



UNIVERSITÀ DEL PIEMONTE ORIENTALE

Dipartimento di Scienze e Innovazione Tecnologica

Corso di Laurea Magistrale in Intelligenza Artificiale e Innovazione Digitale

Relazione per la prova finale

**Identificazione di biomarcatori predittivi nell'Artrite Reumatoide
tramite Graph Neural Networks:**

confronto tra architetture per la stratificazione dei pazienti

Relatore:

Prof.ssa Chiocchetti Annalisa

Correlatore:

Dott. Dossena Marco

Candidato:

Bullano Elia

A handwritten signature in black ink, appearing to read 'Elia Bullano', written in a cursive style.

Anno Accademico 2024/2025

Abstract

L'Artrite Reumatoide è una patologia autoimmune complessa caratterizzata da una profonda eterogeneità spaziale e fenotipica a livello del microambiente sinoviale. Questa complessità si traduce in una marcata variabilità interindividuale nella risposta clinica ai farmaci biologici e sintetici mirati. La stratificazione precoce dei pazienti refrattari alle cure rispetto a quelli responsivi rappresenta, attualmente, una delle sfide centrali della reumatologia di precisione.

Il presente lavoro di tesi propone un *framework* computazionale innovativo volto a superare i limiti dei tradizionali modelli di *Machine Learning*, i quali, trattando le cellule come unità isolate, disperdono la fondamentale informazione topologica e relazionale del tessuto. Sfruttando dati ad alta dimensionalità acquisiti tramite *Imaging Mass Cytometry* su una coorte clinica di cinque pazienti, le biopsie sinoviali sono state modellate come grafi.

Sono state implementate e confrontate tre architetture di *Deep Learning*, ovvero MPNN, GATv2 e SAGEConv. Per arginare il rischio di *overfitting* dovuto alla dimensionalità del dataset, le reti sono state addestrate mediante un paradigma *self-supervised* basato sulla ricostruzione del profilo proteico oscurato, attraverso la tecnica di *Masked Node Feature Reconstruction*.

L'importanza patogenetica di 25 marcatori proteici e 4 *feature* morfologiche è stata calcolata quantificando la difficoltà della rete neurale nel ricostruire il profilo del paziente a seguito della loro permutazione, permettendo l'estrazione di un *consensus ranking* globale.

La successiva validazione predittiva, condotta tramite un algoritmo di *ensemble forward selection* e validata con un protocollo *Leave-One-Patient-Out Cross-Validation* su un classificatore logistico con penalizzazione L2, ha evidenziato l'effetto della *curse of dimensionality*, ovvero come l'impiego del pannello completo saturi lo spazio informativo introducendo rumore vettoriale.

Al contrario, il modello ha isolato un *subset* ottimale di soli quattro marcatori, composto da aSMA, CD45, CD45RO e CD138, in grado di raggiungere una separazione completa nel dataset utilizzato tra le classi cliniche, consentendo una riduzione del pannello originario dell'80%.

L'analisi di interpretabilità biologica rivela che i pazienti *Inadequate Responder* sono caratterizzati da un microambiente tissutale associato a maggiore attivazione miofibroblastica e rimodellamento della matrice extracellulare, nonché da alterazioni morfologiche cellulari, in particolare un aumento delle dimensioni e dell'eccentricità.

Al contrario, i pazienti *Responder* mantengono un'architettura immunitaria adattiva compatibile con una maggiore probabilità di risposta terapeutica.

In conclusione, l'integrazione tra *Spatial Multiplexed Imaging* e *Graph Neural Networks* non fornisce soltanto uno strumento predittivo, ma anche esplorativo, permettendo di identificare una mappa molecolare minima dell'ecosistema sinoviale e aprendo nuove prospettive per la diagnostica e la personalizzazione delle terapie nell'Artrite Reumatoide.

Indice

Abstract	i
1 Introduzione	1
1.1 Obiettivi della Tesi	2
2 Artrite Reumatoide e Microambiente Sinoviale	4
2.1 Strategie Terapeutiche e Target Molecolari	6
2.2 Imaging Mass Cytometry (Hyperion Imaging System)	7
2.3 Modellazione Relazionale del Tessuto Biologico	8
3 Dataset e Pre-processing	10
3.1 Origine dei Dati IMC	10
3.2 Analisi Descrittiva del Dataset	11
3.3 Pipeline di Pre-processing	12
3.4 Costruzione dei Grafi Cellulari	13
3.5 Problema come Task di Classificazione	13
4 Metodologia: Graph Neural Networks	15
4.1 Formalizzazione del Problema	15
4.2 Architettura 1: Message Passing Neural Network (MPNN)	16
4.3 Architettura 2: GATv2 (Graph Attention Network v2)	18
4.4 Architettura 3: SAGEConv (GraphSAGE)	19
4.5 Confronto Strutturale tra le Tre Architetture	21
4.6 Pipeline di Training, Validazione e Ottimizzazione	21
5 Marker Selection e Analisi dell'Importanza delle Feature	23
5.1 Motivazione Clinica della Riduzione dei Marcatori	24
5.2 Metodologia di Feature Selection	25
5.3 Identificazione del Subset Ottimale	27
5.4 Ruolo delle Feature Morfologiche	30

6	Risultati Sperimentali e Interpretabilità del Modello	33
6.1	Confronto tra le tre Architetture GNN	33
6.2	Analisi di Stabilità e Rank Consistency	38
6.3	Analisi Multi-Maschera	39
6.4	Consenso Globale e Fenotipizzazione Clinica	42
6.5	Risultati delle Feature Morfologiche	43
6.6	Confronto Panel Completo vs Subset Ridotto	45
7	Discussione e Conclusioni	50
7.1	Implicazioni Metodologiche	50
7.2	Implicazioni Biologiche nell'Artrite Reumatoide	51
7.3	Limiti dello Studio e Sviluppi Futuri	52
7.4	Conclusioni Finali	53

Capitolo 1

Introduzione

Il presente lavoro si inserisce nel contesto della medicina di precisione, ponendo come oggetto principale di studio l'Artrite Reumatoide, una patologia autoimmune sistemica ad andamento cronico, caratterizzata da una severa infiammazione e dal progressivo rimodellamento del microambiente articolare [1, 2].

L'obiettivo primario della ricerca consiste nell'analizzare il legame tra le caratteristiche del tessuto sinoviale, ottenute tramite biopsie tissutali, e la risposta clinica alle terapie. In particolare, si intende comprendere quali proprietà biologiche del microambiente articolare possano influenzare l'efficacia terapeutica [2].

L'indagine sperimentale si basa su una coorte clinica composta da cinque pazienti, suddivisi in due gruppi distinti: pazienti responsivi al trattamento e pazienti refrattari, ovvero soggetti che sviluppano una resistenza, intrinseca o acquisita, alla terapia [3, 4].

Al fine di elaborare e interpretare dati spaziali ad alta dimensionalità, sono state utilizzate reti neurali basate su grafi, denominate *Graph Neural Networks* (GNN), confrontando tre differenti architetture per analizzarne la capacità di catturare le dinamiche presenti nel tessuto e integrare le informazioni ottenute in una rappresentazione complessiva [5, 6].

La scelta di utilizzare questa specifica tipologia di reti neurali deriva da un presupposto biologico fondamentale: le cellule presenti all'interno di un tessuto non operano come unità isolate, ma costituiscono un ecosistema interconnesso, comunicando e interagendo costantemente con il microambiente circostante. Tali relazioni spaziali possono essere formalizzate attraverso la teoria dei grafi, considerando le cellule come nodi e le interazioni come archi [7, 8].

Le GNN consentono quindi di preservare e analizzare la rete di relazioni cellulari; sfruttando la rappresentazione a grafo, è possibile ricreare computazionalmente un ambiente simile a quello presente all'interno del tessuto biologico [5].

Attraverso questo approccio, è possibile apprendere i *pattern* strutturali ed estrarre le caratteristiche cruciali che determinano la risposta o la resistenza alla terapia, gettando le basi per future applicazioni a supporto del processo decisionale clinico [9].

1.1 Obiettivi della Tesi

Il presente lavoro di tesi si prefigge, come obiettivo primario, di colmare l'attuale divario metodologico nella predizione della risposta al trattamento nell'Artrite Reumatoide, sviluppando un *framework* in grado di identificare biomarcatori potenzialmente predittivi della resistenza terapeutica [10].

In una prima fase, sono state impiegate le *Graph Neural Networks* in sostituzione dei tradizionali algoritmi di *Machine Learning*, con l'obiettivo di modellare le biopsie tissutali come reti complesse, nello specifico come grafi cellulari, preservando l'integrità architettonica originale del tessuto e consentendo l'apprendimento delle dinamiche di comunicazione tra le diverse popolazioni cellulari [11].

Successivamente, al fine di individuare la strategia di aggregazione più idonea, la tesi si propone di implementare e confrontare tre distinte architetture di *Deep Learning*:

- la *Message Passing Neural Network* (MPNN), basata sul paradigma del passaggio di messaggi [12];
- la *Graph Attention Network v2* (GATv2), caratterizzata da meccanismi di attenzione dinamica [13];
- la *GraphSAGE* (SAGEConv), ottimizzata per il campionamento in contesti ad alta densità [14].

Queste architetture differiscono per il modo in cui aggregano l'informazione proveniente dai nodi vicini, offrendo prospettive complementari sull'analisi del microambiente tissutale.

L'apprendimento viene condotto in modalità auto-supervisionata, basata sul mascheramento delle *feature* dei nodi e sulla loro successiva ricostruzione. Questo approccio consente al modello di apprendere le dipendenze biologiche del tessuto senza introdurre *bias* esterni di natura clinica o umana [15].

Successivamente, i risultati ottenuti dalle diverse architetture GNN, ciascuna caratterizzata da *bias* intrinseci, vengono integrati in una classifica globale di consenso, con l'obiettivo di mitigare le instabilità algoritmiche individuali e ottenere una graduatoria robusta dei marcatori.

Infine, tale classifica viene utilizzata per estrarre un *subset* ottimale di marcatori, riducendo la dimensionalità del pannello e individuando un insieme ristretto di biomarcatori chiave per futuri *test* diagnostici [16].

In sintesi, la presente tesi si propone di dimostrare come l'integrazione tra tecniche di *Spatial Multiplexed Imaging* e architetture di *Deep Learning* su grafi possa contribuire all'identificazione di nuovi *network* patogenetici, fornendo strumenti avanzati per la medicina di precisione e la personalizzazione delle terapie [17].

Capitolo 2

Artrite Reumatoide e Microambiente Sinoviale

L'Artrite Reumatoide è una patologia autoimmune sistemica e altamente invalidante, caratterizzata primariamente dall'insorgenza di un'inflammatione cronica a carico delle articolazioni, che evolve progressivamente verso la distruzione irreversibile dei tessuti osteo-articolari [1].

Nonostante l'eziologia sistemica della malattia, il fulcro del processo patologico si sviluppa e si localizza a livello della membrana sinoviale, la quale rappresenta la regione nevralgica della progressione del danno articolare [2].

Per comprendere i meccanismi alla base della patogenesi dell'AR, è necessario analizzare la struttura e la funzione del tessuto sinoviale in condizioni fisiologiche. In condizioni di salute, la membrana sinoviale si presenta come una struttura sottile, composta da pochi strati cellulari distinti in *lining* e *sublining* [8].

Essa svolge un ruolo fondamentale nel mantenimento dell'omeostasi articolare, regolando la produzione di liquido sinoviale, fornendo nutrimento ai condrociti e facilitando la lubrificazione delle superfici cartilaginee [18, 19].

Nell'AR, questo equilibrio viene compromesso e il tessuto evolve nel cosiddetto pannus reumatoide, una struttura iperplastica e fortemente vascolarizzata, caratterizzata da un comportamento invasivo.

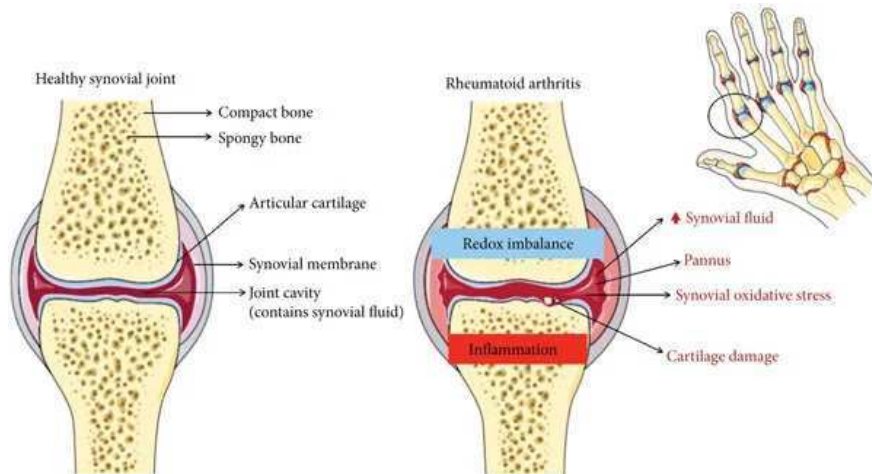


Figura 2.1: Confronto tra articolazione sana e articolazione affetta da Artrite Reumatoide. Nella condizione patologica, la membrana sinoviale iperplastica (pannus) invade la cartilagine e l'osso, determinando erosione articolare e perdita di integrità strutturale (adattata da [8]).

Come illustrato nella Figura 2.1, il pannus invade progressivamente la cartilagine articolare e l'osso subcondrale, risultando responsabile della distruzione osteo-cartilaginea tipica della patologia.

Alla luce di queste trasformazioni, risulta fondamentale analizzare la composizione cellulare del microambiente sinoviale, il quale non costituisce un insieme disorganizzato, ma un ecosistema strutturato e altamente eterogeneo [7, 17].

Tra le principali popolazioni cellulari coinvolte si distinguono i fibroblasti sinoviali, i macrofagi e le cellule del compartimento linfoide [17].

In particolare, i fibroblasti sinoviali rappresentano la componente stromale predominante del pannus. In risposta a stimoli epigenetici e microambientali, tali cellule subiscono una riprogrammazione verso un fenotipo invasivo e resistente all'apoptosi, caratterizzato da elevata proliferazione e secrezione di mediatori pro-infiammatori [8].

Accanto ai fibroblasti, i macrofagi svolgono un ruolo cruciale nella regolazione della risposta infiammatoria [18].

Essi possono derivare sia da popolazioni macrofagiche sia dal reclutamento di monociti circolanti e, all'interno della sinovia, assumono fenotipi eterogenei, producendo citochine pro-infiammatorie che contribuiscono alla cronicizzazione dello stato infiammatorio [20].

Parallelamente, il compartimento linfoide amplifica la risposta immunitaria ed è dominato da linfociti T (CD4+ e CD8+) e linfociti B CD20+, le quali svolgono funzione di presentazione dell'antigene e si differenziano in plasmacellule (CD138+) [21].

Queste popolazioni possono organizzarsi in strutture linfoidi ectopiche, associate a forme di malattia più severe, elevata produzione locale di autoanticorpi e maggiore refrattarietà ai trattamenti [3, 11].

La moderna reumatologia ha quindi superato il concetto di AR come patologia univoca, identificando distinti patotipi infiammatori del tessuto sinoviale: linfoide, mieloide e fibroide.

