UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea in Biologia Molecolare

ELABORATO DI LAUREA

TRASCRITTOMICA SPAZIALE E MECCANISMI DI ATTENZIONE NEL DEEP LEARNING SU IMMAGINI H&E PER L'IDENTIFICAZIONE DI UN PROFILO IMMUNITARIO NELLA PROGNOSI DEL CANCRO OVARICO

Tutor: D.ssa Enrica Calura
Dipartimento di Biologia

Laureanda: Serena Cosma

ANNO ACCADEMICO 2024/2025

INDICE

1. ABSTRACT

2. STATO DELL'ARTE
    a. Introduzione al cancro ovarico
    b. Introduzione dei modelli deep learning
    c. Introduzione all'esperimento

3. APPROCCIO SPERIMENTALE
    a. Addestramento del modello tramite fivefold cross-validation
    b. Valutazione del modello con Kaplan-Meier curve
    c. Struttura del network del modello
    d. Analisi del profilo immunitario

4. RISULTATI E DISCUSSIONE
    a. Validazione e verifica del modello
    b. Correlazione tra l'attention mechanism e l'attività immunitaria

5. BIBLIOGRAFIA

ABSTRACT

Il cancro ovarico in stadio avanzato rappresenta la neoplasia ginecologica con il più elevato tasso di mortalità. Pertanto, l'identificazione di strumenti prognostici affidabili potrebbe contribuire in maniera significativa al miglioramento della gestione clinica della malattia. Nonostante l'importanza di tale obiettivo, le strategie attualmente disponibili per predire il decorso della patologia rimangono limitate.

In questo contesto gli autori propongono un approccio basato sul meccanismo di attenzione nel *deep learning*, applicato all'analisi di immagini istologiche di campioni tumorali colorati con ematossilina ed eosina (H&E). La rete neurale viene addestrata su tali immagini e i risultati ottenuti sono successivamente integrati e interpretati mediante dati di trascrittomica spaziale.

I campioni analizzati appartengono a pazienti con cancro ovarico in stadio III e IV.

Attraverso il meccanismo di attenzione è stato assegnato, a ciascun campione, un punteggio correlato alla probabilità di sopravvivenza del paziente. Tali valori sono stati confrontati con i profili immunitari ricavati dalla trascrittomica spaziale.

L'analisi comparativa evidenzia un'associazione significativa tra i punteggi predittivi di sopravvivenza e specifici profili immunitari presenti nei tessuti tumorali.

Questo suggerisce che il meccanismo di attenzione può rappresentare uno strumento utile per l'identificazione di regioni tissutali caratterizzate da particolari attività immunitarie.

STATO DELL'ARTE

1. Introduzione al cancro ovarico

Il carcinoma epiteliale dell'ovaio, comunemente noto come cancro ovarico, è una neoplasia maligna originata dalle cellule epiteliali della gonade femminile. Dal punto di vista istologico, comprende diversi sottotipi, tra i quali il carcinoma sieroso di alto grado (HGSOC) che rappresenta la forma più comune e aggressiva. Quest'ultimo si manifesta frequentemente in uno stadio avanzato e tende a coinvolgere entrambe le ovaie.

Il carcinoma ovarico è la neoplasia ginecologica con il più alto livello di mortalità, principalmente a causa della sua diagnosi tardiva. Oltre il 75% delle pazienti riceve la diagnosi in una fase avanzata di malattia, poiché nelle fasi iniziali il tumore è generalmente asintomatico. I sintomi emergono solo quando la patologia è già estesa e si presentano in modo aspecifico, rendendo difficile una diagnosi precoce [1].
Tra le manifestazioni cliniche tipiche in stadio avanzato, si riscontrano carcinomatosi peritoneale, ascite, coinvolgimento linfonodale e versamenti pleurici. Queste condizioni determinano sintomi come dolore e distensione addominale, sazietà precoce, disturbi urinari e mancanza di respiro [2].
Tuttavia, tali segni sono frequentemente sottovalutati o attribuiti a condizioni benigne come indigestione, aumento di peso o invecchiamento, contribuendo così al ritardo diagnostico.
La malattia viene comunemente diagnosticata dopo la menopausa, con un picco di incidenza intorno ai 60 anni. Nonostante i trattamenti disponibili, la prognosi rimane severa: solo circa il 50% delle pazienti sopravvive oltre i cinque anni dalla diagnosi [1].

Le cause scatenanti del carcinoma ovarico non sono ancora completamente chiarite.
È noto, tuttavia, che un'importante componente di rischio è rappresentata dalla predisposizione genetica, in particolare mutazioni germinali nei geni *BRCA1* e *BRCA2*, che spiegano la maggior parte dei tumori ovarici ereditari. È importante sottolineare che tali mutazioni possono essere presenti anche in pazienti senza una chiara storia familiare di neoplasie mammarie od ovariche.
Oltre ai fattori genetici, altri elementi di rischio comprendono condizioni come l'endometriosi e variabili ormonali, tra cui l'età della menopausa.

Il trattamento del carcinoma ovarico si basa su un approccio multimodale, con la chirurgia citoriduttiva come prima linea terapeutica. L'obiettivo dell'intervento è la rimozione completa del tumore macroscopicamente visibile; in questo caso comprendente l'utero e tutti i tessuti interessati dal cancro.
A seguire, in base allo stadio del tumore, è indicata una chemioterapia adiuvante a base di composti del platino, altamente efficace nel determinare una remissione clinica nella maggior parte delle pazienti.
Nonostante la buona risposta iniziale alla chemioterapia, la recidiva è molto frequente, il che sottolinea l'urgenza di sviluppare nuove strategie terapeutiche volte a migliorare la sopravvivenza a lungo termine e la qualità della vita delle pazienti [2].

2. Introduzione dei modelli deep learning

Il *deep learning* rappresenta una branca avanzata dell'intelligenza artificiale, in particolare del *machine learning*. Quest'ultimo si occupa della progettazione e dello sviluppo di algoritmi capaci di apprendere automaticamente dai dati, al fine di effettuare previsioni o prendere decisioni senza essere esplicitamente programmati per ogni compito. Tali algoritmi si basano su modelli matematici che, utilizzando dati in ingresso (anche detti dati di *training*) apprendono schemi e relazioni utili all'elaborazione autonoma delle informazioni.

Il *deep learning*, analogamente al *machine learning*, si basa su un insieme di algoritmi progettati per acquisire, elaborare e interpretare i dati. Tuttavia, si differenzia da quest'ultimo per l'impiego di un'architettura gerarchica composta da reti neurali artificiali in grado di gestire autonomamente il processo di analisi dei dati [3].
Queste reti, ispirate all'attività neurologiche del cervello umano, sono composte da milioni di neuroni artificiali interconnessi e distribuiti su molteplici livelli.
Il *deep learning*, quindi, si basa su un'architettura a più livelli organizzata in modo gerarchico, in cui l'elaborazione dei dati avviene in maniera progressiva. Ogni livello della rete neurale riceve in input l'output del livello precedente, permettendo così una trasformazione graduale e gerarchica delle informazioni. Questo processo a cascata consente al sistema di apprendere rappresentazioni complesse dei dati, attraverso un numero elevato di elaborazioni interne, prima che le informazioni siano trasmesse agli strati successivi.

I recenti sviluppi nel campo del *deep learning* hanno avuto un impatto significativo su numerosi settori scientifici.
In particolare, nell'ambito della biologia, il *deep learning* si è rivelato uno strumento estremamente efficace per la gestione e l'analisi di dati genomici complessi. La sua capacità di elaborare grandi volumi di dati permette di ridurre la necessità di selezionare manualmente le caratteristiche rilevanti, automatizzando l'identificazione di pattern e segnali biologici significativi all'interno di vaste banche dati [3].
Anche in ambito oncologico il *deep learning* sta andando ad assumere un ruolo sempre più centrale, soprattutto nell'ambito della prognosi tumorale. In questo contesto, gli algoritmi di *deep learning* possono essere utilizzati per analizzare i dati genomici al fine di valutare la probabilità di sopravvivenza dei pazienti.

Tuttavia, nonostante i progressi significativi, i metodi di *deep learning* e, più in generale, di intelligenza artificiale si trovano ancora in fase di sviluppo. Le previsioni prodotte da questi modelli presentano tuttora limitazioni in termini di accuratezza e affidabilità, specialmente in contesti più complessi e ad alta variabilità.
Ciò nonostante, si prevede che con il continuo avanzamento tecnologico e la ricerca scientifica si possano superare le attuali limitazioni e raggiungere nuovi obiettivi in ambito biomedico e oltre.

3. Introduzione all'esperimento

L'articolo preso in esame ha l'obiettivo di individuare uno strumento predittivo in grado di stimare la prognosi dei pazienti affetti da carcinoma ovarico. Attualmente, i meccanismi alla base della predizione prognostica sono ancora poco studiati e la nostra conoscenza in merito è limitata.
Lo studio ha quindi l'obiettivo di approfondire tali conoscenze, con l'intento di migliorare la gestione clinica del carcinoma ovarico.

Il lavoro è stato pubblicato nel gennaio del 2025 sulla rivista "*Journal of Translational Medicine*" da un gruppo di ricercatori statunitensi con sede in Texas.
I ricercatori hanno utilizzato una rete neurale basata su algoritmi di *deep learning* in ambito oncologico. Il modello è stato addestrato utilizzando immagini istologiche di tessuti tumorali, con l'obiettivo di predire la sopravvivenza dei pazienti. L'addestramento ha incluso un

totale di 773 immagini istologiche ottenute da 335 pazienti affetti da carcinoma sieroso di alto grado (HGSOC).
Al termine della fase di addestramento, le prestazioni del modello predittivo sono state confrontate con i profili immunologici dei pazienti, ottenuti tramite trascrittomica spaziale.
L'analisi ha evidenziato come il modello di *deep learning* sia in grado di riconoscere specifiche caratteristiche del microambiente immunitario tumorale, dimostrandosi efficace nel prevedere la prognosi dei pazienti con carcinoma ovarico.

APPROCCIO SPERIMENTALE

1. Addestramento del modello tramite fivefold cross-validation

La *cross-validation* è una tecnica statistica utilizzata per stimare la capacità predittiva di un modello di *machine learning*, soprattutto quando si dispone di un set di dati limitato. Questo metodo è fondamentale per valutare la qualità di un modello e ottimizzare il rapporto tra accuratezza e costi di raccolta dei dati [4].
In pratica, la *cross-validation* permette di verificare se un modello è adatto all'esperimento specifico, e viene comunemente impiegata per confrontare diversi modelli predittivi.
Può essere definita come una procedura di ricampionamento, che consente di stimare le prestazioni attese del modello su dati non visti, andando quindi a simulare il suo comportamento su dati completamente nuovi rispetto a quelli usati per l'addestramento.

