circolante (ctDNA) a partire da un singolo campione biologico *(Zalis et al., 2024)*. Questa capacità ha reso l'NGS uno strumento diagnostico indispensabile nella pratica oncologica, alla base della profilazione molecolare dei tumori e dell'identificazione delle alterazioni genetiche che ne guidano la crescita *(Ghoreyshi et al., 2025)*.

La quantità e la varietà di informazioni generate dal sequenziamento genomico mostrano numerose sfide notevoli. L'elevato volume di dati prodotti da un singolo test NGS richiede strumenti bioinformatici avanzati per essere analizzato e tradotto. L'interpretazione del significato clinico di ogni variante genetica individuata resta un processo complesso, che deve tenere conto di database di riferimento in continua evoluzione, questioni di consenso informato e problematiche relative alla privacy dei dati genetici *(Ghoreyshi et al., 2025)*.

I dati genomici potrebbero affiancancarsi ad altre tipologie di dati "omici", quali trascrittomica, proteomica, epigenomica e metabolomica,. Sebbene l'analisi computazionale e statistica dei singoli livelli omici sia ormai relativamente consolidata, l'integrazione tra livelli diversi rimane un ambito tutt'altro che standardizzato: le tecnologie di acquisizione dei dati presentano un'elevata variabilità, la crescita del volume dei dataset supera spesso la capacità di elaborazione disponibile, e mancano ancora metodologie condivise per combinare in modo biologicamente coerente informazioni provenienti da diverse piattaforme sperimentali (*Menyhárt & Győrffy, 2021)*.

Come risposta a tutte queste criticità, un numero di studi propone l'impiego di tecniche di intelligenza artificiale e di apprendimento profondo (Deep Learning) capaci di integrare dati genomici e istopatologici in un'unica analisi multimodale, migliorando la capacità predittiva e diagnostica rispetto all'uso dei singoli livelli di dato presi isolatamente *(Unger & Kather, 2024)*. Analogamente, l'applicazione di metodi di apprendimento automatico ai flussi diagnostici molecolari basati su NGS sta emergendo come strumento in grado di velocizzare l'interpretazione delle varianti e alleggerire il lavoro dei *tumor board* molecolari, pur richiedendo una realizzazione e una validazione clinica rigorosa *(Reardon et al., 2026)*.

1.4.3 Imaging medico e patologia digitale

L'imaging medico e la patologia digitale rappresentano due ulteriori pilastri dell'ecosistema dei dati oncologici, entrambi caratterizzati da volumi di informazione particolarmente elevati e da una natura intrinsecamente non strutturata. Le tecniche di imaging radiologico, quali tomografia computerizzata (TC), risonanza magnetica (RM) e tomografia a emissione di positroni (PET), producono immagini tridimensionali ad alta risoluzione utilizzate per la diagnosi e il monitoraggio della risposta al trattamento oncologico. La digitalizzazione, invece, ha dato origine alla patologia digitale, trasformando il vetrino istologico tradizionale in un file ad altissima risoluzione, spesso dell'ordine di diversi gigabyte per singolo caso.

Un importante sviluppo nell'analisi delle immagini radiologiche è rappresentata dalla radiomica, un approccio che consente di estrarre da immagini di uso clinico standard un numero elevato di caratteristiche quantitative (forma, intensità, texture), trasformando così l'immagine medica in un vero e proprio insieme di dati ad alto contenuto informativo. La radiomica è stata proposta come un ponte tra l'imaging medico e la medicina personalizzata, in grado di supportare la diagnosi, la prognosi e la previsione della risposta

terapeutica; ma, la sua piena maturazione come disciplina resta condizionata dalla mancanza di criteri standardizzati di validazione scientifica e clinica delle numerose pipeline di analisi proposte in letteratura *(Lambin et al., 2017).*

La digitalizzazione dei vetrini istologici ha consentito l'applicazione di strumenti di intelligenza artificiale e apprendimento automatico all'analisi dei tessuti tumorali, estraendo fenotipi morfometrici non visibili a occhio nudo e in grado di migliorare la gestione clinica del paziente oncologico *(Bera et al., 2019).* Questa trasformazione digitale della patologia, apre nuove prospettive diagnostiche e di ricerca, ma contemporaneamente pone sfide inerenti alla qualità dell'acquisizione delle immagini, di standardizzazione dei formati e di gestione di volumi di dati *(Niazi et al., 2019).*

Una revisione ha esaminato che, tramite tecniche di Deep Learning, uniscono l'analisi di immagini istopatologiche con dati omici, individuando diverse strategie di fusione multimodale delle reti neurali convoluzionali con i dati genomici. I risultati mostrano che gli approcci multimodali, capaci di integrare congiuntamente informazione visiva e molecolare, tendono a produrre prestazioni superiori rispetto ai modelli basati su una sola tipologia di dato. L'elevata differenza metodologica degli studi esistenti — in termini di algoritmi di fusione, tipologie tumorali ed endpoint clinici considerati — limita ancora la possibilità di trarre conclusioni generalizzabili e sottolinea la necessità di protocolli più uniformi per la validazione di questi modelli *(Schneider et al., 2022).*

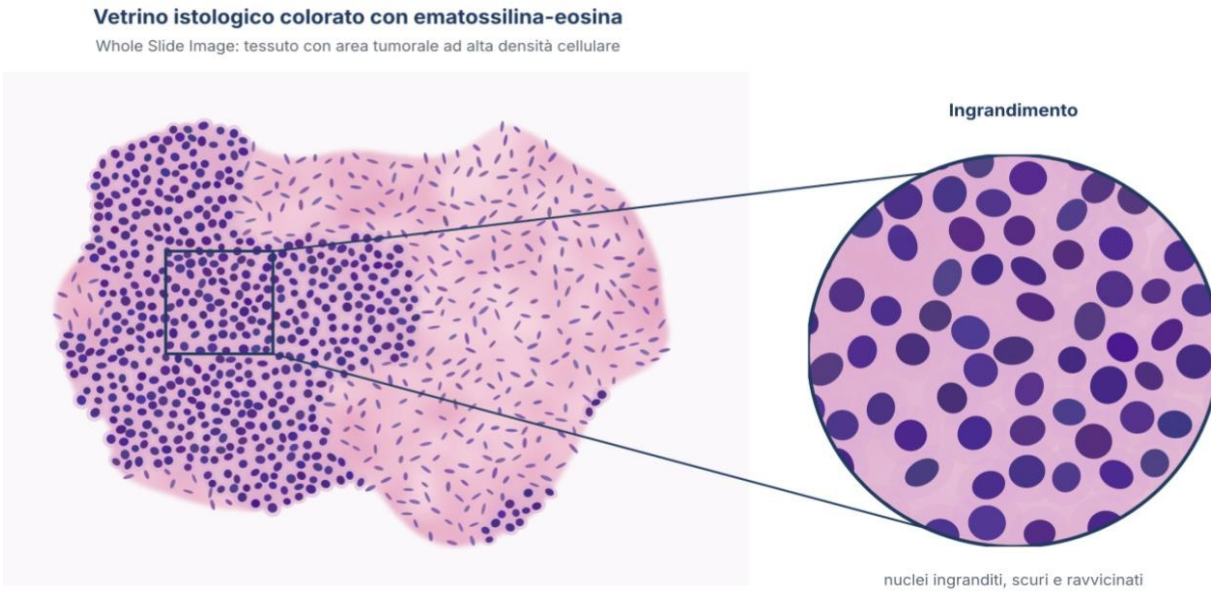


Immagine sintetica generata al computer a scopo illustrativo; non rappresenta un caso clinico reale.

1.5 Limiti degli approcci tradizionali

Alla luce di quanto descritto nei paragrafi precedenti, risulta evidente come i metodi statistici e informatici impiegati in oncologia mostrino limiti di fronte alla mole delle differenze e alla dimensionalità dei dati oggi disponibili. Gli approcci statistici classici, sviluppati per contesti in cui il numero di osservazioni supera notevolmente il numero

di variabili, si trovano spesso a operare nella condizione opposta quando applicati ai dati omici oncologici, in cui il numero di variabili misurate (geni, trascritti, proteine) va a superare il numero di pazienti disponibili: una situazione nota in letteratura come problema "p maggiore di n", che limita l'applicabilità dei metodi di regressione classica e dei test di ipotesi tradizionali *(Rahnenführer et al., 2023).*

I metodi statistici tradizionali, che analizzano tipicamente le variabili in modo indipendente l'una dall'altra, hanno difficoltà a cogliere segnali deboli distribuiti su più geni di uno stesso pathway, segnali che risultano più facilmente rilevabili se analizzati in modo aggregato; questo limite ha reso necessario lo sviluppo di metodi statistici più avanzati, in grado di incorporare esplicitamente la conoscenza biologica pregressa nell'analisi dei dati *(Zhao, Chang, & Long, 2019).*

Anche nel settore degli studi clinici randomizzati, considerati il gold standard della ricerca oncologica, l'introduzione di dati ad alta dimensionalità ha messo in risalto l'inadeguatezza dei metodi di analisi convenzionali.

Tutti questi limiti mostrano come gli strumenti statistici ereditati dalla ricerca clinica, per quanto rigorosi e consolidati, non siano stati concepiti per gestire relazioni non lineari e l'integrazione di più fonti di dati eterogenee, come quelle descritte nei paragrafi precedenti. A differenza dei metodi statistici classici, che si basano tipicamente su ipotesi predefinite circa la struttura dei dati, le tecniche di apprendimento automatico sono in grado di analizzare relazioni non lineari intrinseche ai dati biomedici, risultando adatte ad affrontare le sfide della ricerca oncologica. *(Yadav & Gupta, 2026).* È proprio in risposta a questi limiti strutturali degli approcci tradizionali che, come si vedrà nei capitoli successivi, l'oncologia ha iniziato a rivolgersi in modo sempre più costante a metodologie computazionali avanzate, come ad esempio l'intelligenza artificiale e strumenti in grado di trasformare la complessità dei dati oncologici da ostacolo a risorsa.

1.5.1 Modelli progettati per singoli compiti

La maggior parte delle applicazioni di intelligenza artificiale sviluppate finora in oncologia condivide una caratteristica strutturale: si tratta di modelli progettati per svolgere un singolo compito ben definito (task-specific o narrow AI), come classificare un sottotipo tumorale, o predire un singolo biomarcatore a partire da un vetrino istologico. Questi modelli vengono addestrati su un compito circoscritto e con un set di dati altamente specifico, e proprio per questo che raggiungono molto spesso prestazioni elevate nel contesto per cui sono stati sviluppati (Kakar, 2026).

Questa impostazione *"a compito singolo"* porta a molteplici conseguenze pratiche sul piano dell'implementazione clinica. Un modello addestrato per riconoscere un singolo tipo di alterazione istologica, ad esempio, tende a generalizzare male se applicato a scanner, popolazioni di pazienti o protocolli di acquisizione differenti da quelli usati in fase di addestramento, e richiede in genere un nuovo ciclo di sviluppo e validazione per ogni nuovo compito o contesto clinico. La presenza di questi strumenti nei flussi di lavoro clinici resta alquanto limitata da problemi di qualità dei dati*(Elemento, Khozin, & Sternberg, 2025).*

La frammentazione è abbastanza evidente in patologia digitale, dove la letteratura descrive un vero e proprio cambio di paradigma: dagli algoritmi task-specific, sviluppati per compiti quali la classificazione tumorale, la sottotipizzazione o la predizione di specifici biomarcatori, verso modelli fondazionali più versatili — come UNI, CONCH, GigaPath, mSTAR e Atlas — addestrati su dataset massivi ed eterogenei tramite apprendimento auto-supervisionato e multimodale, capaci di adattarsi a molteplici compiti senza dover essere riaddestrati da zero per ciascuno di essi *(Cheng & Wong, 2025).*

1.5.2 Difficoltà nell'integrazione di fonti eterogenee

Uno dei principali ostacoli all'implementazione clinica dell'intelligenza artificiale in oncologia riguarda la necessità di integrare dati provenienti da sorgenti eterogenee come ad esempio immagini radiologiche e istopatologiche, dati genomici e multi-omici. Come sottolineano *Acosta* e colleghi in una delle review di riferimento sull'IA, la crescente disponibilità di biobanche, cartelle cliniche elettroniche, imaging medico e dati di sequenziamento genomico rappresenta un'opportunità per individuare la complessità della malattia oncologica, ma pone contemporaneamente sfide tecniche e di privacy ancora irrisolte.

Una prima difficoltà è di natura tecnica ed è inerente al la disomogeneità dei formati, delle scale di misura e delle tecnologie di acquisizione. I dati multi-omici, ad esempio,

derivano spesso da piattaforme differenti, con sensibilità, specificità e protocolli di normalizzazione non comparabili tra loro; ciò comporta la necessità di uniformare i dataset prima ancora di poterli utilizzare per l'addestramento di modelli di apprendimento automatico, con il rischio di introdurre distorsioni (bias) legate a una singola fonte di dati, come nel caso della forte dipendenza della letteratura scientifica dal database TCGA *(The Cancer Genome Atlas).* Il problema dei dati mancanti o parziali, invece, va a mettere in risalto il fatto che non tutte le tecnologie omiche vengono applicate a tutti i pazienti, per ragioni di costo o disponibilità del campione biologico, formando dataset incompleti che i modelli di IA devono in qualche modo essere in grado di gestire senza perdere una certa affidabilità.

Le informazioni cliniche, radiologiche e patologiche di uno stesso paziente sono tavolta conservate in sistemi informativi separati e non interoperabili, con formati di archiviazione eterogenei e scarsa standardizzazione tra ospedali e centri di ricerca. *Boehm* e colleghi, in una review pubblicata su *Nature Reviews Cancer*, analizzano come la maggior parte degli approcci di IA in oncologia sia ancora limitata a una singola modalità di dato, mentre l'integrazione multimodale di diagnostica molecolare avanzata rimanga sottosviluppata proprio a causa di queste barriere strutturali. *Huang* e colleghi, allo stesso modo, dimostrano come l'aggregazione di immagini mediche e cartelle cliniche elettroniche mediante Deep Learning richieda scelte non banali, in quanto le due tipologie di dato presentano dimensionalità e livelli di rumore molto diversi tra loro.

CAPITOLO II – FOUNDATION MODELS E LARGE LANGUAGE MODELS

2.1 Definizione di Foundation Models

Sin dalle loro origini, i modelli di intelligenza artificiale sono stati generati per un singolo compito, ognuno con i propri dati annotati. Ogni applicazione necessita di nuove etichette, ossia un nuovo addestramento e una nuova validazione, e il modello non riesce ad adattarsi al meglio a situazioni diverse rispetto a quelle viste durante la prima fase di preparazione.
I modelli di AI però non risultano essere sempre efficaci, ma talvolta possono presentare anche dei limiti specifici come la scarsa capacità di adattamento a compiti nuovi, dipendenza da dati etichettati e scarsa riutilizzabilità. Il superamento di questo limiti è essenzialmente legato ai modelli di grandi dimensioni pre-addestrati su grandi quantità di dati come *Large Language Models* e *Foundation Models* , i quali sono in grado di rispondere a richieste in linguaggio libero anche senza essere stati preparati in modo specifico per quel compito.

Quando si parla di Foundation Models, si fa riferimento a specifici modelli di intelligenza artificiale (AI) addestrati su un enorme quantità di dati, in grado di svolgere numerosi compiti generali e utilizzati come "base" per la creazione di applicazioni più specializzate.
Essi si distinguono dai classici modelli di machine learning per la loro versatilità e dimensioni. Mentre i modelli tradizionali vengono addestrati su sistemi di dati di piccole quantità, i Foundation Models ricorrono al *transfer learning* , ossia al trasferimento di conoscenze assimilate in un compito verso altri compiti. Questa caratteristica ne permette l'utilizzo in ambiti diversi, come la computer vision, l'elaborazione del linguaggio naturale (NPL) e il riconoscimento vocale.

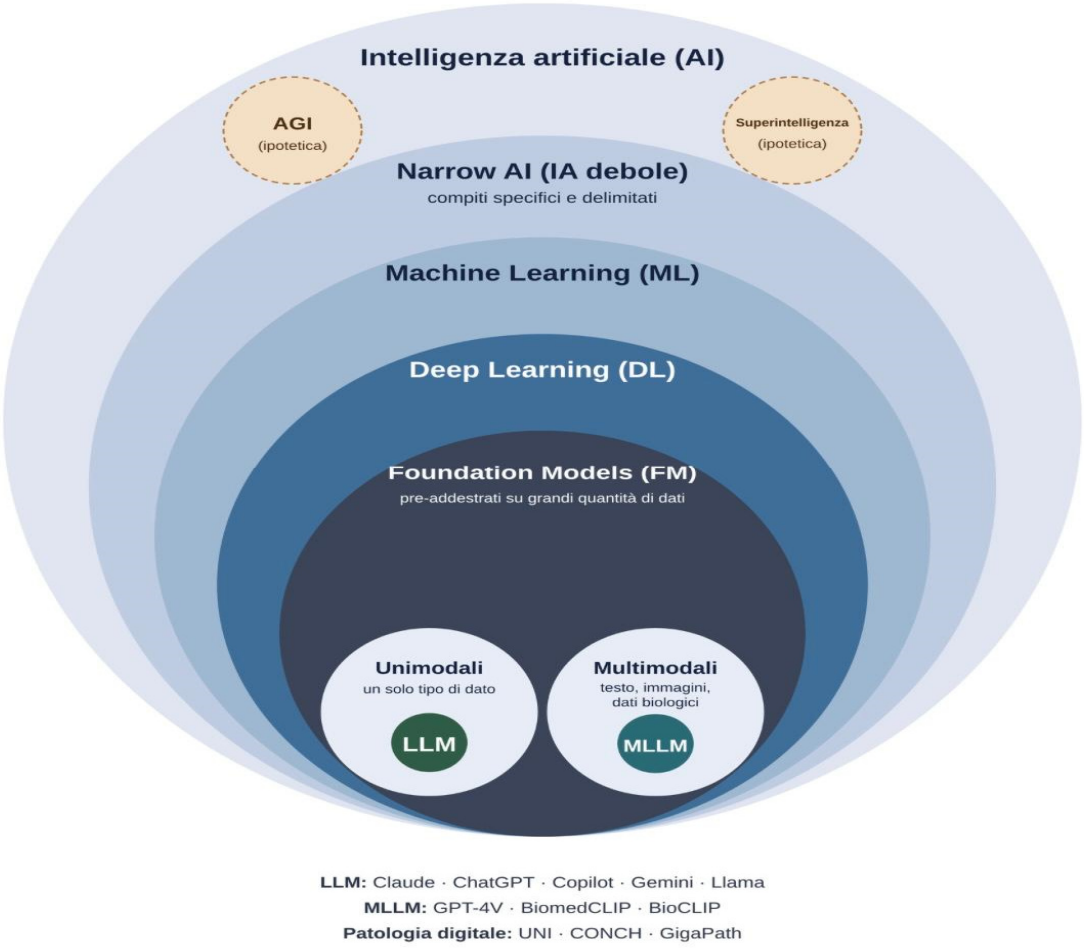
Il termine *"foundation model"* è stato introdotto nel 2021 dai ricercatori del *Center for Research on Foundation Models* (CRFM) e dello *Stanford Institute for Human-Centered Artificial Intelligence* (HAI) dell'Università di Stanford, i quali hanno descritto questa classe di modelli come un vero e proprio "cambiamento di paradigma" (*Bommasani et al., 2021)*. Secondo gli autori, un foundation model è di per sé incompleto, ma costituisce una base comune da cui derivano molteplici modelli per singole applicazioni. La scelta del termine *"foundation"* (fondamenta) vuole mettere in risalto l'importanza della stabilità architetturale e della protezione: fondamenta costruite male compromettono l'intera struttura, mentre fondamenta solide offrono un sostegno affidabile e veritiero per le applicazioni future.

La realizzazione di un Foundation Model segue un processo molto complesso in fasi simili a quelle dello sviluppo di un modello di machine learning tradizionale:

- raccolta dei dati;
- scelta della modalità;
- definizione dell'architettura del modello;
- addestramento;
- valutazione.

La prima fase consiste nel ricavare una massiccia quantità di dati , proveniente da fonti eterogenee. Si tratta di dati non strutturati e privi di etichette, la cui varietà consente al modello di individuare schemi ricorrenti, interpretare il contesto e generalizzare le conoscenze acquisite.

I Foundation Models possono essere **unimodali** o **multimodali**: i primi gestiscono un solo tipo di dato (ad esempio, ricevono testo in input e producono testo in output); i secondi integrano informazioni provenienti da più modalità, come nel caso di un modello che riceve testo e immagini o che genera la trascrizione scritta partendo da una registrazione vocale



Questi ultimi in particolare descrivono in maniera efficiente ed esaustiva il ruolo che occupano i Foundation Models nell'ambito oncologico. Molti di questi modelli si basano su architetture di deep learning, ossia su reti neurali multistrato che mirano a imitare il meccanismo decisionale del cervello umano: il cervello umano tende ad elaborare informazioni attraverso un'ampia rete di miliardi di cellule nervose (neuroni) collegate da sinapsi. Le reti neurali artificiali imitano questa struttura unendo tra loro dati matematici (neuroni artificiali) organizzati in strati profondi (deep learning).

2.1.1 Pre-training su grandi quantità di dati

Secondo lo scrittore Cole Stryker, "un modello pre-addestrato rappresenta un modello di machine learning che è stato precedentemente addestrato su un ampio set di dati per un compito specifico (di solito di carattere generale) e che può essere riutilizzato per un compito diverso ma correlato". Nel caso dei Foundation Models questo principio viene

applicata su larga scala: il pre-training avviene su grandi masse di dati provenienti da fonti differenti come pagine web, libri, immagini ecc...

I modelli linguistici si concentrano maggiormente sull'apprendimento auto-supervisionato, concetto il cui principio si basa sul fatto che il modello impara a prevedere la parola (o il *token*) successiva a partire dal contesto, senza aver bisogno di annotazioni umane.
Dopo questa fase il modello potrà essere adattato a compiti specifici mediante il fine-turning, oppure nel caso di modelli più recenti, tramite istruzioni nel *prompt.*

Il rapporto tra dimensione del modello e quantità di dati rappresenta uno degli aspetti principali in questo ambito, in quanto le cosiddette scaling laws mostrano che le prestazioni migliorano quando questi due fattori crescono insieme, ed è proprio questa caratteristica che rende i Foundation Models così tanto versatili. Non mancano sicuramente numerose criticità come i costo computazionali, la qualità relativa ai dati, i bias che possono essere acquisiti dai corpora di addestramento e le problematiche inerenti a copyright e privacy.

2.1.2 Adattamento a compiti differenti

Uno dei tratti distintivi dei Foundation Models, come detto già nel paragrafo precedente, è la versatilità. Questa caratteristica è una capacità potenziale e non una prestazione garantita, difatti, le rappresentazioni assimilate in fase di pre-addestramento sono ampie e trasferibili, ma non ottimizzate per un obiettivo specifico. È l'adattamento a trasformare questo potenziale in prestazioni affidabili, ed è quindi il meccanismo mediante cui un singolo modello può svolgere compiti differenti senza essere riaddestrato da zero.

Il concetto dei Foundation Models , dove un unico modello pre-addestrato viene adattato a compiti differenti, si è esteso dal linguaggio naturale alla biologia molecolare *(Si et al., 2024*). Come nel linguaggio naturale, un modello pre-addestrato apprende correlazioni dai dati e le rende riutilizzabili in un'ampia gamma di compiti specifici, e questa riutilizzabilità giustifica la denominazione di "modelli di base".

La review di *Si et al. (2024)* divide i modelli in base all'origine dei dati su cui sono costruiti: sequenze di RNA, DNA e proteine, trascrittomica a singola cellula e trascrittomica spaziale. Nei primi tre casi la sequenza biologica viene trattata come un testo, in cui nucleotidi o amminoacidi svolgono il ruolo di token; pre-addestramento, di norma autosupervisionato, permette al modello di assimilare regolarità funzionali e strutturali senza etichette. Nella trascrittomica a singola cellula l'unità di analisi diventa invece il profilo di espressione genica della cellula, mentre i quella spaziale l'espressione è associata anche alla posizione nel tessuto.
Dunque, le strategie descritte precedentemente ,dal prompting al fine-tuning completo fino ai metodi parameter-efficient, trovano in questo settore un terreno di applicazione naturale, in cui la scelta tra di esse dipende dalla quantità di dati annotati disponibili e dalle risorse di calcolo.