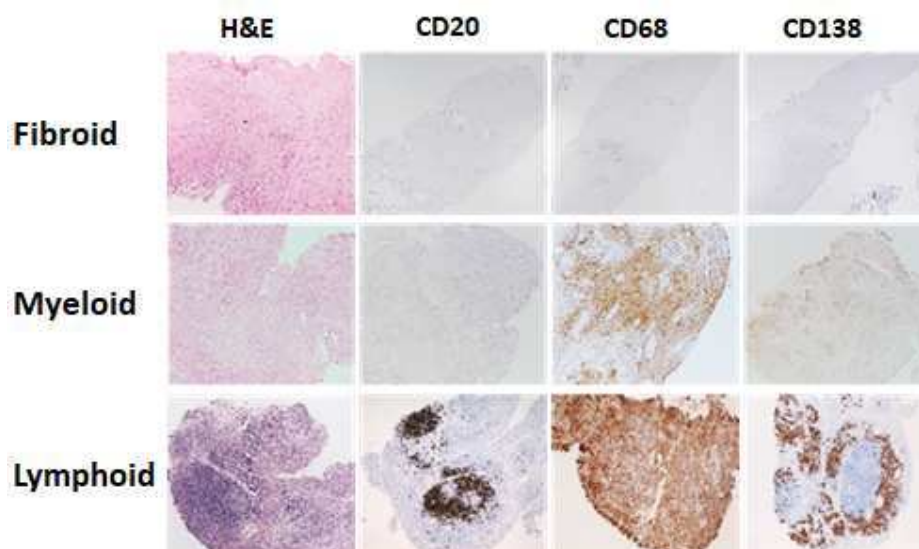


Figura 2.2: Principali patotipi sinoviali nell'Artrite Reumatoide. Il patotipo linfoide è caratterizzato dalla presenza di cellule B e T organizzate, il patotipo mieloide da una predominanza di macrofagi e cellule dell'immunità innata, mentre il patotipo fibroide è dominato da fibroblasti sinoviali e da una componente stromale con scarsa infiltrazione immunitaria (adattata da [10]).

Come illustrato nella Figura 2.2, questi patotipi riflettono diverse organizzazioni cellulari e funzionali del microambiente sinoviale, associate a differenti meccanismi patogenetici e a una risposta terapeutica eterogenea.

È importante sottolineare come tali patotipi non dipendano esclusivamente dalla composizione cellulare, ma anche dalla loro organizzazione spaziale. La posizione relativa delle cellule all'interno del tessuto non è casuale, ma influisce direttamente sul loro comportamento funzionale [2].

Volendo proporre un esempio, i fibroblasti localizzati nello strato di *lining* sono coinvolti nei processi di invasione della cartilagine, mentre sottopopolazioni nel *sublining* contribuiscono al reclutamento immunitario e alla modulazione della risposta infiammatoria [22].

Considerato il ruolo centrale dell'organizzazione cellulare, emerge dunque la necessità di comprendere come tali dinamiche possano essere modulate terapeuticamente.

2.1 Strategie Terapeutiche e Target Molecolari

La moderna gestione terapeutica dell'AR si basa sulla strategia *Treat-to-Target* (T2T) [4], raccomandata dalle linee guida ACR ed EULAR [9].

Questo approccio prevede un adattamento dinamico del trattamento con l'obiettivo di raggiungere la remissione completa o, in alternativa, uno stato di bassa attività di malattia [4].

Nonostante l'introduzione di farmaci biologici e *targeted synthetic DMARDs* abbia migliorato significativamente la prognosi, una quota rilevante di pazienti presenta risposta terapeutica insufficiente [3, 16].

I pazienti vengono quindi classificati in:

- *Responders* (R);
- *Inadequate Responders* (IR).

Questa variabilità suggerisce l'esistenza di sottotipi biologici distinti e rende cruciale l'identificazione di biomarcatori predittivi [10].

Le terapie più recenti mirano non solo al blocco di singole citochine, ma all'interferenza delle interazioni tra popolazioni cellulari, agendo su *hub* di comunicazione tra linfociti, macrofagi e fibroblasti [8, 16].

2.2 Imaging Mass Cytometry (Hyperion Imaging System)

Per analizzare la complessità del microambiente cellulare, sono necessarie tecnologie in grado di preservare l'informazione spaziale.

Le tecniche tradizionali presentano tuttavia importanti limitazioni. L'immunoistochimica consente l'analisi di un numero ristretto di marcatori per campione, riducendo la capacità di caratterizzare in modo esaustivo l'eterogeneità cellulare del tessuto. D'altra parte, approcci basati su *scRNA-seq* permettono una profilazione molecolare ad alta dimensionalità, ma comportano la perdita dell'informazione spaziale, impedendo di ricostruire le relazioni topologiche tra le cellule.

Per superare tali limiti, è stata sviluppata l'*Imaging Mass Cytometry* (IMC), implementata nel sistema *Hyperion* [23, 24].

Questa tecnologia consente la rilevazione simultanea di oltre 40 marcatori, preservando la topologia spaziale del tessuto. Il processo si basa sull'utilizzo di anticorpi coniugati a isotopi metallici, selezionati per evitare sovrapposizioni spettrali tipiche dei fluorofori tradizionali.

Dopodiché, il campione viene sottoposto ad ablazione laser sequenziale, che vaporizza il tessuto in modo altamente controllato a risoluzione subcellulare.

Infine, il materiale ionizzato viene analizzato tramite spettrometria di massa a tempo di volo (*Time-of-Flight*, TOF), permettendo la quantificazione simultanea e multiplexata dei marcatori proteici associati a ciascuna cellula.

Il risultato finale è un *dataset* multidimensionale in cui, a ogni *pixel*, viene associato uno spettro di massa.

Attraverso algoritmi di segmentazione come *CellPose* [25] o *DeepCell* [26], i dati vengono convertiti in rappresentazioni cellulari, permettendo di ricostruire, per ciascuna cellula, le coordinate spaziali, il profilo proteico e le relazioni di vicinato, che costituiscono una approssimazione delle interazioni biologiche all'interno del microambiente tissutale.

2.3 Modellazione Relazionale del Tessuto Biologico

L'avvento dei dati spaziali ha introdotto una nuova prospettiva nello studio dei sistemi biologici [23, 27].

Il tessuto sinoviale può essere interpretato come una rete dinamica di interazioni cellulari, in cui il comportamento di ogni cellula dipende dal contesto microambientale [27].

Questo approccio ha portato allo sviluppo della modellazione relazionale e al concetto di *Digital Twin* [28], che consente simulazioni computazionali del microambiente patologico.

Parallelamente, modelli come i *Variational Autoencoders* (VAE) [29] permettono di identificare *pattern* organizzativi e nicchie cellulari funzionali.

Dal punto di vista computazionale, il tessuto può essere rappresentato come un grafo non orientato:

$$G = (V, E)$$

dove:

- V rappresenta le cellule (nodi);
- E rappresenta le relazioni di vicinato (archi).

Su queste strutture vengono applicate le *Graph Neural Networks* (GNN), basate sul paradigma del *message passing* [12].

L'integrazione tra *imaging* spaziale e GNN rappresenta un approccio avanzato per lo studio del microambiente sinoviale [6, 30], consentendo di identificare *pattern* biologici complessi, migliorando la stratificazione dei pazienti e la personalizzazione delle terapie, rappresentando il punto di partenza per le analisi descritte nel capitolo successivo, in cui tali informazioni vengono trasformate in strutture computazionali per la modellazione del tessuto.

Alla luce delle tecnologie e dei principi biologici discussi nel capitolo precedente, è ora possibile passare dalla descrizione teorica del microambiente sinoviale alla sua rappresentazione operativa all'interno del presente studio.

In particolare, l'integrazione tra dati derivanti da *Imaging Mass Cytometry* e tecniche di analisi spaziale consente di costruire un *dataset* a livello di singola cellula, nel quale le informazioni proteiche e morfologiche vengono organizzate in una forma idonea alla modellazione computazionale.

In conclusione, il capitolo ha illustrato le tecnologie e i principi alla base dell'acquisizione dei dati, ponendo le basi per la costruzione del *dataset* utilizzato nello studio.

Su questa base, nel capitolo successivo viene descritto il *dataset* e la *pipeline* di *pre-processing* adottata, finalizzata a trasformare i dati grezzi in strutture coerenti e idonee alla modellazione mediante *Graph Neural Networks*.

Capitolo 3

Dataset e Pre-processing

Il *pre-processing* rappresenta una fase critica per la validità dell'intero studio, in quanto ha il compito sia di mitigare il rumore tecnico di acquisizione sia di trasformare le informazioni biologiche grezze in strutture dati idonee alla successiva costruzione dei grafi cellulari.

Alla luce di ciò, nelle sezioni seguenti vengono descritti l'origine dei dati, la struttura del *dataset* e la *pipeline* di *pre-processing* adottata.

3.1 Origine dei Dati IMC

Il punto di partenza su cui si basa la *pipeline* di *pre-processing* comprende l'acquisizione di immagini spaziali mediante il sistema *Hyperion Imaging Mass Cytometry*.

A differenza della microscopia ottica, questa tecnologia acquisisce simultaneamente informazioni topologiche e proteiche, vaporizzando il tessuto *pixel* per *pixel* con una risoluzione di circa $1 \mu\text{m}^2$, mediante ablazione laser ad alta frequenza e successiva rilevazione, da parte del *detector*, della quantità di metalli coniugati agli anticorpi tramite spettrometria di massa a tempo di volo [17, 20].

Per ciascuna delle Regioni di Interesse acquisite (*Region of Interest*, ROI), il sistema genera nativamente due file principali:

- file *.mcd*, contenente le matrici dei dati grezzi e le informazioni strutturali dell'immagine;
- file *.txt*, nel quale, per ogni coordinata spaziale del *pixel* (x, y), viene riportata l'intensità del segnale relativa al marcatore, generando una mappa bidimensionale del tessuto.

A partire dai dati acquisiti, risulta ora opportuno descrivere la struttura del *dataset* utilizzato nello studio.

3.2 Analisi Descrittiva del Dataset

Il *dataset* spaziale oggetto di questa tesi è stato strutturato e consolidato in tre *directory* e file principali, che definiscono lo schema dei dati a singola cellula.

In primo luogo, è presente un file denominato `output.csv`, contenente biomarcatori e fenotipi, comprensivo dell'identificativo dei pazienti con le relative ROI, delle intensità di espressione cellulare dei marcatori, dei *cluster_celltype*, ovvero le etichette di appartenenza cellulare derivate da un clustering non supervisionato dei profili di espressione proteica, eseguito mediante algoritmo Phenograph, come descritto nella Sezione 3.3 e infine i marcatori stessi.

Successivamente, è presente una cartella denominata `regionprops`, relativa alla morfologia cellulare, anch'essa organizzata per pazienti e ROI in accordo con il file `output.csv`. Al suo interno, sono riportate sei caratteristiche geometriche e spaziali per ogni singolo oggetto cellulare segmentato:

- **Area**: dimensione complessiva della cellula, quantificata in *pixel*;
- **Centroid-0, Centroid-1**: coordinate spaziali del baricentro cellulare;
- **Axis_major_length**: lunghezza dell'asse maggiore;
- **Axis_minor_length**: lunghezza dell'asse minore;
- **Eccentricity**: grado di allungamento geometrico della cellula.

Infine, vi è la cartella `neighbors`, contenente la lista delle adiacenze fisiche tra cellule, che rappresenta la base per la costruzione della topologia del grafo cellulare.

La coorte clinica oggetto di studio è composta da cinque pazienti affetti da Artrite Reumatoide, dai quali sono state estratte complessivamente 42 ROI:

- RA_016, etichetta *IR*, 8 ROI;
- RA_039, etichetta *Responder*, 10 ROI;
- RA_046, etichetta *Responder*, 10 ROI;
- RA_051, etichetta *IR*, 4 ROI;
- RA_056, etichetta *IR*, 10 ROI.

Il pannello molecolare investigato comprende 25 marcatori proteici:

aSMA, CD14, CD34, CD16, CD163, CD138, CD31, CD45, CD44, OPN, CD4, CD90, CD68, MERTK, CD20, CD8a, VEGF, CD45RA, GRANZYMEB, Ki67, COL1, CD3, CD45RO, HLA-DR, PODOPLANIN

Una volta definita la struttura del *dataset*, è necessario descrivere le operazioni di *pre-processing* applicate per renderlo idoneo all'analisi computazionale.

3.3 Pipeline di Pre-processing

L'estrazione del segnale biologico dalle immagini grezze IMC è stata effettuata utilizzando la *pipeline Steinbock* [31], che consente la segmentazione cellulare, l'estrazione delle intensità e la costruzione delle relazioni spaziali.

I dati sono stati successivamente integrati in ambiente R mediante il pacchetto *SpatialExperiment* [32].

Le intensità dei marcatori sono state trasformate mediante trasformazione *arcsinh*, definita come:

$$x' = \operatorname{arcsinh}(x) = \log\left(x + \sqrt{x^2 + 1}\right) \quad (3.1)$$

Questa trasformazione è ampiamente utilizzata nei dati di citometria, in quanto consente di comprimere i valori estremi mantenendo al contempo la sensibilità nei bassi livelli di espressione, stabilizzando la varianza del segnale biologico [24].

Successivamente, è stata applicata una normalizzazione per correggere i *batch effects* [33], seguita da un *clustering* non supervisionato tramite *Phenograph* [34].

Le *feature* sono state suddivise in:

- **biomarker;**
- **morfologia.**

È stata infine applicata una standardizzazione *Z-score* tramite *StandardScaler*, ottenendo un vettore finale di 29 *feature* (25 biomarcatori + 4 morfologiche).

Dopo aver preprocessato i dati a livello cellulare, è possibile formalizzare il microambiente come una struttura a grafo.

3.4 Costruzione dei Grafi Cellulari

Si giunge, dunque, all'analisi della struttura a grafo cellulare, nella quale il microambiente tissutale è stato modellato come un grafo non orientato $G = (V, E)$, in cui ogni nodo $v \in V$ rappresenta una cellula segmentata e ogni arco $e \in E$ rappresenta una relazione di adiacenza fisica tra due cellule.

A ciascun nodo è associato un vettore di *feature* contenente le informazioni molecolari e morfologiche della cellula.

Le relazioni cellulari sono state costruite a partire dal file `neighbors.csv`, che contiene per ogni cellula la lista delle cellule adiacenti identificate tramite espansione morfologica delle maschere cellulari.

Due nodi vengono connessi da un arco se risultano in contatto fisico diretto, ottenendo così un grafo non pesato che preserva la topologia locale del tessuto.

Questo approccio consente di modellare in modo esplicito le interazioni cellula-cellula, mantenendo la struttura spaziale del microambiente [30].

Una volta definita la rappresentazione del tessuto come grafo, è possibile formalizzare il problema in termini di classificazione.

3.5 Problema come Task di Classificazione

A seguito della definizione del dato, l'obiettivo di questo paragrafo è pertanto quello di definire un *framework* di classificazione binaria.

Il problema è stato formalizzato come un *task* di classificazione.

- *Responder*;
- *Inadequate Responder*.