Tra le varie forme di *cross-validation*, la *fivefold cross-validation* è una delle più semplici ed efficaci. Consiste nel mescolare tutti i dati disponibili e suddividerli casualmente in cinque sottoinsiemi di dimensioni equivalenti. In ciascun ciclo di allenamento, quattro sottoinsiemi (80%) vengono utilizzati per addestrare il modello, mentre il restante 20% viene impiegato per la convalida [5].
Questa operazione viene ripetuta cinque volte, ruotando ogni volta i vari sottoinsiemi, in modo che ciascun gruppo venga impiegato almeno una volta come set di validazione.
Il risultato finale si ottiene calcolando la media delle prestazioni registrate nei cinque cicli, fornendo così una stima più affidabile e generale dell'efficacia del modello.

Per ridurre ulteriormente l'effetto della variabilità dovuta alla casualità nella suddivisione dei dati e nel processo di apprendimento, l'intera procedura può essere ripetuta più volte [4]. In questo modo si ottiene una valutazione ancora più robusta e si minimizza il rischio di *overfitting* legato a una particolare suddivisione dei dati.

In questo studio, la *fivefold cross-validation* è stata applicata per valutare la capacità del modello di distinguere tra due classi: presenza o assenza di cancro.
I dati utilizzati provengono dal database pubblico "*The Cancer Genome Atlas (TCGA)*", dal quale sono stati selezionati pazienti di diverse fasce

di età affetti da cancro ovarico in stadio III e IV. Dopo aver raccolto le immagini, queste sono state mescolate e suddivise casualmente in cinque sottoinsiemi: quattro utilizzati per addestrare il modello e uno riservato alla convalida.
Le immagini sono state ulteriormente pre-elaborate attraverso operazioni di mescolamento, rotazione e capovolgimento, per aumentare la variabilità e ridurre il rischio di *overfitting*.
Per ciascun gruppo di dati, sono stati addestrati 5 modelli distinti, ognuno sottoposto a un processo di *training* e *validation* per 10 volte. Questo approccio ha permesso di selezionare, in modo deterministico, il modello con le migliori prestazioni, identificato tramite la capacità di distinguere efficacemente tra le due classi. In particolare, è stato privilegiato il modello con il miglior rapporto tra veri positivi e falsi positivi.
La validità del modello selezionato è stata successivamente verificata su un set di dati indipendente, non impiegato nelle fasi di *training* e *validation*, per confermare la sua capacità di generalizzazione.
Infine, le prestazioni di tutti i modelli selezionati sono state aggregate calcolando la media dei risultati utilizzando la *curva di Kaplan-Meier* e il *log-rank test*. Questi strumenti statistici hanno permesso di stimare le differenze di sopravvivenza tra i gruppi predetti.

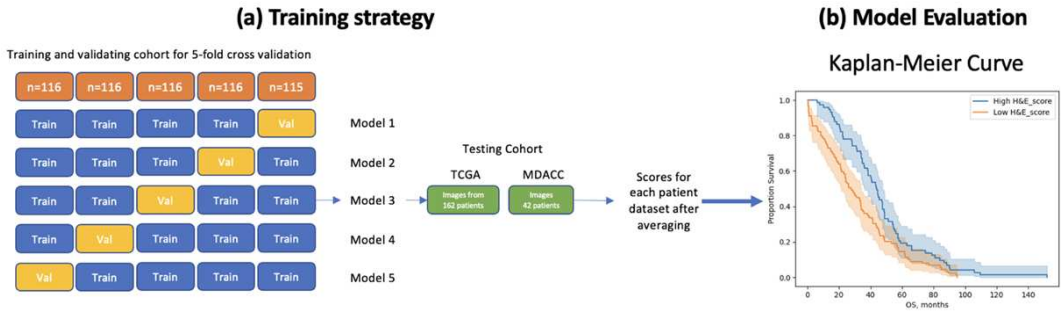

Figura 1: descrizione grafica del modello *fivefold cross-validation* (a); rappresentazione grafica della *curva di Keplan-Meier* (b)

2. Valutazione del modello con Kaplan-Meier Curve

Quando si lavora con i dati di sopravvivenza, uno dei metodi più efficaci per analizzarli è attraverso l'analisi di Kaplan-Meier.
Si tratta di una tecnica statistica che permette di stimare la probabilità che un certo evento si verifichi nel tempo. In altre parole, si può dire che serve a costruire delle curve che mostrano graficamente come varia la

probabilità di sopravvivere rispetto al passare del tempo, facilitando così l'interpretazione del fenomeno clinico osservato [6].

Per comprendere a fondo come viene costruita la curva di Kaplan-Meier, bisogna prima chiarire cosa sono i dati di sopravvivenza.
I dati di sopravvivenza si riferiscono al tempo che intercorre tra un punto di partenza ben definito, che spesso coincide con l'inizio dell'osservazione di un individuo, e il verificarsi di un evento clinico, come ad esempio la morte. Tuttavia, questo evento non deve per forza coincidere con la fine della vita dell'individuo, ma può anche trattarsi di qualsiasi altro evento di interesse clinico o sperimentale [7].
L'obiettivo dell'analisi di sopravvivenza, quindi, è quello di descrivere come cambia la probabilità di osservare un determinato evento nel tempo e di confrontare questa probabilità tra diversi gruppi di individui. Per questo motivo, l'approccio di Kaplan-Meier può essere applicato anche al di fuori dell'ambito medico.

Per realizzare una curva di sopravvivenza di Kaplan-Meier sono necessarie due informazioni: il tempo trascorso fino al momento in cui si verifica l'evento di interesse, e una variabile binaria che indica se l'evento è avvenuto o meno. Sulla base di questi dati, il metodo permette di calcolare la probabilità cumulativa di sopravvivenza in diversi momenti temporali.
La probabilità cumulativa esprime la probabilità che un individuo sopravviva un certo numero di giorni dopo l'inizio dell'osservazione [6].
La curva risultante parte dal valore 1, ovvero dalla certezza che tutti gli individui siano vivi o liberi dall'evento al tempo zero, e decresce progressivamente ogni volta che uno degli individui raggiunge l'evento di interesse. Durante il percorso, si possono incontrare due situazioni: un soggetto può effettivamente raggiungere l'evento, oppure può essere censurato. La censura avviene quando non si conosce con esattezza il tempo di sopravvivenza di un soggetto [7].

Osservando il grafico della curva (Figura 1), le linee orizzontali lungo l'asse delle ascisse indicano quanto dura la sopravvivenza in ciascun intervallo temporale. Quando la linea scende, significa che un individuo ha raggiunto l'evento di interesse. Invece, le distanze verticali sull'asse delle ordinate tra i vari segmenti orizzontali rappresentano quanto cambia la probabilità cumulativa di sopravvivenza ad ogni evento.
Anche la forma complessiva della curva fornisce indicazioni utili. Infatti, una curva con tanti piccoli gradini suggerisce che sono stati inclusi un numero elevato di individui, mentre curve con pochi ma più ampi gradini

implica l'utilizzo di un numero minore di soggetti e, quindi, generalmente offrono una stima meno precisa.

Quando si vogliono confrontare le curve di sopravvivenza di due o più gruppi, si utilizza il *log-rank test*. Questo è un test statistico non parametrico che permette di verificare se le differenze tra le curve siano statisticamente significative [6].

In questo studio la curva di sopravvivenza di Kaplan-Meier è stata utilizzata per osservare in modo immediato la probabilità di sopravvivenza dei singoli campioni analizzati.
In particolare, sono stati confrontati i valori di sopravvivenza predetti dai modelli statistici con i valori reali registrati per ciascun paziente, al fine di valutare quanto effettivamente differissero. Per determinare se le differenze osservate tra i vari gruppi fossero statisticamente significative, ci si è avvalsi del *log-rank test*.
I risultati ottenuti indicano che i modelli hanno raggiunto un buon livello di accuratezza. Questo ci suggerisce che tali modelli sono in grado di predire con una certa affidabilità la sopravvivenza degli individui utilizzando esclusivamente le immagini istologiche come base informativa.

3. Struttura del network del modello

Il network di deep learning utilizzato in questo esperimento è stato realizzato grazie a PyTorch, una libreria di machine learning che unisce efficienza e semplicità d'uso, e che adotta il linguaggio Python come stile di programmazione.
I due modelli neurali adoperati sono ResNet101 e il meccanismo di attenzione.

Le ResNet, ovvero reti neurali residuali, rappresentano una particolare architettura di rete neurale convoluzionale (CNN) profondamente strutturata e ampiamente impiegata in compiti legati all'elaborazione delle immagini.
Nello specifico ResNet101 è composta da 101 strati convoluzionali, progettati per estrarre caratteristiche importanti da un'immagine data in input. Tali strati operano attraverso l'applicazione di filtri convoluzionali i quali generano mappe di caratteristiche che, poi, vengono trasmesse ai livelli successivi della rete.

Una caratteristica distintiva delle ResNet è la presenza di connessioni di salto, o connessioni residue, che permettono al gradiente di saltare determinati strati. Questa strategia consente di attenuare l'incremento dell'errore di training, che spesso si verifica con l'aumentare del numero di livelli presenti nella rete, favorendo così un apprendimento più stabile ed efficace, con risultati prestazionali superiori.

Il meccanismo di attenzione, invece, trae ispirazione dai processi cognitivi umani, secondo i quali l'attenzione si concentra su informazioni rilevanti escludendo quelle superflue. In ambito computazionale questo si traduce nella capacità del modello di assegnare pesi differenti a diverse porzioni dell'input, privilegiando le informazioni più significative [8].
Tale approccio si è rivelato estremamente utile in numerosi contesti applicativi, trovando però particolare affermazione nella medicina di precisione. In questo ambito, l'attenzione ha dimostrato di migliorare l'efficienza e la precisione nell'elaborazione delle informazioni, contribuendo in modo determinante alla comprensione dei meccanismi patologici e allo sviluppo di strategie terapeutiche personalizzate.

Le immagini istologiche utilizzate in questo esperimento sono state sottoposte all'elaborazione da parte dei modelli ResNet101 e del meccanismo di attenzione. La scelta di questi due modelli è stata motivata dal desiderio di verificare la loro capacità di individuare la posizione del tumore all'interno dei tessuti analizzati.