2.2 Large Language Models

I modelli linguistici di grandi dimensioni *(Large Language Models, LLM)* sono sistemi di deep learning addestrati su volumi di dati molto ampi, che permettono loro di interpretare e produrre testo in linguaggio naturale e di svolgere compiti molto diversi tra loro. Dal punto

di vista strutturale si basano sul transformer, un tipo di rete neurale piuttosto efficace nel trattare sequenze di parole e nell' individuare regolarità all'interno dei testi.

Funzionano per mezzo di un approccio statistico : il modello stima, passo dopo passo, quale sia la parola più probabile che segua la sequenza già data, sfruttando le regolarità apprese durante l'addestramento per generare nuovo testo.

La rilevanza degli LLM sta nel fatto che sono tra i primi sistemi di intelligenza artificiale in grado di elaborare su larga scala il linguaggio umano non strutturato, consentendo un'interazione più naturale con le macchine. Diversamente dai motori di ricerca tradizionali, che si limitavano a far corrispondere parole chiave, gli LLM captano il contesto, le sfumature di significato e sono in grado di ragionare. Una volta addestrati, possono essere utilizzati in numerose applicazioni legate al testo, come riassumere un articolo o correggere del codice. Inoltre, se dotati di capacità *"agentiche",* possono eseguire in modo autonomo (non del tutto) attività che di base richiederebbero l'intervento umano.

Gli LLM rappresentano il risultato di decenni di ricerca nell'elaborazione del linguaggio naturale (NLP) e nel machine learning, e il loro sviluppo ha avuto un ruolo centrale nello sviluppo costante dell'intelligenza artificiale tra la fine degli anni 2010 e gli anni 2020. Oggi sono accessibili al grande pubblico per mezzo di servizi come Claude (Anthropic), ChatGPT (OpenAI), Copilot (Microsoft), Gemini (Google) e i modelli Llama (Meta), e trovano un impiego fruttuoso anche in ambito aziendale.

La fase di pre-addestramento parte da una raccolta di dati massiva, composta da miliardi o addirittura trilioni di parole provenienti da libri, articoli, pagine web e altre fonti testuali. Questi materiali vengono in un primo momento ripuliti e preparati, eliminando errori, duplicati e contenuti non desiderati. Successivamente, il testo viene scomposto in unità elementari facilmente riconoscibili dalla macchina, dette *token*, attraverso un processo di tokenizzazione. I token possono corrispondere a parole intere, a parti di parola o a singoli caratteri, in questo modo il linguaggio risulta uniforme e i termini rari o di nuova formazione possono essere gestiti in modo più appropriato.

L'addestramento iniziale avviene tramite apprendimento auto-supervisionato *(self-supervised learning)*. Questa tecnica usufruisce di dati privi di etichette, nonostante l'affinità con l'apprendimento supervisionato, in quanto mira a migliorare le prestazioni rispetto a un riferimento *("ground truth").* La differenza è che, in questo caso, i compiti sono costruiti in maniera tale che la risposta corretta possa essere ricavata dai dati stessi, senza bisogno di annotazioni esterne. Difatti, il modello è chiamato a scoprire schemi, strutture e relazioni nei dati, senza che gli venga indicato l'output esatto per ciascun input.

2.2.1 Pre-turning e fine-turning

Lo sviluppo dei moderni LLM si divide in due fasi principali, il pre-addestramento (pre-training) e il successivo affinamento (fine-tuning) . Nella prima fase il modello apprende la struttura e le regole fondamentali del linguaggio; ne ottiene così una comprensione approfondita e generale, ma con mancanza di specializzazione per scopi specifici. . Questo schema ha avuto una prima dimostrazione con *GPT-1 (OpenAI, 2018).*

Il modello parte da pesi indirizzati in modo casuale e viene addestrato su un grande corpus di testo con l'obiettivo di prevedere in modo estremamente precisione il token successivo.

Di base si tratta di un compito di previsione del token successivo, eseguito su un ampio insieme di dati selezionati. Poiché la risposta corretta è contenuta nel testo stesso, non servono annotazioni manuali, e questo rende possibile l'uso di quantità di dati ingestibili. Il costo in risorse di calcolo è molto elevato, infatti, la maggior parte dell'investimento computazionale nella creazione di un LLM avviene in questa fase. Il risultato di questo processo è un modello di base *(base model)* che è in grado di continuare un testo in modo corretto, ma non è ancora un assistente in grado di seguire tutte le istruzioni.

Per quel che concerne il fine-turning, con questo termine si indica l'adattamento di un modello già pre-addestrato a particolari applicazioni o comportamenti, mediante ulteriore addestramento su un insieme di dati più piccolo e mirato. A differenza del pre-training, in questo caso si usa l'apprendimento supervisionato, cioè esempi di input associati agli output desiderati. Avviene che il modello va a riutilizzare la conoscenza già acquisita e ne modifica i parametri per adattarlo quanto più possibile a uno scopo preciso e mirato. Lo stesso procedimento può servire anche ad aggiungere un dominio specialistico, come quello medico, giuridico o informatico.

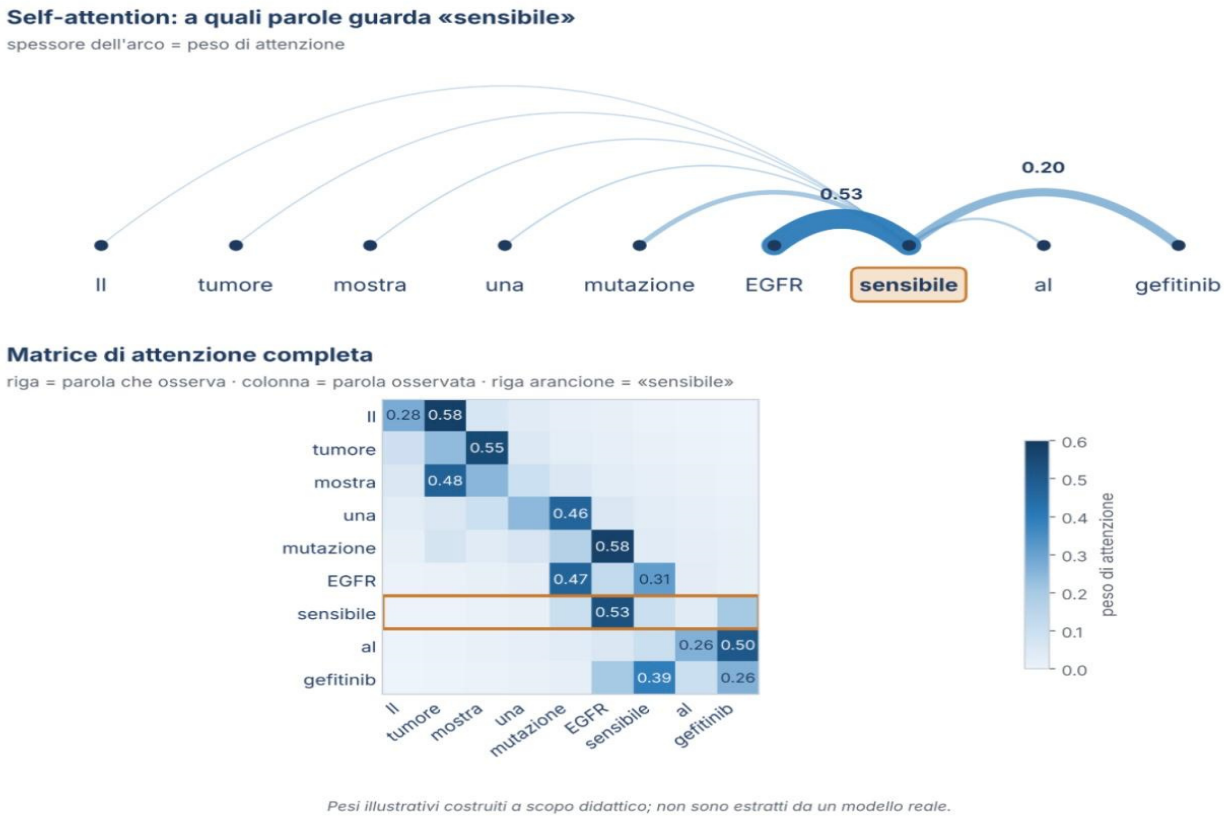
Nella letteratura si distinguono diversi stili di fine-tuning: il primo è il *transfer learning*, in cui il modello pre-addestrato viene adattato ai dati di un compito specifico per migliorarne le prestazioni; il secondo è *l'instruction tuning*, dove il modello viene addestrato su dati in forma di istruzione e coppie input-output per rispondere in modo esaustivo alle richieste degli utenti, migliorando la generalizzazione su compiti mai visti; infine il terzo è *l'alignment tuning,* generato dal fatto che gli LLM possono creare testi falsi o dannosi e e che vanno resi reali e onesti attraverso il check umano.

Per l'allineamento, invece, si utilizza l'apprendimento per rinforzo con feedback umano *(Reinforcement Learning from Human Feedback, RLHF)* . Il fine-tuning supervisionato è stato fondamentale per far seguire al modello le istruzioni, ma la qualità dei dati annotati può essere poco omogenea e la sua capacità di captare preferenze umane è molto limitata. L'RLHF raccoglie giudizi umani sotto forma di preferenze, li usa per addestrare un modello di ricompensa *(reward model)* e, va ad aggiustare pian piano il comportamento del modello. Modelli come GPT-4, Claude e Gemini hanno beneficiato di questi meccanismi, migliorando in numerosi azioni quali ad esempio seguire le istruzioni oppure fornire risposte veritiere ed esaustive.

2.2.2 Architettura Transformer

L'introduzione dell'architettura Transformer, ideata da *Vaswani et al. (2017),* rappresenta il cuore pulsante della tecnologia che ha reso possibile lo sviluppo dei moderni Large Language Models (LLM). Prima del suo avvento, la modellazione del linguaggio si basava su reti neurali sequenziali come RNN, LSTM e GRU. Queste architetture mostravano forti limiti strutturali, ad esempio, l'obbligo di analizzare il testo una parola alla volta impediva di creare un collegamento con il calcolo sulle schede grafiche (GPU) e rendeva difficile per il modello mantenere la memoria di informazioni all'interno di un testo lungo.

Il Transformer ha superato questa criticità eliminando l'elaborazione sequenziale e introducendo il meccanismo di *Self-Attention* (Auto-Attenzione); questo cambio di visuale ha permesso ai modelli di elaborare enormi quantità di testo e di scalare da pochi milioni a centinaia di miliardi di parametri, costruendo le basi per le capacità generative dei modelli linguistici attuali *(Brown et al., 2020).*



Nell'ambito dei Large Language Models, l'architettura si è evoluta in tre gruppi principali:

- *Decoder-Only* (Auto-regressivi), ossia la struttura dominante per gli LLM generativi (come la famiglia GPT, Llama o Mistral). Questi modelli in particolare analizzano il testo dato in input e sviluppano una parola alla volta in modo del tutto causale, basandosi su ciò che è stato scritto in precedenza.

- *Encoder-Only* (Auto-encoding), progettati per la comprensione del linguaggio e la classificazione dei testi (come BERT). Si occupano di leggere l'intera frase in entrambe le direzioni per estrarne il significato.

- *Encoder-Decoder* (Sequence-to-Sequence), sono volte a mantenere la struttura originale a due moduli per compiti di trasformazione diretta da un testo di partenza a uno di arrivo, come la traduzione automatica o il riassunto (come T5 e BART).

2.3 Le capacità dei Large Language Models

Le capacità dei Large Language Model (LLM) derivano da un'architettura e da un metodo di addestramento abbastanza semplici, applicati su scala molto grande. L'architettura è quella del Transformer, che sostituisce la ricorrenza con il meccanismo di self-attention e consente di elaborare le sequenze parallelamente , rendendo praticabile l'addestramento su enormi quantità di testo *(Vaswani et al., 2017).* Quando il modello cresce in parametri, dati e potenza di calcolo, emergono competenze che nessuno aveva prima di allora programmato.

Un primo risultato è che le prestazioni migliorano in modo regolare con la scala: *Kaplan et al. (2020)* mostrano che la perdita segue leggi di potenza rispetto a numero di parametri, quantità di dati e calcolo impiegato. *Hoffmann et al. (2022),* con il modello *Chinchilla,* precisano che per un determinato budget di calcolo risulta più adeguato far crescere in modo proporzionale non solo i parametri, ma anche le dimensioni del modello e il numero di token di addestramento.

Una delle caratteristiche preponderanti è *l'in-context learning*: *Brown et al. (2020),* presentando GPT-3 con 175 miliardi di parametri, dimostrano che un modello può eseguire nuovi compiti a partire da poche istruzioni specifiche forniti nel prompt (few-shot), senza alcun aggiornamento dei pesi. Di conseguenza, la traduzione, la risposta a domande, e il completamento di testi diventano un'attività che il modello affronta senza addestramento .

Un'altra caratteristica principale concerne il ragionamento: *Wei et al. (2022a*) mettono in risalto il concetto che con il chain-of-thought prompting, cioè esempi che includono i passaggi intermedi del ragionamento e le prestazioni dei modelli, migliorano su criticità aritmetiche. *Kojima et al. (2022)* rilevano, invece, che un effetto simile si ottiene anche senza esempi (zero-shot), aggiungendo semplicemente al prompt un incentivo a ragionare passo dopo passo.

Wei et al. (2022b) parlano di abilità emergenti per indicare competenze assenti nei modelli piccoli e che si presentano in modo improvviso oltre una certa scala. A tal proposito, il progetto BIG-bench *(Srivastava et al., 2022)*, ha fornito una base molto vasta per questo tipo di analisi. L'interpretazione è però contestata: *Schaeffer et al. (2023)* sostengono che la discontinuità dipenda dalla scelta di metriche non lineari o discontinue e che, con le prime citate, la crescita delle prestazioni risulti più graduale.

Ouyang et al. (2022) descrivono InstructGPT, ottenuto con fine-tuning supervisionato e apprendimento per rinforzo dal feedback umano (RLHF): le risposte di un modello da 1,3 miliardi di parametri risultavano preferite dai valutatori rispetto a quelle del GPT-3 da 175 miliardi. Questo approccio ha reso gli LLM utilizzabili come assistenti utili alla conversazione quotidiana e, contemporaneamente, ha mostrato la capacità di questi modelli di eseguire le istruzioni in maniera del tutto efficiente.

2.3.1 Comprensione e generazione del linguaggio

Comprensione e generazione sono due proprietà che caratterizzano i modelli che stiamo affrontando : un LLM impara a modellare la struttura statistica del linguaggio, e da questa competenza derivano sia l'abilità di interpretare un testo sia quella di produrne uno.

Per comprensione si intende la capacità di generare rappresentazioni che mostrano significato, contesto e relazioni tra le parti del testo. Modelli come BERT apprendono figure contestuali bidirezionali, in cui lo stesso termine assume vettori diversi a seconda della frase in cui compare, questo consente di chiarire i sensi di una parola e rispondere a domande sul contenuto. Di base non esiste un modulo completamente incentrato sulla comprensione; difatti, studi di probing dimostrano che strutture linguistiche come le relazioni sintattiche emergono nelle rappresentazioni interne come effetto dell'addestramento autosupervisionato, senza essere state insegnate esplicitamente. Visto che la comprensione non è osservabile, viene misurata in modo indiretto, verificando se il modello risolve compiti che richiedono di aver capito. *Benchmark* come *GLUE* uniscono diversi compiti di inferenza e analisi linguistica , mentre *MMLU* valuta conoscenza e ragionamento su 57 discipline diverse. Questi test sono volti a misurare il comportamento, ma non lo stato interno del modello, e questo limite grava sull'interpretazione dei risultati.
La generazione avviene invece in modo autoregressivo: il testo viene prodotto un token alla volta, scegliendo di volta in volta la continuazione più consona data la sequenza precedente. Per prevedere bene la parola successiva il modello deve aver assimilato grammatica, stile e logica del discorso, ed è per questo che il risultato può essere coerente. La qualità dell'output dipende però anche dalla strategia di decodifica, una scelta che produce testo ripetitivo e generico, una in modo troppo casuale lo rende incoerente. Il campionamento nucleus è stato proposto proprio per bilanciare i due estremi.

Nei Foundation Models le due capacità non sono separabili, infatti, per sviluppare bene occorre comprendere il prompt, e la comprensione si presenta attraverso ciò che il modello produce. Con l'aumento di scala, i modelli hanno mostrato di poter eseguire nuovi compiti a partire da istruzioni ed esempi nel prompt, senza modificare i pesi: da qui l'idea di modelli adattabili a un'ampia gamma di applicazioni , con i vantaggi concernenti versatilità e irischi di omogeneizzazione che ne conseguono.

Resta aperta la questione se effettivamente tutti questo porti ad vera comprensione. Secondo una prima analisi, un modello addestrato solo sulla forma linguistica non ha accesso al significato che necessita di un legame con le intenzioni comunicative dei

parlanti. Una posizione opposta ritiene, invece, che il significato possa derivare dalle relazioni interne tra concetti, e quindi emergere anche senza un appoggio diretto al mondo . La letteratura di oggigiorno tende a riconoscere l'utilizzo del termine *"comprensione"* in modi diversi rispetto ai due schieramenti.

2.3.2 Estrazione e sintesi dell'informazione

L'estrazione e la sintesi dell'Informazione *(IE&S)* rappresentano due dei campi di maggiore impatto nell'evoluzione dei modelli di base *(Foundation Models, FM)* e dei modelli linguistici di grande taglia *(Large Language Models, LLM),* in quanto trasformano il modo in cui la conoscenza non strutturata viene organizzata ed elaborata *(Bommasani et al., 2021*l). L'estrazione dell'informazione, che comprende il riconoscimento di entità *(NER),* l'estrazione di relazioni e la costruzione di grafi di conoscenza ,viene oggi svolta dagli LLM in modalità *zero-shot o few-shot,* usufruendo della capacità dei modelli di comprendere contesti articolati senza addestramenti specifici per task *(Brown et al., 2020).* Per quel che concerne la sintesi dell'informazione*, (Vaswani et al., 2017)* permettono ai modelli di superare i limiti dei sistemi estrattivi tradizionali, sviluppando così riassunti coerenti e in grado di evidenziare i punti chiave da testi di grande lunghezza *(Zhang et al., 2020).*

2.3.3 Question answering

Un altro rilevante settore dove operano gli LLM è il *Question Answering* a dominio aperto (Open-Domain QA), dove questi utilizzano la conoscenza acquisita durante la fase di pre-addestramento per rispondere a domande senza la necessità di accedere a documenti esterni. Nei contesti applicativi in cui sono richieste accuratezza, aggiornamento dei dati o tracciabilità delle fonti, la memoria interna del modello presenta criticità inerenti alle *"allucinazioni*", le quali saranno trattate nel capitolo 5. Per baypassare questi limiti, lo stato dell'arte adotta architetture di *Retrieval-Augmented Generation (RAG)* applicate al QA *(Lewis et al., 2020*), cioè un sistema di recupero (retriever) che seleziona i passaggi rilevanti da un database documentale esterno e li fornisce al modello generativo (reader), fornendo risposte a fonti verificate. Infine, l'introduzione di tecniche di ragionamento guidato, come il *Chain-of-Thought (CoT) prompting,* consente agli LLM di affrontare compiti di Multi-hop QA, dove i la risposta finale richiede l'integrazione di molteplici passaggi logici e informativi *(Wei et al., 2022).*
Limiti dei modelli basati sul testo

2.3.4 Limiti dei modelli basati esclusivamente sul testo

Gli LLM vengono addestrati per preannunciare il token successivo su corpora testuali di enorme estensione, e contemporaneamente la scala di dati e parametri ha generato capacità sorprendenti, come l'apprendimento da pochi esempi forniti nel prompt . Questa impostazione ha però un limite strutturale: il modello apprende solo dalla forma linguistica, senza alcun contatto diretto con ciò a cui le parole si riferiscono. *Bender e Koller* infatti sostengono che un sistema addestrato solo sulla forma non possa percepire il significato, perché questo implica un' interconnessione tra le espressioni linguistiche e qualcosa di esterno al linguaggio. Il problema richiama quello del *symbol grounding* formulato da *Harnad* , secondo cui dei simboli definiti solo mediante altri simboli non ottengono un significato intrinseco.

Similmente, *Bisk et al.* posizionanoi modelli addestrati sul testo, osservando così che il linguaggio si sviluppa sulla percezione e sull'azione nel mondo fisico e sociale e che un'ulteriore crescita di scala non basterebbe a colmare questa distanza. A livello pratico non mancano sicuramente conseguenze: un LLM del tutto testuale non può nella fase di ingresso non può ricevere immagini o grafici, e quindi non può svolgere compiti in cui l'informazione risulta essere visiva o sonora. Anche il rapporto tecnico di GPT-4 riconosce che il modello, pur essendo molto più capace rispetto ai precedenti , resta inaffidabile e soggetto a produrre informazioni errate; e proprio quel modello traccia il passaggio a un sistema in grado di accettare immagini oltre al testo.

2.4 Verso i modelli multimodali

I modelli linguistici di grandi dimensioni (LLM) hanno mostrato elevate capacità di comprensione, generazione e ragionamento, ma operano esclusivamente sul testo e non possono percepire il mondo visivo o sonoro in cui l'informazione si manifesta. Questo limite ha portato la ricerca alo sviluppo di altre modalità , generando così modelli linguistici di grandi dimensioni multimodali (Multimodal Large Language Models, MLLM), di cui GPT-4V è l'esempio più noto.

Come osservano *Yin et al. (2024),* gli MLLM utilizzano un LLM potente come *"cervello"* per eseguire compiti multimodali, di base il modello linguistico non viene sostituito, ma diventa il fulcro di ragionamento dove si presentano componenti dedicate alla percezione. In termini di architetttura, un MLLM è formato da un codificatore per la modalità non testuale (ad esempio un codificatore visivo), da un LLM pre-addestrato e di un'interfaccia che collega i due, traducendo le rappresentazioni percettive in uno spazio che il modello linguistico sia in grado di interpretare. L'addestramento va a seguire anch'esso una logica proveniente dagli LLM: una fase di pre-addestramento, seguita da instruction tuning e da un allineamento alle preferenze umane. Il passaggio da LLM a MLLM è rilevante non solo per l'ampliamento e lo sviluppo dell'input, ma anche per le capacità che ne derivano, come la scrittura di storie a partire da immagini oppure il ragionamento matematico senza ricorrere all'OCR, rare nei metodi multimodali tradizionali. Gli autori ci mostrano un possibile percorso indirizzato verso l' intelligenza artificiale generale (AGI) e, questo sviluppo tende a portare anche nuove criticità. In particolare ritroviamo le allucinazioni multimodali, ossia risposte non coerenti con il contenuto visivo effettivamente presente, e nuove tecniche, come l'apprendimento multimodale in-context, che costituiscono il contesto in cui si colloca il presente lavoro.