Le *feature* utilizzate per la classificazione sono ottenute aggregando le rappresentazioni cellulari a livello di ROI e successivamente a livello paziente mediante media delle espressioni, trasformando il problema da scala cellulare a scala clinica.

È stata adottata una strategia *Leave-One-Patient-Out Cross-Validation* (LOPO-CV) [35], in cui a ogni iterazione un paziente viene utilizzato come *test set* e i restanti come *training set*.

Per garantire stabilità e interpretabilità, è stato utilizzato un modello di *Logistic Regression* con penalizzazione L2 (`penalty=L2`) e `class_weight='balanced'`.

Come metrica finale è stata utilizzata l'AUC (*Area Under the ROC Curve*), che misura la capacità del modello di discriminare correttamente tra i due fenotipi clinici.

Al termine della fase di *pre-processing*, il tessuto sinoviale è stato trasformato in una rappresentazione strutturata a livello cellulare, nella quale ogni cellula è descritta da un vettore di *feature* biologiche e morfologiche e le relazioni spaziali tra cellule sono esplicitamente definite.

Pertanto, questa rappresentazione costituisce il punto di partenza per la modellazione computazionale del microambiente, rendendo possibile l'applicazione di architetture in grado di sfruttare la struttura relazionale del tessuto.

Nel capitolo successivo si introduce il *framework* delle *Graph Neural Networks*, utilizzato per apprendere tali relazioni e modellare le dinamiche del microambiente sinoviale.

Capitolo 4

Metodologia: Graph Neural Networks

Come descritto nel Capitolo 3, l'analisi del microambiente sinoviale impone il superamento dei modelli di *Machine Learning* tradizionali, i quali, assumendo in modo semplificato l'indipendenza statistica delle singole osservazioni, falliscono nel considerare e catturare la topologia relazionale del microambiente.

Al fine di preservare l'integrità delle biopsie e superare tale limite, il presente lavoro adotta le *Graph Neural Networks* (GNN), ossia reti neurali che utilizzano come dominio geometrico i grafi [5].

Il microambiente cellulare viene definito come un grafo non orientato $G = (V, E)$, dove V rappresenta l'insieme dei nodi, con $V = N$ cellule, mentre E rappresenta l'insieme degli archi.

A ogni nodo $v \in V$ viene associato un vettore di input contenente le *feature* concatenate dei biomarcatori proteici e dei parametri morfologici, come già descritto nel Capitolo 3.

Dopo aver definito la rappresentazione del tessuto come grafo, al fine di identificare la strategia di aggregazione spaziale più performante per la stratificazione dei fenotipi di risposta terapeutica, sono state implementate tre distinte architetture di GNN.

4.1 Formalizzazione del Problema

Come spiegato in precedenza, a causa dell'eterogeneità spaziale del tessuto sinoviale, sono state adottate le *Graph Neural Networks*, ovvero una classe di architetture progettate per operare su dati strutturati come grafi, con l'obiettivo di recuperare non solo le caratteristiche intrinseche delle singole cellule, ma anche le loro interazioni.

Nel presente studio, le etichette cliniche non sono disponibili a livello di singola cellula, ma a livello di paziente e, indirettamente, di ROI. Di conseguenza, il paradigma di addestramento

è stato strutturato come un problema di apprendimento *self-supervised* [36], attraverso un *task* di ricostruzione stocastica denominato *Masked Node Feature Reconstruction*.

In tale configurazione, il modello riceve in input un grafo con *feature* parzialmente mascherate e deve ricostruire il profilo originale dei nodi sfruttando la struttura del grafo e il contesto locale.

Per ciascun esperimento, è stato applicato un *masking* progressivo, impostando valori di mascheramento pari al 10%, 20%, 30% e 50%, azzerando quindi una determinata percentuale di marcatori all'interno del nodo.

Questa formulazione è coerente con il profilo biologico studiato, poiché il comportamento e il profilo di una cellula non dipendono esclusivamente dalle sue proprietà intrinseche, ma anche dal microambiente in cui essa si trova.

Ne consegue che la capacità di ricostruire *feature* mascherate costituisce una misura indiretta della qualità delle rappresentazioni apprese dal modello.

Inoltre, per garantire robustezza statistica ed escludere *bias* legati all'inizializzazione casuale, sono state eseguite 5 *run* indipendenti utilizzando *seed* differenti.

L'ottimizzazione dei pesi della rete avviene minimizzando una funzione di costo basata sull'Errore Quadratico Medio (*Mean Squared Error*, MSE), valutato sui nodi appartenenti al *set* mascherato.

Dopo aver formalizzato il *task* di apprendimento, è possibile descrivere le architetture considerate nel confronto sperimentale.

4.2 Architettura 1: Message Passing Neural Network (MPNN)

La prima architettura formalizzata e presa in considerazione è la *Message Passing Neural Network* (MPNN) [12], che costituisce il paradigma unificante per la quasi totalità delle moderne architetture GNN.

La sua rilevanza risiede nel fatto che GCN, GAT, GATv2, GraphSAGE e molte altre architetture possono essere ricondotte, con opportune specializzazioni, alla formula generale del *message passing*.

L'utilizzo di questo paradigma è motivato anche da una ragione biologica: nelle reti a grafo basate sul *message passing*, ogni nodo aggiorna iterativamente la propria rappresentazione interna raccogliendo informazioni dal proprio vicinato, analogamente a quanto accade nelle cellule biologiche, le cui caratteristiche dipendono non soltanto da proprietà intrinseche, ma anche dal microambiente circostante [30].

Le fasi del processo, che si ripetono per ogni *layer* della rete, sono tre:

1. **Messaggio:** per ogni arco del grafo viene calcolato un messaggio che codifica l'interazione tra due nodi collegati;
2. **Aggregazione:** i messaggi provenienti dai vicini vengono aggregati tramite una funzione;
3. **Aggiornamento:** la rappresentazione del nodo viene aggiornata combinando lo stato precedente con il messaggio aggregato.

Nel presente lavoro, l'aggregazione è stata implementata tramite media, poiché il tessuto presenta densità cellulare variabile. In zone molto dense, l'aggregazione per somma produrrebbe *embedding* artificialmente più grandi, non per un reale segnale biologico più forte, ma semplicemente per la numerosità dei contributi.

La formulazione matematica utilizzata è la seguente:

$$m_v^{(t+1)} = AGG^{(t)} \left(\left\{ M^{(t)} \left(h_v^{(t)}, h_u^{(t)} \right) \mid u \in N(v) \right\} \right) \quad (4.1)$$

dove $M^{(t)}$ rappresenta la funzione di messaggio e $AGG^{(t)}$ la funzione di aggregazione, implementata mediante media.

Questa scelta è fondamentale per evitare che i nodi appartenenti a regioni molto dense producano rappresentazioni amplificate unicamente a causa del numero di vicini.

Successivamente, la rappresentazione del nodo viene aggiornata combinando lo stato precedente con il messaggio aggregato.

Al fine di individuare la configurazione ottimale del modello, è stata condotta una ricerca degli iperparametri. Lo spazio di ricerca esplorato è riportato nella Tabella 4.1.

Tabella 4.1: Spazio di ricerca degli iperparametri per MPNN.

Iperparametro	Spazio di ricerca
HIDDEN_DIM	{32, 64, 128, 256, 512}
NUM_LAYERS	{2, 3, 4}
DROPOUT	{0.2, 0.3, 0.4}
LEARNING_RATE	{0.001, 0.005}

Attraverso l'utilizzo della libreria *Optuna*, il cui funzionamento verrà descritto nella Sezione 4.6, è stata identificata la combinazione di parametri che massimizza le prestazioni del modello.

I valori ottimali selezionati per l'addestramento della rete neurale sono:

- `HIDDEN_DIM = 256;`
- `NUM_LAYERS = 2;`
- `DROPOUT = 0.2;`
- `LEARNING_RATE = 0.001.`

Dopo aver introdotto la MPNN, è possibile passare a un'architettura più espressiva, capace di attribuire importanza differenziata ai vicini mediante meccanismi di attenzione.

4.3 Architettura 2: GATv2 (Graph Attention Network v2)

La seconda architettura considerata è la *Graph Attention Network* (GAT), nella sua versione evoluta *GATv2*, derivata dalla *Graph Attention Network* originale [37, 38].

Questa architettura introduce un meccanismo di attenzione che consente di pesare in modo differente il contributo dei nodi vicini in base alle loro caratteristiche.

La GAT originale soffre di una limitazione strutturale: la funzione di attenzione è statisticamente lineare separabile, limitando la capacità del modello di catturare relazioni biologiche dipendenti dal contesto cellulare locale. Ciò è espresso dalla seguente formula, che rappresenta il coefficiente di attenzione:

$$e_{ij}^{GAT} = \text{LeakyReLU}(a^T [Wh_i \parallel Wh_j]) \quad (4.2)$$

Questa formulazione descrive un'attenzione statica, poiché il *ranking* dei vicini dipende da una funzione linearmente separabile.

In un contesto biologico, tuttavia, la rilevanza di una cellula vicina può dipendere in modo non lineare dal nodo centrale e dal contesto locale. Per superare tale limite, è stata implementata la *GATv2*, la cui innovazione architetturale risiede nel *Dynamic Attention Mechanism*, capace di ponderare l'importanza specifica di ogni singola interazione cellulare [13].

Successivamente, per rendere confrontabili i coefficienti, lo *score* viene normalizzato mediante funzione *Softmax*, ottenendo il coefficiente di attenzione finale.

Infine, il nodo viene aggiornato calcolando una somma pesata delle *feature* dei vicini. Per massimizzare la stabilità rispetto ai molteplici *pathway* biologici, la rete è stata configurata con un approccio *multi-head*.

Anche in questo caso è stata eseguita una ricerca degli iperparametri. Lo spazio di ricerca esplorato è riportato nella Tabella 4.2.

Tabella 4.2: Spazio di ricerca degli iperparametri per GATv2.

Iperparametro	Spazio di ricerca
HIDDEN_DIM	{32, 64, 128}
NUM_LAYERS	{1, 2, 3}
DROPOUT	{0.2, 0.3, 0.4}
LEARNING_RATE	{0.001, 0.005}
HEADS	{2, 3, 4}

Attraverso la libreria *Optuna*, la combinazione che ha massimizzato le prestazioni di addestramento è risultata la seguente:

- `HIDDEN_DIM = 128`;
- `NUM_LAYERS = 2`;
- `DROPOUT = 0.2`;
- `LEARNING_RATE = 0.005`;
- `HEADS = 4`.

Dopo aver considerato un'architettura basata sull'attenzione, è utile introdurre un modello più leggero e induttivo, utilizzato in questa tesi come ulteriore termine di paragone.

4.4 Architettura 3: SAGEConv (GraphSAGE)

La terza e ultima architettura analizzata è *GraphSAGE*, implementata tramite SAGEConv, abbreviazione di *Graph Sample and Aggregate* [14].

Questa architettura è stata progettata per apprendere una funzione di aggregazione generalizzabile, in grado di produrre *embedding* anche per nodi non osservati durante il *training*, ed è stata introdotta per risolvere vulnerabilità algebriche delle reti esposte a tessuti con densità cellulare altamente variabile.

Tale proprietà risulta particolarmente interessante nel contesto delle biopsie reumatoidi, in cui aree stromali relativamente sparse coesistono con aggregati linfoidi estremamente densi.

Per prevenire che il segnale del nodo venga diluito o eccessivamente influenzato dall'aggregazione di un vicinato troppo popolato, SAGEConv non somma direttamente l'identità del nodo a quella dei vicini, ma effettua una concatenazione tensoriale.

GraphSAGE mantiene così una chiara distinzione algebrica tra il fenotipo intrinseco della cellula *target* e il segnale derivante dal contesto ambientale, riducendo anche la complessità del calcolo dei gradienti nelle zone ad alta densità.

Essa implementa inoltre un algoritmo di campionamento del vicinato, noto come *Neighborhood Sampling*, che seleziona un sottoinsieme fisso di nodi adiacenti. Questa regolarizzazione rende il modello computazionalmente scalabile e robusto nell'interpretazione dei dati.

In definitiva, SAGEConv rappresenta l'architettura più leggera delle tre, priva di *Batch Normalization* e di meccanismi di attenzione, costituendo la *baseline* per valutare se la maggiore complessità delle altre due GNN produca un vantaggio discriminativo reale.

La formulazione matematica dell'aggiornamento è la seguente:

$$h_i^{(l)} = \text{ReLU} \left(W^{(l)} \left[h_i^{(l-1)} \parallel h_{N(i)}^{(l)} \right] + b^{(l)} \right) \quad (4.3)$$

L'elemento distintivo di *GraphSAGE* risiede proprio in questa concatenazione, che consente di mantenere una chiara separazione tra informazione interna e informazione contestuale.

Anche per questa architettura è stata applicata la medesima metodologia di ricerca degli iperparametri. Lo spazio esplorato è riportato nella Tabella 4.3.

Tabella 4.3: Spazio di ricerca degli iperparametri per GraphSAGE.

Iperparametro	Spazio di ricerca
HIDDEN_DIM	{32, 64, 128, 256, 512}
NUM_LAYERS	{1, 2, 3}
DROPOUT	{0.2, 0.3, 0.4}
LEARNING_RATE	{0.001, 0.005}

Tramite ottimizzazione con *Optuna*, la configurazione migliore è risultata:

- HIDDEN_DIM = 512;
- NUM_LAYERS = 2;
- DROPOUT = 0.2;
- LEARNING_RATE = 0.005.

Dopo aver descritto separatamente le tre architetture, è utile evidenziarne le differenze strutturali fondamentali.

4.5 Confronto Strutturale tra le Tre Architetture

Le tre architetture considerate condividono lo stesso obiettivo generale, ossia aggiornare la rappresentazione del nodo sfruttando le informazioni provenienti dal vicinato. Tuttavia, differiscono nel modo in cui definiscono, pesano e integrano tali contributi.

In sintesi:

- la **MPNN** enfatizza la costruzione esplicita del messaggio;
- la **GATv2** introduce una ponderazione adattiva del vicinato;
- la **GraphSAGE** privilegia una struttura più semplice e robusta, adatta a contesti con alta densità cellulare.

Una volta evidenziate le differenze architetturali, è necessario descrivere il protocollo di *training*, validazione e selezione degli iperparametri impiegato nello studio.

4.6 Pipeline di Training, Validazione e Ottimizzazione

Considerata la limitata dimensione della coorte, composta da soli 5 pazienti, un addestramento direttamente supervisionato avrebbe esposto il modello a un rischio elevato di *overfitting*. Per questo motivo, è stato adottato un approccio di apprendimento auto-supervisionato basato sul *Masked Autoencoder* (MAE) [15].

In questo schema, ciascuna GNN viene addestrata a ricostruire le *feature* originali dei nodi a partire da una versione parzialmente mascherata delle stesse. La supervisione proviene quindi direttamente dai dati e non da etichette esterne.

Una volta fissato il *masking*, con valori pari a $\{0.1, 0.2, 0.3, 0.5\}$, vengono selezionati gli indici da mascherare a partire da rumore casuale generato per ciascuna *feature*.