In una prima fase, le immagini sono state introdotte nel modello ResNet101 e fatte transitare attraverso quattro livelli specifici della rete. Da ciascuno di questi livelli è stato estratto un output, ottenendo così un totale di quattro risultati intermedi. Questi sono stati successivamente suddivisi in due gruppi: il primo contenente gli output dei livelli 1 e 2, il secondo quelli dei livelli 3 e 4.
A questo punto è stato integrato il meccanismo di attenzione. Nello specifico, gli output dei livelli 1 e 2 sono stati utilizzati come input per un primo sottosistema basato sull'attenzione, mentre quelli dei livelli 3 e 4 sono stati assegnati a un secondo sottosistema analogo.
L'introduzione del meccanismo di attenzione in questa fase si è rivelata particolarmente importante in quanto consente di incrementare la precisione dei modelli di deep learning, permettendo loro di concentrare l'analisi sulle regioni dell'immagine istologica maggiormente rilevanti ai fini predittivi.

Successivamente, gli output generati dai due sottosistemi di attenzione, insieme all'output finale proveniente da ResNet101, sono stati combinati per formare un unico prodotto. Questo insieme è stato elaborato attraverso tre ulteriori livelli detti "*fully connected layers*", al termine dei quali è stato applicato il metodo SoftMax per individuare la previsione ottimale.
Infine, i risultati ottenuti sono stati analizzati per verificare l'eventuale presenza di correlazioni significative con il profilo immunitario dei soggetti esaminati.

4. Analisi del profilo immunitario

Per comprendere il meccanismo attraverso cui il modello ha generato le sue previsioni, i ricercatori hanno analizzato la relazione tra le regioni evidenziate dal modello e l'attività immunitaria all'interno del tumore.

Una volta ottenuti i risultati dal modello, le immagini sono state suddivise in due gruppi distinti in base ai punteggi di sopravvivenza, determinati attraverso il *p-value* del *test log-rank*: un gruppo ad alto punteggio e un gruppo a basso punteggio.
Per entrambi i gruppi, i profili immunitari sono stati valutati tramite GSEA e la collezione c7 del *database* MSigDB.

GSEA (*Gene Set Enrichment Analysis*) rappresenta un metodo analitico fondamentale per l'interpretazione dei dati di espressione genica, in quanto consente di identificare se l'insiemi di geni predefiniti presentano variazioni coordinate e statisticamente rilevanti tra due condizioni biologiche differenti [9].
Questo tipo di analisi si avvale del supporto del *database* MSigDB (*Molecular Signatures Database*), una risorsa che raccoglie migliaia di set genici eterogenei. MSigDB è considerato uno dei *database* più completi e ampiamente utilizzati per le analisi di arricchimento genico [10].
I geni contenuti in MSigDB sono organizzati in nove collezioni principali, tra cui la collezione c7, che raccoglie geni associati a profili immunitari e che, in questo studio, rappresentano i set utilizzati per l'analisi [9].

I profili immunitari appartenenti alla collezione c7 sono stati confrontati tra i due gruppi mediante test GSEA, e sono stati selezionati i risultati più significativi. La selezione è stata condotta valutando la correlazione tra le mappe di attenzione del modello e i profili immunitari.

Questa correlazione è stata calcolata utilizzando il *coefficiente di correlazione di rango di Spearman*, una misura statistica non parametrica impiegata per valutare l'associazione tra due variabili ordinali.
Attraverso i dati di espressione genica, è stata, quindi, analizzata la risposta immunitaria nei tessuti tumorali, evidenziando il fatto che le aree su cui il modello concentrava la propria attenzione erano associate a specifiche attività immunitarie.
Tali risultati suggeriscono che il modello è in grado di predire la prognosi dei pazienti affetti da carcinoma ovarico grazie alla sua capacità di riconoscere determinati segnali immunitari all'interno del microambiente tumorale.

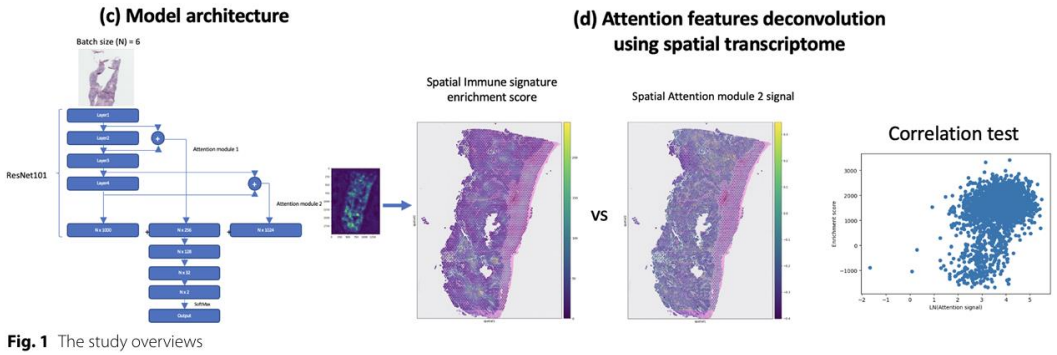

Figura 2: struttura del modello (c); confronto tra il profilo immunitario delle immagini istologiche e la loro analisi data dal modello di attenzione (d)

RISULTATI E DISCUSSIONE

1. Validazione e verifica del modello

In questo studio è stato impiegato un modello di *deep learning* per estrarre e apprendere le caratteristiche contenute nelle immagini istologiche, con l'obiettivo di prevedere la prognosi delle pazienti oncologiche.

Un elemento distintivo di questo approccio è la semplicità dell'architettura del modello. Infatti, grazie all'uso di un meccanismo di attenzione, il sistema è in grado di individuare automaticamente le aree istologiche più rilevanti, analizzando sia le cellule tumorali sia il microambiente circostante. Ciò consente di formulare previsioni senza dover predefinire manualmente una regione di interesse, permettendo al modello di operare in maniera imparziale.
Questa capacità rende possibile effettuare stime prognostiche a partire da immagini con un'elaborazione minima, provenienti sia dalla stessa fonte sia da dataset differenti.
Tale caratteristica potrebbe contribuire a ridurre le barriere tecniche e i costi associati alla patologia digitale.

Nonostante i vantaggi clinici ed economici, il modello presenta alcune limitazioni da superare.
Durante la fase di addestramento, ad esempio, sono stati inclusi esclusivamente pazienti non censurati, ossia solo coloro che possedevano tempi di sopravvivenza noti. Questa scelta, finalizzata a semplificare il processo di training evitando analisi più complesse, ha comunque prodotto buoni risultati: il modello ha mantenuto prestazioni significative anche quando testato su dati misti (censurati e non) provenienti da fonti esterne.
Un'ulteriore limitazione riguarda la composizione demografica del dataset di training. Infatti, la maggior parte delle pazienti coinvolte nell'analisi era di etnia caucasica. Questo potrebbe comportare una maggiore accuratezza del modello su popolazioni simili a quella presa come campione, ma una ridotta precisione verso gruppi etnici differenti, come popolazioni asiatiche o africane.
Per verificarne l'efficacia su scala più ampia, sono quindi necessari ulteriori test su campioni etnicamente diversificati.

Nel contesto della patologia digitale, l'utilizzo dell'intelligenza artificiale e, in particolare, del *machine learning* per applicazioni come la diagnosi oncologica e la previsione della sopravvivenza, è in rapida espansione.

Tuttavia, questi strumenti richiedono ancora approfondimenti, soprattutto per quanto riguarda la trasparenza e l'interpretabilità dei risultati.
L'impiego di modelli basati su reti neurali convoluzionali (CNN) per prevedere la sopravvivenza nelle pazienti con carcinoma ovarico rimane infatti limitato proprio a causa della difficoltà di interpretazione. Per questo motivo, è necessario sviluppare modelli predittivi più accurati e interpretabili, in grado di supportare in maniera più affidabile le decisioni cliniche nella prognosi del carcinoma ovarico.

2. Correlazione tra il meccanismo di attenzione e l'attività immunitaria

Una volta addestrato e stabilizzato, il modello di *deep learning* ha generato risultati che sono stati messi in relazione con i profili immunitari ottenuti tramite trascrittomica spaziale.
Questa correlazione è resa possibile dal fatto che il meccanismo di attenzione è in grado di apprendere informazioni relative alla composizione cellulare e all'attività di segnalazione cellulare, entrambi fattori chiave nella progressione del cancro.
Sfruttando questa capacità, lo studio è riuscito a identificare profili immunitari associati ai punteggi prognostici prodotti dal modello. In particolare, è emersa una correlazione tra i punteggi di sopravvivenza istologica e specifici marcatori immunitari, come la presenza di neutrofili e di linfociti infiltranti il tumore (TIL).
Questi ultimi sono di particolare rilievo nella prognosi del carcinoma ovarico, poiché la loro presenza e distribuzione può contribuire alla classificazione del tumore in base al grado di aggressività.

Nel complesso, i risultati indicano che alcune caratteristiche immunitarie presenti nei campioni tumorali possono essere apprese dal modello di *machine learning* e utilizzate per migliorare la previsione della prognosi nelle pazienti affette da carcinoma ovarico.
Tuttavia, sebbene lo studio dimostri la fattibilità di questa correlazione tra analisi istologica con meccanismo di attenzione e profili immunitari derivati dalla trascrittomica spaziale, sono necessarie ulteriori ricerche e sperimentazioni per confermarne la sua solidità.

In questo studio è stato dimostrato che un modello di *machine learning* dopo essere stato addestrato, validato e testato su immagini istologiche

H&E, è in grado di predire la sopravvivenza di pazienti affette da carcinoma sieroso di alto grado dell'ovaio (HGSOC).
Le previsioni generate dal modello risultano correlate con l'immuno-attività tumorale, evidenziando una coerenza con i dati ottenuti tramite analisi del trascrittoma spaziale.
Questa osservazione suggerisce che il modello è in grado di utilizzare le informazioni immunologiche presenti nelle immagini istologiche per prevedere l'esito clinico delle pazienti.

Il lavoro ha previsto l'addestramento, la validazione e il test di un nuovo modello di *deep learning* con meccanismo di attenzione, applicato a immagini complete di sezioni tumorali H&E.
L'integrazione di questo approccio con le più recenti piattaforme omiche spaziali, come la trascrittomica spaziale, potrebbero consentire la creazione di modelli predittivi ancora più accurati, capaci di fornire approfondimenti sui meccanismi morfologici e immunologici che collegano le caratteristiche del tessuto tumorale al fenotipo maligno della malattia.