2.4.1 Potenzialità dei Foundation Models multimodali

Il progresso dei Foundation Models non si limita in concetti tipici del testo e delle immagini, ma si estende a domini in cui l'informazione biologica è distribuita su larga scala, come quello delle scienze omiche. Per omica si intende l'insieme delle discipline che studiano in maniera globale le molecole presenti in una cellula, in un tessuto o in un organismo (genoma, trascrittoma, proteoma ecc…). Le tecnologie omiche, difatti, hanno generato dataset molecolari di importanti dimensioni e complessità, ricchi di informazione biologica ma al tempo stesso ad alta dimensionalità, caratteristica che spesso limita l'efficacia delle tecniche di apprendimento automatico. I FM, a tal proposito mostrano

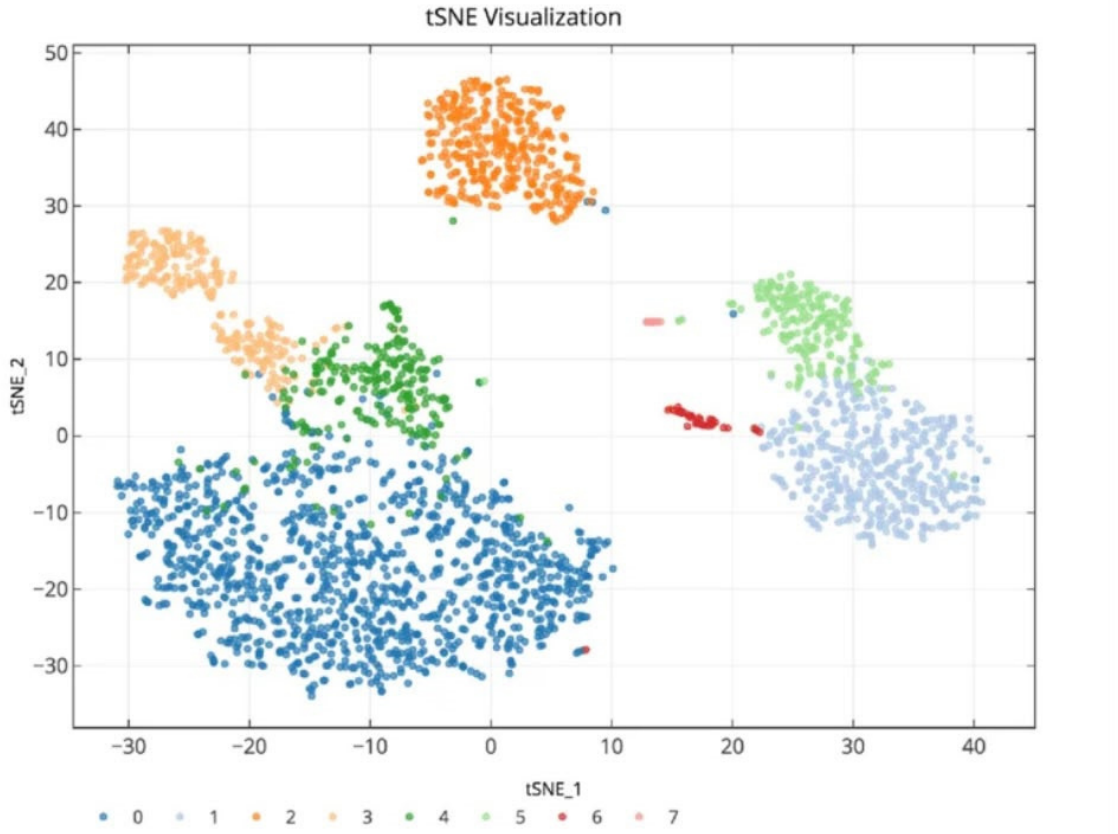
un’alternativa: assimilando rappresentazioni generalizzabili direttamente dai dati biologici, senza richiedere un'etichettatura manuale estesa dei dati, con una logica molto simile a quella che ha reso gli LLM e gli MLLM adattabili a compiti molto diversi tra loro.

È proprio in questa capacità di generalizzazione che si colloca gran parte del potenziale dei FM in ambito omico: una volta pre-addestrato su grandi quantità di dati biologici grezzi, un singolo modello può essere adattato a molteplici compiti, anziché richiedere lo sviluppo di un modello dedicato per ciascun problema. Liu et al. (2026) ne valutano l'utilità pratica in attività che mettono in risalto l'identificazione di biomarcatori in risposta alle numerose criticità .
Per favorire il confronto tra approcci diversi, *Liu et al.* Propongono un catalogo di dataset e framework, rappresentando un contributo che risponde a uno dei limiti più volte segnalato in questo campo, ossia la mancanza di benchmark standardizzati. Infine, gli autori propongono numerosi sviluppi futuri relativi alla generazione di FM, auspicando lo sviluppo di modelli capaci di andare oltre la semplice correlazione statistica per integrare ragionamento causale, dinamiche temporali e validazione sperimentale autonoma . Si tratta di un nuovo progetto volto a indicare come le potenzialità dei FM in ambito omico non si esauriscano nelle applicazioni già presenti , ma dimostrino la loro capacità di non limitarsi più a descrivere correlazioni nei dati osservati, bensì contribuiscono a formulare e verificare ipotesi biologiche. Questo percorso mostra come l'evoluzione dei FM non riguardi soltanto testo e immagini, ma si estenda fino al punto di coinvolgere diversi livelli di informazione biologica.

2.4.2 Integrazione di testo, immagini e dati biologici attraverso i Foundation Models

L'integrazione di modalità eterogenee , in particolare testo naturale, dati omici/strutturali e visualizzazioni biologiche ,rappresenta la frontiera più promettente nell'applicazione dei Large Language Models (LLM) e dei Foundation Models (FM) alle scienze della vita. In un contesto biomedico, i soli modelli linguistici testuali presentano molteplici limitazioni i quando devono interpretare la complessità tridimensionale e funzionale dei sistemi biologici. Per superare tale ostacolo, l'architettura dei moderni FM multimodali sfrutta spazi di rappresentazione latente condivisi *(latent embedd*ings): attraverso tecniche di allineamento cross-modale e apprendimento autosupervisionato *(self-supervised learning),* il modello riesce a collegare descrizioni testuali di articoli scientifici o referti clinici a sequenze genomiche, strutture proteiche 3D e immagini microscopiche o istopatologiche.

Come documentato negli archivi di *PubMed Central (NCBI),* i modelli di base si stanno evolvendo da sistemi unimodali a framework completamente integrati. Da un lato, l'elaborazione del linguaggio naturale crea un collegamento con la visione artificiale tramite architetture come BioCLIP o BiomedCLIP; dall'altro, l'integrazione dei dati omici e molecolari è guidata da *Protein Language Models (PLM*) come ProtT5 ed ESM-2, oltre che da modelli genomici *(Gene-LLMs come DNABERT e Nucleotide Transformer).* Questi ultimi in particolare interpretano il DNA, l'RNA e le sequenze aminoacidiche come *"linguaggi biologici"* .L’interazione che si genera tra queste modalità all'interno di un singolo spazio multimodale consente l’identificazione degli effetti inerenti alle varianti genetiche e l'interpretazione contestuale di bioimmagini, trasformando i grandi modelli di linguaggio in vere e proprie macchine utilizzate per la biologia dei sistemi e la medicina di precisione.



t-SNE grafico che visualizza le assegnazioni dei cluster delle cellule.

Analisi ed espolarazione sui dati trascrittomici di una singola cellula (Clarence K. Mah. Et al., 2018)

CAPITOLO III – FOUNDATION MODELS E LARGE LANGUAGE MODELS IN ONCOLOGIA

3.1 Il problema dell'eterogeneità dei dati oncologici

In oncologia lo stato del paziente è descritto da una vastità di modalità, che comprende radiologia, istologia, genomica e cartelle cliniche elettroniche *(Lipkova et al., 2022*). I modelli di intelligenza artificiale attuali agiscono maggiormente su una sola modalità, trascurando il contesto clinico più ampio, e questo ne riduce il potenziale. È l'eterogeneità dei dati a rendere questo limite così difficile da superare: le fonti disponibili includono biobanche, cartelle cliniche elettroniche, imaging medico, biosensori e sequenziamento genomico (*Acosta et al., 2022),* e si differenziano in ambito statistico, scala e struttura. Le sfide principali dell'integrazione sono la complessità di combinare tipi di dato diversi, l'allineamento tra modalità e la mancanza di intere modalità o di campioni *(Waqas et al., 2023).*

A queste criticità si aggiunge la scarsa presenza di dati utilizzabili: secondo *Boehm et al. (2022)* "la maggior parte dei dataset medici è ancora troppo sparsa per addestrare i moderni modelli di apprendimento automatico, e per colmare il divario servono ingegneria dei dati, metodi computazionali per dati eterogenei e modelli di dati sinergici". Tutto ciò si trasferisce in maniera del tutto concreta ; difatti uno studio sul carcinoma polmonare riporta che solo il 21% dei pazienti aveva i risultati dei test molecolari disponibili alla prima visita oncologica, e nel 19% dei casi la chemioterapia veniva avviata prima di ottenere la diagnosi molecolare.

I FM nascono come risposta a questa situazione, proprio perché *Bommasani et al. (2021)* li definiscono come "modelli addestrati su dati diversificati e adattabili a molti compiti." Pur basandosi su deep learning e transfer learning standard, la loro scala sviluppa capacità emergenti, e la loro efficacia favorisce l'omogeneizzazione. In ambito medico, *Moor et al. (2023)* mettono in risalto il concetto dell'IA medica generalista: modelli costruiti con apprendimento autosupervisionato su grandi dataset diversificati, in grado di interpretare combinazioni flessibili di modalità (imaging, cartelle cliniche, esami di laboratorio, genomica, testo) con pochi o nessun dato annotato per il compito specifico. Questo approccio riduce la dipendenza dall'annotazione manuale, ma non elimina del tutto il problema dei dati: il problema verterà sulla qualità e sulla validazione dei dati di addestramento, dove *Moor et al.* ne indicano la verifica come criticità principale.

3.1.1 Frammentazione delle informazioni

La riuscita integrazione dei Foundation Models (FM) e dei Large Language Models (LLM) nella pratica e nella ricerca oncologica deve dar conto alla marcata frammentazione e dispersione delle informazioni biomediche. In campo oncologico, il quadro clinico di ciascun paziente è distribuito all'interno di silos informativi disomogenei ;quali ad esempio referti anatomo-patologici, note cliniche non strutturate, immagini radiologiche, pannelli di profilazione genomica e registri di sperimentazioni. Questa diversità dei dati ostacola l'addestramento di modelli di intelligenza artificiale realmente multimodali, capaci di estrarre e sintetizzare in modo coerente la complessa storia naturale del tumore.

Parallelamente si individua una forte frammentazione metodologica nella letteratura scientifica di riferimento: le valutazioni sulle prestazioni, sulla sicurezza clinica e sulla

propensione dei modelli a generare "allucinazioni" informative derivano per lo più da studi condotti su singoli task o specifiche aree di patologia, privi di benchmark standardizzati ed eticamente condivisi . Il superamento di queste barriere e la creazione di ecosistemi di dati integrati (clinico-omici) costituiscono pertanto il requisito fondamentale per passare dall'uso sperimentale degli LLM a un loro impiego clinico sicuro e personalizzato .

3.1.2 Dati strutturati e non strutturati

Nel settore dell'informatica biomedica applicata all'oncologia , una delle sfide di maggiore rilievo per l'implementazione efficace dei Foundation Models (FM) e dei Large Language Models (LLM) risiede nella divisione tra dati strutturati e dati non strutturati. I dati strutturati, tipicamente rappresentati da codici diagnostici ICD, tabelle di laboratorio, dosaggi farmacologici e registri di stadiazione, costituiscono la parte più accessibile ed elaborabile dai classici algoritmi . Tuttavia, questa componente rappresenta una porzione limitata del patrimonio informativo relativo al paziente oncologico, caratterizzata da una rigidità che spesso tralascia le sfumature evolutive e contestuali della patologia.

Al contrario, si stima che oltre l'80% delle informazioni di interesse clinico e prognostico sia presente sotto forma di dati non strutturati : in ambito oncologico, questo importante volume informativo comprende i diari di decorso clinico, i verbali delle discussioni dei team multidisciplinari *(Tumor Board*), le descrizioni macro e microscopiche dei referti di anatomia patologica e le annotazioni sulle tossicità dei trattamenti. Nonostante questi testi in linguaggio naturale contengano dettagli inerenti alla risposta terapeutica e sulla qualità di vita del paziente, la loro natura eterogenea ha storicamente impedito un'estrazione automatizzata e sistematica delle conoscenze .

La generazione dei Foundation Models e degli LLM rappresenta un punto di svolta fondamentale: grazie alle capacità di comprensione del linguaggio naturale (*Natural Language Processing)* e al pre-addestramento su corpora di letteratura biomedica, questi modelli consentono di trasformare narrazioni cliniche complesse in rappresentazioni strutturate ed elaborabili.

La vera sfida metodologica della ricerca attuale non consiste dunque nel sostituire i dati strutturati con quelli narrativi, bensì nel realizzare un'integrazione armonica e multimodale; e solamente l'interazione tra la precisione quantitativa delle variabili strutturate e la corposità semantica dei testi narrativi permetterà di costruire profili di rischio dinamici, perfezionare la stratificazione prognostica e guidare con la massima accuratezza le decisioni terapeutiche personalizzate.

3.2 Cartelle cliniche elettroniche

Le Cartelle Cliniche Elettroniche (conosciute a livello internazionale come Electronic Health Records o EHR), come già accennato nel 1 capitolo, rappresentano la digitalizzazione sistematica della storia sanitaria del paziente, integrando all'interno di un'unica piattaforma informatica dati demografici.
Nell'ambito della ricerca oncologica, gli EHR costituiscono un patrimonio informativo di fondamentale importanza *(Real-World Data),* in grado di fornire una visione reale ed esaustiva dell'evoluzione della patologia tumorale; in oncologia, invece, le EHR comprendono i diari di decorso, i referti narrativi di anatomia patologica, le valutazioni radiologiche e i verbali dei team multidisciplinari *(Tumor Board)*. Come mostrato dalla letteratura recente, la vestità semantica di queste narrazioni testuali contiene dettagli

cruciali sulla stadiazione, sulla risposta ai trattamenti e sulle tossicità farmacologiche, che restano escluse dai i sistemi basati unicamente su dati codificati . L'accesso e la valorizzazione di questa enorme massa di informazioni non strutturate presenti all'interno degli EHR richiedono molto spesso l'utilizzo di tecniche avanzate di Natural Language Processing (NLP), capaci di estrarre e riportare ad un formato analizzabile la variabilità del linguaggio medico.

3.2.1 Referti e documentazione medica

La documentazione clinica e la refertazione specialistica rappresentano il nucleo narrativo da cui dipende l'intera caratterizzazione fenotipica, diagnostica e prognostica del paziente . I referti anatomo-patologici costituiscono, in questo ambito , il *"gold standard"* diagnostico: in essi vengono analizzate le formazioni biologiche di cruciale importanza, quali la tipizzazione istologica, il grado di differenziazione cellulare, la presenza di invasione vascolare, ematica o perineurale, lo stato dei margini di resezione chirurgica e l'espressione di marcatori immunoistochimici o alterazioni molecolari targetizzabili .

A questa fondamentale componente ritroviamo la refertazione radiologica (relativa ad indagini quali TC, PET-TC e Risonanza Magnetica), che fornisce una valutazione morfologica e funzionale dinamica della patologia tumorale. I referti di imaging tendono a descrivere la localizzazione della massa primaria, la quantificazione del carico metastatico, il coinvolgimento linfonodale e la risposta ai trattamenti sistemici sulla base di criteri semi-quantitativi espressi tuttavia in linguaggio naturale; e grazie a questi si mette il risalto la capacità di aggiungere una modesta quantità di testi clinici differenti che guidano il paziente lungo l'intero patient journey.

Nonostante l'importanza di questi dati per l'oncologia, la loro natura strutturale pone sfide particolarmente complesse: la documentazione clinica si caratterizza infatti per un'elevata frammentazione linguistica, determinata dalla presenza di abbreviazioni gergali specifiche di singole istituzioni, sintassi ellittiche ed espressioni discorsive fortemente dipendenti dallo stile individuale del medico.

Questa eterogeneità ha trasformato i referti clinici in veri e propri *"silos trasparenti",* dove tutte le informazioni ad altissimo valore scientifico restano immagazzinate all'interno di documenti isolati, rendendone quasi impossibile l' estrazione manuale su larga scala.

L'introduzione delle moderne tecniche di Natural Language Processing (NLP) e dei Large Language Models (LLM) consente ad oggi di superare i limits dei sistemi informatici, fornendo strumenti in grado di comprendere il contesto semantico del testo medico e di estrarre variabili complesse quali ad esempio lo stadio di malattia. La valorizzazione della documentazione clinica non strutturata costituisce quindi il passaggio fondamentale per trasformare il dato narrativo in informazione computabile, creando il presupposto per un'effettiva integrazione tra la dimensione clinico-patologica e i dati di profilazione genomica.

3.2.2 Estrazione automatica dell'informazione

Nel contesto dell'elaborazione del linguaggio naturale applicata alla ricerca oncologica, l'estrazione automatica dell'informazione *(Information Extraction, IE)* rappresenta l'insieme delle metodologie computazionali volte a trasformare il testo narrativo non strutturato in rappresentazioni di dati relazionali e strutturate. A differenza della semplice

classificazione di documenti, il processo di IE si divide in sottomoduli specialistici sequenziali e interconnessi, volti a estrapolare entità cliniche di interesse biomedico. Il primo livello operativo è rappresentato dal riconoscimento delle entità nominate *(Named Entity Recognition, NER*), il cui compito è individuare all'interno delle note cliniche o dei referti specifici concetti oncologici, quali tipi di neoplasie, biomarcatori molecolari (es. espressione di PD-L1, mutazioni di EGFR o KRAS), stadi di malattia (TNM), regimi farmacologici e dosaggi.

L'isolamento dei singoli termini, però, non basta per ricostruire il quadro clinico del paziente; è necessario applicare tecniche avanzate di estrazione delle relazioni *(Relation Extraction, RE*) per stabilire nessi causali, anatomici o temporali tra le entità individuate. L'algoritmo di RE consente, ad esempio, di associar in modo univoco una specifica mutazione genetica a una sede tumorale , oppure di collegare un evento avverso a un determinato agente chemioterapico somministrato. Una delle sfide critiche della IE in ambito oncologico risiede nella gestione della contestualizzazione semantica: il sistema deve possedere una spiccata capacità di identificazione delle negazioni *(negation detection*) e di risoluzione delle incertezze *(hedging/speculation),* per distinguere accuratamente un'ipotesi diagnostica o una lesione esclusa (es. "assenza di metastasi epatiche") da un dato di fatto confermato.Allo stesso modo , l'analisi temporale *(temporal information extraction)* risulta indispensabile per ordinare diacronicamente gli eventi, permettendo di ricostruire la progressione di malattia e la risposta ai trattamenti nel tempo.

L'evoluzione delle architetture di apprendimento automatico, dal passaggio dai sistemi basati su regole euristiche e modelli classici (come i Conditional Random Fields) ai modelli di tipo Transformer e ai Large Language Models (LLM) , ha migliorato l'accuratezza e la sensibilità di queste estrazioni. Nonostante ciò, la generalizzabilità di questi strumenti su diverse istituzioni resta condizionata dalla variabilità espressiva dei clinici e dalla presenza di acronimi ambigui, rendendo necessaria la continua validazione di questi modelli attraverso benchmark clinici standardizzati e la standardizzazione dei concetti estratti tramite ontologie biomediche (quali UMLS, SNOMED CT e NCI Thesaurus).

3.2.3 Integrazione dei dati clinici

La medicina pubblica, in Europa come negli Stati Uniti, si sta orientando verso l'integrazione dei dati del singolo paziente in un'unica storia clinica, consultabile tra le diverse aziende ospedaliere e sanitarie di uno stesso territorio regionale. Questo processo comporta la raccolta di una quantità crescente di informazioni appartenenti a esami di routine e da controlli periodici, e conferisce rilevanza clinica anche ad analisi ad ampio spettro, come quelle omiche (genomica, trascrittomica, epigenomica, metabolomica). La disponibilità di dati così diversi apre la strada a una medicina personalizzata e integrativa, nella quale il risultato di un qualsiasi esame può contribuire a interpretare sintomi osservati in un secondo momento. In questo modo si riduce il rischio che patologie complesse e multisettoriali restino non diagnosticate a causa dell mancanza di collegamento tra manifestazioni scorrelate.

La raccolta dei dati non è però sufficiente: se non proviene da un'analisi adeguata, rischia di mostrarsi come un accumulo fine a se stesso. Il trattamento di questi dati richiede infatti competenze specializzate, in particolare per quanto riguarda i dati di *Next Generation Sequencing (NGS),* che svolgono un ruolo preponderante sia nello studio delle malattie

complesse, attraverso l'identificazione di varianti genetiche germinali o somatiche, sia nella valutazione della salute e dello stile di vita, ad esempio mediante la metagenomica. Il percorso comprende l'allineamento delle librerie al genoma di riferimento, l'identificazione di mutazioni e polimorfismi, la valutazione dell'espressione genica e della metilazione, e l'integrazione dei dati omici tra diverse sorgenti e con database pubblici come TCGA, GEO e NCBI. A questo si aggiunge l'associazione genotipo-fenotipo, che vede l'utilizzo di tecniche statistiche, di machine learning e di metodi di inferenza causale, e la gestione di pipeline automatizzate e scalabili, applicabili a un numero piuttosto elevato di pazienti. Su questo fronte risultano fondamentali anche le competenze di gestione dei *Big Data* (ad esempio Cloudera, Hadoop, Elastic Net) e di intelligenza artificiale applicata all'ambito biomedico, in particolare reti neurali e deep learning.
Possiamo dire inoltre che la formazione di figure professionali in grado di muoversi tra bioinformatica e clinica consente di costruire una rete di collaborazione tra università e aziende ospedaliere, favorendo la condivisione di dati e informazioni utili a progetti di ricerca. Questo collegamento rappresenta un elemento chiave sia per lo sviluppo di nuovi metodi di analisi ,sia per l'integrazione di queste competenze nelle strutture sanitarie e nelle imprese regionali che operano nel settore biomedico.

3.3 Dati genomici

Il nucleo della medicina di precisione consiste nel migliorare il beneficio terapeutico per ciascun paziente, ricorrendo alla profilazione genomica e molecolare del tumore; questa attività viene definita, in modo analogo, anche *"patologia di precisione"*. Il cambiamento non riguarda soltanto gli strumenti diagnostici, ma anche il ruolo stesso della disciplina: la patologia si è infatti sviluppata da disciplina paraclinica a settore medico, parte integrante dei consulti oncologici multidisciplinari *(tumour board)*. Il suo contributo è considerato critico in tutto il percorso del paziente oncologico, cioè nella prevenzione, nella diagnosi, nella terapia e nel follow-up. Di conseguenza, anche la figura del patologo è chiamata a evolversi, per rendere la propria professione adeguata alle prospettive future dell'oncologia di precisione.

In questo ambito, i dati genomici assumono un'importanza particolare: essi derivano dallo sviluppo dei metodi molecolari, le cosiddette tecnologie *"omiche"*, che nell'ultimo decennio sono cresciuti molto velocemente e hanno offerto nuove prospettive alla ricerca sul cancro. I dati ottenuti analizzano le alterazioni molecolari, le quali evidenziano ciascun tumore e includono anche l'informazione morfologica fornita dall'esame istologico dei vetrini colorati con ematossilina ed eosina, che costituisce la base della diagnosi. Le due fonti di informazione sono quindi complementari: l'immagine istologica documenta l'aspetto morfologico del tessuto, mentre i dati omici ne descrivono le caratteristiche molecolari.

La diffusione delle tecnologie omiche e gli approcci di deep learning hanno rivelato un enorme potenziale nell'analisi delle immagini mediche, in particolare nel settore della patologia digitale. La combinazione dei dati delle immagini con quelli omici, può consentire sia la scoperta di nuovi biomarcatori del cancro sia una previsione più precisa della prognosi del paziente.

Su questo tema è stata condotta una revisione della letteratura, che ha esaminato diversi metodi di fusione multimodale tra l'analisi di immagini basata su reti neurali convoluzionali e i dati omici, concentrandosi in particolare sull'impatto di questa

combinazione sulle prestazioni di classificazione. Gli autori hanno consultato *PubMed,* selezionando gli articoli sottoposti a revisione e pubblicati in inglese tra gennaio 2015 e giugno 2021, e hanno creato un collegamento tra termini di ricerca relativi all'apprendimento profondo, alla patologia digitale, all'omica e alla fusione multimodale. Al termine della selezione, sono stati individuati 11 studi che si avvicinavano di più ai criteri di inclusione, ovvero lavori che utilizzavano reti neurali convoluzionali per l'analisi di immagini di pazienti oncologici colorate con ematossilina ed eosina, in combinazione con dati omici integrati.

Le pubblicazioni sono state poi classificate in base ai rispettivi endpoint; sette studi si sono concentrati sull'analisi della sopravvivenza, mentre i restanti quattro si sono concentrati sulla previsione di sottotipi di cancro e di malignità. Per quel che concerne i risultati, i classificatori basati sulle immagini mostrano già prestazioni elevate nella diagnostica prognostica e predittiva del cancro; difatti l 'integrazione dei dati omici ha portato ad un miglioramento ni in tutti gli studi descritti.

Gli stessi autori precisano che si tratta dei primi studi sull'argomento e che necessitano ancora di un controllo indispensabile per dimostrarne la generalizzabilità e la robustezza. Sono inoltre auspicabili ulteriori studi, più completi e basati su campioni di dimensioni maggiori, per valutare con maggiore precisione le prestazioni dei modelli e determinarne i benefici clinici.