La funzione di perdita utilizzata è la *Mean Squared Error*, calcolata sui nodi di *training*:

$$L(\theta) = \frac{1}{|T_{\text{train}}| \cdot F} \sum_{i \in T_{\text{train}}} \sum_{f=1}^F (\hat{z}_{i,f} - x_{i,f})^2 \quad (4.4)$$

I parametri vengono ottimizzati tramite l'algoritmo *Adam* [39]. Il *learning rate* viene adattato mediante *scheduler* *ReduceLRonPlateau*, che lo dimezza nel caso in cui la perdita di validazione non migliori per tre valutazioni consecutive.

Per la selezione degli iperparametri è stata impiegata la libreria *Optuna* [40], che adotta il campionatore TPE (*Tree-structured Parzen Estimator*). Inoltre, per ridurre il costo computazionale, il *pruning* dei *trial* viene effettuato tramite *MedianPruner*, con `n_warmup_steps = 2`.

A seguito del *training*, l'importanza delle *feature f* viene stimata tramite *permutation importance* [41, 42]. Le importanze vengono infine normalizzate.

Per ciascuna architettura, il *training* viene ripetuto su 5 *run* con semi diversi per ciascuno dei 4 valori di *masking*, producendo 20 modelli indipendenti per architettura. Il punteggio globale delle *feature* viene quindi calcolato tramite media pesata, con pesi inversamente proporzionali al tasso di *masking*.

Per quanto riguarda il protocollo *Leave-One-Patient-Out Cross-Validation* [35], esso esegue 5 iterazioni. Al termine del *training*, gli *embedding* dei nodi estratti all'ultimo *layer* vengono aggregati per ROI tramite *Global Mean Pooling* e successivamente aggregati per media a livello paziente.

Su queste rappresentazioni viene infine applicata una regressione logistica con penalizzazione L2 e `class_weight = balanced`.

La metrica finale di valutazione è l'AUC-ROC, che misura la capacità del modello di assegnare un punteggio maggiore ai pazienti *responder* rispetto agli *inadequate responder*. Nel protocollo LOPO-CV, un valore pari a 1 indica separabilità perfetta tra le classi, mentre un valore pari a 0.5 indica classificazione casuale.

Cosicché, il *framework* descritto nel presente capitolo, consente di affrontare in modo sistematico tre differenti strategie di aggregazione, valutandone sia la capacità di ricostruzione *self-supervised*, sia il potenziale rispetto alla risposta terapeutica.

Dunque, definite le architetture e la *pipeline* di apprendimento, è possibile di seguito utilizzare le rappresentazioni apprese per stimare l'importanza dei biomarcatori e guidare la selezione delle *feature*.

Capitolo 5

Marker Selection e Analisi dell'Importanza delle Feature

Il pannello molecolare analizzato in questo studio è composto da 25 marcatori proteici, ai quali si aggiungono 4 *feature* morfologiche, per un totale di 29 dimensioni del vettore costruito per ciascuna cellula segmentata nelle ROI.

Sebbene tale dimensionalità sia relativamente contenuta rispetto ad altri domini dell'analisi multi-omica, come ad esempio la trascrittomica a singola cellula, essa pone un problema di primaria importanza nel contesto clinico dell'Artrite Reumatoide. In particolare, la disponibilità pratica di un pannello diagnostico risulta inversamente proporzionale alla sua complessità e al suo costo di esecuzione.

A questa considerazione si aggiunge una seconda esigenza di natura metodologica. L'addestramento di architetture neurali, come in questo caso delle *Graph Neural Networks* applicate a dati ad alta dimensionalità spaziale, solleva il problema della cosiddetta *black box*. In ambito clinico, infatti, sebbene i modelli di *Deep Learning* possano garantire performance elevate, la sola accuratezza non è sufficiente, poiché risulta necessario comprendere quali variabili risultino realmente importanti nelle decisioni dell'algoritmo.

Alla luce di queste considerazioni, il presente capitolo ha il compito di illustrare l'applicazione delle tecniche adottate per quantificare il peso di ciascun marcatore, effettuando una procedura di *feature selection* finalizzata all'identificazione di un *subset* ottimale [42].

Per chiarire la logica della procedura adottata, si introducono dapprima le motivazioni cliniche e computazionali alla base della riduzione dei marcatori, per poi descrivere la metodologia di *feature selection* e la definizione del *subset* ottimale.

5.1 Motivazione Clinica della Riduzione dei Marcatori

L'utilizzo della tecnologia di *Imaging Mass Cytometry* consente l'esplorazione simultanea di un ampio pannello di marcatori, offrendo una visione estremamente estesa del tessuto sinoviale. Tuttavia, tale vantaggio è accompagnato da un costo elevatissimo per singola analisi e dalla necessità di una competenza tecnica altamente specializzata, sia nella fase di acquisizione sia in quella di *pre-processing*.

Questa osservazione rende evidente un primo punto: un eventuale *test* diagnostico basato su biopsia sinoviale può diventare concretamente trasferibile nella pratica clinica solo se supportato da un pannello ridotto di marcatori.

All'interno del sistema ospedaliero tradizionale, infatti, la fenotipizzazione tissutale per la stratificazione dei pazienti reumatoidi si affida a tecniche di immunoistochimica o immunofluorescenza multiplexata, ottimizzate per analisi rapide e a basso costo, ma limitate a un numero ristretto di *target*. Per questo motivo, un *test* diagnostico basato su un numero elevato di marcatori risulterebbe economicamente dispendioso e incompatibile con i tempi richiesti dalla strategia terapeutica *Treat-to-Target* [4].

Accanto a questa motivazione clinica, emerge una motivazione di natura matematica, riconducibile alla *curse of dimensionality* [43]. In un *dataset* con numero limitato di pazienti, ma caratterizzato da elevata parametrizzazione e da *feature* ridondanti, che spesso forniscono la medesima informazione biologica, il rischio di *overfitting* aumenta drasticamente. In tali condizioni, il modello può attribuire importanza al rumore di fondo invece che al reale segnale patogenetico.

L'approccio convenzionale alla *feature selection* nei contesti di biomarcatori tissutali si basa tipicamente su metodologie statistiche, come il *t-test*, applicate direttamente alle intensità di espressione dei singoli marcatori, oppure su modelli ad albero come la *Random Forest* [41]. Tali approcci, sebbene consolidati, trascurano un'informazione che, nel contesto tissutale, è fondamentale: la struttura spaziale del microambiente.

Nel caso del tessuto sinoviale, infatti, un marcatore può essere espresso ad alta intensità in uno specifico tipo cellulare, ma il suo valore biologico non dipende soltanto dal livello di espressione. Esso dipende anche da fattori quali le cellule che lo circondano, il *pattern* topologico del tessuto e le interazioni che la cellula intrattiene con le cellule adiacenti.

Da questo punto di vista, tali informazioni risultano inaccessibili alle metodologie che trattano le cellule come unità indipendenti. Al contrario, le GNN apprendono le relazioni di vicinato tra le cellule e consentono quindi di comprendere l'importanza del marcatore nel contesto, e non soltanto nella singola cellula [30].

Questo rende la *feature selection* guidata da GNN superiore agli approcci tradizionali nel contesto del presente studio, con la finalità non di identificare semplicemente i marcatori con maggiore variabilità espressiva tra pazienti, ma quelli la cui rimozione dal vettore delle *feature* provoca il maggiore deterioramento delle prestazioni della rete neurale [30, 42].

La riduzione del pannello non deve quindi essere interpretata come una semplificazione tecnica, ma come un obiettivo scientifico e clinico. Isolare una firma molecolare minima significa infatti ridurre il rumore, aumentare la robustezza del modello e facilitare la futura generalizzazione su nuovi pazienti [35].

Come conseguenza di queste motivazioni, il presente capitolo persegue tre obiettivi principali:

1. identificare, tra i 25 marcatori del pannello, quelli che le reti neurali a grafo considerano maggiormente informativi per caratterizzare in modo differenziato il microambiente tissutale sinoviale tra pazienti *Responder* e *Inadequate Responder* [10];
2. costruire un *subset* diagnostico ottimale che conservi la capacità discriminativa del pannello completo e sia, al tempo stesso, di dimensione compatibile con future validazioni cliniche [3];
3. dimostrare la robustezza dei marcatori selezionati, verificando che essi emergano in modo coerente attraverso le tre architetture GNN, i quattro tassi di *masking* e i cinque semi casuali di inizializzazione, escludendo che il risultato sia un artefatto di una singola configurazione sperimentale [35, 44].

Una volta chiarita la motivazione della riduzione dei marcatori, è ora possibile descrivere il criterio con cui l'importanza delle *feature* è stata stimata e aggregata.

5.2 Metodologia di Feature Selection

Al fine di identificare e classificare l'importanza biologica dei marcatori, superando i *bias* intrinseci di ogni singola architettura neurale, è stata implementata una *pipeline* di *ensemble feature selection*, basata sull'algoritmo di *Permutation Feature Importance* [41, 42].

Questo approccio è *model-agnostic*, non richiede modifiche all'architettura della rete neurale ed è applicabile in modo identico a tutte e tre le architetture implementate.

Il suo funzionamento si basa sulla perturbazione di una *feature* all'interno del calcolo, causando un abbassamento delle performance del modello. Il risultato finale indica quanto una *feature* sia importante in funzione della riduzione di performance osservata.

L'algoritmo implementato può essere riassunto nel seguente primo passaggio, che consiste nel calcolare la perdita del modello sul *test set* con i dati originali non modificati:

$$L_{\text{base}} = \text{MSE}(f_{\theta}(X, E)_{T_{\text{test}}}, X_{T_{\text{test}}}) \quad (5.1)$$

Successivamente, per ciascuna *feature* vengono eseguite cinque permutazioni indipendenti, andando a permutare casualmente i valori di quella colonna in tutti i nodi del grafo.

Per ogni permutazione si ricalcola la perdita e la si confronta con quella originale, assegnando il valore più elevato alla *feature* più critica, mentre valori vicini a zero indicano *feature* scarsamente informative.

Dal punto di vista interpretativo, per ogni marcatore il vettore delle espressioni associato ai nodi del grafo viene permutato casualmente, alterando in modo irreversibile la corretta localizzazione spaziale della proteina nel tessuto. Successivamente, la GNN pre-addestrata viene interrogata utilizzando un *set* di dati corrotto [42].

Se il marcatore è fondamentale per comprendere l'architettura patologica, la sua permutazione causerà un elevato incremento dell'Errore Quadratico Medio nella fase di ricostruzione del tessuto. Al contrario, se il marcatore risulta ininfluenza o non particolarmente indispensabile, la *loss* rimarrà pressoché invariata [42].

Dopo il calcolo dell'importanza per permutazione, è stata introdotta una seconda fase di aggregazione multicomponente, che combina quattro diverse fonti di evidenza indipendenti [45].

Per ciascun marcatore, il punteggio assegnato è una combinazione lineare pesata delle seguenti quattro componenti:

- **Importanza differenziale** (peso 40%), che misura la differenza assoluta tra l'importanza media nei pazienti *Responder* e *Inadequate Responder*, normalizzata tra tutti i marcatori [10];
- **Significatività statistica** (peso 30%), che converte il *p-value* del *permutation test* a quattro livelli;
- **Effect size** (peso 20%), basata sul coefficiente di Cohen's *d* [46], che quantifica l'entità della separazione pratica tra i due gruppi;
- **Importanza ML** (peso 10%), integrata tramite *Random Forest* e *Gradient Boosting*, addestrate sulle importanze GNN come vettori di *feature* per la classificazione.

Il punteggio finale del marcatore è definito come:

$$S(i) = 0.4 \cdot S_{\text{diff}}(i) + 0.3 \cdot S_{\text{sig}}(i) + 0.2 \cdot S_{\text{effect}}(i) + 0.1 \cdot S_{\text{ML}}(i) \quad (5.2)$$

I marcatori vengono successivamente ordinati in base al punteggio e assegnati a quattro livelli gerarchici di rilevanza, denominati *tier*, come segue:

- **Tier 1:** *critical*;
- **Tier 2:** *important*;
- **Tier 3:** *useful*;
- **Tier 4:** *optional*.

L'assegnazione viene eseguita indipendentemente per ciascun *mask ratio*, con la finalità di ottenere una classifica globale consensuale. A tal fine, i risultati delle tre architetture vengono dapprima normalizzati tramite *Z-score* e successivamente combinati con un voto di maggioranza, ottenendo una *consensus ranking* finale [45].

Questa graduatoria rappresenta l'esito conclusivo della fase di interpretabilità e costituisce il punto di partenza per la selezione del *subset* ottimale.

Per verificare che i marcatori individuati non siano soltanto interpretabili, ma anche utili dal punto di vista predittivo, è stata quindi introdotta una fase di validazione discriminativa.

La validazione finale del ranking è stata condotta utilizzando un classificatore lineare di *Logistic Regression* con penalizzazione L2, valutato mediante protocollo *Leave-One-Patient-Out Cross-Validation*. L'uso di questo modello garantisce una discriminazione globale stabile, limita il rischio di *overfitting* e consente di verificare se il *subset* ottimale possieda una capacità predittiva superiore rispetto all'utilizzo del *set* completo di marcatori [35].

Tale scelta è stata preferita a modelli più complessi, come la *Random Forest*, in quanto, in scenari con numerosità campionaria molto ridotta, la RF rischierebbe di generare regole spurie, memorizzando il rumore di fondo e producendo indicazioni fuorvianti su pazienti futuri.

Una volta ottenuta la classifica globale di consenso, è quindi possibile analizzare come le prestazioni cambino al crescere del numero di marcatori inclusi.

5.3 Identificazione del Subset Ottimale

Dopo aver stilato la classifica globale di consenso, l'obiettivo si è spostato sulla determinazione del *Subset Ottimale*, definito come il minor numero di marcatori strettamente necessario per ottenere una classificazione perfetta in grado di distinguere i fenotipi *Responder* e *Inadequate Responder*, senza incorrere in perdita di informazione o in ridondanza.

Questo passaggio non si basa sul risultato di una singola architettura, ma, come descritto in precedenza, sulla convergenza dei risultati ottenuti attraverso le tre reti neurali considerate:

- la **MPNN** [12], che interroga le interazioni asimmetriche tra coppie di cellule adiacenti;
- la **GATv2** [13], che apprende quali vicini siano informativamente più rilevanti per ciascuna cellula;
- la **SAGEConv** [14], che estrae *pattern* strutturali di vicinato in modo induttivo.

La finalità è fare emergere marcatori considerati importanti da tutti e tre i meccanismi di aggregazione, piuttosto che marcatori rilevanti soltanto per una singola architettura [45].

Per selezionare il *subset* ottimale, è stata eseguita una procedura di *Forward Feature Selection*, utilizzando la classifica di consenso prodotta dalle GNN. Questo approccio riduce la complessità computazionale e rende la procedura eseguibile senza costi aggiuntivi rilevanti, a differenza di una ricerca esaustiva su tutti i possibili sottoinsiemi, che sarebbe risultata computazionalmente proibitiva.

La scelta metodologica di tale approccio si basa sull'ipotesi che, se la classifica GNN è effettivamente informativa, l'aggiunta dei marcatori più rilevanti debba produrre un rapido incremento dell'AUC nelle prime iterazioni, seguito da una stabilizzazione o da una riduzione della performance quando iniziano a essere aggiunti marcatori ridondanti o rumorosi.