Nonostante l'accuratezza metodologica, è necessario condurre ulteriori indagini per valutare l'efficacia del modello in contesti clinici reali, addestrandolo e testandolo su dataset più ampi e diversificati. Solo in questo modo sarà possibile definire con precisione i limiti e le potenzialità di un modello di *deep learning* con meccanismo di attenzione, non solo in ambito oncologico ma anche in altre applicazioni mediche.

BIBLIOGRAFIA

1. Doubeni CA, Doubeni AR, Myers AE. Diagnosis and Management of Ovarian Cancer. *Am Fam Physician*. 2016;93(11):937-944.
2. Konstantinopoulos PA, Matulonis UA. Clinical and translational advances in ovarian cancer therapy. *Nat Cancer*. 2023;4(9):1239-1257. doi:10.1038/s43018-023-00617-9
3. Lee M. Deep Learning Techniques with Genomic Data in Cancer Prognosis: A Comprehensive Review of the 2021-2023 Literature. *Biology (Basel)*. 2023;12(7):893. Published 2023 Jun 21. doi:10.3390/biology12070893
4. Chang-Xue Jack Feng, Zhi-Guang Samuel Yu, Unnati Kingi, M. Pervaiz Baig, Threefold vs. fivefold cross validation in one-hidden-layer and two-hidden-layer predictive neural network modeling of machining surface roughness data, *Journal of Manufacturing Systems*, 2005, ISSN 0278-6125, https://doi.org/10.1016/S0278-6125(05)80010-X.
5. Xin Yang, Ning Zhang, Paul Schrader, A study of brain networks for autism spectrum disorder classification using resting-state functional connectivity, *Machine Learning with Applications*, 2022, ISSN 2666-8270, https://doi.org/10.1016/j.mlwa.2022.100290.
6. Graziella D'Arrigo, Daniela Leonardis, Samar Abd ElHafeez, Maria Fusaro, Giovanni Tripepi, Stefanos Roumeliotis, Methods to Analyse Time-to-Event Data: The Kaplan-Meier Survival Curve, Oxidative Medicine and Cellular Longevity, 2021, https://doi.org/10.1155/2021/2290120
7. Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg*. 2010;143(3):331-336. doi:10.1016/j.otohns.2010.05.007
8. Zhaoyang Niu, Guoqiang Zhong, Hui Yu, A review on the attention mechanism of deep learning, *Neurocomputing*, 2021, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2021.03.091.
9. UC San Diego and Broad Institute, Overview, Gene Set Enrichment Analysis (GSEA), https://www.gsea-msigdb.org/gsea/index.jsp (visitato il 25/07/2025)

10. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417-425. doi:10.1016/j.cels.2015.12.004

Ng et al. Journal of Translational Medicine (2025) 23:113
https://doi.org/10.1186/s12967-024-06007-8

Journal of
Translational Medicine

RESEARCH

Open Access

Spatial transcriptome reveals histology-correlated immune signature learnt by deep learning attention mechanism on H&E-stained images for ovarian cancer prognosis

Chun Wai Ng[1], Kwong-Kwok Wong[1], Barrett C. Lawson[2], Sammy Ferri-Borgogno[1*] and Samuel C. Mok[1*]

Abstract

Background The ability to predict the prognosis of patients with ovarian cancer can greatly improve disease management. However, the knowledge on the mechanism of the prediction is limited. We sought to deconvolute the attention feature learnt by a deep learning convolutional neural networks trained with whole-slide images (WSIs) of hematoxylin-and-eosin (H&E)–stained tumor samples using spatial transcriptomic data.

Methods In this study, 773 WSIs of H&E-stained tumor sections from 335 patients with treatment naïve high-grade serous ovarian cancer who were included in The Cancer Genome Atlas (TCGA) Pan-Cancer study were used to train, and validate, and to test a ResNet101 CNN model modified with attention mechanism. WSIs from patients in an independent cohort were used to further evaluate the model.

Results The prognostic value of the predicted H&E-based survival scores from the trained model on patient survival was evaluated. The attention signals learnt by the model were then examined their correlation with immune signatures using spatial transcriptome. After validating the model with the testing datasets, pathway enrichment analysis showed that the H&E—based survival score significantly correlated with certain immune signatures and this was validated spatially using spatial transcriptome data generated from ovarian cancer FFPE samples by correlating the selected signature and attention signal.

Conclusions In conclusion, attention mechanism might be useful to identify regions for their specific immune activities. This could guide future pathological study for the useful immunological features that are important in modulating the prognosis of ovarian cancer patients.

Keywords Attention, Deep learning, H&E, Immune signature, Ovarian cancer, Prognosis, Spatial transcriptome

*Correspondence:
Sammy Ferri-Borgogno
sferri@mdanderson.org
Samuel C. Mok
scmok@mdanderson.org
Full list of author information is available at the end of the article

© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Background

Advanced ovarian cancer, which has a 5-year survival rate of less than 30%, is the deadliest among gynecologic cancers. Most ovarian cancer is high-grade, serous ovarian cancer (HGSOC), and the poor survival rate among patients with the disease is mainly due to the fact that the disease is usually diagnosed at a late stage [1, 2]. The standard treatment for HGSOC is cytoreductive surgery before platinum-based chemotherapy or after neoadjuvant chemotherapy [3]. However, patient treatment response and survival duration decrease with ages and disease stages [4–6]. In addition, most of the HGSOC patients relapse because they develop resistance to taxane-based chemotherapy. Nevertheless, 15% of patients diagnosed with advanced HGSOC have overall survival (OS) durations of more than 10 years despite the development of recurrent diseases. Discovering predictive markers for prognosis could help researchers to identify therapeutic targets for the disease, thus improving patient survival rates and disease management.

Models that can predict survival in patients with advanced HGSOC have been described recently [7–11], but they have limited performance and usually require transcriptomic analysis or tedious image processing, which can be time consuming and costly, especially in regions with limited resources. Thus, a cost-effective and interpretable method for the prediction of ovarian cancer survival is urgently needed for both patients and clinicians. Such a method would improve treatment decision-making and disease management, especially for patients with elevated risks of poor outcomes [12].

Recent advancements in machine-learning algorithms for computer vision have created an interest in their applicability in digital pathology [13, 14]. Deep learning models trained on data such as histological and computed tomography images have been used to predict signaling activity, mutation and prognosis [15–18]. Although ovarian tumor histological images have been employed to predict patient prognosis, the mechanism of the prediction is not fully understood. The understanding the features learned by the model would enhance the confidence in applying image-based predictive model in clinical setting, and also aid the pathological research of the disease in the future.

In this study, WSIs of H&E-stained tumor sections and clinical data from The Cancer Genome Atlas (TCGA) ovarian cancer dataset were employed to train a model that can be used to predict the survival of HGSOC patients. Validation studies using WSIs of H&E-stained tumor sections obtained from an independent patient cohort from The University of Texas MD Anderson Cancer Center (MDACC) were also performed to further determine the performance of the predictive model. The immunological signatures correlated with the model were determined and validated using spatial transcriptome data of ovarian cancer FFPE samples.

Methods

Image and clinical data preparation for model training and validation

WSIs of H&E-stained tumor sections from treatment naïve advanced HGSOC in the TCGA-OV dataset were downloaded from the Genomic Data Commons (GDC) portal using GDC client [19]. TCGA's clinical data for patients with HGSOC were downloaded from the GDC portal and from the cBioportal (TCGA-OV Pan-Cancer) dataset (Accessed on October 1, 2023) [20, 21]. The cBioportal and GDC data were matched using TCGA patient IDs and then merged. Images from patients of all age and stage III-IV (i.e., images from patients of stage I-II were excluded) were included for training. All the patients were female. The characteristics and demographics of the included patients are shown in Table 1. The median OS and Progression-free survival (PFS) durations were 35 and 15 months, respectively. Most patients were white, and the median age at diagnosis was 60. Only patients with clinical stage III or IV disease were selected. The patients without this information were classified as

Table 1 The clinical characteristics of the patients included in the TCGA dataset

Variable	N = 335
Overall survival in months, median	35
Progression-free survival in months, median	15
Race, n	
American Indian or Alaskan Native	3
Asian	6
Black or African American	19
Native Hawaiian or other Pacific Islander	1
White	301
Not available	5
Age at diagnosis in years, median	60
Residual disease, n	
No Macroscopic disease	39
1–10 mm	171
11–20 mm	24
> 20 mm	73
Not available	28
Clinical stage, n	
I	0
II	0
III	267
IV	68

TCGA, The Cancer Genome Atlas

having stage III disease so that they could be included in this study.

Model training with the fivefold cross-validation process and testing

TCGA images included in this study were first separated into training and validation (n = 579), and testing (n = 194) datasets (Fig. 1a). The training and validation dataset was then separated into 5 folds for training (n = 463/464) and validation (n = 116/115) [22]. Models for each fold were trained 5 times for each of the 10 epochs to select the model with the best area under the receiver operating characteristic (AUROC) values, which were then evaluated. After the top models in each fold were selected, the models were then evaluated with the blind testing TCGA and MDACC datasets (Fig. 1b). WSIs of H&E-stained treatment-naïve tumor sections and clinicopathological characteristics from the MDACC dataset were obtained from the ovarian cancer repository of the Department of Gynecologic Oncology and Reproductive Medicine under protocols approved by the University of Texas MD Anderson's Institutional Review Board. Written informed consent from the patients were obtained by front desk personnel, and the studies were conducted in accordance with recognized ethical guidelines. TCGA data were obtained from public repository and did not require ethical approval.

The fivefold cross-validation process of model training was described (a). The training images were separated into 5 training and validation sets, each with different validation images. From each fold, a model was selected with the best AUROC evaluated using validation images. The models trained were then evaluated with datasets from the testing cohort by averaging the score outputs from the 5 models using Kaplan–Meier curve and log-rank test (b). Training and prediction were performed by feeding images (1024 × 1024 resolution; batch size [N] = 6) into the pretrained ResNet101 CNN model for training (c). The output of layers 1 and 2 of the pretrained ResNet101 model was extracted, as was the output of layers 3 and 4, and the resultant information was used as the attention mechanism. The output of the ResNet101 model was concatenated with the output of attention modules 1 and 2, and was followed by 3 additional layers of fully connected neural networks for a final output (N × 2) after SoftMax processing. The attention features generated by attention module 2 were investigated for their correlation with immune signature enrichment score (d). Abbreviations: MDACC, The University of Texas MD Anderson Cancer Center; TCGA, The Cancer Genome Atlas; Train, training; Val, validation.