3.3.1 Mutazioni e biomarcatori

Il sequenziamento del DNA ad alta capacità ha cambiato in modo radicale lo studio del cancro, perché permette di misurare con precisione le mutazioni somatiche che caratterizzano lo sviluppo del tumore *(Raphael et al., 2014)*. Il primo problema che i ricercatori si trovano ad affrontare è che ogni tumore ne contiene moltissime, ma non tutte hanno lo stesso peso; occorre infatti distinguere le mutazioni driver, responsabili dello sviluppo e della progressione della malattia, dalla grande quantità di mutazioni *passenger (Raphael et al., 2014).* Individuare la prima tipologia di mutazioni è una delle sfide principali della genomica oncologica, e un'ulteriore difficoltà è che per la maggior parte dei tumori non si conoscono ancora mutazioni ricorrenti per cui esista un farmaco specifico *(Raphael et al., 2014).*

Quando una mutazione risulta essere particolarmente rilevante può diventare un biomarcatore.: il *National Cancer Institute* statunitense definisce il biomarcatore come una molecola biologica presente nel sangue o nei tessuti, che individua un processo normale o anomalo oppure una condizione o una malattia *(Mishra e Verma, 2010*). Nella pratica oncologica i biomarcatori si differenziano in base alla funzione che svolgono: quelli diagnostici servono a identificare e confermare la presenza della malattia come ad esempio l'HPV, che è considerato un marcatore diagnostico dei tumori dell'utero e della cervice perché è presente in più del 90% delle lesioni *(Mishra e Verma, 2010).*

La distinzione più importante per la clinica è però quella tra biomarcatori prognostici e predittivi; in quanto i marcatori prognostici rappresentano misure effettuate prima del trattamento che mostrano l'esito ottenuto inerente al paziente *(Simon, 2013).* I marcatori predittivi, invece, indicano quali pazienti hanno maggiore o minore probabilità di trarre

beneficio da uno specifico trattamento *(Simon, 2013)*. Due esempi tratti da questo stesso contributo descrivono meglio la differenza: l'amplificazione di HER2 predice il beneficio del trastuzumab, mentre nei tumori del colon-retto la presenza di mutazioni di KRAS identifica i pazienti poco adatti alle terapie con anticorpi anti-EGFR *(Simon, 2013)*. Nel secondo caso il biomarcatore ha dunque la funzione di escludere una terapia e non di indicarla.

Un biomarcatore genomico di tipo diverso, che non riguarda un singolo gene ma l'insieme delle alterazioni, è il carico mutazionale tumorale (tumor mutational burden, TMB), cioè il numero totale di mutazioni presenti in un campione di tumore *(Chalmers et al., 2017)*. Un TMB elevato è considerato un indicatore emergente di sensibilità agli inibitori dei checkpoint immunitari, e in un'analisi su oltre 100.000 genomi tumorali è risultato associato alla risposta al blocco di PD-1 e PD-L1 più strettamente della sola espressione di queste proteine misurata con immunoistochimica *(Chalmers et al., 2017*). La stessa analisi mostra inoltre che una quota di pazienti presenta un TMB alto in quasi tutti i tipi di cancro, comprese molte forme rare (*Chalmers et al., 2017).*

I tumori con un sistema di riparazione del DNA per appaiamento errato *(mismatch repair)* accumulano migliaia di mutazioni, soprattutto nei microsatelliti, e vengono perciò definiti MSI (*Luchini et al., 2019).* Secondo le questioni riportate dalla *Società Europea di Oncologia Medica (ESMO)*, l'MSI, insieme al TMB e all'espressione di PD-1/PD-L1, ha un ruolo di biomarcatore predittivo per l'immunoterapia *(Luchini et al., 2019)*. Nel carcinoma del colon-retto la differenza è reale: studi clinici riportano un tasso di sopravvivenza fino al 78% nei pazienti MSI-high contro l'11% in quelli stabili, anche se la classificazione MSI effettivamente non va spiegare il meccanismo di risposta legato al carico di neoantigeni *(Fabrizio et al., 2018).*

Questi biomarcatori hanno un legame diretto con il tema della tesi: come visto nel paragrafo precedente, tra gli studi che integrano immagini istologiche e dati omici alcuni hanno come obiettivo proprio la previsione dell'instabilità dei microsatelliti, della malignità o dei sottotipi di cancro *(Schneider et al., 2022).* Se un modello di deep learning riesce a colpire un vetrino colorato con ematossilina ed eosina una caratteristica molecolare come l'MSI, il patologo può selezionare i pazienti da avviare ai test genomici, e la patologia digitale diventa un collegamento verso la medicina di precisione.

3.3.2 Caratterizzazione molecolare del tumore

Conoscere il profilo molecolare di un tumore è considerato fondamentale per proporre il trattamento più adatto al singolo paziente, perché consente di tener conto non solo della sua storia clinica e delle sue condizioni generali, ma anche delle caratteristiche biologiche specifiche della sua neoplasia *(AIMAC, 2022).* L'attuale classificazione di neoplasie affianca ai criteri istologici quelli molecolari, e ne deriva una diagnosi cosiddetta *"integrata",* più accurata nella stima della prognosi e della risposta alle terapie *(AIMAC, 2022).*

Quando si parla di profilo molecolare ci si riferisce alla presenza o all'assenza dei biomarcatori, che vengono individuati con tecniche di laboratorio avanzate in grado di rilevare le alterazioni del DNA estratto dalle cellule tumorali, responsabili dei cambiamenti della loro struttura e della loro funzione *(AIMAC, 2022).* Uno dei metodi più usati è il sequenziamento genico, cioè l'insieme di tecniche che consentono di analizzare parallelamente milioni di frammenti di DNA *(AIMAC, 2022).* Si tratta della stessa famiglia

di tecnologie ad alto rendimento descritta nei paragrafi precedenti, che qui trova una forte ricaduta nella pratica diagnostica.

L'esempio più chiaro risiede nei gliomi, in quanto per questi tumori il profilo molecolare è essenziale già per formulare la diagnosi *(AIMAC, 2022)*. Qui due marcatori hanno un ruolo particolare: il primo è la mutazione del gene IDH, dove la sua presenza o assenza divide i gliomi in due gruppi; e quello con la mutazione ha la prognosi migliore *(AIMAC, 2022).* Il secondo è la metilazione del gene MGMT, associata a una risposta migliore alla chemioterapia e a una prognosi più favorevole *(AIMAC, 2022).* Alla luce della distinzione vista nel paragrafo 3.3.1, IDH si comporta quindi come un marcatore prevalentemente prognostico, mentre la metilazione di MGMT ha anche un valore nel rilevare la risposta a un trattamento.

3.3.3 Integrazione dei dati genomici e molecolari

Nessun tipo di dato, preso da solo, descrive in modo completo un tumore, e per questo la ricerca oncologica si è spostata verso l'analisi integrata. Lo studio del *Cancer Genome Atlas,* che ha analizzato circa 10.000 tumori di 33 tipi diversi usando dati su aneuploidia, metilazione del DNA, espressione di mRNA e miRNA e proteine (*Hoadley et al., 2018*); ha constatato che quasi tutte queste piattaforme formano gruppi di tumori soprattutto in base all'istologia o al tessuto *(Hoadley et al., 2018).* Le caratteristiche molecolari, quindi, non sostituiscono l'informazione morfologica, ma vanno ad interagire direttamente con quest'ultima.

Lo stesso principio è entrato nella pratica clinica, difatti, nella quinta edizione della classificazione OMS dei tumori del sistema nervoso centrale *(2021*), molte caratteristiche molecolari sono diventate criteri diagnostici essenziali, e questi tumori necessitano di una diagnosi integrata "istomolecolare" *(Sahm et al., 2023).* Le linee guida europee di neuro-oncologia indicano per determinare i marcatori il sequenziamento di nuova generazione di DNA e RNA, il profilo di metilazione, l'immunoistochimica e, nel glioblastoma IDH-wildtype, l'analisi del promotore di MGMT, importante come marcatore predittivo *(Sahm et al., 2023).*

Le tecnologie ad alto rendimento producono informazioni molto numerose e servono strategie di integrazione adeguate, che comprendono anche metodi per cercare di ridurne il più possibile la dimensionalità *(Mol. Omics, 2022)*. Gli approcci principali si dividono in: non supervisionati, i quali ricavano informazioni senza etichette note, e supervisionati, che usano fenotipi già conosciuti per migliorare l'accuratezza *(Mol. Omics, 2022).*

3.4 Imaging radiologico

L'imaging radiologico comprende modalità come radiografia, TC, risonanza magnetica e PET; tutte queste tecniche sono sono preponderanti in pratica clinica per la diagnosi, la pianificazione del trattamento e il monitoraggio del paziente nel tempo *(Wu et al., 2025).* Le modalità si differenziano in base alla natura dei dati: TC e RM producono volumi tridimensionali, mentre radiografia ed ecografia generano immagini bidimensionali *(Ma et*

al., 2024). Similmente, l'imaging 3D offre un quadro più dettagliato dello stato del paziente rispetto al 2D *(Hamamci et al., 2026),* mentre la radiografia del torace resta l'esame di imaging più frequente nella pratica clinica *(Chen et al., 2024).* Questa varietà di modalità rappresenta un limite per gli approcci tradizionali; difatti, *Ma et al. (2024)* sostengono che i metodi esistenti sono spesso limitati all'uso su una singola modalità o patologia e si generalizzano poco ad altri compiti.

In questo settore i Foundation Models si posizionano come una risposta alquanto promettente: *Moor et al. (2023)* analizzano l'eventuale utilizzo di modelli in grado di interpretare le immagini diagnostiche insieme ad altri dati clinici e di produrre output come annotazioni sulle immagini. Nella segmentazione, *MedSAM (Ma et al., 2024)* è stato addestrato su 1.570.263 coppie immagine-maschera, con 10 modalità di imaging e oltre 30 tipi di tumore; in 86 compiti di validazione interna e 60 esterni ha mostrato un' accuratezza superiore ai modelli specialistici per singola modalità. Per la radiografia del torace viene preso in considerazione *RAD-DINO (Pérez-García et al., 2025),* ossia un codificatore addestrato solo su immagini, senza supervisione testuale, il quale ottiene prestazioni uguali o superiori ai modelli supervisionati dal linguaggio e migliora all'aumentare di quantità e varietà dei dati. *CheXagent (Chen et al., 2024)* unisce invece un codificatore di immagini a un modello linguistico: in una simulazione con otto radiologi, la bozza di referto prodotta ha fatto risparmiare agli specializzandi il 36% del tempo, senza perdita di qualità.

Lo sviluppo si estende pian piano a più modalità, ma soprattutto al 3D. *RadFM (Wu et al., 2025)* è stato costruito su 16 milioni di scansioni 2D e 3D che coprono oltre 5000 patologie; su cinque compiti di valutazione ha superato modelli multimodali come OpenFlamingo, Med-Flamingo, MedVInT e GPT-4V. CT-CLIP *(Hamamci et al., 2026)* ed è stato addestrato su 25.692 TC del torace di 21.304 pazienti.

Non mancano però numerose criticità: un confronto tra RAD-DINO, CheXagent e BiomedCLIP mostra che la strategia di pre-addestramento condiziona i risultati, perché i modelli senza supervisione testuale lavorano meglio nella segmentazione, quelli con testo nella classificazione, e BiomedCLIP risulta mutevole *(Li et al., 2025).* Sono state inoltre osservate differenze di prestazione per sesso, etnia ed età *(Chen et al., 2024);* infatti, molti dataset provengono da un solo centro, come CT-RATE *(Hamamci et al., 2026*), e la verifica dei risultati resta, secondo *Moor et al. (2023)*, il problema principale. Per questo tali strumenti vanno identificati come supporto al radiologo e non come sostituti.

3.4.1 Immagini istopatologiche

Le immagini istopatologiche costituiscono immagini tessutali osservate al microscopio, in genere su sezioni colorate con ematossilina-eosina (H&E). Da oltre un secolo, dopo i principi della patologia cellulare di *Rudolf Virchow*, la diagnosi definitiva di tumore si basa sull'esame microscopico di vetrini istologici. Con l'introduzione commerciale degli scanner per *whole-slide imaging (WSI)*, l'istopatologia è entrata in una dimensione digitale, dove il miglioramento di risoluzione e qualità degli scanner, la trasmissione dei dati, la capacità di archiviazione e i costi in calo ne hanno favorito la diffusione, rendendo l'istologia una risorsa ricca di dati per l'analisi computazionale *(Sali et al., 2026).* Una singola scansione ha dimensioni gigapixel e può essere composta da decine di migliaia di tessere *(tile)*, per questo molti modelli hanno dovuto analizzarne solo una parte, perdendo la base dell'intero vetrino *(Xu et al., 2024).*

I primi sistemi di intelligenza artificiale per la patologia digitale erano basati su reti neurali convoluzionali, progettate per compiti specifici come il riconoscimento del tumore, la gradazione e il conteggio cellulare; e la loro applicazione clinica era fortemente limitata dalla fragilità in condizioni reali. Le differenze tra scanner, protocolli di colorazione e flussi di laboratorio possono infatti ridurre le prestazioni, mentre il costo dell'annotazione grava specialmente sui tumori rari *(Sali et al., 2026).* Di conseguenza, solo tre prodotti di patologia con intelligenza artificiale hanno ricevuto un'autorizzazione dalla *FDA statunitense*: due servono a individuare le aree sospette nel carcinoma prostatico, il terzo stratifica il rischio prognostico sempre nella prostata *(Sali et al., 2026).*

I Foundation Models patologici, vengono pre-addestrati con apprendimento auto-supervisionato su milioni di WSI, e le rappresentazioni apprese sono poi adattate a compiti come rilevazione del tumore, sottotipizzazione, gradazione, predizione di biomarcatori e prognosi (*Sali et al., 2026).* Questi ultimi citati, li classificano in modelli immagine, visione-linguaggio e multimodali. Tra i primi, UNI è stato addestrato su oltre 100 milioni di immagini estratte da più di 100.000 WSI colorati con H&E, relative a 20 tipi di tessuto, e valutato su 34 compiti diagnostici *(Chen et al., 2024).* Prov-GigaPath, pre-addestrato su 1,3 miliardi di tessere da 171.189 vetrini, modella l'intero vetrino e ottiene i risultati migliori in 25 compiti su 26 *(Xu et al., 2024).* Virchow, addestrato su circa 1,5 milioni di vetrini, raggiunge un'AUC di 0,95 nel rilevare tumori comuni e rari, e con meno dati di addestramento eguaglia modelli clinici specifici per tessuto *(Vorontsov et al., 2024).* Tra i modelli visione-linguaggio, CONCH è stato addestrato su oltre 1,17 milioni di coppie immagine-didascalia ed è applicabile a classificazione, segmentazione, descrizione automatica e ricerca immagine-testo *(Lu et al., 2024).*

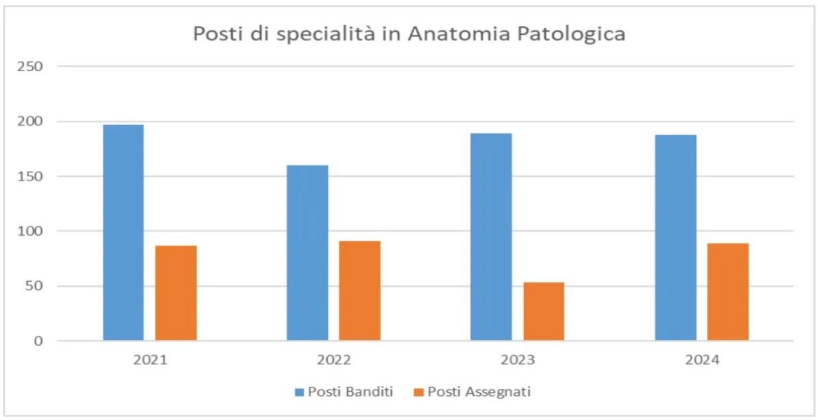
3.4.2 Integrazione di immagini e patologia digitale

Affinchè radiologia e anatomia patologica possano interagire sul piano computazionale, il vetrino deve diventare prima di tutto un'immagine. Nella patologia digitale uno scanner converte il preparato istologico in una whole slide image (WSI), la quale viene archiviata e consultata dal patologo a schermo al posto del microscopio. Gli scanner per WSI sono entrati in commercio all'inizio degli anni 2000, e il loro progresso di risoluzione, trasmissione dei dati e capacità di archiviazione, insieme alla riduzione dei costi, ha poi favorito la diffusione di questa pratica *(Sali et al., 2026). MarketsandMarkets* prevede che il mercato mondiale passi da circa 1,05 a 2,01 miliardi di dollari tra il 2024 e il 2029, con una crescita annua del 13,1% (*MarketsandMarkets, s.d.).* Altre società di analisi indicano valori molto più alti, da 7,8 a 13,7 miliardi (BCC Research, s.d.), perché dipendono da ciò che includono nel conteggio.



Il parere di uno specialista proveniente da altro ospedale può arrivare in tempi molto più brevi rispetto alla spedizione dei vetrini: un primario dell'ASL di Biella ricorda di aver atteso una settimana per un parere su un linfoma inviato a Torino *(Quotidiano Sanità, s.d..)*

Le immagini digitali possono essere ingrandite, misurate ed elaborate, e contemporaneamente preparano il terreno all'intelligenza artificiale. Questo supporto è importante perché nel concorso di specializzazione del 2025 una quota consistente dei contratti non è stata assegnata, e Anatomia Patologica ha mostrato percentuali analoghe (*Quotidiano Sanità, 2025)*



Sul piano normativo, in Italia le linee guida ministeriali del 2015 regolano la conservazione e l'archiviazione del materiale fisico di anatomia patologica *(Ministero della Salute, 2015)*; difatti, l 'archivio digitale viene sommato al costo del campione da conservare, e a livello europeo i dispositivi con intelligenza artificiale rientrano nei regolamenti MDR e IVDR e nell'AI Act *(Sali et al., 2026).* Intanto il *PNRR* sta accelerando la diffusione: la Regione Veneto ha assegnato una gara da 5,9 milioni di euro per un software di anatomia patologica che gestirà oltre 2,5 milioni di vetrini l'anno, con archivio delle immagini virtualizzate, visualizzatore dedicato e collegamento al Fascicolo Sanitario Elettronico *(Finanza Online, 2023). L'ARNAS Brotzu* ha completato la digitalizzazione del proprio laboratorio, collocando la Sardegna tra le prime regioni a farlo su scala aziendale *(L'Unione Sarda, s.d.).*

Il vetrino una volta digitalizzato, può essere affiancato alle immagini radiologiche, in particolare la radiologia descrive il tumore a livello macroscopico, la patologia a livello cellulare, e combinarle può migliorare diagnosi e prognosi, anche se gli studi ancora non realizzano sistemi completi di fusione *(Sali et al., 2026).* Un esempio è *RAPIDS,* che unisce la RM pre-trattamento e le biopsie colorate con H&E per predire la risposta alla chemioradioterapia nel tumore del retto, con prestazioni migliori dei modelli a singola modalità *(Feng et al., 2022).*

3.5 Approcci multimodali in oncologia

La pratica oncologica produce dati di natura molto diversa, con presenza di variabili cliniche strutturate (dati anagrafici, stadiazione, esami di laboratorio), testi non strutturati (referti di patologia e di radiologia, note dei medici), immagini (radiologiche e vetrini digitalizzati) e i profili molecolari. Per *Tripathi et al. (2025),* integrare queste fonti offre grandi opportunità: stratificare meglio i pazienti, predire la risposta alle terapie e modellare l'evoluzione della malattia. Eppure i dati restano frammentati, e manca un metodo per introdurli in rappresentazioni uniche del paziente *(Tripathi et al., 2025).* L'idea alla base, già formulata da *Boehm et al. (2022),* è che dati ottenuti con tecniche diverse presentino più contenuto informativo rispetto a ciascuna modalità presa da sola. Il progetto *The Cancer Genome Atlas* ha caratterizzato a livello molecolare migliaia di tumori *(Weinstein et al., 2013)*, e analogamente, archivi come The Cancer Imaging Archive raccolgono le immagini radiologiche corrispondenti *(Clark et al., 2013).*

3.5.1 Fusione delle diverse modalità

Le strategie di integrazione sono state classificate da *Waqas et al. (2024),* i quali passano in rassegna i principi dell'apprendimento multimodale e le architetture più usate, in particolare le reti neurali su grafo e i Transformer. Un esempio con la radiologia è lo studio di *Vanguri et al. (2022*): su 247 pazienti con carcinoma polmonare non a piccole cellule hanno combinato TC, vetrini digitalizzati di immunoistochimica e dati genomici, e il modello multimodale ha mostrato la risposta all'immunoterapia in maniera superiore rispetto ai singoli biomarcatori. L'approccio più recente usa i Foundation Models per trasformare ogni modalità in un vettore compatto, chiamato *embedding,* e poi unire i vettori: *HONeYBEE* ,infatti, genera gli embedding con modelli specifici per modalità, cioè un modello linguistico clinico per i testi, UNI per i vetrini, RadImageNet per le immagini radiologiche e SeNMo per i dati molecolari, per poi unitli tra loro con tre strategie come la concatenazione, la media e il prodotto di Kronecker *(Tripathi et al., 2025).*

Gli embedding clinici sono la modalità più forte, con il 98,5% di accuratezza nella classificazione del tumore, il 96,4% di precisione nel recupero di pazienti simili e indici di concordanza superiori a 0,80 per la sopravvivenza in 27 tipi su 33. Gli autori *Tripathi et al.,* attribuiscono il primato dei dati clinici alla cura con cui sono compilati in TCGA, e ritengono che la fusione conti di più nella pratica, dove la documentazione clinica è incompleta.

L'integrazione di dati provenienti da fonti eterogenee rappresenta una delle sfide centrali e più complesse nell'ambito dell'oncologia di precisione guidata dai Foundation Models (FM) e dal Deep Learning multimodale. Nell'analisi oncologica, la natura dei dati spazia da vettori tabulari di parametri clinici e fenotipici, a matrici ad alta dimensionalità tipiche dei dati genomici ed omici, fino ad arrivare a volumi e mappe spaziali complessi, quali le immagini di radiologia *(TAC, RMN, PET)* e gli esami istopatologici ad alta risoluzione *(Whole Slide Images, WSI)*.

La fusione delle modalità *(multimodal fusion)* mira a combinare tali flussi informativi eterogenei in una rappresentazione condivisa, superando i limiti dei singoli approcci unimodali. Nel contesto dei Foundation Models, le principali strategie di fusione vengono classificate in base alla fase dell'architettura in cui avviene l'integrazione dei dati: fusione precoce *(early fusion)*, fusione intermedia o latente *(intermediate/joint fusion)* e fusione tardiva *(late fusion*), a cui si aggiungono approcci ibridi e basati su meccanismi di attenzione.

- Fusione Precoce (Early Fusion / Data-Level Fusion)

La fusione precoce prevede la combinazione delle modalità direttamente allo stadio di input o immediatamente dopo una prima estrazione di feature di basso livello. In questa strategia, i dati grezzi o le caratteristiche grezze estratte (es. caratteristiche radiomiche unite a covariate cliniche o profili di espressione genica) vengono collegati in un unico vettore o matrice prima di essere passati all'architettura di apprendimento principale.

Questa stategìa, permette al modello di rilevare interazioni tra diverse modalità fin dai primi livelli di elaborazione, proponendo una struttura semplice ed economica dal punto di vista della progettazione di rete. Presenta una forte vulnerabilità per quel che concerne discrepanze di dimensionalità, di scala e alle differenti frequenze di campionamento tra le modalità (es. confronto tra una matrice genomica ed un'immagine 3D); inoltre, risulta poco robusta in presenza di dati mancanti (missing modalities).

- Fusione Intermedia o Latente (Intermediate / Joint / Feature-Level Fusion)

La fusione intermedia rappresenta la metodologia d'elezione nella maggior dei Foundation Models oncologici. Ciascuna modalità viene inizialmente elaborata da un encoder dedicato e specializzato (ad esempio, architetture Vision Transformer per immagini e istopatologia, e reti Fully Connected o Transformers per dati clinico-omici) per generare uno spazio di rappresentazione latente ad alta astrazione. Successivamente, queste rappresentazioni intermedie vengono fuse all'interno di uno spazio vettoriale tramite operazioni matematiche, quali l'operatore di concatenazione profonda, il tensor pooling bilineare o, più frequentemente, i meccanismi di cross-attention.

I meccanismi di cross-attention permettono a una modalità (es. l'immagine istopatologica) di guidare l'attenzione del modello verso elementi rilevanti di un'altra modalità (es. la mutazione genica o la nota clinica), rilevando complesse correlazioni inter-modali non lineari.