La procedura di *forward selection* valuta, per ciascun valore k , le prestazioni classificatorie del *subset* composto dai primi k marcatori in cima alla classifica GNN. Tali prestazioni vengono misurate tramite protocollo *Leave-One-Patient-Out Cross-Validation* a livello ROI, utilizzando una regressione logistica con penalizzazione L2 come classificatore finale [35].

Per ogni iterazione del protocollo, il paziente p_i costituisce il *test set* e i restanti pazienti costituiscono il *training set*. Inoltre, la standardizzazione viene applicata separatamente sul *training set* e solo successivamente trasferita al *test set*, in modo da prevenire il *data leakage*.

La formulazione del classificatore è la seguente:

$$\hat{y}_{p_i}^{(k)} = \sigma \left(w_k^\top \bar{x}_{(p_i, S_k)} + b_k \right) = \frac{1}{1 + \exp \left(- \left(w_k^\top \bar{x}_{(p_i, S_k)} + b_k \right) \right)} \quad (5.3)$$

dove $\bar{x}_{(p_i, S_k)}$ rappresenta il vettore delle medie di espressione per ROI del paziente p_i , limitato ai marcatori contenuti in S_k e calcolato sull'intero profilo cellulare delle ROI del *test set*.

Successivamente, l'AUC-ROC viene aggregata su tutte le iterazioni del protocollo LOPO.

L'analisi dell'AUC, ottenuta aggiungendo cumulativamente i marcatori, ha confermato l'impatto della *curse of dimensionality* [43]. Utilizzando l'intero pannello di marcatori, il modello ha registrato un'AUC pari a 0.667, dimostrando che l'eccesso di proteine ha saturato lo spazio informativo, generando rumore e mascherando il reale segnale clinico.

Al contrario, il *subset* individuato dall'algoritmo tramite l'*elbow point* ha identificato la miglior configurazione con 4 marcatori, corrispondenti ai primi quattro biomarcatori della classifica spaziale, raggiungendo una classificazione diagnostica pari a **AUC = 1.000**.

La progressione delle prestazioni al crescere del numero di marcatori è riportata nella Tabella 5.1.

Tabella 5.1: Andamento dell'AUC nella *forward feature selection*.

k	Marcatore aggiunto	AUC
1	aSMA	0.833
2	CD45RO	0.667
3	CD45	0.833
4	CD138	1.000
5	CD68	0.833
6	CD44	1.000
7	CD163	1.000
8	VEGF	0.833
9	CD90	0.833
10	PODOPLANIN	0.667
11	CD45RA	0.500
12	MERTK	0.500
13	CD3	0.500
14	HLA-DR	0.667
15	CD4	0.667
16	COL1	0.833
17	Ki67	0.500
18	CD31	0.500
19	CD8a	0.500
20	CD16	0.667
21	OPN	0.500
22	CD34	0.667
23	CD14	0.500
24	CD20	0.667
25	GRANZYMEB	0.667

I marcatori selezionati all'interno del *subset* ottimale sono quindi:

$$\{aSMA, CD45RO, CD45, CD138\}$$

Dal punto di vista biologico, questo *subset* mostra una notevole complementarità informativa [3, 8]. In particolare:

- **aSMA** quantifica il rimodellamento stromale e i miofibroblasti invasivi;
- **CD45** rappresenta un marcatore pan-leucocitario utile per misurare la densità dell'infiltrato globale;
- **CD45RO** identifica cellule adattive della memoria;
- **CD138** segnala l'infiltrazione di plasmacellule secernenti anticorpi.

Formalmente, il valore ottimale è definito come:

$$k^* = \min \left\{ k \in \{1, \dots, 25\} \mid AUC(k) = \max_{j=1, \dots, 25} AUC(j) \right\} = 4 \quad (5.4)$$

L'insieme di questi quattro marcatori suggerisce che la separazione tra *Responder* e *Inadequate Responder* emerga dall'interazione tra componente stromale, infiltrato immunitario globale e asse dell'immunità adattiva.

L'osservazione che l'aggiunta di ulteriori marcatori non migliori, ma, in molti casi, riduca drasticamente la performance del modello, raggiungendo valori di AUC pari a 0.5, quindi compatibili con una classificazione casuale, indica che i marcatori a basso ranking non sono semplicemente meno utili, ma introducono rumore e ostacolano la separazione lineare dei due gruppi clinici.

Dopo aver discusso la riduzione del pannello proteico, è opportuno considerare il contributo delle *feature* morfologiche, che rappresentano una seconda sorgente di informazione integrata nel grafo cellulare.

5.4 Ruolo delle Feature Morfologiche

Parallelamente all'esplorazione dei marcatori proteici, l'analisi di interpretabilità ha evidenziato anche il contributo delle quattro metriche spaziali e morfologiche fornite alle reti neurali durante la costruzione dei grafi.

La scelta di mantenere le *feature* biologiche separate dai *biomarker* nell'analisi dell'importanza riflette una distinzione biologica fondamentale. I *biomarker* codificano infatti lo stato funzionale e fenotipico della cellula, ossia quante proteine esprime e il suo stato di attivazione, mentre le *feature* morfologiche codificano il suo stato fisico e geometrico, ossia quanto è grande, quanto è allungata e quanto è circolare [47].

Le due classi di *feature* sono state mantenute separate poiché risultano biologicamente indipendenti e non sovrapponibili in modo diretto, pur potendo essere considerate complementari, nel senso che entrambe contribuiscono a caratterizzare il fenotipo cellulare all'interno del microambiente tissutale.

Per questa ragione, pur risultando meno predittive dei marcatori proteici, le *feature* morfologiche sono state utilizzate all'interno delle GNN come modulatori contestuali, poiché l'inclusione della geometria cellulare rende possibile il disaccoppiamento di profili proteici simili, ma associati a stati cellulari differenti [8].

Le quattro *feature* considerate, come già descritto nel Capitolo 3, sono:

- **Area;**
- **Axis_major_length;**
- **Axis_minor_length;**
- **Eccentricity.**

La Tabella 5.2 riassume i risultati ottenuti per le *feature* morfologiche, riportando i valori medi osservati nei due gruppi clinici e il *rank* medio risultante dall'integrazione delle tre architetture.

Tabella 5.2: Ranking delle *feature* morfologiche.

Feature	Valore Responder	Valore IR	Rank medio
eccentricity	0.8191	0.8772	1.6667
area	0.4070	0.4325	1.8333
axis_minor_length	0.4334	0.4470	3.0000
axis_major_length	0.5266	0.5431	3.5000

I risultati mostrano che, in media, le cellule appartenenti ai pazienti *Inadequate Responder* tendono a essere più grandi e più eccentriche rispetto a quelle dei *Responder*.

In particolare, l'eccentricità emerge come la *feature* morfologica più rilevante, suggerendo che l'allungamento cellulare possa riflettere stati di attivazione o rimodellamento associati a un microambiente patologico più aggressivo.

Sebbene queste *feature* non sostituiscano i marcatori proteici nella discriminazione dei fenotipi clinici, il loro contributo appare coerente con l'ipotesi che la risposta terapeutica dipenda non solo dalla composizione molecolare del tessuto, ma anche dalla sua organizzazione strutturale.

Pertanto, dall'analisi condotta nel presente capitolo si evince che la selezione delle *feature* guidata da GNN consente di identificare una firma ridotta, ma altamente informativa, nella quale le informazioni molecolari e morfologiche contribuiscono in modo complementare alla caratterizzazione del microambiente sinoviale.

Orbene, i risultati ottenuti dalla procedura di selezione costituiscono la base per l'analisi sperimentale, presentata nel capitolo successivo, in cui viene valutata la capacità discriminativa del *subset* identificato.

Capitolo 6

Risultati Sperimentali e Interpretabilità del Modello

Il presente capitolo descrive in dettaglio i risultati ottenuti dall'applicazione della *pipeline* al *dataset* dei pazienti affetti da AR, con l'obiettivo di identificare, attraverso un approccio privo di assunzioni a priori, un insieme di biomarcatori proteici e caratteristiche morfologiche cellulari capaci di discriminare con affidabilità i pazienti con risposta positiva al trattamento dai pazienti resistenti.

Come affrontato nei capitoli precedenti, il *dataset* è composto da cinque pazienti, due dei quali *Responder* (RA_039 e RA_046) e tre *Inadequate Responder* (RA_016, RA_051 e RA_056).

Da essi sono state estratte complessivamente 42 ROI, 20 appartenenti al gruppo *Responder* e 22 al gruppo *inadequate responder*, per un totale di 209 652 cellule processate, ciascuna descritta da 25 biomarcatori e 4 caratteristiche morfologiche, con un vettore finale di 29 *feature* per ogni nodo del grafo.

Ciascuna delle tre architetture analizzate è stata addestrata con un meccanismo di mascheramento e ricostruzione, eseguendo 5 *run* indipendenti per ciascun *mask ratio*, per un totale di 20 modelli distinti per architettura.

I risultati finali comprendono sia quelli ottenuti a livello individuale per ogni singola architettura sia quelli legati all'aggregazione delle tre in un consenso globale.

Dopo aver definito il contesto sperimentale, è ora possibile confrontare il comportamento delle tre architetture considerate.

6.1 Confronto tra le tre Architetture GNN

La prima fase della valutazione ha previsto un'analisi comparativa della *performance* dei diversi modelli di apprendimento utilizzati, ovvero delle tre architetture di *deep learning*

geometrico implementate: MPNN, GATv2 e *GraphSAGE*.

Tutte e tre le reti sono state sottoposte al medesimo *task Self-Supervised Masked Node Reconstruction*, con l'obiettivo di minimizzare l'Errore Quadratico Medio nella previsione di un profilo oscurato [15].

Sebbene condividano lo stesso obiettivo di apprendimento, le tre architetture differiscono per il modo in cui aggregano l'informazione proveniente dal vicinato cellulare. Per questa ragione, il confronto tra i ranking prodotti dai modelli consente di valutare se i biomarcatori più importanti emergano in modo consistente oppure se alcune architetture risultino più sensibili a specifiche configurazioni del microambiente.

L'architettura MPNN implementa nel modo più diretto il paradigma generale del passaggio di messaggi, nel quale, a ogni *layer*, ciascun nodo aggrega i messaggi ricevuti dai nodi vicini tramite una somma ponderata con trasformazione lineare, aggiornando poi la propria rappresentazione tramite una funzione di *update* [12].

I risultati del ranking dei biomarcatori ottenuti tramite l'architettura MPNN sono riportati nella Tabella 6.1.

Tabella 6.1: Ranking dei biomarcatori ottenuto tramite architettura MPNN.

Consensus Rank	Marker	Consensus Tier	Phenotype	Avg Score
1	aSMA	TIER 1 (Critical)	IR-ENRICHED	0.8323
2	CD45RO	TIER 1 (Critical)	RESPONDER	0.8239
3	CD68	TIER 1 (Critical)	IR-ENRICHED	0.7890
4	CD163	TIER 2 (Important)	IR-ENRICHED	0.7513
5	CD138	TIER 2 (Important)	RESPONDER	0.7392
6	CD45	TIER 1 (Critical)	RESPONDER	0.6571

Dalla Tabella 6.1 emerge che i marcatori più rilevanti siano aSMA, CD45RO, CD68, CD163 e CD45, con un *focus* particolare su aSMA, che occupa la prima posizione e suggerisce come la componente stromale e il rimodellamento tissutale rappresentino un asse centrale nel fenotipo *Inadequate Responder*.

Parallelamente, la presenza nelle prime posizioni di CD45RO, CD138 e CD45 suggerisce che il modello riconosca anche un asse immunitario associato ai pazienti *Responder*, legato alla presenza e all'organizzazione di cellule immunitarie.

Nel complesso, il ranking MPNN mostra una struttura bilanciata tra marcatori stromali, mieloidi e immunitari, confermando la capacità del modello di catturare il microambiente come rete di interazioni tra componenti differenti.

L'architettura GATv2 rappresenta un'evoluzione della classica GAT, introducendo un meccanismo di attenzione dinamica nel quale il peso assegnato a ciascun nodo non dipende solo dalle *feature* del nodo stesso, ma anche da quelle dei nodi vicini, rendendo il modello capace di apprendere relazioni asimmetriche e dipendenti dal contesto nel grafo cellulare [13].

I risultati del ranking dei biomarcatori ottenuti tramite l'architettura GATv2 sono riportati nella Tabella 6.2.

Tabella 6.2: Ranking dei biomarcatori ottenuto tramite architettura GATv2.

Consensus Rank	Marker	Consensus Tier	Phenotype	Avg Score
1	CD45	TIER 1 (Critical)	RESPONDER	0.8371
2	CD45RO	TIER 1 (Critical)	RESPONDER	0.8253
3	aSMA	TIER 1 (Critical)	IR-ENRICHED	0.8253
4	CD138	TIER 2 (Important)	RESPONDER	0.8094
5	CD44	TIER 2 (Important)	RESPONDER	0.7722
6	CD3	TIER 3 (Useful)	RESPONDER	0.7286

Dalla Tabella 6.2 si osserva che i primi quattro marcatori coincidono con quelli presenti nella *subset* ottimale, ossia CD45, CD45RO, CD138 e aSMA.

A differenza della MPNN, qui emerge con maggiore forza il compartimento immunitario adattativo, con CD45 e CD45RO in cima alla classifica, suggerendo che il meccanismo di attenzione riesca a enfatizzare *pattern* relazionali che coinvolgono cellule immunitarie attivate.

La presenza di aSMA conferma comunque l'importanza della componente stromale anche in questo modello.

Nel complesso, GATv2 fornisce quindi una prospettiva complementare rispetto alla MPNN, maggiormente orientata a catturare relazioni contestuali e differenze locali.

L'architettura *GraphSAGE* adotta invece un approccio basato sul campionamento locale dei vicini e sull'aggregazione tramite una media pesata, nella quale il modello apprende come combinare la rappresentazione del nodo stesso con un campione aggregato dei vicini, risultando scalabile su grafi di grandi dimensioni [14].

Nel contesto della biologia cellulare, essa tende a catturare proprietà di volume e densità locale del tessuto, ossia il *pattern* medio di espressione proteica del microambiente di ciascuna cellula.

I risultati del ranking dei biomarcatori ottenuti tramite l'architettura *GraphSAGE* sono riportati nella Tabella 6.3.

Tabella 6.3: Ranking dei biomarcatori ottenuto tramite architettura GraphSAGE.

Consensus Rank	Marker	Consensus Tier	Phenotype	Avg Score
1	aSMA	TIER 1 (Critical)	IR-ENRICHED	0.8312
2	CD45RO	TIER 1 (Critical)	RESPONDER	0.8198
3	CD138	TIER 2 (Important)	RESPONDER	0.7826
4	CD45	TIER 2 (Important)	RESPONDER	0.6581
5	CD68	TIER 1 (Critical)	IR-ENRICHED	0.6415
6	CD44	TIER 1 (Critical)	RESPONDER	0.6264

Dalla Tabella 6.3 si può notare come le prime posizioni siano occupate da aSMA, CD45RO, CD138, CD45 e CD68, facendo emergere anche in questo caso due assi principali: da un lato la componente stromale associata a refrattarietà terapeutica, dall'altro la componente immunitaria associata ai *responder*.