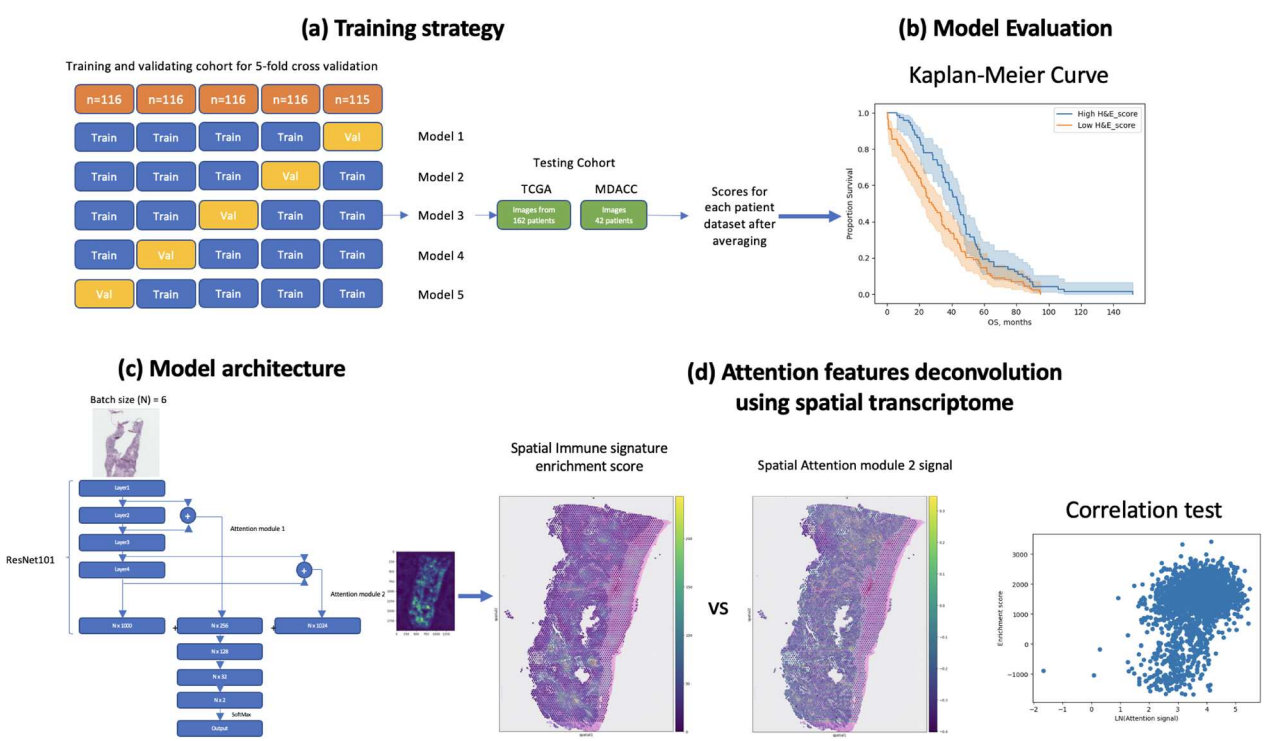

Fig. 1 The study overviews

The architecture of the deep learning model with an attention mechanism

The deep learning network was constructed with the PyTorch [23] framework in Python. An overview of the model is shown in Fig. 1c. The input training batch size (N) was 6 images. The Layers 1 and 2 of the pretrained ResNet101 model from PyTorch were interpolated, as were layers 3 and 4. The interpolated layers from the 2 attention modules were processed with AvgPool2d to create a flattened layer (N × 256 and N × 1024, respectively) and were fed into the fully connected neural network layers together with the output of the model (N × 1000) for attention mechanism. The N × 2280 layer was followed by 3 fully connected layers with 128, 32, and 2 perceptrons, respectively. Each of the layers (except the final layer) was followed by a dropout with rate of 0.2 and was activated with a rectified linear unit. The final layer was processed with SoftMax to produce the final output. The losses of between the target and predicted value during training were determined using FocalLoss, and the Adam algorithm was used as an optimizer. The learning rate was 0.0002, and the learning-rate decay was 0.1 for every 7 epochs.

Spatial immune signature enrichment analysis

The predicted H&E—based survival scores of the images of the TCGA testing cohort were separated into 2 groups (low and high score) by step-wise experiment for the lowest log-rank p-value. The 2 groups were then determined for statistics for the MSigDB c7 immune signatures by GSEA. The most significant signature was selected for the correlation study with the attention signal using spatial transcriptome data of ovarian cancer FFPE samples (Fig. 1d) (https://www.10xgenomics.com/datasets?query=&page=1&configure%5BhitsPerPage%5D=50&configure%5BmaxValuesPerFacet%5D=1000&refinementList%5Bspecies%5D%5B0%5D=Human&refinementList%5BanatomicalEntities%5D%5B0%5D=Ovary&refinementList%5Bplatform%5D%5B0%5D=Visium%20Spatial&refinementList%5BpreservationMethods%5D%5B0%5D=FFPE, accessed on October 1, 2023).

Image processing and augmentation

The original images were divided into 2 square images (resolution, 1024 × 1024) if their initial widths and lengths were not the same (Supplementary Fig S1a). Both images were fed into model training, whereas only 1 of the images was used for validation and testing. During training, the order of the images was shuffled, normalized, augmented with vertical or horizontal flipping, rotated, and affined randomly. Two examples of image augmentation are shown in Supplementary Fig S1b.

Statistical methods

The AUROC was determined using the methods in Python's scikit-learn library (version 1.2.2) [24, 25]. Spearman correlation test was determined by Python library Scipy [26]. Kaplan–Meier survival curves and log-rank test were drawn using the methods in Python's library lifelines [27]. GSEA statistics and ssGSEA enrichment score were determined by GSEApy (version 1.1.1) [28]. Hazard ratio and Chi-square test were determined with GraphPad Prism (version 10.1.1). Spatial cell clustering was performed by Scanpy [29] and Squidpy [30]. The *p*-value below 0.05 and false discovery rate (FDR) below 0.25 were considered statistically significant.

Results

The training, validation, and testing process for the fivefold cross-validation process

WSIs of H&E-stained HGSOC tumor sections were downloaded from TCGA. The OS data for patients with stage III or IV cancer were used to train the model (target = 1 for patients with OS durations ≥ 36 months; target = 0 for patients with OS durations < 36 months). To obtain the accurate OS durations required for training, and to simplify the training process, uncensored patients from TCGA were selected for this study (773 images from 335 patients). The 773 images of the 335 patients were randomly split into a training and validation dataset for the fivefold cross-validation process and a testing dataset as described in methods. The metadata of the training, validating, and testing TCGA images are shown in Supplementary Table S1. To take into account the randomness of the training process, and to select the best possible model, 5 trainings were done for each fold, each fold with 10 epochs. The models with the best AUROCs, as evaluated with the validation datasets, were selected for each fold. The final H&E-based survival scores were calculated from the average scores of the selected 5 models from the 5 folds. The H&E-based survival scores, together with patient ages, were then evaluated using the TCGA training dataset, the TCGA testing dataset, and the MD Anderson Cancer Center testing dataset for prognosis prediction.

The evaluation of the deep learning model using TCGA data

After training the model with the fivefold cross-validation process, the best AUROC values for each fold, which were obtained using validation images, were 0.705, 0.687, 0.628, 0.703, and 0.759. First, for the patients in the training/validating and testing dataset, we evaluated the effects of the ages at diagnosis and H&E-based survival scores on OS and PFS durations.

The effects of the patients' ages at diagnosis and H&E-based survival scores on survival were evaluated using Kaplan–Meier curves and the log-rank test (Fig. 2). For the patients in the training/validation dataset, the age at diagnosis was not prognostic for both OS and PFS duration (Fig. 2a, OS, p-value = 0.169; Fig. 2c, PFS, p-value = 0.1303), but it was prognostic for the patients in the testing dataset (Fig. 2e, OS, p-value = 0.0009; Fig. 2g, PFS, p-value = 0.0108). The H&E-based survival scores from the trained model were significantly prognostic for both the patients in the training and validation dataset (Fig. 2b, OS, p-value < 0.0001; Fig. 2d, PFS, p-value = 0.0032) and those in the testing dataset (Fig. 2f, OS, p-value = 0.0045; Fig. 2h, PFS, p-value = 0.0048).

We also evaluated if the model predicted patient prognosis based on other well-known prognostic covariate such as debulking status. We demonstrated that the AUROC of the model in predicting less optimal debulked (> 10 mm residual tumor) patients were 0.52 (data not shown). This suggests that the model did not predict patient prognosis based on debulking status.

The prognostic values of the age at diagnosis and H&E-based scores for OS and PFS duration were compared for patients with HGSOC. Kaplan–Meier curves and log-rank test results are shown for both the TCGA training/validation dataset (a, b, for OS; c, d, for PFS) and the testing dataset (e, f, for OS; g, h, for PFS). The panels show the results for the age at diagnosis (a, c, e, g) and the predicted H&E-based survival scores (b, d, f, h). Abbreviations: H&E, hematoxylin and eosin; HGSOC, advanced, high-grade, serous ovarian cancer; OS, overall survival; PFS, progression-free survival; TCGA, The Cancer Genome Atlas.

Evaluation of the model using the MDACC dataset

The MDACC dataset, which consisted of 42 patients with HGSOC, was used for further evaluation of the deep learning model with patient characteristics shown in Supplementary Table S2. Images of H&E-stained patient tumors were scanned from H&E slides made from formalin-fixed, paraffin-embedded ovarian tumor blocks prepared from treatment naïve patients. As was done with the TCGA datasets, the MDACC dataset was used to evaluate the deep learning model. AUROC, Kaplan–Meier curves and log-rank tests were used to correlate the output scores and the patients' 5-year OS durations.

Using the AUROC method, we determined the performance of the model to predict patients with 5-year overall survival. The fivefold models had the AUROC results of 0.720, 0.686, 0.711, 0.736, and 0.707, and with AUROC 0.73 after averaging the 5 scores (Fig. 3a), which indicates that the model could predict the prognosis of the MDACC ovarian cancer patients. The Kaplan–Meier curve was shown (Fig. 3b) with the most significant log-rank test result (p-value = 0.0047) with a cut-off score of 0.448.

We also evaluated the performance of the model with or without stage I–II patients and similar results were obtained (data not shown), suggesting that the model did not predict prognosis based on stage information associated with the images.