Richiede, inoltre, elevate risorse computazionali, grandi volumi di dati per l'addestramento e un tuning accurato dell'architettura per prevenire l'overfitting.

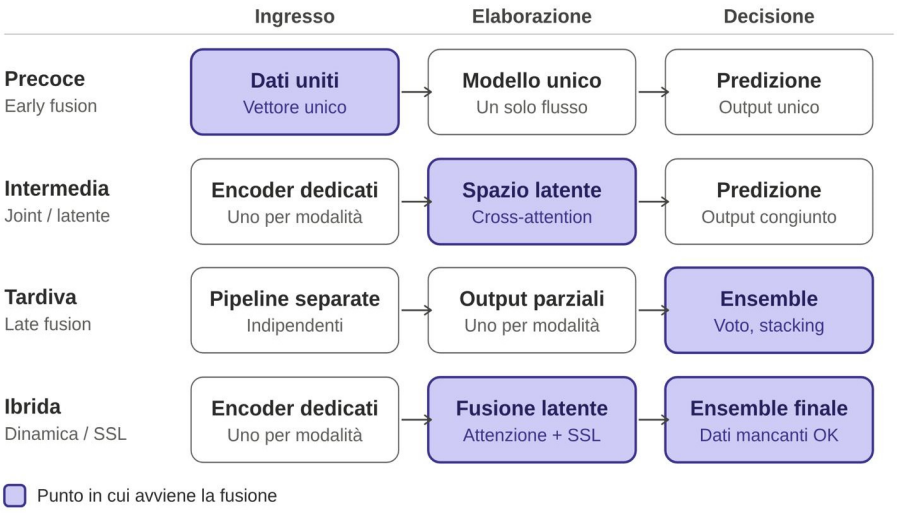
- Fusione Tardiva (Late / Decision-Level Fusion)

La fusione tardiva mantiene pipeline di elaborazione completamente separate e indipendenti per ciascuna modalità sensoriale. Ogni modalità produce un proprio output o una decisione predittiva autonoma (ad esempio, una probabilità di sopravvivenza o un punteggio di rischio). Soltanto nella fase finale, queste decisioni vengono collegate attraverso tecniche di ensemble, quali la media pesata, il voto di maggioranza (voting) o metamodelli di stacking.

Essa mantiene una forte modularità e tolleranza ai guasti: se una modalità risulta assente o sbagliata in un paziente, il sistema riesce comunque formulare una previsione basandosi sui restanti rami. Permette inoltre di riutilizzare modelli unimodali pre-addestrati senza dover riaddestrare la rete da capo. Non è in grado però di modellare le interazioni sottili o le sinergie complesse tra i dati prima della decisione finale, perdendo l'informazione contestuale incrociata tra le diverse fonti.

- Fusione Ibrida e Dinamica nei Foundation Models

Negli approcci più recenti applicati all'oncologia, si sta evolvendo l'uso di strategie ibride che combinano la fusione intermedia e la fusione tardiva per bilanciare la ricchezza espressiva dei modelli con la loro robustezza clinica. L'impiego di tecniche di *Self-Supervised Learning (SSL)* applicate a vettori di fusione dinamica consente ai Foundation Models oncologici di analizzare allineamenti spazio-temporali tra biopsie liquide, imaging e dati clinici, consentendo stime prognostiche accurate ed interpretabili anche di fronte a scenari clinici complessi e incompleti.



3.5.2 Vantaggi dell'analisi integrata

L'integrazione di dati multimodali permette di evidenziare caratteristiche causali e biologiche nascoste, che rimarrebbero invisibili se valutate singolarmente, offrendo una comprensione olistica della complessità della patologia. Un esempio di questo approccio è rappresentato dal quadro di sottotipizzazione a fusione multimodale *(MOFS, Multimodal Fusion Subtyping),* il quale crea interazioni tra dati radiologici, patologici, genomici, trascrittomici e proteomici per identificare sottotipi integrati di glioma adulto IDH-wildtype. Tale integrazione ha permesso di distinguere tre sottotipi fenotipicamente e clinicamente distinti:

1. MOFS1 (Proneurale): caratterizzato da una prognosi favorevole, un'elevata attività di sviluppo neurologico e da un'abbondante infiltrazione di neurociti.
2. MOFS2 (Proliferativo): associato alla prognosi deteriore, un'elevata attività proliferativa e marcata instabilità genomica. In questo contesto, la proteina STRAP è emersa come biomarcatore prognostico chiave e potenziale bersaglio terapeutico correlato al fenotipo proliferativo.
3. MOFS3 (Ricco di microambiente tumorale - TME): caratterizzato da una prognosi intermedia, con un'abbondante presenza di componenti stromali e immunitarie, e dimostrato essere particolarmente sensibile alla terapia immunologica anti-PD-1.

L'approccio multimodale non solo affina la classificazione biologica del tumore, ma si traduce in strumenti clinici diretti e non invasivi. Attraverso lo sviluppo di classificatori basati su reti neurali profonde (DNN, Deep Neural Network) allenati su caratteristiche radiologiche, è possibile prevedere in modo non invasivo il sottotipo multimodale del paziente, migliorando la traducibilità clinica e sviluppando decisioni terapeutiche personalizzate (come la selezione di candidati per l'immunoterapia o terapie target specifiche).

A livello metodologico, l'integrazione di dati eterogenei è stata fortemente accelerata dall'evoluzione delle architetture di Deep Learning: sebbene i primi modelli basati su reti neurali convoluzionali (CNN) mostrassero capacità limitate nel generalizzare le prestazioni su compiti o set di dati differenti, l'avvento dei modelli foundation basati sull'architettura Transformer ha rnotevolmente migliorato il settore della patologia digitale.
Questi modelli permettono: l'apprendimento di rappresentazioni altamente scalabili e robuste tra coorti diverse, l'inferenza efficace in modalità few-shot e persino zero-shot in svariate applicazioni oncologiche e l'elaborazione e la fusione coerente di dati istopatologici, radiologici, molecolari e clinico-testuali.

Si può costatare quindi che la fusione multimodale guidata da modelli di intelligenza artificiale avanzati permette in questo modo di superare la frammentazione delle informazioni mediche, fornendo diagnosi e prognosi più accurate e ponendo le basi per una gestione personalizzata e mirata del paziente oncologico.

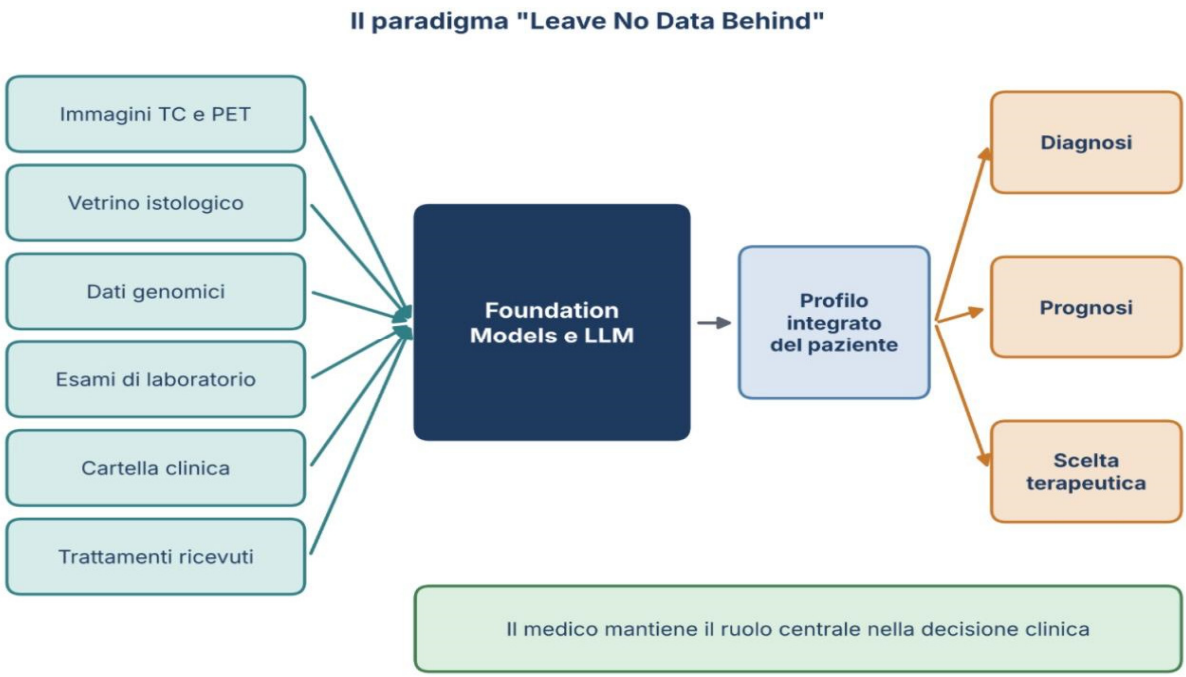
3.6 Il paradigma "Leave no Data Behind"

Il paradigma *"Leave No Data Behind",* letteralmente *"non lasciare indietro alcun dato"*, è stato proposto da *Corso et al. (2026)* in una rassegna dedicata all'impiego dei Foundation Models (FM) e dei Large Language Models (LLM) in oncologia. Il punto di partenza prevede una constatazione: i dati di un paziente oncologico sono molto numerosi, ma vengono spesso analizzati separatamente. Le immagini di TC e PET, il vetrino istologico, i dati genomici, gli esami di laboratorio, la cartella clinica e le informazioni sui trattamenti ricevuti costituiscono, infatti, mole di informazioni distinte. Secondo gli autori, l'obiettivo è quello di integrarli invece di considerarli indipendenti, in modo che nessun dato disponibile sul paziente resti inutilizzato *(Corso et al., 2026).*

I Foundation Models rappresentano un'evoluzione rispetto ai modelli di intelligenza artificiale tradizionali, perché non sono progettati per un unico compito, ma possono essere addestrati su quantità enormi e diverse di dati e poi adattati a molte applicazioni. In oncologia questo li rende molto interessanti, perché permette di pensare a modelli capaci di combinare testo, immagini, dati patologici e informazioni molecolari *(Corso et al., 2026),* insieme alle le strategie di fusione multimodale descritte in precedenza. Gli LLM, in particolare, possono analizzare il linguaggio presente in cartelle cliniche, referti e letteratura scientifica, ad esempio estraendo informazioni rilevanti dalla storia clinica, riassumendo cartelle complesse o aiutando il medico a recuperare informazioni utili. Gli autori ricordano tuttavia che questi sistemi possono produrre informazioni errate o inventate, le cosiddette *hallucinations,* e che non possono quindi essere considerati del tutto affidabili per prendere decisioni cliniche *(Corso et al., 2026).*

Un'ulteriore applicazione concerne le immagini radiologiche e la patologia digitale. Attraverso TC, RM, PET e immagini istologiche, l'intelligenza artificiale può imparare a riconoscere le caratteristiche del tumore e supportare diagnosi, segmentazione, stadiazione e valutazione della risposta al trattamento. Le immagini possono presentare informazioni biologiche non immediatamente visibili all'occhio umano, e l'IA potrebbe contribuire a identificare nuovi biomarcatori *(Corso et al., 2026).* A queste informazioni si aggiungono i dati genomici e multi-omici: il tumore non è formato solo dal suo aspetto al microscopio, ma anche da mutazioni, espressione genica e proteine; per cui integrarli con immagini e dati clinici potrebbe permettere di comprenderne meglio la biologia e di individuare il trattamento più adatto al singolo paziente *(Corso et al., 2026).*

Si raggiunge, quindi, un'oncologia sempre più personalizzata, in cui un sistema di IA considera contemporaneamente storia clinica, imaging, istologia, genomica e risposta alle terapie precedenti, per supportare il medico nella diagnosi, nella prognosi e nella scelta terapeutica *(Corso et al., 2026).* Gli autori sono però molto trasparenti sui limiti: le dimensioni di un modello o le buone prestazioni su un dataset non garantiscono che funzioni nella pratica clinica; in quanto i dati possono essere incompleti, diversi da un ospedale all'altro e affetti da bias, e molti risultati derivano da studi retrospettivi, per cui servono validazioni esterne e studi prospettici che dimostrino un reale beneficio per i pazienti. Restano inoltre aperte questioni di privacy, sicurezza, trasparenza e responsabilità: non basta che l'IA fornisca una risposta, occorre poterne valutare l'affidabilità, e il medico deve mantenere un ruolo centrale nella decisione *(Corso et al., 2026).*



3.6.1 Superamento della frammentazione dei dati

I dati di salute sono per natura multimodali, ma la maggior parte dei modelli sviluppati finora lavora su una sola modalità alla volta, ignorando in questo modo gran parte dell'informazione disponibile sul paziente *(Acosta et al., 2022).* In oncologia il problema è molto evidente: radiologia, istopatologia, genomica e dati clinici vengono generati da processi diversi, conservati in sistemi separati e analizzati con strumenti distinti, mentre la loro integrazione potrebbe offrire una descrizione più completa della malattia rispetto a ciascuna fonte presa singolarmente (*Boehm et al., 2022).*

A rendere difficile il superamento di questa frammentazione non è solo la separazione organizzativa dei dati, ma anche la loro eterogeneità tecnica: le diverse modalità si distinguono per struttura, dimensionalità e scala, e nella pratica clinica è frequente che per un paziente alcune di esse manchino del tutto. *Lipkova et al. (2022)* discutono questi ostacoli, insieme alla disponibilità limitata di dati annotati e alla necessità di rendere interpretabili i modelli, come le principali sfide dell'integrazione multimodale in oncologia. Sono gli stessi limiti già emersi nell'analisi delle strategie di fusione, dove ad esempio la fusione precoce risulta fragile rispetto a dati mancanti e a discrepanze di scala.

Un contributo ulteriore viene dai Large Language Models, che consentono di includere nell'analisi anche i dati testuali non strutturati, come cartelle cliniche, referti e letteratura scientifica, spesso esclusi dai modelli tradizionali. È su questa integrazione che si fonda il paradigma *"Leave No Data Behind"* descritto in precedenza, secondo cui nessun dato disponibile sul paziente dovrebbe restare inutilizzato *(Corso et al., 2026).* Gli stessi autori e la letteratura citata ricordano però che il superamento della frammentazione non è ancora dimostrato nella pratica clinica: dati incompleti, differenze tra ospedali e bias rendono necessarie validazioni esterne e studi prospettici (*Corso et al., 2026; Moor et al., 2023).*

CAPITOLO IV – FOUNDATION MODELS E MEDICINA PERSONALIZZATA

4.1 Dalla medicina tradizionale alla medicina di precisione

Per lungo tempo la pratica clinica si è basata sull'esperienza del medico, integrata dalle prove ottenute in studi condotti su gruppi di pazienti *(Sackett et al., 1996).* In oncologia questo significava classificare il tumore in base alla sede, all'aspetto istologico e allo stadio, e assegnare lo stesso trattamento a pazienti con la stessa diagnosi, aspettandosi che la risposta potesse variare molto da un individuo all'altro. A mettere in discussione questo schema sono stati i progressi della genomica, che hanno aperto la strada a trattamenti adattati al singolo paziente anziché identici per tutti *(Hamburg & Collins, 2010).*

Il passaggio si è concretizzato grazie ad alcuni risultati clinici decisivi: l'impiego di imatinib nella leucemia mieloide cronica ha mostrato che un farmaco diretto contro l'anomalia molecolare responsabile della malattia può produrre risposte efficaci *(Druker et al., 2001).* Nel tumore al polmone, la scoperta che la risposta a gefitinib era legata alla presenza di mutazioni del gene EGFR ha dimostrato che la scelta della terapia poteva dipendere dal profilo molecolare del tumore e non più solamente dal suo aspetto al microscopio *(Lynch et al., 2004; Paez et al., 2004).* Da quel momento il profilo molecolare è entrato pian piano nella pratica oncologica accanto ai criteri clinici e istologici tradizionali.

Il cambiamento ha avuto anche un riconoscimento istituzionale, con l'annuncio nel 2015 di un'iniziativa nazionale statunitense dedicata a questo ambito *(Collins & Varmus, 2015).* Nonostante ciò, bisogna però considerare, che la diffusione del sequenziamento ha prodotto una quantità di dati superiore rispetto a quella utilizzabile con i metodi tradizionali, e interpretare queste informazioni per portarle nella pratica clinica è diventata una sfida da superare (*Ashley, 2016).*

4.1.1 Definizione di medicina personalizzata e di precisione

Il termine medicina personalizzata è entrato nell'uso comune in seguito ai progressi della genomica, e indica l'impiego delle informazioni genetiche e molecolari di un individuo per orientare le decisioni di prevenzione, diagnosi e trattamento della malattia *(Hamburg & Collins, 2010).* L'idea alla base è che la risposta a un farmaco, o il rischio di sviluppare una patologia, dipendano dalle caratteristiche del singolo paziente e che queste possano essere misurate e utilizzate nella scelta clinica.

L'espressione medicina di precisione si è affermata più di recente e viene riferita a strategie di prevenzione e di trattamento che tengono conto della variabilità individuale, cioè delle differenze nei geni, nell'ambiente e nello stile di vita *(Collins & Varmus, 2015).* Rispetto alla definizione precedente, il concettosi sposta dal profilo genetico, a un insieme più vasto di informazioni sul paziente e sul suo contesto.

La scelta del termine *"precisione"* ha una motivazione precisa: Il *National Research Council (2011)*, nel documento dedicato a una nuova tassonomia delle malattie fondata sulla biologia molecolare, ha preferito questa espressione in quanto*"personalizzata"* può far pensare che si sviluppino trattamenti unici per ogni singola persona. Ciò che si intende, invece, è classificare i pazienti in sottogruppi che si differenziano per suscettibilità alla malattia, per le caratteristiche biologiche e la prognosi, o per la risposta ad uno specifico

trattamento. Nella pratica i due termini sono spesso usati come sinonimi, e la letteratura segnala che per la medicina di precisione manca una definizione condivisa: viene spesso confusa con concetti vicini come la medicina stratificata e la terapia mirata, e viene proposta una lettura più ampia che la considera un processo continuo di costruzione e applicazione di nuove classificazioni dei pazienti *(König et al., 2017).*

In questo elaborato si utilizzerà l'espressione oncologia di precisione, coerentemente con la letteratura più recente, intendendo con essa l'applicazione in ambito oncologico delle strategie sopra descritte.

4.1.2 Eterogeneità dei tumori

I tumori non rappresentano entità uniformi, e la loro eterogeneità si manifesta tra pazienti con la stessa diagnosi, ma anche all'interno dello stesso paziente, tra regioni diverse di una singola massa e tra tumore primario e metastasi, e cambia nel tempo con l'evoluzione della malattia e sotto la pressione dei trattamenti *(Dagogo-Jack & Shaw, 2018).* Questa variabilità è uno dei motivi per cui pazienti con la stessa diagnosi possono rispondere in modo molto diverso alla stessa terapia.

Uno dei lavori che ha reso evidente l'importanza di questo fenomeno è lo studio di *Gerlinger et al. (2012),* che ha evidenziato più regioni dello stesso tumore in carcinomi renali. Gran parte delle alterazioni genetiche non era presente in tutte le regioni analizzate, e il tumore mostrava un'evoluzione ramificata in sottopopolazioni cellulari distinte. La conseguenza pratica è che una singola biopsia non può rappresentare l'intero mondo genetico della malattia. *Burrell et al. (2013)* hanno discusso le cause di questa variabilità, legate all'instabilità genomica e ai processi evolutivi che agiscono sulle popolazioni di cellule tumorali, e le sue conseguenze sull'evoluzione del cancro.

L'eterogeneità pone dunque un limite alle strategie basate su una sola misura o su un solo campione, e rende necessario descrivere la malattia a più livelli: molecolare, tissutale e dell'intero organismo. Le diverse modalità di dati colgono aspetti fondamentali di questa complessità, ed è su questa considerazione che si fonda l'interesse per l'integrazione multimodale in oncologia *(Lipkova et al., 2022),* tema ripreso nelle sezioni precedenti a proposito dei Foundation Models.

4.1.3 Importanza delle caratteristiche individuali sul paziente

Se l'eterogeneità riguarda il tumore, un secondo livello di variabilità riguarda il paziente che lo ospita: due persone con la stessa diagnosi, lo stesso stadio e lo stesso trattamento possono avere esiti molto diversi, perché la risposta alla terapia dipende anche da caratteristiche individuali, genetiche e non.

Il miglior caso documentato è quello della farmacogenomica; dove le varianti genetiche ereditarie influenzano il modo in cui l'organismo metabolizza i farmaci e, di conseguenza, l'efficacia e la tossicità del trattamento. *Relling e Evans (2015)* descrivono come queste informazioni siano già entrate nella pratica clinica per portare alla scelta e al dosaggio di diversi farmaci. In oncologia un esempio riguarda le fluoropirimidine, per le quali le linee guida del *Clinical Pharmacogenetics Implementation Consortium* raccomandano di adattare la dose in base al genotipo del gene DPYD, in modo da ridurre il rischio di tossicità grave nei pazienti con ridotta attività dell'enzima (*Amstutz et al., 2018).*

Oltre a queste caratteristiche inerenti alla variabilità genica, ne emergono ulteriori anche nello studio dell'immunoterapia. Difatti, una revisione sistematica con meta-analisi degli studi sugli inibitori dei checkpoint immunitari ha confrontato l'efficacia del trattamento in uomini e donne, rilevando differenze tra i due sessi nell'entità del beneficio *(Conforti et al., 2018)*. Anche il microbiota intestinale è stato associato alla risposta all'immunoterapia anti-PD-1 nei tumori epiteliali *(Routy et al., 2018)*

4.2 Definizione di stratificazione

Nella medicina di precisione, per stratificazione si intende la suddivisione di una popolazione di pazienti in sottogruppi per caratteristiche biologiche, cliniche o di rischio, allo scopo di orientare le decisioni terapeutiche *(König et al., 2017; Hingorani et al., 2013). König et al. (2017)* osservano che le definizioni più diffuse di medicina di precisione ruotano proprio attorno alla stratificazione: questa deriva da dati su larga scala (clinici, di stile di vita, genetici, biomarcatori) e supera l'approccio basato su segni e sintomi. Gli autori sostengono però che la stratificazione non sia il punto d'arrivo, ma una componente di un processo più ampio, che comprende anche il suo contributo in un miglioramento dell'assistenza. *Hingorani et al. (2013*) definiscono la medicina stratificata, invece, come l'indirizzamento dei trattamenti in base alle caratteristiche biologiche o di rischio condivise da sottogruppi di pazienti, in contrapposizione alla medicina *"empirica"* o all comer. L'obiettivo è individuare chi trarrà il maggior beneficio, o il minor danno, da uno specifico intervento. La stessa fonte distingue due componenti: l'identificazione di differenze clinicamente rilevanti nel rischio assoluto di esito (stratificazione prognostica) e l'identificazione di fattori che orientano alla risposta individuale al trattamento (stratificazione predittiva).