Pertanto, *GraphSAGE* appare vicino alla MPNN nella struttura del ranking, ma con una maggiore enfasi sui *pattern* medi di vicinato.

Le tabelle riportate per ciascuna architettura contengono, in linea generale, il *rank* del marcatore, il livello gerarchico di importanza, denominato *Tier*, il fenotipo prevalente e uno *score* medio aggregato.

Quest'ultimo non rappresenta la semplice abbondanza del marcatore, ma una misura composta che integra informazione statistica e informazione appresa dal modello. Nel dettaglio, il ranking finale deriva dalla combinazione di quattro componenti:

- significatività statistica;
- *effect size*;
- importanza stimata con modelli classici;
- *permutation feature importance*, cioè l'incremento della *loss* quando il marcatore viene corrotto.

Di conseguenza, il fenotipo associato a ciascun marcatore non deve essere interpretato come una semplice differenza di espressione grezza, ma come una direzione biologica del contributo del marcatore alla struttura del grafo.

Pertanto, un marcatore associato agli *Inadequate Responder* indica che la sua alterazione compromette maggiormente la coerenza del tessuto in quel gruppo, suggerendo un ruolo rilevante nella resistenza terapeutica.

Per completezza, i *plot* associati a ciascun modello riassumono in maniera grafica:

- il ranking di consenso dei biomarcatori;
- la stabilità del ranking tra *run* diverse;
- il ranking delle *feature* morfologiche;
- i marcatori *top-ranked*;
- le differenze medie tra *Responder* e *Inadequate Responder*.

Dopo aver analizzato separatamente i tre modelli, è utile confrontare anche la configurazione ottimale e il grado di convergenza dei ranking ottenuti.

La Tabella 6.4 riassume gli iperparametri selezionati per le tre architetture tramite ottimizzazione con *Optuna*.

Tabella 6.4: Configurazione ottimale delle tre architetture.

Caratteristica	GATv2	SAGEConv	MessagePassing
Tipo di aggregazione	Attenzione multi-head (heads=4)	Campionamento + media pesata	Somma con trasformazione lineare
Hidden dim	128	512	256
Numero di layer	2	2	2
Dropout	0.20	0.20	0.20
Learning rate	0.005	0.005	0.001
Heads	4	N/A	N/A
R^2 validazione	0.8957	0.9011	0.8913
Ottimizzazione	Optuna, 50 trial	Optuna, 50 trial	Optuna, 50 trial

La Tabella 6.4 mostra che le *performance* di validazione dei tre modelli risultano molto simili. In particolare, *GraphSAGE* ottiene il valore di R^2 più elevato, seguito da GATv2 e MPNN. Tuttavia, le differenze sono contenute, rafforzando l'idea di adottare un approccio *ensemble* nella successiva integrazione dei ranking.

Per quantificare la convergenza tra i tre modelli, è stata calcolata la correlazione di Spearman tra i vettori di *rank* prodotti da ciascuna architettura, riportata nella Tabella 6.5.

Tabella 6.5: Correlazione di Spearman tra i ranking delle tre architetture GNN.

	SAGEConv	GATv2	MPNN
SAGEConv	1.000	0.684	0.888
GATv2	0.684	1.000	0.709
MPNN	0.888	0.709	1.000

I valori di correlazione della Tabella 6.5 mostrano che SAGEConv e MPNN condividono la correlazione più elevata, indicando che catturano segnali biologici simili, mentre GATv2 presenta correlazioni più basse, suggerendo che il meccanismo di attenzione introduca una prospettiva complementare e non ridondante.

Questo risultato valida l'ipotesi che l'*ensemble* dei tre modelli catturi aspetti biologici diversi del tessuto sinoviale e che il consenso globale sia più robusto rispetto all'*output* di un singolo modello [45].

Dato tale scenario, la sezione successiva presenta l'integrazione dei ranking in una classifica globale di consenso, che costituirà la base per l'identificazione dei biomarcatori più rilevanti.

6.2 Analisi di Stabilità e Rank Consistency

Nel contesto della *biomarker discovery*, l'analisi di stabilità rappresenta un criterio fondamentale, poiché un marcatore identificato come rilevante in un singolo esperimento, ma che perde tale rilevanza al variare del *seed* o del mascheramento, non può essere considerato affidabile [35].

Per questa ragione, oltre al ranking medio, è stata condotta un'analisi di stabilità volta a quantificare la consistenza dei risultati ottenuti dalle tre architetture.

In particolare, per ciascun marcatore sono stati calcolati:

- il *rank* minimo e il *rank* massimo osservati tra le diverse *run*;
- il *rank range*, definito come differenza tra *rank* massimo e minimo;
- il *rank* medio;
- la deviazione standard del *rank*;
- il *coefficient of variation* dello *score*;
- una categoria di stabilità, distinta tra *High*, *Medium* e *Low*.

Questi indicatori permettono di distinguere i marcatori che emergono in modo consistente da quelli la cui importanza dipende maggiormente dalla specifica configurazione sperimentale [42].

Dopo aver definito i criteri di valutazione, è possibile analizzare i risultati ottenuti per ciascuna architettura.

Nel caso di GATv2, i marcatori più stabili risultano aSMA, CD138 e CD45RA. In particolare, aSMA presenta una variabilità contenuta e mantiene posizioni costantemente elevate, confermando il ruolo centrale della componente stromale anche nel modello basato su attenzione dinamica.

Nel complesso, GATv2 mostra una buona stabilità per i marcatori di vertice associati al compartimento immunitario adattativo e stromale.

Per *GraphSAGE*, i marcatori di maggiore interesse biologico, come aSMA, CD45, CD45RO e CD138, mostrano una stabilità intermedia, indicando che il modello tende a riconoscerli in modo consistente, pur con una certa variabilità tra *run*.

Anche in questo caso, marcatori associati al compartimento macrofagico e ad alcune componenti stromali secondarie risultano più sensibili alla struttura locale del grafo e quindi meno uniformemente catturati dalle diverse esecuzioni.

Nel caso della MPNN, il quadro è simile a quello della *GraphSAGE*: i marcatori di maggiore interesse si collocano nella fascia intermedia di stabilità, mentre marcatori secondari come PODOPLANIN, CD34 e GRANZYMEB mostrano un'elevata stabilità.

Considerando congiuntamente le tre architetture, emerge un risultato importante: i marcatori biologicamente più rilevanti non sono sempre quelli con la stabilità assoluta più alta, ma tendono comunque a collocarsi in fasce medio-alte in modo consistente.

Più precisamente, aSMA, CD45RO, CD45 e CD138 mostrano una buona consistenza complessiva tra le *run*, pur con variazioni moderate, rafforzando la loro interpretazione come segnali robusti del microambiente cellulare.

Pertanto, l'analisi di stabilità conferma che i marcatori posti ai vertici del *consensus ranking* non emergono come artefatti di singole *run*, ma rappresentano segnali riproducibili all'interno del *framework* sperimentale.

Questo comportamento suggerisce che la MPNN sia in grado di individuare segnali biologici forti, ma che parte della gerarchia interna tra i marcatori più importanti risulti meno stabile rispetto a quella osservata nei marcatori secondari.

6.3 Analisi Multi-Maschera

Per valutare la robustezza dei biomarcatori rispetto a differenti livelli di informazione disponibile, è stato adottato un approccio di *Multi-Masking Ensemble Voting*, che rappresenta il valore informativo di un biomarcatore non solo sulla base di un singolo mascheramento del grafo cellulare, ma attraverso un insieme di condizioni progressivamente più aggressive [15].

In particolare, sono stati considerati quattro livelli di mascheramento, nei quali ogni *mask ratio* simula uno scenario progressivamente più difficile: si parte dal 10%, ovvero una modifica lieve, in cui solo questa percentuale di *feature* viene azzerata prima della ricostruzione, passando poi al 20%, 30% fino al 50% di *feature* azzerate, situazione nella quale la rete deve processare il tessuto con metà dei canali proteici mancanti.

Lo scopo di questo approccio è verificare se il segnale di un biomarcatore sia fondamentale per la rappresentazione del microambiente tissutale solo in contesti di alta informazione o anche in condizioni di informazione ridotta [30].

Dopo aver definito la logica dell'analisi multi-maschera, è possibile discutere i risultati ottenuti per ciascuna architettura.

Nel caso di GATv2, il numero di marcatori classificati come *Tier 1* rimane costantemente pari a quattro per tutti i livelli di *masking*, come riportato nella Tabella 6.6.

Tabella 6.6: Analisi multi-masking GATv2.

Mask Ratio	Num Tier1	Tier1 Markers	Avg Score
10%	4	CD68, CD45, CD45RO, aSMA	0.8324
20%	4	CD45, CD68, CD45RO, aSMA	0.8298
30%	4	CD45, CD68, aSMA, CD45RO	0.8306
50%	4	CD45, CD44, aSMA, HLA-DR	0.8283

La Tabella 6.6 mostra che CD45 e aSMA compaiono in tutti e quattro i *mask ratio*, mentre CD68 e CD45RO vengono in rilievo nelle prime tre condizioni di mascheramento, ma non nel mascheramento massimo del 50%, indicando che il livello di informazione in tale scenario non è sufficiente a farli emergere come marker critici. In questa situazione compaiono invece marcatori quali CD44 e HLA-DR.

Per *GraphSAGE*, anche in questo caso il numero di marcatori classificati come *Tier 1* rimane costante e pari a quattro, come riportato nella Tabella 6.7.

Tabella 6.7: Analisi multi-masking GraphSAGE.

Mask Ratio	Num Tier1	Tier1 Markers	Avg Score
10%	4	CD68, aSMA, CD45, CD45RO	0.8255
20%	4	CD68, aSMA, CD45RO, CD44	0.8290
30%	4	CD68, CD44, aSMA, CD45RO	0.8277
50%	4	aSMA, CD45RO, CD138, CD44	0.8211

Nella Tabella 6.7 si osserva che aSMA e CD45RO rimangono marcatori informativi in tutte le condizioni di *masking*, mentre CD68 compare solo nelle prime tre, CD44 non è informativo solo nella percentuale di *masking* più bassa, ma lo è nelle restanti tre, e CD138 emerge soltanto alla massima percentuale di mascheramento.

Infine, nel caso della MPNN, sono sempre quattro i marcatori presenti nel *Tier 1*, come riportato nella Tabella 6.8.

Tabella 6.8: Analisi multi-*masking* MPNN.

Mask Ratio	Num Tier1	Tier1 Markers	Avg Score
10%	4	CD68, CD45, aSMA, CD45RO	0.8312
20%	4	CD68, aSMA, CD45, CD45RO	0.8299
30%	4	aSMA, CD45RO, CD68, CD44	0.8269
50%	4	aSMA, CD45RO, VEGF, CD138	0.8128

Nella Tabella 6.8 si osserva che aSMA e CD45RO compaiono in tutti i tassi di mascheramento, CD68 nei primi tre, CD45 nei primi due, CD44 solo al 30%, mentre VEGF e CD138 compaiono solo al 50%.

Il risultato fondamentale di questa analisi consiste nel fatto che, in tutte le configurazioni e per tutte le architetture, il numero di marker classificati come *Tier 1* è sempre pari a quattro. Questo fattore non è stato imposto a priori, ma emerge naturalmente dalla distribuzione degli *score*, indicando un divario netto tra i primi quattro marcatori e il quinto, indipendentemente dall'architettura e dal mascheramento.

Inoltre, alcuni biomarcatori mostrano una persistenza trasversale attraverso *mask ratio* e modelli diversi: in particolare, aSMA compare sistematicamente in tutte le architetture e in tutte le condizioni di *masking*.

Nel complesso, questa analisi rafforza l'ipotesi che il segnale biologico in grado di distinguere *Responder* e *Inadequate Responder* sia concentrato in un numero ristretto di marcatori.

Pur considerando la limitata dimensione della coorte, la persistenza di questo *pattern* anche in condizioni di *masking* estremo suggerisce che un pannello ridotto possa essere sufficiente a catturare i marcatori principali della stratificazione terapeutica.

Dopo aver verificato la robustezza dei biomarcatori rispetto a differenti livelli di mascheramento, è possibile integrare i ranking prodotti dalle tre architetture in una classifica globale di consenso.

6.4 Consenso Globale e Fenotipizzazione Clinica

Dopo aver analizzato separatamente i ranking prodotti dalle tre architetture GNN, emerge la necessità di integrare tali risultati in una rappresentazione unificata, al fine di ottenere una stima più robusta e meno sensibile ai *bias* specifici di ciascun modello.

Il punto centrale dell'analisi consiste quindi nella definizione di un consenso globale, ottenuto aggregando i ranking derivanti dalle tre architetture neurali a grafo.

La posizione finale di ciascun biomarcatore è calcolata come media aritmetica delle posizioni nei ranking individuali, assegnando a ciascun modello un peso uniforme del 33%.

Questa scelta riflette il principio dell'*ensemble learning*: in assenza di un singolo modello sistematicamente superiore, un'aggregazione unanime riduce la varianza del ranking finale e produce una classifica più robusta [45].

La Tabella 6.9 riporta il consenso globale ottenuto aggregando i ranking delle tre architetture mediante media aritmetica.

Tabella 6.9: Classifica globale di consenso dei biomarcatori.

Rank	Phenotype	Marker	Rank SAGE	Rank GAT	Rank MPNN	Mean Rank
1	IR	aSMA	1	3	1	1.67
2	RESPONDER	CD45RO	2	2	2	2.00
3	RESPONDER	CD45	4	1	6	3.67
4	RESPONDER	CD138	3	4	5	4.00
5	IR	CD68	5	7	3	5.00
6	RESPONDER	CD44	6	5	12	7.67
7	IR	CD163	8	11	4	7.67
8	RESPONDER	VEGF	9	15	8	10.67
9	IR	CD90	16	8	9	11.00
10	IR	PODOPLANIN	7	20	7	11.33

Il ranking della Tabella 6.9 evidenzia come aSMA rappresenti il marcatore più rilevante nei pazienti *inadequate responder*, mentre CD45RO, CD45 e CD138 caratterizzano il fenotipo *Responder*, suggerendo l'esistenza di due assi biologici distinti.

Per interpretare correttamente questa tabella, è importante sottolineare che le architetture GNN non hanno quantificato il valore espressivo della cellula, ovvero la sua abbondanza proteica, ma hanno calcolato la dipendenza topologica del tessuto. In altri termini, esse rappresentano il *Delta Loss* di permutazione, cioè il collasso predittivo che la simulazione del tessuto subisce quando il marcatore viene oscurato [42], definito come:

$$\text{importance}[i] = \max(0, MSE_{\text{permuted}} - MSE_{\text{baseline}}) \quad (6.1)$$

successivamente normalizzato nell'intervallo $[0, 1]$.

Di conseguenza, un *rank* elevato indica che quel marcatore è particolarmente importante per preservare la coerenza strutturale del microambiente appresa dalla rete.

Per quanto concerne il fenotipo *inadequate responder*, l'algoritmo ha assegnato il primo posto assoluto ad aSMA, seguito al quinto posto da CD68. Tuttavia, la vera divergenza biologica associata agli *inadequate responder* è rappresentata dal primo marcatore, la cui centralità è coerente con un microambiente stromale che altera la vascolarizzazione e ostacola la penetrazione dei farmaci [8].