Images of H&E-stained tumor sections obtained from the MDACC tumor bank were predicted for H&E-based

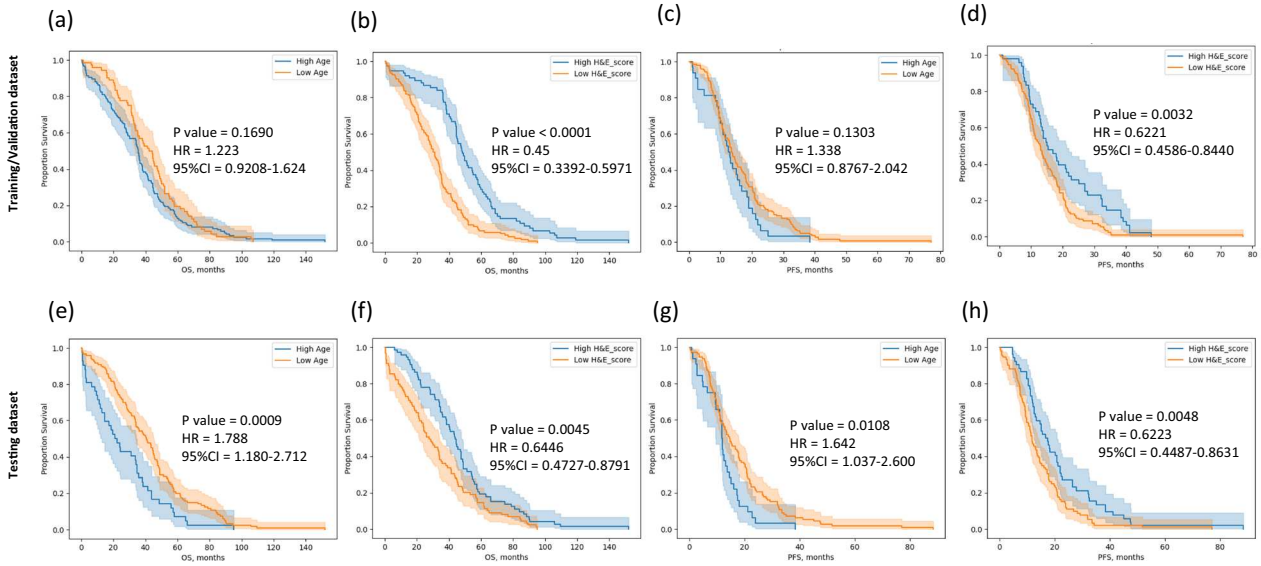

Fig. 2 Kaplan–Meier curves demonstrating that the deep learning model was prognostic for OS and PFS durations in patients with HGSOC

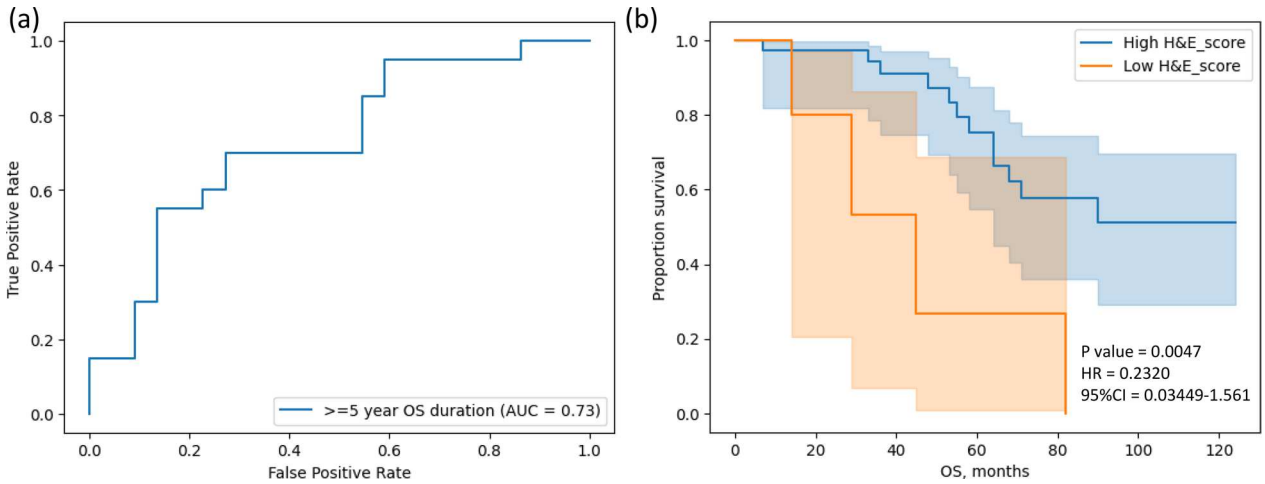

Fig. 3 The evaluation of the deep learning model with the MDACC testing dataset

survival score. The predicted H&E—based survival scores were evaluated for the 5-year survival prediction using AUROC (a) and OS durations using Kaplan–Meier curves and the log-rank test (b). Abbreviations: H&E, hematoxylin and eosin; MDACC, The University of Texas MD Anderson Cancer Center; OS, overall survival.

Image features emphasized by the attention mechanism of the deep learning model

As described, the model trained on WSIs had an attention mechanism; such mechanism could improve the accuracy of the deep learning model, greatly enhancing its interpretability to help researchers better understand its decision-making and underlining mechanisms used to determine disease progression through both cancer cells and tumor microenvironment. However, the attention features learnt in pathological images are usually uninterpretable. We therefore employed spatial transcriptome to deconvolute the attention features.

The output of attention module 2 was extracted and overlaid with original images to form density maps. Red and blue coloring indicated regions with higher and lower importance, respectively, for the decision-making of the model (Fig. 4). Notably, the red regions fell mainly on the tumor tissue in both training and testing images instead of the blank areas. This indicated that the model had been trained well and performed its predictions using the features of the tumor region. Notably, immune

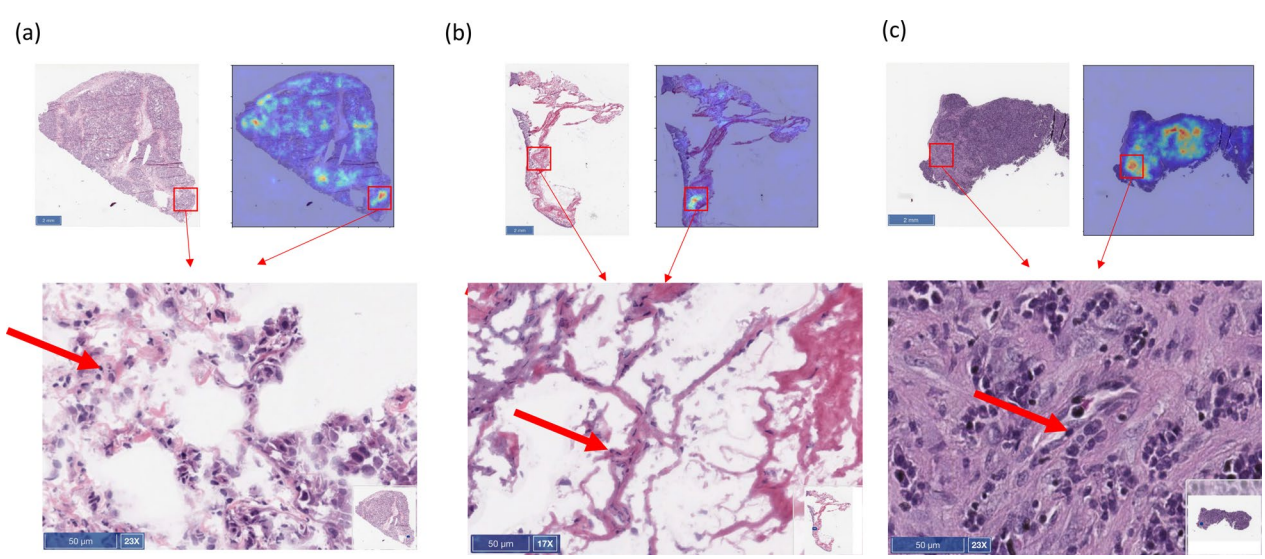

Fig. 4 Image features emphasized by the attention mechanism of the deep learning model

infiltrates were seen in the area with high attention signal. We then interrogated the correlation between the attention signal and the immune signature.

The interpretability of the model was demonstrated by the intensity of the regions using the attention mechanism to examine the (a) TCGA training and validation, (b) TCGA testing, and (c) MDACC H&E-stained tumor images used in this study with red arrow highlighting immune cell infiltrations. Abbreviations: H&E, hematoxylin and eosin; MDACC, The University of Texas MD Anderson Cancer Center; TCGA, The Cancer Genome Atlas.

Immuno-signature enrichment analysis reveals correlation between attention signal and immune activity

To explore the underlying mechanism by which the signatures that the model learnt can predict the prognosis of the patient samples, we performed a pathway enrichment analysis to evaluate the differential immunological pathway activation between TCGA testing samples with low and high H&E—based survival score. By employing the gene expression data from cBioportal, GSEApy and Molecular Signature Database (MSigDB) [31–33], pathway enrichment analyses on testing TCGA dataset using the c7 immunological pathway gene set collection was performed. The samples were first labelled as high if their scores were higher than the median of the predicted H&E—based survival scores. The c7 immune signatures were then compared using GSEA test between the two groups, and the most significant signatures were selected for further analysis, and the results are shown in Supplementary Table S3. The enrichment scores of the 9 significant signatures were shown (Fig. 5a). The top signature GSE37416_0H_VS_48H_F_TULARENSIS_LVS_NEUTROPHIL_UP (Fig. 5b) was further validated for its relationship with the attention signal. The heatmap for the genes of the GSE37416_0H_VS_48H_F_TULARENSIS_LVS_NEUTROPHIL_UP immunological signature is shown in Supplementary Fig S2. Spatial transcriptome data of an ovarian cancer FFPE samples downloaded from 10X genomic were employed to investigate the relationship between the attention signal and immune signature. By using the enrichment scores for each spot in the samples as determined by the ssGSEA method, and integrating the attention signal detected from the whole H&E image of the spatial transcrioptome samples as spatial prognostic information of the samples (Fig. 5c), their correlation was determined by Spearman correlation test. We focused on the tumor cell cluster regions as these regions should have more prognostic information. Results showed that the attention signal in the tumor regions of the two ovarian cancer spatial