I progressi nelle tecnologie di profilazione tumorale hanno ampliato l'accesso alla diagnostica molecolare, e le tecnologie omiche sono oggi utilizzate per approfondire la biologia del tumore e migliorare la stratificazione dei pazienti verso terapie mirate *(Teleanu et al., 2025)*: il programma MASTER *(Molecularly Aided Stratification for Tumor Eradication Research),* ne è un esempio. La necessità di stratificare nasce dall'eterogeneità tumorale, uno dei principali ostacoli al successo terapeutico. Essa si presenta all'interno dello stesso paziente, come eterogeneità intratumorale, intermetastatica e intrametastatica, e incide su diagnosi, previsione prognostica, scelta del trattamento e resistenza *("Tumor heterogeneity: a great barrier in the age of cancer immunotherapy").*

Nel contesto di questa tesi è utile precisare come la stratificazione venga operazionalizzata quando si usano i Foundation Models (FM)multimodali. Qusesti, estendono il concetto dal rischio alla risposta al trattamento. Il modello oFM, sviluppato su una coorte reale di 1,67 milioni di pazienti oncologici che integra traiettorie cliniche, DNA, RNA e patologia H&E, ha migliorato l'AUC prognostica per la sopravvivenza globale (0,774 contro 0,563 delle feature di riferimento); e ha inoltre migliorato la classificazione del beneficio terapeutico in 9 coorti comparative su 11 *(Vorontsov et al., 2026).* Le revisioni sul tema evidenziano che gli FM hanno elevati requisiti computazionali e pipeline di validazione complesse *("The rise of generalist foundation models and quantum computing in oncology").*

Per gli LLM, il quadro è meno consolidato; difatti le revisioni disponibili individuano come uso principale il supporto alle decisioni cliniche, accanto all'educazione di pazienti e professionisti e al supporto alla *ricerca ("A narrative review on the application of large language models to support cancer care and research").* La letteratura sull'impiego degli

LLM nelle decisioni oncologiche è eterogenea, con una revisione sistematica che include 56 studi *("Large language model integrations in cancer decision-making*"); di conseguenza il loro contributo alla stratificazione è in parte indiretto, ad esempio come encoder di testo clinico dentro pipeline multimodali *(Song et al., 2025).*

4.2.1 Identificazione di sottogruppi

L'identificazione di sottogruppi rappresenta il passaggio che precede la stratificazione. Prima di assegnare un paziente a una classe di rischio o di risposta, occorre stabilire quali gruppi esistano e su quali caratteristiche si distinguano. Nella classificazione molecolare dei tumori si contrappongono due approcci: la classificazione supervisionata sfrutta dati etichettati per scoprire interazioni genotipo-fenotipo o stratificare il rischio dei pazienti; mentre il clustering non supervisionato individua nuovi sottotipi a partire da dati molecolari privi di etichette. I sottotipi così ottenuti possono così differenziarsi e definiti soprattutto da caratteristiche patologiche *(Wang & Gu, 2016).* Alcuni sottotipi sono già entrati nella pratica: i recettori ER, HER2 e PR definiscono i sottotipi del carcinoma mammario, l'elevato tasso di mutazione definisce il sottotipo con instabilità dei microsatelliti e l'ipermetilazione delle isole CpG definisce il fenotipo CIMP *(Wang & Gu, 2016).* La necessità di individuare sottogruppi deriva dall'eterogeneità dei tumori; difatti, essi si sviluppano a causa di pressioni di natura trofica, metabolica, immunologica e terapeutica, e mostrano diversità genetica, fenotipica e comportamentale in tutte le componenti del microambiente. Ciò influisce sulla storia naturale della malattia e sulla sensibilità ai trattamenti (*Vitale et al., 2021).*

Per il clustering multi-omico, *Wang e Gu (2016)*distinguono tre metodi specifici: il clustering integrativo diretto, il clustering di cluster e il clustering integrativo regolatorio. Il loro scopo è mostrare il passaggio necessario all'analisi post-clustering, che comprende la caratterizzazione molecolare dei gruppi e la verifica del loro valore prognostico con l'analisi di sopravvivenza. L'esempio di riferimento su larga scala è l'analisi integrativa del *PanCancer Atlas. Hoadley et al. (2018)* hanno analizzato circa 10.000 tumori di 33 tipi con dati da più piattaforme e hanno identificato 28 sottotipi molecolari, la cui struttura è dominata dalla cellula di origine. Stimano inoltre che almeno un paziente su dieci verrebbe classificato e trattato in modo diverso rispetto alla classificazione istopatologica. Su scala più ristretta, *Suter et al. (2022)* hanno sperimntato un modello di miscela di reti bayesiane per il clustering multi-omico del carcinoma epatocellulare; e hanno identificato tre sottotipi principali, alcuni dei quali associati alla sopravvivenza anche dopo aggiustamento per stadio clinico. Questo è il criterio che un sottogruppo deve soddisfare per essere utile alla stratificazione: non basta che sia separabile statisticamente, deve portare informazione prognostica aggiuntiva rispetto alle variabili cliniche.

Gli FM sono addestrati con apprendimento auto-supervisionato su dataset prevalentemente privi di etichette *(Tarhini et al., 2025*). A livello di singola cellula, scGPT è stato preaddestrato su oltre 33 milioni di cellule e adattato a compiti come l'annotazione dei tipi cellulari e l'integrazione di dati multi-batch e multi-omici *(Cui et al., 2024).*

Per gli LLM, il contributo riguarda soprattutto il dato testuale: *CancerLLM* è stato preaddestrato su circa 2,7 milioni di note cliniche e oltre 515.000 referti di anatomia patologica relativi a 17 tipi di cancro, ed è stato sviluppato per estrarre fenotipi e generare diagnosi *(Li et al., 2024). Pelossof et al. (2025)* ad esempio, descrivono un LLM addestrato su note cronologiche costruite da dati strutturati di 580.000 pazienti con carcinoma mammario. Il clustering dei suoi embedding ha rivelato 10 gruppi nel carcinoma

mammario metastatico, con livelli di rischio prognostico diversi. Gli autori propongono di usare questi gruppi come base per cercare ulteriori stratificazioni legate alla risposta ai trattamenti. Le revisioni sistematiche segnalano inoltre che gli LLM in oncologia sono valutati soprattutto sulla capacità di rispondere a domande cliniche, con forte variabilità di risultati e assenza di protocolli di reporting standardizzati *(Carl et al., 2024).* Tra i limiti principali si indicano la conoscenza di dominio limitata e la dipendenza dalla supervisione umana *(Hao et al., 2025).*

4.2.2 Caratteristiche cliniche e molecolari

I sottogruppi si definiscono combinando due tipi di caratteristiche: caratteristiche cliniche e molecolari. Le più importanti sono lo stadio TNM, il grado di differenziazione istologica e il performance status ECOG, che indicano il carico di malattia e la condizione funzionale del paziente *(Li et al., 2026).* Nel carcinoma mammario, ad esempio, i marcatori prognostici sono caratteristiche del tumore, dello stato linfonodale, del grado e dell'invasione linfovascolare *(ESMO, s.d.).* In contrapposizione però, sono variabili statiche, che colgono poco l'evoluzione del tumore nel tempo *(Li et al., 2026).*

Nel caso delle caratteristiche molecolari, per l'immunoterapia, solo tre biomarcatori hanno ottenuto l'approvazione della FDA e sono ampiamente usati: l'espressione di PD-L1, il carico mutazionale (TMB) e l'instabilità dei microsatelliti o deficit di riparazione del mismatch (MSI/dMMR). Il TMB si basa sull'ipotesi che molte mutazioni somatiche producano più neoantigeni, riconoscibili dai linfociti T CD8+ *(Fumet et al., 2020).*
Gli LLM estraggono variabili cliniche dai referti: *Tripathi et al. (2025)* hanno ottenuto sul dataset TCGA un accordo con gli esperti del 98,5% per l'istologia e dell'87,6% per lo stadio.

4.3 Clinical Decision Support Systems

I sistemi di supporto alle decisioni cliniche *(clinical decision support systems, CDSS)* servono ad accompagnare i clinici nei processi decisionali complessi. Sono in uso dagli anni Ottanta e oggi vengono spesso erogati tramite le cartelle cliniche elettroniche e i flussi di lavoro informatizzati. Restano però incertezze sul loro effetto sui professionisti, sugli esiti per i pazienti e sui costi, e accanto a casi di successo esistono insuccessi che mostrano come i CDSS non siano privi di rischi *(Sutton et al., 2020).* Nel settore oncologico, l'AI è ormai diffusa in molti ambiti: diagnostica per immagini, classificazione patologica, scelta dei trattamenti e analisi prognostica. Nella diagnostica per immagini ad esempio,la collaborazione uomo-macchina ha raggiunto un livello di maturità consolidato. I CDSS basati su AI non hanno invece ancora realizzato la stessa interazione nella pratica clinica, anche perché gli aspetti etici legati all'uso di una tecnologia emergente nelle decisioni cliniche non sono ancora stati del tutto definiti *(Wang et al., 2023).* Tra i sistemi ad oggi pi1u utilizzati a livello globale emergono *Watson for Oncology (WFO)* e il sistema *CSCO AI della Chinese Society of Clinical Oncology.* Quest'ultimo, a differenza di WFO, è costruito sulla piattaforma CSCO e sul suo database, e organizza in mappe di conoscenza gli schemi terapeutici attraverso le sue sue linee guida *(Wang et al., 2023).*

Nello studio *di Somashekhar et al. (2018)* su 638 carcinomi mammari, WFO ha mostrato una concordanza del 93% con il tumor board; dove l'età e lo stadio influenzavano la concordanza, mentre lo stato recettoriale da solo no. Una meta-analisi di 9 studi su 2.463 pazienti ha stimato una concordanza dell'81,52%, considerando i livelli *"raccomandato"* e *"da considerare"*. Il valore più alto riguarda il carcinoma mammario e il più basso quello

gastrico, e la concordanza è maggiore negli stadi I-III che nello stadio IV, con l'eccezione del polmone *(Jie et al., 2021).* I risultati dipendono molto da come si definisce la concordanza: in uno studio su 65 pazienti con carcinoma gastrico avanzato si è passati dal 41,5% al livello *"raccomandato"* all'87,7% includendo anche *"da considerare" (Choi et al., 2019).*

Con i large language model (LLM) i CDSS si stanno evolvendo in modo sempre più graduale: Med-PaLM mostra prestazioni incoraggianti nelle domande di conoscenza clinica, ma resta inferiore ai clinici *(Singhal et al., 2023).* In oncologia di precisione, *Benary et al. (2023)* hanno sottoposto 10 pazienti fittizi con tumore avanzato a quattro LLM. Le opzioni terapeutiche proposte non hanno raggiunto la qualità e la credibilità degli esperti umani, sebbene offrano spunti utili e complementari alle procedure consolidate. Questo mostra che il ruolo più realistico degli LLM, per ora, sia di supporto alla selezione della letteratura più che di decisione autonoma *(Benary et al., 2023).*

4.3.1 Ruolo dell'AI nel percorso decisionale

Nel percorso decisionale oncologico l'AI interviene in più punti: nella diagnostica per immagini e patologica, nell'integrazione dei dati del paziente e nella selezione dei trattamenti, dove i large language model (LLM) sono oggi la novità principale. Nella pratica, però, l'integrazione nell'ambito di lavoro di natura clinica resta limitata da problemi di qualità dei dati, accuratezza dei modelli e mancanza di validazione con studi clinici. *Rajpurkar et al. (2022)* sostengono che gli studi prospettici e i progressi nell'analisi delle immagini hanno ridotto le criticità tra ricerca e applicazione. Indicano, oltretutto, come direzioni promettenti le fonti di dati non immagine e la collaborazione uomo-AI, accanto a sfide che comprendono scarsità di dati e bias razziali.

In uno studio prospettico su 1.000 casi (mammella, polmone e colon-retto), l'équipe multidisciplinare ha riesaminato la propria decisione dopo aver visto le raccomandazioni di *Watson for Oncology* e l'ha cambiata in 136 casi (13,6%). Nel 55% di questi, *Watson* aveva mostrato evidenze su trattamenti più nuovi, nel 30% un'alternativa più personalizzata e nel 15% nuove informazioni dai dati genotipici e fenotipici (*Somashekhar et al., 2018b)*. Il sistema di base tende ad integrare quindi la decisione umana senza sostituirla. Questo ruolo è però esposto al rischio di *automation bias*, la tendenza ad investire troppo sull'automazione: una revisione sistematica di 74 studi ha mostrato che dipende dallo stile cognitivo e dall'esperienza dell'utente, da fiducia e carico di lavoro, dalla complessità del compito e dalla pressione del tempo. Tra i fattori protettivi ci sono la formazione, la responsabilità esplicita dell'utente e la progettazione del sistema, ad esempio fornire informazioni invece che raccomandazioni e mostrare livelli di confidenza avanzati *(Goddard et al., 2012).*

Affinchè l'AI entri davvero nel percorso decisionale servono prove di efficacia e condizioni di adozione; difatti, *Kelly et al. (2019)* ricordano che gli esempi di AI implementata con successo sono ancora pochi. Indicano in particolare gli studi randomizzati come standard di riferimento per l'evidenza e raccomandano metriche che riflettano l'utilità clinica reale, oltre all'accuratezza tecnica. Individuano inoltre dei rischi di dataset shift, di fattori di confondimento appresi per errore, di bias discriminatori e di scarsa generalizzazione a nuove popolazioni. *Shortliffe e Sepúlveda (2018)* sottolineano, invece, che il supporto decisionale ha sia benefici sia barriere, e che complessità del sistema e comunicazione carente ne ostacolano un uso affidabile; per questo servono modelli di ricerca e di valutazione che lo portino in modo sicuro nel lavoro quotidiano.

CAPITOLO V – LIMITI DELL'INTELLIGENZA ARTIFICIALE IN ONCOLOGIA

5.1 Limiti e affidabilità dei Large Language Models

Nonostante le importanti potenzialità applicative dei modelli linguistici di grandi dimensioni (Large Language Models, LLM) nella pratica clinica e nella ricerca oncologica, l'adozione di questi sistemi incontra numerose criticità inerenti alla loro affidabilità, accuratezza e stabilità decisionale. In un campo complesso come l'oncologia, l'errore algoritmico o la distorsione informativa possono comportare conseguenze cliniche rilevanti per i pazienti.

Recenti studi di benchmarking pubblicati in letteratura e indicizzati su banche dati biomediche internazionali *(PubMed/NCBI)* evidenziano che, sebbene i LLM mostrino un'elevata accuratezza in specifici ambiti terapeutici, come le indicazioni per la terapia endocrina adiuvante e le terapie mirate, presentano margini di variabilità e performance disomogenee. In particolare, le ricerche dimostrano una tendenza dei modelli a sovrastimare la necessità di trattamenti specifici (come la radioterapia adiuvante) o a mostrare tassi di errore quando si tratta di definire protocolli complessi o test genomici. Valutazioni condotte sulla comparazione tra i suggerimenti dell'intelligenza artificiale e le raccomandazioni di esperti oncologi *(Molecular Tumor Boards)* indicano che la concordanza non è uniforme e risente fortemente dell'architettura del modello utilizzato *(es. differenze riscontrate tra modelli proprietari come GPT-4, Claude e modelli open-source come LLaMA).*

Un altro limite strutturale riguarda la gestione del contesto clinico e la prevenzione degli errori di accuratezza, comunemente noti come *"allucinazioni"* o risposte incoerenti rispetto alle linee guida aggiornate. Per alleviare queste criticità e preservare la sicurezza del paziente, la ricerca oncologica più recente evidenzia la necessità di superare i modelli aperti standard mediante l'impiego di architetture potenziate dal contesto, come i sistemi basati su Retrieval-Augmented Generation (RAG) e flussi di lavoro agentici (agentic workflows) in grado di ancorare le risposte a evidenze cliniche tracciabili.

Permangono sicuramente sfide metodologiche connesse alla trasparenza dei processi decisionali (black-box), alla protezione della privacy dei dati sanitari e alla necessità di validazioni prospettiche rigorose prima di una reale integrazione dei LLM nei percorsi decisionali multidisciplinari dell'oncologia di precisione.

5.1.1 Cosa sono le allucinazioni

Le allucinazioni rappresentano uno dei limiti strutturali più rilevanti dell'intelligenza artificiale generativa. Si tratta di istanze in cui un modello, come un grande modello linguistico (LLM) o un generatore di immagini, produce output apparentemente coerenti, autorevoli e plausibili, ma che in realtà risultano errati, irrilevanti o interamente inventati.

Questo fenomeno si verifica perché i sistemi di IA non possiedono una percezione intrinseca della verità o della falsità; la loro operatività si basa solo su un esercizio di pattern-matching statistico, volto a prevedere il token o l'elemento successivo più probabile in base ai dati di addestramento osservati. Di conseguenza, il modello ottimizza per la plausibilità sintattica e stilistica, non per la correttezza oggettiva. Le principali cause che scatenano le allucinazioni includono:

- dati di addestramento difettosi o insufficienti; come ad esempio errori, bias o lacune presenti nei dati vengono appresi dal modello come modelli regolari e accettabili.

- limiti intrinseci di progettazione; come l'incapacità di codificare direttamente aspetti complessi come la validità legale, scientifica o etica.

- overfitting (sovraadattamento) e correlazioni spurie: una sintonizzazione eccessiva sui dati di addestramento che porta il modello a considerare predittivi elementi irrilevanti.

- attacchi avversari e prompt injection: input manipolati o trappole logiche progettate per aggirare i guardrail di sicurezza e forzare il modello a generare risposte fuorvianti.

Trattandosi di una caratteristica intrinseca al funzionamento probabilistico dell'IA generativa, le allucinazioni possono essere pian piano ridotte, ma non completamente eliminate, evidenziando una sfida critica specialmente nei contesti applicativi ad alto rischio come quelli legale, medico e finanziario.

5.1.2 Privacy, sicurezza e aspetti etici

L'utilizzo dei Large Language Models (LLM) e dei Foundation Models (FM) ha sviluppato in modo graduale il panorama tecnologico e industriale, generando questioni di natura giuridica, etica e di sicurezza informatica. Poiché queste architetture vengono pre-addestrate su corpus di dati di dimensioni imponenti mediante tecniche di web scraping, esse includono informazioni riservate e dati personali *(Personally Identifiable Information, PII).* A questo proposito, la letteratura scientifica ha dimostrato come i modelli linguistici soffrano del fenomeno della memorizzazione, arrivando a *"rigurgitare" (regurgitation)* porzioni esplicite dei dati di addestramento quando sollecitati da prompt specifici o da attacchi mirati (Training Data Extraction Attacks) . Inoltre, l'uso continuo delle interazioni degli utenti per il raffinamento dei modelli (fine-tuning) e per l'integrazione contestuale espone il sistema a vulnerabilità di tipo *Membership Inference*, mettendo a dura prova la conformità alle normative sulla protezione dei dati personali, quali il GDPR europeo e il recente AI Act.

Tra le minacce più critiche figurano le tecniche di *Prompt Injection* e *Jailbreaking,* attraverso le quali un attaccante può manipolare l'input fornito al modello per eludere le barriere di sicurezza *(guardrails),* forzando l'esecuzione di istruzioni non autorizzate o l'esfiltrato di informazioni sensibili. A ciò si aggiungono i rischi legati all'avvelenamento dei dati (*Data Poisoning)* e alle *Backdoor Attacks* sia nella fase di pre-addestramento sia nei meccanismi di Retrieval-Augmented Generation (RAG), i quali permettono di alterare il comportamento logico dell'agente generativo o di facilitare attacchi di estrazione del modello *(Model Inversion).*

5.1.3 Responsabilità clinica

L'impiego dei LLM nel settore clinico sta vivendo una forte accelerazione, offrendo numerose opportunità che vanno dall'assistenza al triage alla stesura di documentazione clinica. Ciononostante, il dibattito attorno alla loro attendibilità viene penalizzato da criteri di misurazione obsoleti e decontestualizzati.

Una ricerca apparsa su *npj Digital Medicine* introduce un paradigma valutativo, fondato sulla netta distinzione tra sicurezza ed efficacia dell'operato medico. La riflessione dei ricercatori concerne un presupposto chiaro: la capacità di un sistema di generare risposte coerenti ed esaustive non garantisce l'assenza di rischi, soprattutto quando il modello trascura fattori di rischio, elementi critici o scenari clinici articolati.

Allo stato attuale, la maggior parte dei test di verifica per gli LLM in ambito sanitario sfrutta quiz a risposta multipla; tali prove valutano l'attitudine del modello nell'individuare l'opzione corretta, ma falliscono nel replicare la complessità del processo decisionale medico, che richiede un'analisi aperta, adattabile alle circostanze e spesso esposta a forti margini di incertezza. Ragion per cui, queste metriche finiscono per esaltare la proprietà di linguaggio e l'erudizione teorica, mettendo in ombra aspetti cruciali quali:

- l'identificazione tempestiva di condizioni ad alto rischio;
- il controllo delle incompatibilità terapeutiche e delle reazioni avverse ai farmaci;
- l'aderenza ai protocolli clinici nelle situazioni più intricate;
- la prevenzione di scelte terapeutiche dannose nei soggetti maggiormente vulnerabili.

Per un'effettiva e sicura integrazione dell'intelligenza artificiale in medicina, diventa imperativo sviluppare standard di verifica basati su scenari reali, in grado di testare la prudenza clinica tanto quanto la competenza nozionistica. Solo così sarà possibile trasformare i LLM in alleati davvero affidabili per i professionisti della salute e sicuri per i pazienti.

CONCLUSIONI

Questo lavoro ha analizzato il ruolo dei Foundation Models e dei Large Language Models in oncologia, con particolare attenzione all'integrazione di dati eterogenei e alle prospettive per la medicina di precisione.

In sintesi, il cancro è un insieme eterogeneo di malattie, descritto da dati di natura diversa: clinici, genomici, di imaging e di patologia digitale. I modelli tradizionali di Machine Learning e Deep Learning hanno dato buoni risultati in diagnosi, prognosi e predizione della risposta ai trattamenti; ma vengono progettati per singoli compiti, generalizzano poco e faticano a integrare fonti diverse. I Foundation Models, sono addestrati su grandi quantità di dati e adattabili a compiti differenti, mentre i Large Language Models sono basati sull'architettura Transformer, nascono come risposta a questi limiti.

Di rilevante importanza è il paradigma "Leave No Data Behind", il quale indica l'integrazione come chiave per descrivere il paziente in modo completo: combinare clinica, imaging, istologia e genomica permette di stratificare i pazienti in sottogruppi di rischio e di risposta, andando oltre la sola classificazione istologica. Nei sistemi di supporto alle decisioni cliniche questi strumenti affiancano il medico senza sostituirlo.

Il passaggio dalla ricerca alla pratica clinica richiederà alcuni passi; come validazioni esterne e studi prospettici che dimostrino un beneficio reale per i pazienti e benchmark standardizzati, anche per valutare gli LLM in scenari clinici realistici e non solo con quiz a risposta multipla. Inoltre, servono un quadro normativo adeguato (MDR, IVDR, AI Act) e la formazione dei professionisti, per un uso consapevole ed evitare l'eccessiva fiducia nell'automazione. Dal punto di vista scientifico, la direzione indicata dalla letteratura è quella di modelli capaci di andare oltre la correlazione statistica, integrando ragionamento causale e dinamiche temporali *(Liu et al., 2026).*

Restano di conseguenza criticità importanti, che riguardano sia gli aspetti tecnici sia quelli etici e clinici. Le allucinazioni, cioè risposte coerenti e plausibili ma errate o inventate, derivano dal funzionamento probabilistico dei Large Language Models: il modello ottimizza la plausibilità del testo e non la sua correttezza. Queste, possono essere ridotte, ad esempio con architetture di Retrieval-Augmented Generation che ancorano le risposte a fonti verificabili, ma non eliminate. Le valutazioni disponibili mostrano inoltre prestazioni disomogenee: le raccomandazioni terapeutiche degli LLM non sono ancora paragonabili a quelle degli esperti *(Benary et al., 2023)* e la concordanza con i tumor board varia in base al modello e al tipo di tumore.

Non mancano sicuramente limiti inerenti alla privacy e alla sicurezza dei dati, dove i modelli possono memorizzare e restituire informazioni presenti nei dati di addestramento e sono esposti ad attacchi come la prompt injection, un problema serio quando si trattano dati sanitari e genomici, tutelati dal GDPR e dall'AI Act. A ciò si aggiunge la scarsa trasparenza dei modelli, spesso *"black box"*, e i bias legati a dati incompleti, diversi da un ospedale all'altro o provenienti da un solo centro. Resta infine aperta la questione della responsabilità clinica: non basta che il sistema fornisca una risposta, occorre poterne valutare l'affidabilità, e la decisione finale deve restare al medico. Anche per questo servono validazioni esterne, studi prospettici e metodi di valutazione basati su scenari clinici reali, che oggi spesso mancano.

In conclusione, dal lavoro emerge che il valore dell'intelligenza artificiale in oncologia non sta nella potenza dei singoli modelli, ma nella qualità dei dati su cui si basano e nella capacità di integrarli. Un modello molto grande non garantisce risultati affidabili se i dati sono incompleti, distorti o provenienti da una sola istituzione. Per questo la sfida non è solo tecnologica, ma anche organizzativa ed etica, e richiede la collaborazione tra bioinformatici, biotecnologi, medici e patologi. Gli LLM e i Foundation Models vanno considerati strumenti di supporto e non sostituti del medico, in quanto possono rendere più rapida l'analisi di grandi quantità di informazioni, ma la responsabilità della decisione clinica resta umana. Usati con spirito critico, in modo trasparente e con validazioni rigorose, possono contribuire a rendere l'oncologia più precisa e più vicina al singolo paziente.