Per il fenotipo *Responder*, invece, emergono nelle posizioni 2, 3 e 4 tre marcatori centrali, ossia CD45RO, CD45 e CD138. Essi non indicano uno stato di salute, ma suggeriscono che il tessuto responsivo collassi se privato di questa tipologia cellulare, indicando che l'architettura del microambiente non è ancora irrigidita dalla fibrosi stromale [17].

Questo risultato è particolarmente interessante, poiché suggerisce che l'esito favorevole della terapia possa essere associato a un microambiente ancora vulnerabile alla modulazione farmacologica, nel quale la componente immunitaria non è ancora stata sostituita da una rigidità stromale dominante [17].

Nel complesso, la classifica di consenso evidenzia quindi una chiara polarizzazione biologica:

- marcatori stromali e mieloidi associati al fenotipo *Inadequate Responder*;
- marcatori immunitari adattivi e plasmacellulari associati al fenotipo *Responder*.

Questa distinzione costituisce il principale risultato interpretativo del ranking globale e rappresenta la base per la successiva selezione del *subset* ottimale di marcatori.

6.5 Risultati delle Feature Morfologiche

Parallelamente all'analisi dei biomarcatori proteici, è stata condotta una valutazione specifica del contributo delle *feature* morfologiche.

Mentre i marcatori misurano la presenza di molecole specifiche e il *Delta Loss* misura la dipendenza topologica del grafo da tali molecole, le *feature* morfologiche, derivate da un file differente denominato *regionprops*, descrivono lo stato geometrico delle cellule nel tessuto, ovvero forma, dimensione e organizzazione spaziale [47].

L'inclusione di queste *feature* è rilevante, poiché anche la morfologia rappresenta un indicatore dello stato funzionale della cellula: diversi stati cellulari possono infatti esibire *pattern* geometrici distinti e biologicamente significativi.

Le *feature* morfologiche sono incluse nel vettore di input della GNN come colonne aggiuntive concatenate al profilo proteomico standardizzato:

$$\mathbf{X}_{\text{combined}} = \left[\mathbf{X}_{\text{bio}}^{\text{scaled}} \parallel \mathbf{X}_{\text{morph}}^{\text{scaled}} \right]$$

In questo modo, ogni nodo del grafo possiede un vettore misto tra marcatori e morfologia, utilizzato dalla GNN per apprendere le correlazioni tra morfologia e proteomica nel contesto del vicinato spaziale, cosa che un'analisi tradizionale non sarebbe in grado di fare.

Il *Delta Loss* morfologico misura quindi quanto la coerenza spaziale della forma cellulare sia necessaria alla simulazione del tessuto, e non soltanto la presenza di una specifica quantità di proteina.

Le Tabelle 6.10 e 6.11 riportano rispettivamente il ranking globale e i valori medi osservati per le quattro *feature* morfologiche.

Tabella 6.10: Ranking globale delle *feature* morfologiche.

Feature	Rank GAT	Rank SAGE	Rank MPNN	Res MPNN	Rank Medio
eccentricity	1.5	1.75	1.75	0.8106	1.6667
area	2.5	1.75	1.25	0.4240	1.8333
axis_minor_length	2.5	3.0	3.5	0.4430	3.0000
axis_major_length	3.5	3.5	3.5	0.5257	3.5000

Tabella 6.11: Valori medi e fenotipo associato delle *feature* morfologiche.

Feature	Val Resp	Val IR	Fenotipo
eccentricity	0.8191	0.8772	IR +7.1%
area	0.4070	0.4325	IR +6.3%
axis_minor_length	0.4334	0.4470	IR +3.1%
axis_major_length	0.5266	0.5431	IR +3.1%

Il risultato più rilevante che emerge dalle Tabelle 6.10 e 6.11 riguarda la *feature eccentricity*, che occupa la prima posizione nel ranking globale, con una differenza tra i due gruppi clinici del 7.1%, indicando che le cellule dei pazienti *inadequate responder* presentano una forma più allungata rispetto a quelle dei *responder*.

Dal punto di vista biologico, un'eccentricità più alta nei tessuti *inadequate responder* è coerente con l'attivazione dei fibroblasti sinoviali, poiché tali cellule, nei pazienti con

AR resistente, acquisiscono un fenotipo invasivo e migratorio, con processi citoplasmatici allungati e morfologia fusiforme [8, 47].

La GNN rileva questa informazione attraverso la valutazione di quanto la forma cellulare sia necessaria alla coerenza topologica del grafo.

Anche la *feature area* occupa una posizione elevata nel ranking, evidenziando una differenza significativa tra i pazienti *Responder* e *Inadequate Responder*, con un incremento del 6.3% nei pazienti refrattari. Questo suggerisce che anche la dimensione cellulare differisca tra i gruppi e che cellule più grandi, quindi potenzialmente associate a maggiore densità di macrofagi e plasmacellule, rappresentino una caratteristica del fenotipo refrattario [18].

Più in generale, tutte e quattro le *feature* morfologiche mostrano valori mediamente superiori nei pazienti *inadequate responder*. Questo dato suggerisce che il fenotipo refrattario non sia caratterizzato soltanto da differenze molecolari, ma anche da una riorganizzazione strutturale del tessuto, osservabile nella geometria delle cellule che compongono il microambiente.

Al contrario, i pazienti *responder* presentano valori più contenuti per le metriche considerate, in modo coerente con un tessuto meno deformato e con una morfologia cellulare globalmente meno aggressiva.

Dopo aver discusso il contributo delle *feature* morfologiche, è possibile passare al risultato conclusivo dello studio, ossia il confronto tra *panel* completo e *subset* ridotto.

6.6 Confronto Panel Completo vs Subset Ridotto

La parte conclusiva dell'analisi sperimentale riguarda il confronto tra il pannello completo di marcatori e il pannello ridotto identificato tramite la procedura di *forward feature selection* guidata dal *consensus ranking* delle GNN.

Come descritto nel capitolo precedente, la *forward selection* segue l'ordine del ranking globale di consenso: a ogni iterazione viene aggiunto un nuovo marcatore al *subset* corrente e la capacità discriminativa del pannello risultante viene valutata tramite regressione logistica con penalizzazione L2, utilizzando il protocollo *Leave-One-Patient-Out Cross-Validation*.

La scelta di utilizzare il ranking GNN come *baseline*, invece di eseguire una ricerca esaustiva, deriva da due considerazioni principali:

- una motivazione computazionale, poiché analizzare tutte le combinazioni di 25 marcatori sarebbe proibitivo in presenza di LOPO-CV;
- una motivazione epistemologica, poiché il ranking GNN non è arbitrario, ma riflette la struttura del tessuto appresa dalle tre reti neurali, costituendo la miglior stima

disponibile della rilevanza biologica di ciascun marcatore nel contesto spaziale del grafo cellulare.

Per valutare in che misura il consensus ranking derivato dalle GNN fosse effettivamente informativo ai fini predittivi, è stata applicata una procedura di *forward selection*, nella quale i marcatori sono stati aggiunti progressivamente secondo l'ordine del ranking globale. A ogni iterazione, il *subset* corrente è stato valutato mediante regressione logistica con penalizzazione L2 in validazione Leave-One-Patient-Out.

La Tabella 6.12 mostra l'andamento della performance nei primi step della procedura, che sono sufficienti a evidenziare il risultato principale dell'analisi.

Tabella 6.12: Andamento della forward selection guidata dal consensus ranking GNN nei primi step di selezione.

k	Marker	AUC LOPO-CV	Δ AUC
1	aSMA	0.833	–
2	CD45RO	0.667	-0.166
3	CD45	0.833	+0.166
4	CD138	1.000	+0.167
5	CD68	0.833	-0.167
6	CD44	1.000	+0.167
7	CD163	1.000	0.000
8	VEGF	0.833	-0.167
9	CD90	0.833	0.000
10	PODOPLANIN	0.667	-0.166

Dalla Tabella 6.12 emerge chiaramente che la massima capacità discriminativa viene raggiunta già nei primi step della procedura, in particolare per $k = 4$, corrispondente al *subset*:

$$\{aSMA, CD45RO, CD45, CD138\}$$

Con questo *subset*, il classificatore raggiunge un'AUC pari a 1.000, indicando una separazione completa tra le due classi nel protocollo LOPO-CV. Al contrario, l'utilizzo del pannello completo di 25 marcatori determina una riduzione della performance fino a un valore di AUC pari a 0.667.

Questo risultato evidenzia che l'aumento del numero di *feature* non corrisponde automaticamente a un miglioramento della classificazione. In una coorte di dimensioni ridotte, l'aggiunta di marcatori meno informativi o ridondanti introduce rumore e può degradare la capacità discriminativa del modello.

L'andamento osservato nei primi passi della *forward selection* è particolarmente istruttivo. Con il solo marcatore aSMA, infatti, il modello raggiunge già un'AUC pari a 0.833, indicando che la componente stromale contiene un segnale discriminativo rilevante.

L'aggiunta di CD45RO determina una riduzione della performance, suggerendo che due segnali biologici distinti non siano sufficienti, da soli, a produrre una separazione stabile nel contesto di una coorte così limitata.

L'inserimento di CD45 ripristina l'AUC al valore iniziale, introducendo una nuova dimensione informativa che contribuisce alla separazione.

La svolta si verifica con l'aggiunta del quarto marcatore, CD138, che stabilizza il classificatore e porta l'AUC a 1.000, corrispondente a una separazione netta tra i pazienti.

Un ulteriore aspetto rilevante è che l'utilizzo di subset più ampi non comporta un miglioramento delle prestazioni. Ad esempio, con 10 marcatori l'AUC risulta pari a 0.667, valore analogo a quello ottenuto con il pannello completo. Questo comportamento evidenzia un effetto di saturazione dello spazio informativo e la presenza di ridondanza tra i marcatori.

L'utilizzo di questi quattro marcatori copre quindi lo spazio informativo necessario alla discriminazione fenotipica, configurandosi come il *subset* ottimale per questa ricerca e riducendo il pannello dell'80%.

Di seguito è riportata la figura riassuntiva della selezione del *subset* ottimale.

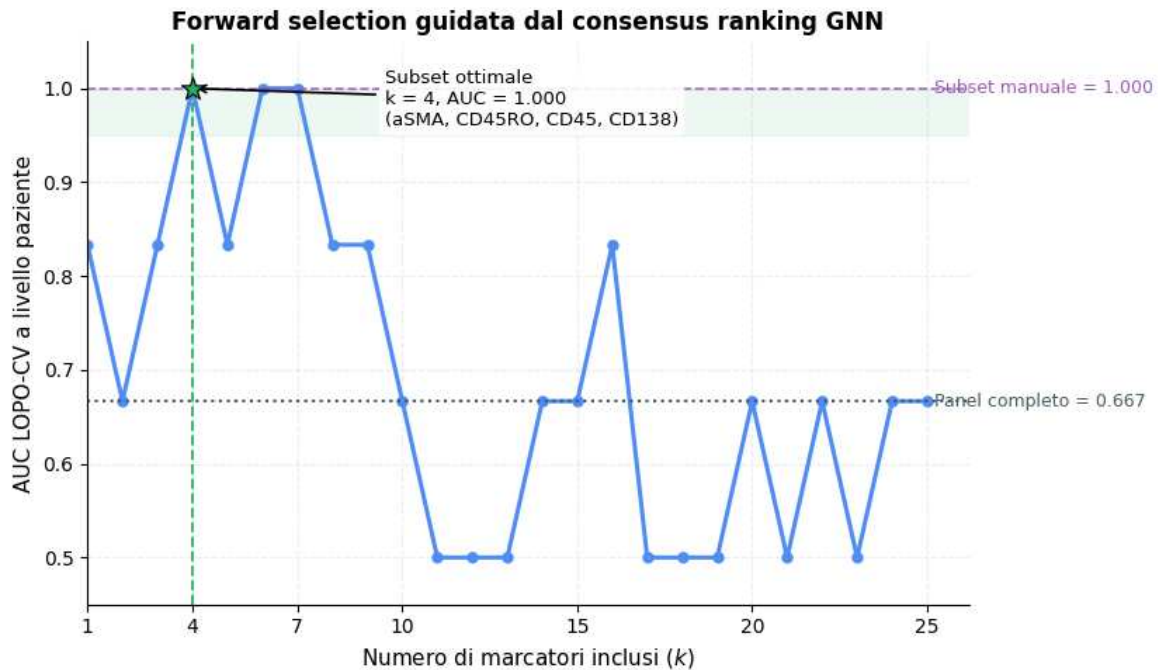


Figura 6.1: Curva AUC in validazione LOPO-CV in funzione del numero di marcatori inclusi, ottenuta tramite forward selection guidata dal consensus ranking GNN. Il primo massimo di AUC pari a 1.000 viene raggiunto con $k = 4$, corrispondente al subset minimo $\{aSMA, CD45RO, CD45, CD138\}$. Il pannello completo (25 marcatori) mostra una riduzione della performance (AUC = 0.667), evidenziando un effetto di ridondanza informativa.

La Figura 6.1 mostra l'andamento della AUC in funzione del numero di marcatori inclusi, evidenziando come il primo massimo venga raggiunto per $k = 4$, corrispondente al *subset* ottimale individuato.

Questa rappresentazione conferma che il ranking derivato dalle GNN è informativo nel guidare la selezione del punto ottimale senza ricorrere a una ricerca esaustiva.

In particolare, il *subset* ottimale individuato integra:

- una componente stromale (*aSMA*);
- una componente immunitaria globale (*CD45*);
- una componente della memoria (*CD45RO*);
- una componente plasmacellulare (*CD138*).

Questa combinazione suggerisce che la differenza tra i due fenotipi clinici non dipenda da un singolo *pathway*, ma dall'interazione tra rimodellamento stromale e organizzazione immunitaria del microambiente sinoviale.

I risultati presentati in questo capitolo evidenziano come l'approccio proposto sia in grado di identificare un insieme ristretto di biomarcatori altamente informativi, capaci di discriminare in modo netto i fenotipi clinici considerati.

Tuttavia, in aggiunta alle *performance* ottenute, si pone l'obiettivo di interpretare tali evidenze nel contesto biologico e metodologico dell'Artrite Reumatoide, al fine di comprenderne il significato e le implicazioni.

Dunque, nel capitolo successivo si propone una discussione critica dei risultati, analizzandone le implicazioni cliniche, i limiti dello studio e le possibili direzioni future.

Capitolo 7

Discussione e Conclusioni

Il presente capitolo sintetizza e interpreta le evidenze emerse durante l'intero percorso sperimentale, contestualizzandole all'interno del panorama scientifico della medicina di precisione.

Nel complesso, i risultati ottenuti mostrano che un approccio basato sulla modellazione relazionale del tessuto mediante *Graph Neural Networks* è in grado di estrarre, a partire da biopsie sinoviali, una firma molecolare ridotta, stabile e biologicamente interpretabile.

Questo risultato differenzia il *framework* proposto dai tradizionali approcci di *feature selection*, che tendono a trattare le cellule come osservazioni indipendenti e a trascurare la struttura spaziale del microambiente.

Dopo aver presentato i risultati sperimentali, è opportuno discuterne le implicazioni metodologiche, biologiche e applicative, evidenziando al tempo stesso i limiti dello studio e le possibili direzioni future.