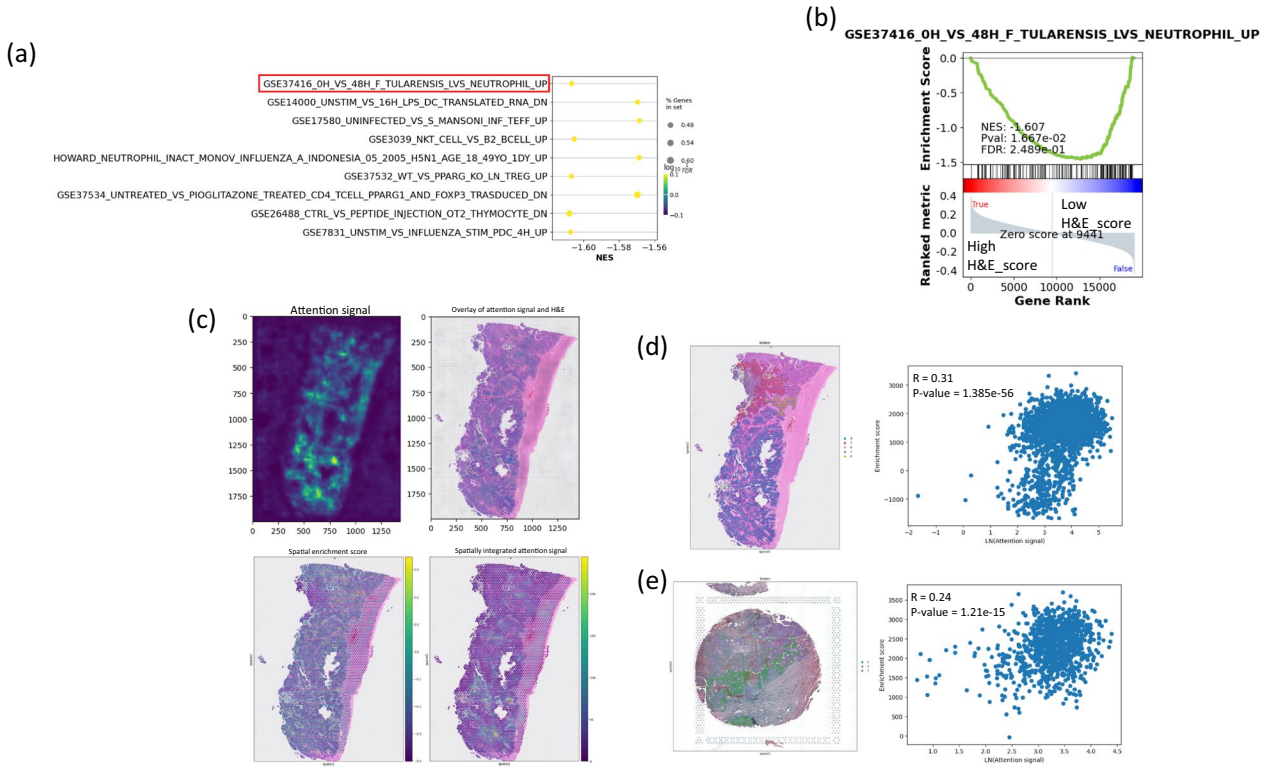

Fig. 5 Immuno-signature enrichment analysis reveals correlation between attention signal and immune activity

transcriptome samples (Fig. 5d and e) were significantly correlated with the enrichment score (R = 0.31/0.24, P-value = 1.385e−56/1.21e−15). This suggests that the model predict the prognosis of the ovarian cancer patients by detecting specific types of immune activity.

The pathway enrichment analysis of the testing samples with low and high H&E—based survival score was done using GSEApy and c7 immunological signature gene sets (n = 5219) from MSigDB. The significant gene sets that were most significantly enriched are shown with the first 3 most significant signatures highlighted (a). The statistical results of the highlighted pathways GSE37416_0H_VS_48H_F_TULARENSIS_LVS_NEUTROPHIL_UP (b). The attention signal integration and GSVA enrichment score of the signature GSE37416_0H_VS_48H_F_TULARENSIS_LVS_NEUTROPHIL_UP of two spatial transcriptomic ovarian cancer FFPE samples were downloaded from 10X genomics. An example of attention signal and spatial transcriptome integration is shown (c), and the correlation of the enrichment score and the natural logarithm of attention signal of the tumor cell region the spatial FFPE sample (d, e) are shown.

Discussion

The use of machine learning for applications such as cancer diagnosis and outcome prediction is growing in the field of pathology; however, the application of CNN models using H&E WSIs developed for outcome prediction in ovarian cancer patients has been limited due to their interpretability. There is an unmet urgent need in developing more efficient and interpretable prediction models for prognosis in patients with ovarian cancer. Such models would also form a base for the future research (e.g., the development of multimodality prediction models). In this study, we demonstrated that a machine-learning model trained, validated, and tested on H&E WSIs can predict survival in patients with HGSOC. The results of the model were found to be correlated with immunoactivity by integrating with spatial transcriptome analysis. This finding suggests that our model predicts clinical outcome with immunological information contained in images of H&E-stained tumor sections.

Deep learning model has been employed for learning the histopathological features associated with prognosis of cancer patients. However, the interpretability of deep learning is challenging especially when it is applied in clinical settings. In this study, to interpret the features that have been learnt by the model, an analytical method was performed to unravel the features the model trained to differentiate patient survival. As cell composition and cell signaling activity play crucial parts in cancer progression, and it is known that they can be learnt by deep learning model [15, 34], we tried to reverse this process by identifying the immunological signatures that are correlated with the predicted scores of the model. By this method, we identified immunological signatures that are related to the predicted histological survival scores.

Tumor-infiltrating lymphocytes have been found to be important in the prognosis of ovarian cancer patients. With histological images, ovarian cancer can be separated into groups with different risk by identifying TILs [35]. Different subtypes of TILs in ovarian cancer were also found to differentially affect the progression of ovarian cancer [36, 37]. From the pathway enrichment results in this study together with the spatial transcriptomes, we did a novel analysis to determine their relationship. We found that the histology-based output score of our model correlated with enriched immune signature associated with neutrophil [38], and other immune cell types as shown in Fig. 5a. Among them, the association with neutrophils is the most significant. The top gene in the enriched pathway associated with neutrophil is OASL, which has been shown to play a role in neutrophil recruitment..Studies also revealed its role in chemoresistance through T cell suppression [39–41]. Taken together, these findings suggest that certain immune features of the tumor sample could be learnt by the deep learning model to predict the prognosis of ovarian cancer patients.

One special aspect of the model described in this study is its simplicity. Via its attention mechanism, the model could select the most relevant tumor histology, including both the cancer cells and the tumor microenvironment, to make a prediction in an unbiased manner without being taught to look for a specific region of interest. As a result, this model predicted prognostic outcomes with minimal image processing. It is also noteworthy that, although the H&E images from the MDACC dataset and public spatial sample were inevitably slightly different from the images from the TCGA, this model generalized well when presented with images from different sources. This characteristic could allow the model to greatly reduce the technical barriers and costs of digital pathology.

Although this study has presented a model that may assist in predicting the prognosis of patients with HGSOC, it has several limitations. First, only uncensored patients were included. Since the model was trained using patients' confirmed OS durations to simplify the model-training process, FocalLoss was used to calculate the resultant errors. While some would argue that this would lead to erroneous results due to the loss of information, the percentage of censored patients in the TCGA- ovarian cancer (OV Pan-Cancer) dataset was relatively low (37%), and the use of censored data could also potentially lead to biases due to the uncertain information within them [42]. Nevertheless, the samples of the

MDACC dataset included both censored and uncensored data and showed significant results. Second, the majority of patients included in the training data were white, this means that the model could potentially perform better on white patients. Further investigation should be performed with dataset from patients with different ethnic background. Third, even though the attention mechanism highlights areas that might be linked to adaptive immune signatures, the correlation between those area and immune cell subtypes needs further investigation. Finally, the TCGA H&E images were split by images rather than patients in order to maximize the available images for training and testing.. Although this might result in information leakage, our evaluation results together with our own dataset and independent spatial transcriptome sample showed consistent results for both prognosis and pathway enrichment prediction.

In conclusion, we trained, validated, and tested a novel deep learning model with an attention mechanism using WSIs of H&E-stained tumor sections from patients with HGSOC. With the advancements of spatial omics platforms such as spatial transcriptomics [43], H&E-based predictive model can be integrated with these platforms to generate a prediction model with higher performance, and to provide insights into the morphological and immunological mechanism by which immunological features in tumor tissue link to the malignant phenotype of the disease. Further investigation into the clinical application of the model will need to be done by training and evaluating a full model with the whole dataset from more diverse patients.

Abbreviations

FDR	False discovery rate
GDC	Genomic Data Commons
GSEA	Gene set enrichment analysis
H&E	Hematoxylin and eosin
HGSOC	High-grade serous cancer
MDACC	The University of Texas MD Anderson Cancer Center
MSigDB	Molecular signature database
OS	Overall survival
PFS	Progression-free survival
ssGSEA	Single-sample Gene Set Enrichment Analysis
TCGA	The Cancer Genome Atlas
WSI	Whole slide image

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-024-06007-8.

Additional file 1. Figure S1. Overview of image augmentation for training of the deep learning model. Rectangular images were converted into 2 JPEG files, 1 of which was flipped horizontally or vertically. Examples from 2 images are shown in. Images were normalized, vertically or horizontally flipped, and randomly affined before being fed into the model for training. Examples from 2 images are shown in. Figure S2. The gene heatmap of the top immune pathway of the pathway enrichment analysis. The gene heatmap of the most significant pathways between the predicted low and high H&E—based survival scores from the GSEA pathway enrichment analysis, which are GSE3039_ALPHAALPHA_CD8_TCELL_VS_B2_BCELL_UP, GSE14026_TH1_VS_TH17_UP, and GSE23114_PERITONEAL_CAVITY_B1A_BCELL_VS_SPLEEN_BCELL_IN_SLE2C1_MOUSE_UP are shown respectively.

Additional file 2. Table S1. TCGA images metadata for the 5-fold cross-validation process, training, and testing

Additional file 3. Table S2. The clinical characteristics of the patients included in the MDACC dataset

Additional file 4. Table S3. Immune signature enrichment results between TCGA testing samples with low and high H&E predicted score

Acknowledgements

The results here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. We thank Laura L. Russell, scientific editor, Research Medical Library of the University of Texas MD Anderson Cancer Center, for editing this article.

Author contributions

CWN contributed on the conception, methodology, data curation, analysis, data interpretation, and preparation of the manuscript. KKW contributed to the critical review of the manuscript. BCL contributed on the data acquisition, analysis, interpretation, and critical review of the manuscript. SFB contributed on the conception, data curation, and critical review of the manuscript. SCM contributed on the conception, data curation, critical review and manuscript revision.