BIBLIOGRAFIA

1. National Cancer Institute (s.d.). Sito ufficiale del National Cancer Institute (NCI). https://www.cancer.gov/

2. Zhao, L., Lee, V. H. F., Ng, M. K., Yan, H., & Bijlsma, M. F. (2019). Molecular subtyping of cancer: current status and moving toward clinical applications. Briefings in Bioinformatics, 20(2), 572–584. https://doi.org/10.1093/bib/bby026

3. Russell, S. J., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson. https://www.pearson.com/en-us/subject-catalog/p/artificial-intelligence-a-modern-approach/P200000003500/9780134610993

4. Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., Hirschberg, J., Kalyanakrishnan, S., Kamar, E., Kraus, S., Leyton-Brown, K., Parkes, D., Press, W., Saxenian, A., Shah, J., Tambe, M., & Teller, A. (2016). Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence. Stanford University. https://ai100.stanford.edu/2016-report (Section I: What is Artificial Intelligence?, https://ai100.stanford.edu/2016-report/section-i-what-artificial-intelligence; arXiv: https://doi.org/10.48550/arXiv.2211.06318)

5. OECD (2019). Artificial Intelligence in Society. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/11/hello-world_7734f030/726fd39d-en.pdf

6. Mitchell, T. M. (1997). Machine Learning. McGraw-Hill. https://www.cs.cmu.edu/afs/cs.cmu.edu/user/mitchell/ftp/mlbook.html

7. Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer. https://doi.org/10.1007/978-0-387-31073-2

8. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines, 28, 689–707. https://doi.org/10.1007/s11023-018-9482-5

9. Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

10. Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

11. Szeliski, R. (2022). Computer Vision: Algorithms and Applications (2nd ed.). Springer. https://link.springer.com/book/10.1007/978-3-030-34372-9

12. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105. https://doi.org/10.1145/3065386

13. Russakovsky, O., Deng, J., Su, H., et al. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115, 211–252. https://arxiv.org/abs/1409.0575 (sito ufficiale: https://image-net.org/challenges/LSVRC/)

14. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

15. Dosovitskiy, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/2010.11929

16. Siciliano, B., & Khatib, O. (Eds.). (2016). Springer Handbook of Robotics (2nd ed.). Springer. https://www.springer.com/9783319325507

17. Thrun, S., Burgard, W., & Fox, D. (2005). Probabilistic Robotics. MIT Press. https://mitpress.mit.edu/9780262201629/probabilistic-robotics/

18. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press. http://incompleteideas.net/book/the-book-2nd.html

19. Mnih, V., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518, 529–533. https://doi.org/10.1038/nature14236

20. Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529, 484–489. https://doi.org/10.1038/nature16961

21. Yudkowsky, E. (2022). AGI Ruin: A List of Lethalities. Machine Intelligence Research Institute (MIRI). https://intelligence.org/2022/06/10/agi-ruin/

22. Eraslan, G., Avsec, Ž., Gagneur, J., & Theis, F. J. (2019). Deep learning: new computational modelling techniques for genomics. Nature Reviews Genetics, 20, 389–403. https://doi.org/10.1038/s41576-019-0122-6

23. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

24. Libbrecht, M. W., & Noble, W. S. (2015). Machine learning applications in genetics and genomics. Nature Reviews Genetics, 16, 321–332. https://doi.org/10.1038/nrg3920

25. Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. Journal of Machine Learning Research, 17. https://www.jmlr.org/papers/v17/15-522.html

26. Brohan, A., et al. (2023). RT-2: Vision-language-action models transfer web knowledge to robotic control. arXiv:2307.15818. https://arxiv.org/abs/2307.15818

27. Handelman, G. S., Kok, H. K., Chandra, R. V., Razavi, A. H., Lee, M. J., & Asadi, H. (2018). eDoctor: machine learning and the future of medicine. Journal of Internal Medicine, 284(6), 603–619. https://doi.org/10.1111/joim.12822

28. Swanson, K., Wu, E., Zhang, A., Alizadeh, A. A., & Zou, J. (2023). From patterns to patients: advances in clinical machine learning for cancer diagnosis, prognosis, and treatment. Cell, 186(8), 1772–1791. https://doi.org/10.1016/j.cell.2023.01.035

29. Tran, K. A., Kondrashova, O., Bradley, A., Williams, E. D., Pearson, J. V., & Waddell, N. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. Genome Medicine, 13, 152. https://doi.org/10.1186/s13073-021-00968-x

30. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.). Springer. https://doi.org/10.1007/978-0-387-84858-7

31. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://www.deeplearningbook.org/

32. Murphy, K. P. (2022). Probabilistic Machine Learning: An Introduction. MIT Press. https://probml.github.io/pml-book/book1.html

33. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998–6008. https://arxiv.org/abs/1706.03762

34. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

35. Géron, A. (2022). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media. https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967/

36. Lotter, W., et al. (2024). Artificial intelligence in oncology: current landscape, challenges, and future directions. Cancer Discovery. https://doi.org/10.1158/2159-8290.CD-23-1199

37. Lipková, J., Chen, R. J., Chen, B., et al. (2022). Artificial intelligence for multimodal data integration in oncology. Cancer Cell, 40(10), 1095–1110. https://doi.org/10.1016/j.ccell.2022.09.012

38. Lin, R., Zhao, Z., Liu, Z., Kang, J., Zhang, K., Huang, X., & Yu, Y. (2026). Artificial intelligence in clinical oncology: multimodal integration and translational development. Cancer Letters, 649, 218493. https://doi.org/10.1016/j.canlet.2026.218493

39. Elemento, O., Leslie, C., Lundin, J., & Tourassi, G. (2021). Artificial intelligence in cancer research, diagnosis and therapy. Nature Reviews Cancer, 21, 747–752. https://doi.org/10.1038/s41568-021-00399-1

40. Bulten, W., Pinckaers, H., van Boven, H., et al. (2022). Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nature Medicine, 28, 154–163. https://doi.org/10.1038/s41591-021-01620-2

41. Shmatko, A., Ghaffari Laleh, N., Gerstung, M., & Kather, J. N. (2022). Artificial intelligence in histopathology: enhancing cancer research and clinical oncology. Nature Cancer, 3, 1026–1038. https://doi.org/10.1038/s43018-022-00436-4

42. Spratt, D. E., et al. (2023). Artificial intelligence predictive model for hormone therapy use in prostate cancer. NEJM Evidence. https://doi.org/10.1056/EVIDoa2300023

43. Guo, J., et al. (2024). Cross-modal deep learning model for predicting pathologic complete response to neoadjuvant chemotherapy in breast cancer. npj Precision Oncology. https://doi.org/10.1038/s41698-024-00678-8

44. Jiang, P., et al. (2022). Big data in basic and translational cancer research. Nature Reviews Cancer. https://www.nature.com/articles/s41568-022-00502-0

45. Liu, B., Shang, Q., Li, J., Yao, S., Ouyang, M., Wang, Y., Luo, S., & Ouyang, Q. (2026). AI and big data in oncology: a physician-centered perspective on emerging clinical and research applications. Cancer Innovation, 5(1), e70047. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12855167/

46. Silva, M. C., Eugénio, P., Faria, D., & Pesquita, C. (2022). Ontologies and knowledge graphs in oncology research. Cancers, 14(8), 1906. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9029532/

47. The potential use of big data in oncology (2019). Oral Oncology. https://www.sciencedirect.com/science/article/pii/S1368837519303021

48. Big data and artificial intelligence in cancer research (2023). Review. https://www.sciencedirect.com/science/article/abs/pii/S2405803323002170

49. Electronic records, registries, and the development of "big data": crowd-sourcing quality toward knowledge (2017). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5277016/

50. A review of big data technology and its application in cancer care (2024). Computers in Biology and Medicine. https://www.sciencedirect.com/science/article/abs/pii/S0010482524006620

51. Kokkotou, E., Anagnostakis, M., Evangelou, G., Syrigos, N. K., & Gkiozos, I. (2024). Real-world data and evidence in lung cancer: a review of recent developments. Cancers, 16(7), 1414. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11010882/

52. Benedum, C. M., Sondhi, A., Fidyk, E., Cohen, A. B., Nemeth, S., Adamson, B., Estévez, M., & Bozkurt, S. (2023). Replication of real-world evidence in oncology using electronic health record data extracted by machine learning. Cancers, 15(6), 1853. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10046618/

53. Tait, K., Cronin, J., Wiper, O., Wallis, J., Davies, J., & Dürichen, R. (2025). ArcTEX—a novel clinical data enrichment pipeline to support real-world evidence

oncology studies. Frontiers in Digital Health. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12098606/

54. Casagni, M., Llewellyn, N., Kokolus, M., et al. (2024). Integrated electronic health record tools to access real-world data in oncology research. JAMIA Open, 7(4), ooae144. https://academic.oup.com/jamiaopen/article/7/4/ooae144/7921520

55. George, S., Campbell, N., Hillman, S. L., et al. (2024). Feasibility of structuring electronic health record data to facilitate real-world data research: ICAREdata methods applied to multicenter cancer clinical trials. Cancer. https://doi.org/10.1002/cncr.35528

56. Castellanos, E. H., Wittmershaus, B. K., & Chandwani, S. (2024). Raising the bar for real-world data in oncology: approaches to quality across multiple dimensions. JCO Clinical Cancer Informatics, 8, e2300046. https://ascopubs.org/doi/10.1200/CCI.23.00046

57. Ru, B., Sillah, A., Desai, K., Chandwani, S., Yao, L., & Kothari, S. (2024). Real-world data quality framework for oncology time to treatment discontinuation use case: implementation and evaluation study. JMIR Medical Informatics, 12, e47744. https://doi.org/10.2196/47744

58. Zalis, M., Viana Veloso, G. G., Aguiar Jr., P. N., Gimenes, N., Reis, M. X., Matsas, S., & Ferreira, C. G. (2024). Next-generation sequencing impact on cancer care: applications, challenges, and future directions. Frontiers in Genetics, 15, 1420190. https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2024.1420190/full

59. Ghoreyshi, N., Heidari, R., Farhadi, A., Chamanara, M., Farahani, N., Vahidi, M., & Behroozi, J. (2025). Next-generation sequencing in cancer diagnosis and treatment: clinical applications and future directions. Discover Oncology, 16(1), 578. https://pmc.ncbi.nlm.nih.gov/articles/PMC12009796/

60. Menyhárt, O., & Győrffy, B. (2021). Multi-omics approaches in cancer research with applications in tumor subtyping, prognosis, and diagnosis. Computational and Structural Biotechnology Journal, 19, 949–960. https://doi.org/10.1016/j.csbj.2021.01.009

61. Unger, M., & Kather, J. N. (2024). Deep learning in cancer genomics and histopathology. Genome Medicine, 16, 44. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10976780/

62. Reardon, B., et al. (2026). Convergence of machine learning and genomics for precision oncology. Nature Reviews Cancer, 26(3), 217–229. https://www.nature.com/articles/s41568-025-00897-6

63. Lambin, P., Leijenaar, R. T. H., Deist, T. M., Peerlings, J., de Jong, E. E. C., van Timmeren, J., Sanduleanu, S., Larue, R. T. H. M., Even, A. J. G., Jochems, A., van Wijk, Y., Woodruff, H., van Soest, J., Lustberg, T., Roelofs, E., van Elmpt, W., Dekker, A., Mottaghy, F. M., Wildberger, J. E., & Walsh, S. (2017). Radiomics: the bridge between medical imaging and personalized medicine. Nature Reviews Clinical Oncology, 14(12), 749–762. https://doi.org/10.1038/nrclinonc.2017.141

64. Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V., & Madabhushi, A. (2019). Artificial intelligence in digital pathology — new tools for diagnosis and precision

oncology. Nature Reviews Clinical Oncology, 16(11), 703–715. https://pmc.ncbi.nlm.nih.gov/articles/PMC6880861/

65. Niazi, M. K. K., Parwani, A. V., & Gurcan, M. N. (2019). Digital pathology and artificial intelligence. The Lancet Oncology, 20(5), e253–e261. https://doi.org/10.1016/S1470-2045(19)30154-8

66. Schneider, L., Laiouar-Pedari, S., Kuntz, S., Krieghoff-Henning, E., Hekler, A., Kather, J. N., Gaiser, T., Fröhling, S., & Brinker, T. J. (2022). Integration of deep learning-based image analysis and genomic data in cancer pathology: a systematic review. European Journal of Cancer, 160, 80–91. https://doi.org/10.1016/j.ejca.2021.10.007

67. Rahnenführer, J., De Bin, R., Benner, A., Ambrogi, F., Lusa, L., Boulesteix, A.-L., Migliavacca, E., Binder, H., Michiels, S., Sauerbrei, W., McShane, L., & topic group "High-dimensional data" (TG9) of the STRATOS initiative (2023). Statistical analysis of high-dimensional biomedical data: a gentle introduction to analytical goals, common approaches and challenges. BMC Medicine, 21(1), 182. https://doi.org/10.1186/s12916-023-02858-y

68. Zhao, Y., Chang, C., & Long, Q. (2019). Knowledge-guided statistical learning methods for analysis of high-dimensional -omics data in precision oncology. JCO Precision Oncology, 3, 1–9. https://pmc.ncbi.nlm.nih.gov/articles/PMC9797232/

69. Yadav, K., & Gupta, T. (2026). Machine learning and AI for cancer research and care: a review of applications, limitations, and future directions. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC13476218/

70. Cherlin, S., Bigirumurame, T., Grayling, M. J., Nsengimana, J., Ouma, L., Santaolalla, A., Wan, F., Williamson, S. F., & Wason, J. M. S. (2024). Utilising high-dimensional data in randomised clinical trials: a review of methods and practice. Clinical Trials. https://doi.org/10.1177/26320843231186399

71. Kakar, M. (2026). The rise of generalist foundation models and quantum computing in oncology. Technology in Cancer Research & Treatment. https://doi.org/10.1177/15330338261476208 (PMC: https://pmc.ncbi.nlm.nih.gov/articles/PMC13451690/)

72. Elemento, O., Khozin, S., & Sternberg, C. N. (2025). The use of artificial intelligence for cancer therapeutic decision-making. NEJM AI, 2(5). https://doi.org/10.1056/AIra2401164
73. Cheng, C. H., & Wong, C. C. (2025). The role of artificial intelligence-based foundation models and "copilots" in cancer pathology: potential and challenges. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12763834/

74. Marra, A., Morganti, S., Pareja, F., Campanella, G., Bibeau, F., Fuchs, T., Loda, M., Parwani, A., Scarpa, A., Reis-Filho, J. S., Curigliano, G., Marchiò, C., & Kather, J. N. (2025). Artificial intelligence entering the pathology arena in oncology: current applications and future perspectives. Annals of Oncology. https://doi.org/10.1016/j.annonc.2025.03.006

75. Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. (2022). Multimodal biomedical AI. Nature Medicine, 28(9), 1773–1784. https://doi.org/10.1038/s41591-022-01981-2

76. Boehm, K. M., Khosravi, P., Vanguri, R., Gao, J., & Shah, S. P. (2022). Harnessing multimodal data integration to advance precision oncology. Nature Reviews Cancer, 22(2), 114–126. https://doi.org/10.1038/s41568-021-00408-3

77. Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. npj Digital Medicine, 3, 136. https://doi.org/10.1038/s41746-020-00341-z

78. Steyaert, S., Pizurica, M., Nagaraj, D., Khandelwal, P., Hernandez-Boussard, T., Gentles, A. J., & Gevaert, O. (2023). Multimodal data fusion for cancer biomarker discovery with deep learning. Nature Machine Intelligence, 5(4), 351–362. https://doi.org/10.1038/s42256-023-00633-5

79. Fountzilas, E., Pearce, T., Baysal, M. A., Chakraborty, A., & Tsimberidou, A. M. (2025). Convergence of evolving artificial intelligence and machine learning techniques in precision oncology. npj Digital Medicine, 8, 75. https://doi.org/10.1038/s41746-025-01471-5

80. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... Liang, P. (2021). On the opportunities and risks of foundation models. arXiv:2108.07258. https://arxiv.org/abs/2108.07258

81. IBM (s.d.). Foundation models. IBM Think. https://www.ibm.com/it-it/think/topics/foundation-models

82. Stryker, C. (s.d.). Pre-trained models. IBM Think. https://www.ibm.com/it-it/think/topics/pretrained-model

83. Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. Journal of Machine Learning Research, 21(140), 1–67. https://arxiv.org/abs/1910.10683

84. Si, Y., Zou, J., Gao, Y., Chuai, G., Liu, Q., & Chen, L. (2024). Foundation models in molecular biology. Biophysics Reports, 10(3), 135–151. https://pmc.ncbi.nlm.nih.gov/articles/PMC11252241/

85. IBM (s.d.). Large language models. IBM Think. https://www.ibm.com/it-it/think/topics/large-language-models

86. Radford, A., et al. (2018). Improving language understanding by generative pre-training. OpenAI. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

87. Aalto University (s.d.). Pre-training and fine-tuning. In Introduction to Large Language Models, CS Foundations, part 4. https://csfoundations.cs.aalto.fi/en/courses/introduction-to-large-language-models/part-4/2-pre-training-and-fine-tuning

88. Isambard-AI Documentation, University of Bristol (s.d.). Pre-training. In Tutorials: LLMs. https://docs.isambard.ac.uk/user-documentation/tutorials/llms/pre-training/

89. NVIDIA (2025). Next steps. In NeMo Framework User Guide (v25.04), Data Curation. https://docs.nvidia.com/nemo-framework/user-guide/25.04/datacuration/nextsteps.html

90. Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2023). A comprehensive overview of large language models. arXiv:2307.06435 (v10, 17 ottobre 2024). https://arxiv.org/pdf/2307.06435

91. Tie, G., et al. (2025). A survey on post-training of large language models. arXiv:2503.06072. https://arxiv.org/pdf/2503.06072

92. Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems (NeurIPS 2020). https://arxiv.org/abs/2005.14165

93. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805. https://arxiv.org/abs/1810.04805

94. Kaplan, J., et al. (2020). Scaling laws for neural language models. arXiv:2001.08361. https://arxiv.org/abs/2001.08361

95. Hoffmann, J., et al. (2022). Training compute-optimal large language models. arXiv:2203.15556. https://arxiv.org/abs/2203.15556

96. Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Le, Q., & Zhou, D. (2022a). Chain-of-thought prompting elicits reasoning in large language models. Advances in Neural Information Processing Systems (NeurIPS 2022). https://arxiv.org/abs/2201.11903

97. Kojima, T., et al. (2022). Large language models are zero-shot reasoners. arXiv:2205.11916. https://arxiv.org/abs/2205.11916

98. Wei, J., et al. (2022b). Emergent abilities of large language models. arXiv:2206.07682. https://arxiv.org/abs/2206.07682

99. Srivastava, A., et al. (2022). Beyond the imitation game: quantifying and extrapolating the capabilities of language models (BIG-bench). arXiv:2206.04615. https://arxiv.org/abs/2206.04615

100. Schaeffer, R., Miranda, B., & Koyejo, S. (2023). Are emergent abilities of large language models a mirage? arXiv:2304.15004. https://arxiv.org/abs/2304.15004

101. Ouyang, L., et al. (2022). Training language models to follow instructions with human feedback. arXiv:2203.02155. https://arxiv.org/abs/2203.02155

102. Manning, C. D., Clark, K., Hewitt, J., Khandelwal, U., & Levy, O. (2020). Emergent linguistic structure in artificial neural networks trained by self-supervision. Proceedings of the National Academy of Sciences, 117(48). https://www.pnas.org/doi/10.1073/pnas.1907367117

103. Wang, A., et al. (2018). GLUE: a multi-task benchmark and analysis platform for natural language understanding. arXiv:1804.07461. https://arxiv.org/abs/1804.07461

104. Hendrycks, D., et al. (2021). Measuring massive multitask language understanding. arXiv:2009.03300. https://arxiv.org/abs/2009.03300

105. Holtzman, A., et al. (2020). The curious case of neural text degeneration. arXiv:1904.09751. https://arxiv.org/abs/1904.09751

106. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: can language models be too big? Proceedings of FAccT '21, ACM. https://dl.acm.org/doi/10.1145/3442188.3445922

107. Piantadosi, S. T., & Hill, F. (2022). Meaning without reference in large language models. arXiv:2208.02957. https://arxiv.org/abs/2208.02957

108. Mitchell, M., & Krakauer, D. C. (2023). The debate over understanding in AI's large language models. Proceedings of the National Academy of Sciences, 120(13). https://www.pnas.org/doi/10.1073/pnas.2215907120

109. Zhang, J., Zhao, Y., Saleh, M., & Liu, P. (2020). PEGASUS: pre-training with extracted gap-sentences for abstractive summarization. International Conference on Machine Learning (ICML 2020). https://arxiv.org/abs/1912.08777

110. Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems (NeurIPS 2020). https://arxiv.org/abs/2005.11401

111. Bender, E. M., & Koller, A. (2020). Climbing towards NLU: on meaning, form, and understanding in the age of data. Proceedings of ACL 2020. https://aclanthology.org/2020.acl-main.463/

112. Harnad, S. (1990). The symbol grounding problem. Physica D, 42, 335–346. https://www.sciencedirect.com/science/article/pii/0167278990900876

113. Bisk, Y., Holtzman, A., Thomason, J., et al. (2020). Experience grounds language. Proceedings of EMNLP 2020. https://arxiv.org/abs/2004.10151

114. OpenAI (2023). GPT-4 technical report. arXiv:2303.08774. https://arxiv.org/abs/2303.08774

115. Yin, S., Fu, C., Zhao, S., Li, K., Sun, X., Xu, T., & Chen, E. (2024). A survey on multimodal large language models. National Science Review. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11651817/

116. Liu, H., Cai, W., Sun, Y., Liu, H., Zou, Z., Zhao, Q., Georgaka, S., Zhou, H., & Sun, J. (2026). Foundation models in omics research: a comprehensive survey. Briefings in Bioinformatics. https://pubmed.ncbi.nlm.nih.gov/42533816/

117. Large language models in bioinformatics: applications and perspectives. PubMed Central, PMC10802675. https://pmc.ncbi.nlm.nih.gov/articles/PMC10802675/

118. Biomedical multimodal large language models: from model-centric development to clinically grounded evaluation and integration. PubMed Central, PMC13329790. https://pmc.ncbi.nlm.nih.gov/articles/PMC13329790/

119. Large language models for bioinformatics. PubMed Central, PMC12806030. https://pmc.ncbi.nlm.nih.gov/articles/PMC12806030/

120. Gene-LLMs: a comprehensive survey of transformer-based genomic large language models. PubMed Central, PMC12558637. https://pmc.ncbi.nlm.nih.gov/articles/PMC12558637/

121. Omics-based large language models: a new engine for drug discovery. PubMed Central, PMC12827895. https://pmc.ncbi.nlm.nih.gov/articles/PMC12827895/

122. Waqas, A., Tripathi, A., Ramachandran, R. P., Stewart, P. A., & Rasool, G. (2024). Multimodal data integration for oncology in the era of deep neural networks: a review. Frontiers in Artificial Intelligence, 7, 1408843. https://doi.org/10.3389/frai.2024.1408843

123. Moor, M., Banerjee, O., Abad, Z. S. H., Krumholz, H. M., Leskovec, J., Topol, E. J., & Rajpurkar, P. (2023). Foundation models for generalist medical artificial intelligence. Nature, 616(7956), 259–265. https://pubmed.ncbi.nlm.nih.gov/37045921/

124. AI-assisted multimodal data integration for precision oncology (2026). Frontiers in Artificial Intelligence. https://doi.org/10.3389/frai.2026.1917442

125. Chen, D., et al. (2025). Large language models in oncology: a review. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12164365/

126. Gould, A., et al. (2025). Use of large language models in clinical cancer research. https://pubmed.ncbi.nlm.nih.gov/40388683/

127. Clusmann, J., et al. (2023). The future landscape of large language models in medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10564948/

128. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6322008/

129. Kearney, S. P., et al. (2024). Can generalist foundation models learn medical reasoning? Cell Reports Medicine. https://www.cell.com/cell-reports-medicine/fulltext/S2666-3791(24)00213-9

130. Bilal, M., Hamza, A., & Malik, N. (2025). NLP for analyzing electronic health records and clinical notes in cancer research: a review. https://pubmed.ncbi.nlm.nih.gov/39894080/

131. Savova, G. K., et al. (2010). Mayo Clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. Journal of the American Medical Informatics Association. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2995661/