7.1 Implicazioni Metodologiche

Il presente studio introduce un paradigma rilevante nell'applicazione dell'intelligenza artificiale alla patologia digitale e, nello specifico, nell'AR.

Il *framework* proposto supera i limiti delle tecniche tradizionali basate su immunistochimica e citometria di massa, analizzate mediante modelli di *Machine Learning* classici, i quali comportano una perdita di informazione spaziale, ignorando la struttura del microambiente tissutale e trattando le cellule come unità indipendenti e decontestualizzate [32].

Questa scelta metodologica risulta particolarmente significativa nel contesto dell'AR, in cui il significato biologico di una cellula non dipende esclusivamente dal suo profilo molecolare, ma anche dal contesto locale in cui essa è inserita, definito dalle relazioni di vicinato con le altre cellule del microambiente.

Un secondo elemento di rilievo riguarda la formalizzazione del *Delta Loss* di permutazione come misura di importanza dei biomarcatori in contesti di analisi spaziale.

A differenza degli approcci tradizionali di *biomarker discovery*, che si basano su *t-test* e *fold-change* per misurare differenze di espressione oppure su modelli supervisionati come *Random Forest*, addestrati su profili medi per paziente [42], nel presente studio l'importanza di una *feature* è definita in funzione della perdita di ricostruzione del modello.

Questo cambio di prospettiva è metodologicamente rilevante: un marcatore può risultare abbondante senza essere cruciale, mentre *feature* meno espresse possono avere un ruolo centrale nella stabilità del microambiente. In questo senso, il *Delta Loss* non misura la quantità, ma il contributo strutturale della *feature*.

Un ulteriore contributo consiste nell'utilizzo congiunto di tre architetture GNN distinte — GATv2, SAGEConv e MPNN — per costruire un consenso globale che mitiga i *bias* algoritmici individuali.

Questa scelta si fonda sul principio dell'*ensemble learning*: modelli caratterizzati da *bias* differenti, se opportunamente aggregati, producono stime più robuste rispetto a singoli modelli. Le correlazioni di Spearman osservate tra i ranking confermano una convergenza sostanziale, ma non completa, indicando che le architetture catturano aspetti complementari del microambiente.

Infine, il lavoro evidenzia un aspetto metodologico cruciale nei contesti a bassa numerosità campionaria: le performance di un classificatore lineare raggiungono un massimo per un numero limitato di *feature* e degradano sistematicamente all'aumentare della dimensionalità [43].

Questo risultato ha importanti implicazioni pratiche, mostrando che l'aggiunta di marcatori non garantisce un miglioramento delle performance e può, anzi, introdurre rumore e *overfitting*. Pertanto, una selezione guidata da ranking informativi come quello derivato dalle GNN risulta fondamentale, soprattutto in contesti clinici in cui la raccolta dei dati è costosa e invasiva.

7.2 Implicazioni Biologiche nell'Artrite Reumatoide

Dal punto di vista biologico, i risultati ottenuti permettono di ridefinire il concetto di farmaco-resistenza nell'AR.

In particolare, emerge come la refrattarietà ai trattamenti non sia semplicemente attribuibile a un aumento dell'infiammazione, ma a una profonda riorganizzazione strutturale del tessuto [8].

Il *subset* ottimale identificato — costituito da {aSMA, CD45RO, CD45, CD138} — distingue in modo netto i pazienti *responder* dagli *inadequate responder*.

Nei pazienti *Responder* emergono CD45RO, CD45 e CD138, che identificano un microambiente dominato da componenti immunitarie adattative e plasmacellulari. La prevalenza di queste popolazioni suggerisce che l'organizzazione del tessuto sia ancora guidata da meccanismi immunitari attivi, piuttosto che da una riorganizzazione stromale irreversibile, rendendo il microambiente più suscettibile alla modulazione farmacologica. [3, 17].

Al contrario, nei pazienti *inadequate responder*, aSMA occupa la posizione dominante, indicando un microambiente caratterizzato da attivazione miofibroblastica e fibrosi stromale. In questo contesto, i fibroblasti contribuiscono alla formazione di una matrice densa che altera la permeabilità vascolare e ostacola la diffusione dei farmaci [8, 47].

Anche l'analisi delle *feature* morfologiche supporta questa interpretazione: parametri come eccentricità e area cellulare risultano aumentati nei pazienti *inadequate responder*, suggerendo una trasformazione non solo molecolare, ma anche strutturale del tessuto [47].

Nel complesso, questi risultati indicano che i due fenotipi clinici non differiscono semplicemente per livelli di espressione, ma per l'organizzazione complessiva del microambiente sinoviale.

Dopo aver discusso le varie implicazioni biologiche dei risultati, è necessario considerare i principali limiti dello studio.

7.3 Limiti dello Studio e Sviluppi Futuri

Nonostante i risultati ottenuti siano promettenti, il presente studio presenta alcune limitazioni intrinseche.

Il limite principale è rappresentato dalla dimensione della coorte: sebbene il *dataset* includa 42 ROI e 209 652 cellule, esso deriva da soli cinque pazienti. Questo limita la generalizzabilità dei risultati e richiede validazioni su coorti indipendenti più ampie.

L'utilizzo del protocollo *Leave-One-Patient-Out Cross-Validation* riduce il rischio di *data leakage*, ma non può compensare completamente la variabilità biologica interindividuale [35].

Un ulteriore limite riguarda l'utilizzo della regressione logistica L2 come classificatore finale. Sebbene questa scelta sia metodologicamente conservativa e appropriata per piccoli campioni, essa non è in grado di catturare relazioni non lineari tra i biomarcatori.

Inoltre, l'aggregazione delle ROI a livello paziente comporta una perdita di informazione sulla variabilità intra-paziente, che potrebbe contenere segnali biologici rilevanti.

Alla luce di queste considerazioni, i principali sviluppi futuri includono:

- validazione del *subset* a quattro marcatori su coorti indipendenti e più ampie;

- estensione della *pipeline* ad altre tecnologie di *imaging* e omiche;
- utilizzo di classificatori più espressivi in presenza di *dataset* più grandi;
- introduzione di etichette a livello cellulare per apprendimento supervisionato.

Questi sviluppi potrebbero trasformare il *framework* da strumento esplorativo a sistema predittivo clinicamente utilizzabile.

Successivamente, dopo aver concluso e discusso i limiti e sviluppi futuri, è possibile formulare le conclusioni finali del lavoro.

7.4 Conclusioni Finali

Questa tesi ha affrontato uno dei problemi più rilevanti della reumatologia moderna: identificare, a partire da biopsie sinoviali, le caratteristiche biologiche in grado di predire la risposta alla terapia biologica nei pazienti con AR.

L'approccio proposto integra *Imaging Mass Cytometry*, *Graph Neural Networks*, apprendimento *self-supervised* basato su *masking* e una procedura di *ensemble ranking* seguita da *forward selection*.

I risultati mostrano che, tale approccio, consente di identificare una firma ridotta e interpretabile in grado di discriminare efficacemente i due fenotipi clinici. In particolare, il subset ottimale è composto da quattro marcatori proteici — aSMA, CD45RO, CD45 e CD138 — mentre le *feature* morfologiche contribuiscono in modo complementare alla caratterizzazione del microambiente, permettendo di cogliere differenze strutturali tra i due gruppi di pazienti.

In particolare, il lavoro suggerisce che:

- i pazienti *Responder* sono caratterizzati da un microambiente dominato da circuiti immunitari adattativi;
- gli *Inadequate Responder* presentano una maggiore centralità della componente stromale e fibrotica, accompagnata da variazioni morfologiche cellulari che riflettono un rimodellamento del tessuto.

Il contributo principale della tesi risiede nell'aver dimostrato che la risposta terapeutica deve essere studiata non solo in termini di abbondanza molecolare, ma anche in termini topologici.

Per concludere il presente studio di tesi, si evince come l'analisi del tessuto sinoviale mediante *Graph Neural Networks* rappresenti un passo significativo verso una ridefinizione del problema della risposta terapeutica nell'Artrite Reumatoide, spostando il *focus* dalla singola molecola alla struttura relazionale del microambiente.

In questa prospettiva, la medicina di precisione non si limita all'identificazione di biomarcatori isolati, ma si estende alla comprensione dell'intero ecosistema cellulare che determina la suscettibilità del tessuto all'azione farmacologica.

Bibliografia

- [1] J. S. Smolen, D. Aletaha, and I. B. McInnes. Rheumatoid arthritis. *The Lancet*, 388(10055):2023–2038, October 2016.
- [2] I. B. McInnes and G. Schett. The pathogenesis of rheumatoid arthritis. *New England Journal of Medicine*, 365(23):2205–2219, December 2011.
- [3] F. Humby et al. Rituximab versus tocilizumab in anti-tnf inadequate responder patients with rheumatoid arthritis (r4ra): 16-week outcomes of a stratified, biopsy-driven, multicentre, open-label, phase 4 randomised controlled trial. *The Lancet*, 397(10271):305–317, January 2021.
- [4] J. S. Smolen et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Annals of the Rheumatic Diseases*, 75(1):3–15, January 2016.
- [5] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.
- [6] S. Li, H. Hua, and S. Chen. Graph neural networks for single-cell omics data: a review of approaches and applications. *Briefings in Bioinformatics*, 26(2):bbaf109, March 2025.
- [7] Y. Lei et al. Synovial microenvironment-influenced mast cells promote the progression of rheumatoid arthritis. *Nature Communications*, 15(1):113, January 2024.
- [8] A. P. Croft et al. Distinct fibroblast subsets drive inflammation and damage in arthritis. *Nature*, 570(7760):246–251, June 2019.
- [9] J. S. Smolen et al. Eular recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2022 update. *Annals of the Rheumatic Diseases*, 82(1):3–18, January 2023.
- [10] M. J. Lewis et al. Molecular portraits of early rheumatoid arthritis identify clinical and treatment response phenotypes. *Cell Reports*, 28(9):2455–2470.e5, August 2019.

- [11] W. L. Hamilton, R. Ying, and J. Leskovec. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*, 2017.
- [12] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. *arXiv preprint arXiv:1704.01212*, 2017.
- [13] S. Brody, U. Alon, and E. Yahav. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491*, 2021.
- [14] W. L. Hamilton, R. Ying, and J. Leskovec. Inductive representation learning on large graphs. *arXiv preprint arXiv:1706.02216*, 2017.
- [15] Z. Hou et al. Graphmae: Self-supervised masked graph autoencoders. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 594–604, Washington, DC, USA, 2022. ACM.
- [16] B. I. Wallace, L. Cooney, and D. A. Fox. New molecular targets in the treatment of rheumatoid arthritis. *Current Opinion in Rheumatology*, 36(3):235–240, May 2024.
- [17] Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Consortium et al. Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nature Immunology*, 20(7):928–942, July 2019.
- [18] K. Knab, D. Chambers, and G. Krönke. Synovial macrophage and fibroblast heterogeneity in joint homeostasis and inflammation. *Frontiers in Medicine*, 9:862161, April 2022.
- [19] M. Canavan et al. Rheumatoid arthritis synovial microenvironment induces metabolic and functional adaptations in dendritic cells. *Clinical and Experimental Immunology*, 202(2):226–238, October 2020.
- [20] S. Alivernini et al. Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis. *Nature Medicine*, 26(8):1295–1306, August 2020.
- [21] F. Humby et al. Ectopic lymphoid structures support ongoing production of class-switched autoantibodies in rheumatoid synovium. *PLoS Medicine*, 6(1):e1, January 2009.
- [22] J. De Lima et al. Spatial mapping of rheumatoid arthritis synovial niches reveals a lyve1+ macrophage network associated with response to therapy. *Annals of the Rheumatic Diseases*, 84(12):1955–1967, December 2025.
- [23] V. Marx. Method of the year: spatially resolved transcriptomics. *Nature Methods*, 18(1):9–14, January 2021.

- [24] S. C. Bendall et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science*, 332(6030):687–696, May 2011.
- [25] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods*, 18(1):100–106, January 2021.
- [26] N. F. Greenwald et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature Biotechnology*, 40(4):555–565, April 2022.
- [27] G. Palla et al. Squidpy: a scalable framework for spatial omics analysis. *Nature Methods*, 19(2):171–178, February 2022.
- [28] on behalf of the Swedish Digital Twin Consortium et al. Digital twins to personalize medicine. *Genome Medicine*, 12(1):4, December 2020.
- [29] R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, and N. Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15(12):1053–1058, December 2018.
- [30] M. Ali, S. Richter, A. Ertürk, D. S. Fischer, and F. J. Theis. Graph neural networks learn emergent tissue properties from spatial molecular profiles. *Nature Communications*, 16(1):8419, September 2025.
- [31] J. Windhager et al. An end-to-end workflow for multiplexed image processing and analysis. *Nature Protocols*, 18(11):3565–3613, November 2023.
- [32] D. Righelli et al. Spatialexperiment: infrastructure for spatially-resolved transcriptomics data in r using bioconductor. *Bioinformatics*, 38(11):3128–3131, May 2022.
- [33] J. Čuklina, P. G. A. Pedrioli, and R. Aebersold. Review of batch effects prevention, diagnostics, and correction approaches. In R. Matthiesen, editor, *Mass Spectrometry Data Analysis in Proteomics*, pages 373–387. Springer, New York, NY, 2020.
- [34] J. H. Levine et al. Data-driven phenotypic dissection of aml reveals progenitor-like cells that correlate with prognosis. *Cell*, 162(1):184–197, July 2015.
- [35] A. Vabalas, E. Gowen, E. Poliakoff, and A. J. Casson. Machine learning algorithm validation with a limited sample size. *PLOS ONE*, 14(11):e0224365, November 2019.
- [36] W. Hu et al. Strategies for pre-training graph neural networks. *arXiv preprint arXiv:1905.12265*, 2019.
- [37] L. Kuett et al. Three-dimensional imaging mass cytometry for highly multiplexed molecular and cellular mapping of tissues and the tumor microenvironment. *Nature Cancer*, 3(1):122–133, December 2021.

- [38] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2018.
- [39] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [40] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631, Anchorage, AK, USA, 2019. ACM.
- [41] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, October 2001.
- [42] A. Altmann, L. Tološi, O. Sander, and T. Lengauer. Permutation importance: a corrected feature importance measure. *Bioinformatics*, 26(10):1340–1347, May 2010.
- [43] G. Hughes. On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63, January 1968.
- [44] S. Finzel, P. Aegerter, G. Schett, and M.-A. D’Agostino. Identification, localization and differentiation of erosions and physiological bone channels by ultrasound in rheumatoid arthritis patients. *Rheumatology*, 59(12):3784–3792, December 2020.
- [45] T. G. Dietterich. Ensemble methods in machine learning. In *Multiple Classifier Systems*, pages 1–15. Springer, Berlin, Heidelberg, 2000.
- [46] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Psychology Press, New York, NY, 2 edition, 2009.
- [47] C. D. Buckley, D. Pilling, J. M. Lord, A. N. Akbar, D. Scheel-Toellner, and M. Salmon. Fibroblasts regulate the switch from acute resolving to chronic persistent inflammation. *Trends in Immunology*, 22(4):199–204, April 2001.