Funding

This work was supported in part by National Institutes of Health grant U01CA294459 and Debra Blum Endowment for Ovarian Cancer Research.

Availability of data and materials

The TCGA data included in this study were downloaded from the GDC portal and cBioportal. The data from MDACC dataset are not publicly available due to patient privacy but are available upon reasonable request. The Jupyter notebooks used to generate the results is available as supplementary file.

Declarations

Ethics approval and consent to participate/Consent for publication

WSIs of H&E-stained tumor sections and clinicopathological characteristics from the MDACC dataset were obtained from the ovarian cancer repository of the Department of Gynecologic Oncology and Reproductive Medicine under protocols approved by the University of Texas MD Anderson's Institutional Review Board. Written informed consent from the patients were obtained by front desk personnel, and the studies were conducted in accordance with recognized ethical guidelines. TCGA data were obtained from public repository and did not require ethical approval.

Competing interests

The authors have declared that no competing interests exist.

Author details

[1]Department of Gynecologic Oncology and Reproductive Medicine, Unit 1362, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA. [2]Department of Anatomical Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

Received: 29 March 2024 Accepted: 18 December 2024
Published online: 24 January 2025

References

1. Lheureux S, Gourley C, Vergote I, Oza AM. Epithelial ovarian cancer. Lancet. 2019;393(10177):1240–53. https://doi.org/10.1016/S0140-6736(18)32552-2.

2. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. CA Cancer J Clin. 2023;73(1):17–48. https://doi.org/10.3322/caac.21763.
3. Kuroki L, Guntupalli SR. Treatment of epithelial ovarian cancer. BMJ. 2020;371: m3773. https://doi.org/10.1136/bmj.m3773.
4. Deng F, Xu X, Lv M, Ren B, Wang Y, Guo W, et al. Age is associated with prognosis in serous ovarian carcinoma. J Ovarian Res. 2017;10(1):36. https://doi.org/10.1186/s13048-017-0331-6.
5. Schuurman MS, Kruitwagen R, Portielje JEA, Roes EM, Lemmens V, van der Aa MA. Treatment and outcome of elderly patients with advanced stage ovarian cancer: a nationwide analysis. Gynecol Oncol. 2018;149(2):270–4. https://doi.org/10.1016/j.ygyno.2018.02.017.
6. van Walree IC, van Soolingen NJ, Hamaker ME, Smorenburg CH, Louwers JA, van Huis-Tanja LH. Treatment decision-making in elderly women with ovarian cancer: an age-based comparison. Int J Gynecol Cancer. 2019;29(1):158–65. https://doi.org/10.1136/ijgc-2018-000026.
7. Boehm KM, Aherne EA, Ellenson L, Nikolovski I, Alghamdi M, Vazquez-Garcia I, et al. Multimodal data integration using machine learning improves risk stratification of high-grade serous ovarian cancer. Nat Cancer. 2022;3(6):723–33. https://doi.org/10.1038/s43018-022-00388-9.
8. Fan Z, Jiang Z, Liang H, Han C. Pancancer survival prediction using a deep learning architecture with multimodal representation and integration. Bioinform Adv. 2023;3(1):vbad006. https://doi.org/10.1093/bioadv/vbad006.
9. Huang Z, Johnson TS, Han Z, Helm B, Cao S, Zhang C, et al. Deep learning-based cancer survival prognosis from RNA-seq data: approaches and evaluations. BMC Med Genomics. 2020;13(Suppl 5):41. https://doi.org/10.1186/s12920-020-0686-1.
10. Lee M. Deep learning techniques with genomic data in cancer prognosis: a comprehensive review of the 2021–2023 literature. Biology (Basel). 2023;12(7):893. https://doi.org/10.3390/biology12070893.
11. Xiao Y, Bi M, Guo H, Li M. Multi-omics approaches for biomarker discovery in early ovarian cancer diagnosis. EBioMedicine. 2022;79: 104001. https://doi.org/10.1016/j.ebiom.2022.104001.
12. Parikh RB, Basen-Enquist KM, Bradley C, Estrin D, Levy M, Lichtenfeld JL, et al. Digital health applications in oncology: an opportunity to seize. J Natl Cancer Inst. 2022;114(10):1338–9. https://doi.org/10.1093/jnci/djac108.
13. Jahn SW, Plass M, Moinfar F. Digital pathology: advantages, limitations and emerging perspectives. J Clin Med. 2020. https://doi.org/10.3390/jcm9113697.
14. Niazi MKK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. Lancet Oncol. 2019;20(5):e253–61. https://doi.org/10.1016/S1470-2045(19)30154-8.
15. Ng CW, Wong KK. Deep learning-enabled breast cancer endocrine response determination from H&E staining based on ESR1 signaling activity. Sci Rep. 2023;13(1):21454. https://doi.org/10.1038/s41598-023-48830-x.
16. Chaunzwa TL, Hosny A, Xu Y, Shafer A, Diao N, Lanuti M, et al. Deep learning classification of lung cancer histology using CT images. Sci Rep. 2021;11(1):5471. https://doi.org/10.1038/s41598-021-84630-x.
17. Khairi SSM, Bakar MAA, Alias MA, Bakar SA, Liong CY, Rosli N, et al. Deep learning on histopathology images for breast cancer classification: a bibliometric analysis. Healthcare (Basel). 2021. https://doi.org/10.3390/healthcare10010010.
18. Meirelles ALS, Kurc T, Kong J, Ferreira R, Saltz JH, Teodoro G. Building efficient cnn architectures for histopathology images analysis: a case-study in tumor-infiltrating lymphocytes classification. Front Med (Lausanne). 2022;9: 894430. https://doi.org/10.3389/fmed.2022.894430.
19. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. N Engl J Med. 2016;375(12):1109–12. https://doi.org/10.1056/NEJMp1607591.
20. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401–4. https://doi.org/10.1158/2159-8290.CD-12-0095.
21. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioportal. Sci Signal. 2013;6(269):l1. https://doi.org/10.1126/scisignal.2004088.
22. Ojalam M, Garriga GC. Permutation tests for studying classifier performance. J Mach Learn Res. 2010;11(6):1833.
23. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: an imperative style, high-performance deep learning library. In: Paszke A, editor., et al., Proceedings of the international conference on neural information processing systems. Red Hook: Curran Associates Inc; 2019. p. 721.
24. Pölsterl S. Scikit-survival: a library for time-to-event analysis built on top of scikit-learn. J Mach Learn Res. 2020;21(1):8747–52.
25. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–30.
26. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. Scipy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72. https://doi.org/10.1038/s41592-019-0686-2.
27. Davidson-Pilon C. Lifelines: survival analysis in Python. J Open Source Softw. 2019;4(40):1317.
28. Fang Z, Liu X, Peltz G. Gseapy: a comprehensive package for performing gene set enrichment analysis in Python. Bioinformatics. 2023. https://doi.org/10.1093/bioinformatics/btac757.
29. Wolf FA, Angerer P, Theis FJ. Scanpy: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15. https://doi.org/10.1186/s13059-017-1382-0.
30. Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, et al. Squidpy: a scalable framework for spatial omics analysis. Nat Methods. 2022;19(2):171–8. https://doi.org/10.1038/s41592-021-01358-2.
31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102(43):15545–50. https://doi.org/10.1073/pnas.0506580102.
32. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (Msigdb) 3.0. Bioinformatics. 2011;27(12):1739–40. https://doi.org/10.1093/bioinformatics/btr260.
33. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (Msigdb) hallmark gene set collection. Cell Syst. 2015;1(6):417–25. https://doi.org/10.1016/j.cels.2015.12.004.
34. Fu Y, Jung AW, Torne RV, Gonzalez S, Vöhringer H, Shmatko A, et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer. 2020;1(8):800–10. https://doi.org/10.1038/s43018-020-0085-8.
35. Hwang C, Lee SJ, Lee JH, Kim KH, Suh DS, Kwon BS, et al. Stromal tumor-infiltrating lymphocytes evaluated on h&e-stained slides are an independent prognostic factor in epithelial ovarian cancer and ovarian serous carcinoma. Oncol Lett. 2019;17(5):4557–65. https://doi.org/10.3892/ol.2019.10095.
36. Hudry D, Le Guellec S, Meignan S, Becourt S, Pasquesoone C, El Hajj H, et al. Tumor-infiltrating lymphocytes (tils) in epithelial ovarian cancer: heterogeneity, prognostic impact, and relationship with immune checkpoints. Cancers (Basel). 2022. https://doi.org/10.3390/cancers14215332.
37. Gupta P, Chen C, Chaluvally-Raghavan P, Pradeep S. B cells as an immune-regulatory signature in ovarian cancer. Cancers (Basel). 2019. https://doi.org/10.3390/cancers11070894.
38. Schwartz JT, Bandyopadhyay S, Kobayashi SD, McCracken J, Whitney AR, Deleo FR, et al. Francisella tularensis alters human neutrophil gene expression: Insights into the molecular basis of delayed neutrophil apoptosis. J Innate Immun. 2013;5(2):124–36. https://doi.org/10.1159/000342430.
39. Liu Y, Yang R, Zhang M, Yang B, Du Y, Feng H, et al. Multi-omics landscape of interferon-stimulated gene OASL reveals a potential biomarker in pan-cancer: from prognosis to tumor microenvironment. Front Immunol. 2024;15:1402951. https://doi.org/10.3389/fimmu.2024.1402951.
40. Huang X, Nepovimova E, Adam V, Sivak L, Heger Z, Valko M, et al. Neutrophils in cancer immunotherapy: friends or foes? Mol Cancer. 2024;23(1):107. https://doi.org/10.1186/s12943-024-02004-z.
41. Chen S, Sun Z, Zhao W, Meng M, Guo W, Wu D, et al. Oligoadenylate synthetases-like is a prognostic biomarker and therapeutic target in pancreatic ductal adenocarcinoma. Ann Transl Med. 2022;10(3):138. https://doi.org/10.21037/atm-21-6618.

42. Coemans M, Verbeke G, Dohler B, Susal C, Naesens M. Bias by censoring for competing events in survival analysis. BMJ. 2022;378: e071349. https://doi.org/10.1136/bmj-2022-071349.
43. Lin JR, Wang S, Coy S, Chen YA, Yapp C, Tyler M, et al. Multiplexed 3D atlas of state transitions and immune interaction in colorectal cancer. Cell. 2023;186(2):363-81 e19. https://doi.org/10.1016/j.cell.2022.12.028.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.