132. Uzuner, Ö., et al. (2011). 2010 i2b2/VA workshop on natural language processing challenges for clinical records: assertion status and relation extraction. Journal of the American Medical Informatics Association. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3168328

133. Università di Bologna (s.d.). Fondo Sociale Europeo 2014–2020: progetti e iniziative. https://www.unibo.it/it/ricerca/progetti-e-iniziative/fondo-sociale-europeo-2014-2020/99640/769/8765

134. Sharma, S., George, P., & Waddell, N. (2021). Precision diagnostics: integration of tissue pathology and genomics in cancer. Pathology, 53(7), 809–817. https://doi.org/10.1016/j.pathol.2021.08.003

135. Raphael, B. J., Dobson, J. R., Oesper, L., & Vandin, F. (2014). Identifying driver mutations in sequenced cancer genomes: computational approaches to enable precision medicine. Genome Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3978567/

136. Mishra, A., & Verma, M. (2010). Cancer biomarkers: are we ready for the prime time? Cancers, 2(1), 190–208. https://doi.org/10.3390/cancers2010190

137. Simon, R. (2013). Actionable prognostic biomarkers. In Genomic Clinical Trials and Predictive Medicine. Cambridge University Press. https://www.cambridge.org/core/books/genomic-clinical-trials-and-predictive-medicine/actionable-prognostic-biomarkers/E5B4C634E63147815318B36FAB7CE1A8

138. Chalmers, Z. R., et al. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. Genome Medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC5395719

139. Luchini, C., Bibeau, F., Ligtenberg, M. J. L., et al. (2019). ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. Annals of Oncology, 30(8), 1232–1243. https://doi.org/10.1093/annonc/mdz116

140. Fabrizio, D. A., et al. (2018). Beyond microsatellite testing: assessment of tumor mutational burden identifies subsets of colorectal cancer who may respond to immune checkpoint inhibition. Journal of Gastrointestinal Oncology, 9(4), 610–617. https://doi.org/10.21037/jgo.2018.05.06

141. AIMAC (2022). La caratterizzazione molecolare. In I tumori cerebrali. Associazione Italiana Malati di Cancro. https://www.aimac.it/libretti-tumore/tumore-cervello/caratterizzazione-molecolare

142. Hoadley, K. A., Yau, C., Hinoue, T., et al. (2018). Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer. Cell, 173(2), 291–304.e6. https://doi.org/10.1016/j.cell.2018.03.022

143. Sahm, F., et al. (2023). Molecular diagnostic tools for the World Health Organization (WHO) 2021 classification of gliomas, glioneuronal and neuronal tumors; an EANO guideline. Neuro-Oncology, noad100. https://doi.org/10.1093/neuonc/noad100

144. Multi-omics data integration approaches for precision oncology (2022). Molecular Omics, 18, 469–479. https://doi.org/10.1039/D1MO00411E

145. Wu, C., et al. (2025). Towards generalist foundation model for radiology by leveraging web-scale 2D&3D medical data. Nature Communications. https://doi.org/10.1038/s41467-025-62385-7

146. Ma, J., et al. (2024). Segment anything in medical images. Nature Communications, 15, 654. https://doi.org/10.1038/s41467-024-44824-z

147. Hamamci, I. E., et al. (2026). Generalist foundation models from a multimodal dataset for 3D computed tomography. Nature Biomedical Engineering. https://doi.org/10.1038/s41551-025-01599-y

148. Chen, Z., et al. (2024). CheXagent: towards a foundation model for chest X-ray interpretation. arXiv:2401.12208. https://arxiv.org/abs/2401.12208

149. Pérez-García, F., et al. (2025). Exploring scalable medical image encoders beyond text supervision (RAD-DINO). Nature Machine Intelligence, 7. https://arxiv.org/abs/2401.10815

150. Li, F., et al. (2025). Evaluating vision language models (VLMs) for radiology: a comprehensive analysis. arXiv:2504.16047. https://arxiv.org/abs/2504.16047

151. Sali, R., Aschenaki, Y., Leveillee, R., Baba, F., Tessema, L., Dixon, C., Bachhamba, K., & Zhang, D. Y. (2026). Foundation models in computational pathology: methods, applications and clinical implications. BMJ Oncology, 5(1), e001102. https://doi.org/10.1136/bmjonc-2026-001102

152. Xu, H., et al. (2024). A whole-slide foundation model for digital pathology from real-world data. Nature, 630(8015), 181–188. https://doi.org/10.1038/s41586-024-07441-w

153. Chen, R. J., et al. (2024). Towards a general-purpose foundation model for computational pathology (UNI). Nature Medicine, 30, 850–862. https://doi.org/10.1038/s41591-024-02857-3

154. Vorontsov, E., et al. (2024). A foundation model for clinical-grade computational pathology and rare cancers detection. Nature Medicine, 30(10), 2924–2935. https://doi.org/10.1038/s41591-024-03141-0

155. Lu, M. Y., et al. (2024). A visual-language foundation model for computational pathology. Nature Medicine. https://doi.org/10.1038/s41591-024-02856-4

156. MarketsandMarkets (s.d.). Digital Pathology Market – Global Forecast to 2029 (comunicato stampa). https://marketsandmarkets.com/PressReleases/pathology.asp

157. BCC Research (s.d.). Digital Pathology: Technologies and Global Markets. https://www.bccresearch.com/market-research/healthcare/digital-pathology-tech-markets-report.html

158. Quotidiano Sanità (s.d.). Tumori. Asl Biella. Quando la tecnologia scende in campo per migliorare la qualità delle diagnosi. https://www.quotidianosanita.it/?p=79943

159. Quotidiano Sanità (2025). Borse specializzazione medicina (contratti non assegnati, concorso 2025). https://www.quotidianosanita.it/?p=145846

160. Ministero della Salute (2015). Linee guida per la tracciabilità, raccolta, trasporto, conservazione e archiviazione di cellule e tessuti per indagini diagnostiche di Anatomia Patologica. https://sa-fvg.cultura.gov.it/fileadmin/materiali/Linee_guida_gestione_campioni_biologici_-_Ministero_della_Salute__2015_.pdf

161. Finanza Online (2023, 30 gennaio). Gpi: vince gara da 5,9 mln per software di Anatomia Patologica in Veneto. https://www.finanzaonline.com/notizie/gpi-vince-gara-da-59-mln-per-software-di-anatomia-patologica-in-veneto

162. L'Unione Sarda (s.d.). Anatomia Patologica completamente digitale: il Brotzu guida l'innovazione in Sardegna. https://www.unionesarda.it/news-sardegna/cagliari/anatomia-patologica-completamente-digitale-il-brotzu-guida-linnovazione-in-sardegna-qdhc6idn

163. Feng, L., et al. (2022). Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. The Lancet Digital Health, 4(1), e8–e17. https://doi.org/10.1016/S2589-7500(21)00215-6

164. Dimitriou, N., Arandjelović, O., & Caie, P. D. (2019). Deep learning for whole slide image analysis: an overview. Frontiers in Medicine, 6, 264. https://doi.org/10.3389/fmed.2019.00264

165. Philips (s.d.). The nuances of digital pathology integration and interoperability (white paper). https://www.usa.philips.com/healthcare/white-paper/interoperability-digital-pathology

166. AuntMinnie (2020). How to integrate pathology into an enterprise imaging system. https://www.auntminnie.com/industry-news/market-analysis/article/15632201/how-to-integrate-pathology-into-an-enterprise-imaging-system

167. Tripathi, A., Waqas, A., Schabath, M. B., Yilmaz, Y., & Rasool, G. (2025a). HONeYBEE: enabling scalable multimodal AI in oncology through foundation model-driven embeddings. npj Digital Medicine, 8(1), 622. https://doi.org/10.1038/s41746-025-02003-4

168. Weinstein, J. N., et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. Nature Genetics, 45(10), 1113–1120. https://doi.org/10.1038/ng.2764

169. Clark, K., et al. (2013). The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. Journal of Digital Imaging, 26, 1045–1057. https://doi.org/10.1007/s10278-013-9622-7

170. Vanguri, R. S., et al. (2022). Multimodal integration of radiology, pathology and genomics for prediction of response to PD-(L)1 blockade in patients with non-small cell lung cancer. Nature Cancer, 3, 1151–1164. https://doi.org/10.1038/s43018-022-00416-8

171. Xiang, J., et al. (2025). A vision–language foundation model for precision oncology (MUSK). Nature, 638(8051). https://doi.org/10.1038/s41586-024-08378-w

172. Yang, H., Yang, M., Chen, J., Yao, G., Zou, Q., & Jia, L. (2025). Multimodal deep learning approaches for precision oncology: a comprehensive review. Briefings in Bioinformatics, 26(1), bbae699. https://academic.oup.com/bib/article/26/1/bbae699/7942793

173. Emergent Mind (2025). Multimodal fusion strategies in deep learning. https://www.emergentmind.com/topics/multimodal-fusion-strategies

174. Zhang, B., et al. (2024). Multimodal integration strategies for clinical application in oncology. PMC12405423. https://pmc.ncbi.nlm.nih.gov/articles/PMC12405423/

175. Haghighat, F., et al. (2025). Multimodal deep learning and data fusion in precision breast cancer. ISJ Trend, 235364. https://www.isjtrend.com/articles/235364

176. Liu, Z., Wu, Y., Xu, H., Wang, M., Weng, S., Pei, D., et al. (2025). Multimodal fusion of radio-pathology and proteogenomics identify integrated glioma subtypes with prognostic and therapeutic opportunities. Nature Communications, 16(1), 3510. https://pubmed.ncbi.nlm.nih.gov/40222975/ (testo completo: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11994800/)

177. Jang, H. J., & Lee, S. H. (2025). AI-driven digital pathology: deep learning and multimodal integration for precision oncology. International Journal of Molecular Sciences, 27(1), 379. https://pubmed.ncbi.nlm.nih.gov/41516254/ (testo completo: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12785522/)

178. Corso, F., Zec, A., Favali, M., Ligero, M., Hilgers, L., Invernizzi, L. M., … Prelaj, A. (2026). Leave No Data Behind: exploring a new paradigm in oncology with foundation models and large language models. Cell Reports Medicine, 102966. https://doi.org/10.1016/j.xcrm.2026.102966

179. Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. BMJ, 312(7023), 71–72. https://doi.org/10.1136/bmj.312.7023.71

180. Hamburg, M. A., & Collins, F. S. (2010). The path to personalized medicine. New England Journal of Medicine, 363(4), 301–304. https://doi.org/10.1056/NEJMp1006304

181. Druker, B. J., Talpaz, M., Resta, D. J., et al. (2001). Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. New England

Journal of Medicine, 344(14), 1031–1037. https://doi.org/10.1056/NEJM200104053441401

182. Lynch, T. J., Bell, D. W., Sordella, R., et al. (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non–small-cell lung cancer to gefitinib. New England Journal of Medicine, 350(21), 2129–2139. https://doi.org/10.1056/NEJMoa040938

183. Paez, J. G., Jänne, P. A., Lee, J. C., et al. (2004). EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. Science, 304(5676), 1497–1500. https://doi.org/10.1126/science.1099314

184. Collins, F. S., & Varmus, H. (2015). A new initiative on precision medicine. New England Journal of Medicine, 372(9), 793–795. https://doi.org/10.1056/NEJMp1500523

185. Ashley, E. A. (2016). Towards precision medicine. Nature Reviews Genetics, 17(9), 507–522. https://doi.org/10.1038/nrg.2016.86

186. National Research Council (2011). Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, DC: The National Academies Press. https://doi.org/10.17226/13284

187. König, I. R., Fuchs, O., Hansen, G., von Mutius, E., & Kopp, M. V. (2017). What is precision medicine? European Respiratory Journal, 50(4), 1700391. https://doi.org/10.1183/13993003.00391-2017

188. Dagogo-Jack, I., & Shaw, A. T. (2018). Tumour heterogeneity and resistance to cancer therapies. Nature Reviews Clinical Oncology, 15(2), 81–94. https://doi.org/10.1038/nrclinonc.2017.166

189. Gerlinger, M., Rowan, A. J., Horswell, S., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. New England Journal of Medicine, 366(10), 883–892. https://doi.org/10.1056/NEJMoa1113205

190. Burrell, R. A., McGranahan, N., Bartek, J., & Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. Nature, 501(7467), 338–345. https://doi.org/10.1038/nature12625

191. Relling, M. V., & Evans, W. E. (2015). Pharmacogenomics in the clinic. Nature, 526(7573), 343–350. https://pmc.ncbi.nlm.nih.gov/articles/PMC4711261

192. Amstutz, U., Henricks, L. M., Offer, S. M., et al. (2018). Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing: 2017 update. Clinical Pharmacology & Therapeutics, 103(2), 210–216. https://pubmed.ncbi.nlm.nih.gov/29152729/

193. Conforti, F., Pala, L., Bagnardi, V., et al. (2018). Cancer immunotherapy efficacy and patients' sex: a systematic review and meta-analysis. The Lancet Oncology, 19(6), 737–746. https://doi.org/10.1016/S1470-2045(18)30261-4

194. Routy, B., Le Chatelier, E., Derosa, L., et al. (2018). Gut microbiome influences efficacy of PD-1–based immunotherapy against epithelial tumors. Science, 359(6371), 91–97. https://doi.org/10.1126/science.aan3706

195. Hingorani, A. D., van der Windt, D. A., Riley, R. D., et al. (2013). Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ, 346, e5793. https://doi.org/10.1136/bmj.e5793

196. Teleanu, M. V., Schneider, A., Ball, C. R., et al. (2025). Celebrating Ulrik Ringborg: multi-omics-based patient stratification for precision cancer treatment. Biomolecules, 15(5), 693. https://doi.org/10.3390/biom15050693

197. [Autori da completare] (anno da completare). Tumor heterogeneity: a great barrier in the age of cancer immunotherapy. https://pmc.ncbi.nlm.nih.gov/articles/PMC7918981

198. Vorontsov, E., Wang, Y. K., Bozkurt, A., et al. (2026). A multimodal foundation model for longitudinal patient representation and scalable insight generation in oncology. arXiv:2608.24688 (preprint). https://papers.cool/arxiv/2608.24688

199. [Autori da completare] (2024). A narrative review on the application of large language models to support cancer care and research. Yearbook of Medical Informatics. https://doi.org/10.1055/s-0044-1800726

200. Hao, Y., Qiu, Z., Holmes, J., et al. (2025). Large language model integrations in cancer decision-making: a systematic review and meta-analysis. npj Digital Medicine. https://doi.org/10.1038/s41746-025-01824-7

201. Song, S., Borjigin-Wang, M., Madejski, I., & Grossman, R. L. (2025). Multimodal cancer modeling in the age of foundation model embeddings. arXiv:2505.07683. https://arxiv.org/abs/2505.07683

202. Kaczmarzyk, J. R., Van Alsten, S. C., Cozzo, A. J., et al. (2026). Towards interpretable prediction of recurrence risk in breast cancer using pathology foundation models. npj Digital Medicine. https://doi.org/10.1038/s41746-025-02334-2

203. Ramanathan, V., Xu, T., Pati, P., Ahmed, F., Goubran, M., & Martel, A. L. (2025). ModalTune: fine-tuning slide-level foundation models with multi-modal information for multi-task learning in digital pathology. arXiv:2503.17564. https://arxiv.org/abs/2503.17564

204. Wang, D., & Gu, J. (2016). Integrative clustering methods of multi-omics data for molecule-based cancer classifications. Quantitative Biology, 4(1), 58–67. https://doi.org/10.1007/s40484-016-0063-4

205. Vitale, I., Shema, E., Loi, S., & Galluzzi, L. (2021). Intratumoral heterogeneity in cancer progression and response to immunotherapy. Nature Medicine, 27(2), 212–224. https://doi.org/10.1038/s41591-021-01233-9

206. Suter, P., Dazert, E., Kuipers, J., et al. (2022). Multi-omics subtyping of hepatocellular carcinoma patients using a Bayesian network mixture model. PLoS Computational Biology, 18(9), e1009767. https://doi.org/10.1371/journal.pcbi.1009767

207. Tarhini, A. A., Dave, P., & El Naqa, I. (2025). General artificial intelligence for the diagnosis and treatment of cancer: the rise of foundation models. BJR|Artificial Intelligence, 2(1), ubaf015. https://doi.org/10.1093/bjrai/ubaf015

208. Cui, H., Wang, C., Maan, H., et al. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nature Methods. https://doi.org/10.1038/s41592-024-02201-0

209. Li, M., Zhan, Z., Huang, J., et al. (2024). CancerLLM: a large language model in cancer domain. arXiv:2406.10459. https://arxiv.org/abs/2406.10459

210. Pelossof, R., Carty, M., Ahmed, T., Lauly, S., Purpura, A., Mueller, E., & Guinney, J. (2025). Multi-modal large language models for metastatic breast cancer prognosis [abstract]. AACR Annual Meeting 2025. https://doi.org/10.1158/1538-7445.AM2025-5006

211. Carl, N., Schramm, F., Haggenmüller, S., et al. (2024). Large language model use in clinical oncology. npj Precision Oncology, 8, 240. https://doi.org/10.1038/s41698-024-00733-4

212. Kedzierska, K. Z., Crawford, L., Amini, A. P., & Lu, A. X. (2023). Assessing the limits of zero-shot foundation models in single-cell biology. bioRxiv. https://doi.org/10.1101/2023.10.16.561085

213. Wang, J., Hong, R., Demicco, E. G., et al. (2023). Deep learning integrates histopathology and proteogenomics at a pan-cancer level. Cell Reports Medicine, 4(9), 101173. https://doi.org/10.1016/j.xcrm.2023.101173

214. Li, M. F., Du, S. N., Bao, P. T., & Li, Y. G. (2026). Multidimensional integration: a novel breakthrough in prognostic prediction for immunochemotherapy in human epidermal growth factor receptor-2-negative advanced gastric cancer. World Journal of Gastrointestinal Oncology, 18(9), 118614. https://doi.org/10.4251/wjgo.118614

215. ESMO (s.d.). Prognostic biomarkers. In ESMO Handbook of Translational Research, Chapter 1. https://oncologypro.esmo.org/education-library/esmo-handbooks/translational-research/prognostic-biomarkers

216. Fumet, J.-D., Truntzer, C., Yarchoan, M., & Ghiringhelli, F. (2020). Tumour mutational burden as a biomarker for immunotherapy: current data and emerging concepts. https://pmc.ncbi.nlm.nih.gov/articles/PMC9473693

217. Tripathi, A., Waqas, A., Venkatesan, K., et al. (2025b). Employing consensus-based reasoning with locally deployed LLMs for enabling structured data extraction from surgical pathology reports. medRxiv. https://www.medrxiv.org/node/1100144

218. [Autori da completare] (anno da completare). Potential predictive biomarkers in antitumor immunotherapy: navigating the future of antitumor treatment and immune checkpoint inhibitor efficacy. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC11625675/

219. Kather, J. N., Heij, L. R., Grabsch, H. I., et al. (2020). Pan-cancer image-based detection of clinically actionable genetic alterations. Nature Cancer, 1(8), 789–799. https://eprints.whiterose.ac.uk/159675

220. Menon, T. P., Mahajan, A., & Powell, D. (2026). Foundation model embeddings for multimodal oncology data integration. npj Digital Medicine, 9(1), 131. https://doi.org/10.1038/s41746-025-02312-8

221. Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. npj Digital Medicine, 3, 17. https://doi.org/10.1038/s41746-020-0221-y

222. Wang, L., Chen, X., Zhang, L., Li, L., Huang, Y., Sun, Y., & Yuan, X. (2023). Artificial intelligence in clinical decision support systems for oncology. International Journal of Medical Sciences, 20(1), 79–86. https://doi.org/10.7150/ijms.77205

223. Somashekhar, S. P., Sepúlveda, M.-J., Puglielli, S., et al. (2018). Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. Annals of Oncology, 29(2), 418–423. https://doi.org/10.1093/annonc/mdx781

224. Jie, Z., Zhiying, Z., & Li, L. (2021). A meta-analysis of Watson for Oncology in clinical application. Scientific Reports, 11, 5792. https://doi.org/10.1038/s41598-021-84973-5

225. Choi, Y. I., Chung, J.-W., Kim, K. O., et al. (2019). Concordance rate between clinicians and Watson for Oncology among patients with advanced gastric cancer: early, real-world experience in Korea. Canadian Journal of Gastroenterology and Hepatology, 2019, 8072928. https://doi.org/10.1155/2019/8072928

226. Singhal, K., Azizi, S., Tu, T., et al. (2023). Large language models encode clinical knowledge. Nature, 620(7972), 172–180. https://doi.org/10.1038/s41586-023-06291-2

227. Benary, M., Wang, X. D., Schmidt, M., et al. (2023). Leveraging large language models for decision support in personalized oncology. JAMA Network Open, 6(11), e2343689. https://doi.org/10.1001/jamanetworkopen.2023.43689

228. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38. https://doi.org/10.1038/s41591-021-01614-0

229. Somashekhar, S. P., Sepúlveda, M.-J., et al. (2018b). A prospective blinded study of 1000 cases analyzing the role of artificial intelligence: Watson for Oncology and change in decision making of a multidisciplinary tumor board (MDT) from a tertiary care cancer center [abstract]. Journal of Clinical Oncology, ASCO Annual Meeting 2018. https://ace.asco.org/content/253859/abstract/177645

230. Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: a systematic review of frequency, effect mediators, and mitigators. Journal of the American Medical Informatics Association, 19(1), 121–127. https://doi.org/10.1136/amiajnl-2011-000089

231. Kelly, C. J., Karthikesalingam, A., Suleyman, M., Corrado, G., & King, D. (2019). Key challenges for delivering clinical impact with artificial intelligence. BMC Medicine, 17, 195. https://doi.org/10.1186/s12916-019-1426-2

232. Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. JAMA, 320(21), 2199–2200. https://doi.org/10.1001/jama.2018.17163

233. Ah-Thiane, L., et al. (anno da completare). Large language models as decision-making tools in oncology: comparing artificial intelligence suggestions and expert recommendations. JCO Clinical Cancer Informatics. https://pmc.ncbi.nlm.nih.gov/articles/PMC11949217/

234. Kaufmes, K., et al. (anno da completare). Evaluating medium scale, open-source large language models: towards decision support in a precision oncology care delivery context. Studies in Health Technology and Informatics. https://pubmed.ncbi.nlm.nih.gov/40899530/

235. Jun, H., et al. (anno da completare). A context-augmented large language model for accurate precision oncology medicine recommendations. Cancer Cell. https://pmc.ncbi.nlm.nih.gov/articles/PMC12811686/

236. IBM (s.d.). What are AI hallucinations? IBM Think. https://www.ibm.com/think/topics/ai-hallucinations

237. Carlini, N., et al. (2021). Extracting training data from large language models. USENIX Security Symposium. https://arxiv.org/abs/2012.07805

238. Shokri, R., et al. (2017). Membership inference attacks against machine learning models. IEEE Symposium on Security and Privacy. https://doi.org/10.1109/SP.2017.41

239. Wei, A., Haghtalab, N., & Steinhardt, J. (2023). Jailbroken: how does LLM safety training fail? NeurIPS. https://arxiv.org/abs/2307.02483

240. OWASP Foundation (2023). OWASP Top 10 for Large Language Model Applications. https://owasp.org/www-project-top-10-for-large-language-model-applications/

241. Das, B. C., Amini, M. H., & Wu, Y. (2025). Security and privacy challenges of large language models: a survey. ACM Computing Surveys. https://arxiv.org/abs/2402.00888

242. Zou, A., et al. (2023). Universal and transferable adversarial attacks on aligned language models. arXiv:2307.15043. https://arxiv.org/abs/2307.15043

243. Ji, Z., et al. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys. https://dl.acm.org/doi/10.1145/3571730

244. [Autori da completare] (2025). npj Digital Medicine. https://www.nature.com/articles/s41746-025-0227

245. Mah, C. K., Tabor, T., & Mesirov, J. P. (2018). An accessible, interactive GenePattern Notebook for analysis and exploration of single-cell transcriptomic data. https://doi.org/10.12688/f1000research.15830.1

A nonno Gianni, perché tu sei
come il vento: non ti vedo, ma
ti percepisco.

Mi manchi immensamente.

E come amavi ripetermi,
forse oggi un po'di più
"ad maiora semper".

Ogni atomo del tuo corpo viene da una stella che è esplosa, e gli atomi nella tua mano sinistra vengono da una stella differente da quella corrispondente alla mano destra.
È la cosa più poetica che conosco della fisica:
"tu sei polvere di stelle".

Lawrence M. Krauss