

La radiomica come elemento fondante della medicina di precisione in ambito oncologico

Sara Colantonio¹, Emanuele Carlini¹, Claudia Caudai¹, Danila Germanese¹, Paolo Manghi¹, Maria Antonietta Pascali¹, Andrea Barucci², Daniele Farnesi², Nicola Zoppetti², Valentina Colcelli², Roberto Pini², Roberto Carpi³, Marco Esposito³, Emanuele Neri⁴, Chiara Romei⁵, Mariaelena Occhipinti⁶

¹Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Consiglio Nazionale delle Ricerche, Pisa

²Istituto di Fisica Applicata "N. Carrara", Consiglio Nazionale delle Ricerche, Firenze; ³USL Toscana Centro, Firenze; ⁴Dip. Ricerca Translazionale e delle Nuove Tecnologie in Medicina e Chirurgia, Univ. di Pisa; ⁵Azienda Ospedaliera Universitaria Pisana, Pisa; ⁶Dipartimento di scienze Biomediche Sperimentali e Cliniche "Mario Serio", Università di Firenze, Firenze

Abstract

Questo documento introduce e inquadra le attività che un gruppo interdisciplinare di ricercatori e clinici sta portando avanti grazie a tecniche di analisi di immagini, *machine learning* e intelligenza artificiale, a supporto della medicina di precisione in ambito oncologico. Partendo dalla comprensione del fenomeno fisico e dalla caratterizzazione dei processi biologici che sottendono alla formazione delle immagini biomedicali, attraverso tecniche di analisi *radiomica* dei dati radiologici e di *mining* di dati complessi, eterogenei e multisorgente, le soluzioni studiate mirano a supportare i clinici nel continuum dei processi diagnostici, prognostici e terapeutici in ambito oncologico.

1 Introduzione

La pietra miliare della medicina di precisione è rappresentata da un approccio personalizzato ai processi di prevenzione, diagnosi, prognosi e follow-up terapeutico. La realizzazione di un approccio di questo tipo si basa fortemente sulla modellazione delle caratteristiche dei singoli casi e della loro variabilità, in contrasto con l'approccio classico, denominato "one size fits all", basato su protocolli generalizzati derivati da valutazioni (medie) su intere popolazioni. La personalizzazione diventa requisito fondamentale in ambito oncologico, data la notevole intra- e inter-variabilità delle forme neoplastiche. In questo ambito, il raggiungimento dell'approccio personalizzato presuppone l'integrazione di dati eterogenei dei pazienti secondo un approccio olistico, che tenga conto dell'eterogeneità genotipica e fenotipica del tumore. L'*imaging* biomedico è fondamentale in questa integrazione, giacché fornisce informazioni morfologiche e funzionali multiparametriche preziose per una valutazione predittiva e prognostica basata sul fenotipo tumorale, nonché per approfondimenti sui meccanismi alla base delle risposte dei pazienti alla terapia [ESR 2015].

È ormai noto che i tumori solidi non si presentano come un'entità omogenea, ma sono al contrario formati da popolazioni multiclonali di cellule cancerose. Queste esibiscono

una notevole variabilità spaziale e temporale che può fornire preziose informazioni sull'aggressività del tumore stesso [Gigin Lin *et al*, 2017]. La modalità di analisi comunemente adottata nella pratica clinica per la caratterizzazione molecolare del neoplasma si basa sulla biopsia ad ago, la quale, tuttavia, rappresenta una procedura invasiva che non sempre riesce a campionare l'intera gamma di potenziali variazioni biologiche all'interno di un tumore. I dati di *imaging* permettono di ottenere in maniera non invasiva dei campioni dell'intera area tumorale e possono evidenziare caratteristiche fenotipiche delle zone tumorali a livello cellulare e genetico. Tuttavia, le caratteristiche di *imaging* sono spesso valutate visivamente e descritte qualitativamente dai radiologi o dai clinici, dando luogo a una descrizione soggettiva dei fenotipi di *imaging* del tumore, con notevole variabilità intra- e inter-osservatore [Trattign 2017]. Vi è, dunque, la necessità di una quantificazione oggettiva e riproducibile di parametri estratti da dati di *imaging*.

1.1 Le potenzialità della radiomica

La radiomica è emersa come una disciplina omica di analisi di immagini con l'obiettivo di porre rimedio a questo problema. Basandosi su un approccio quantitativo, l'analisi radiomica ha la caratteristica di essere oggettiva e ripetibile. L'estrazione di parametri quantitativi dai dati di *imaging* medicale non è, come noto, un processo nuovo; nondimeno la radiomica è nata come un approccio innovativo, emerso tra le scienze omiche perché prevede di rappresentare le informazioni intrinseche al dato immagine attraverso una quantità massiva di caratteristiche o *feature* (da 700 fino a qualche migliaio) estratte con funzioni matematiche anche complesse (e.g., analisi wavelet). Le *feature* analizzate descrivono intensità, forma, dimensioni, volume e struttura tissutale della neoplasia e dei tessuti circostanti e possono essere esplorate con tecniche di *mining* per identificare i tratti fenotipici e micro-ambientali di una neoplasia, in maniera distinta o complementare ai dati forniti dai test di laboratorio e dai test genomici o proteomici. Grazie a una maggiore digitalizzazione dei processi clinici e alla diffusione dei sistemi PACS, l'analisi radiomica può essere realizzata su vaste popolazioni di pazienti oncologici, permettendo di

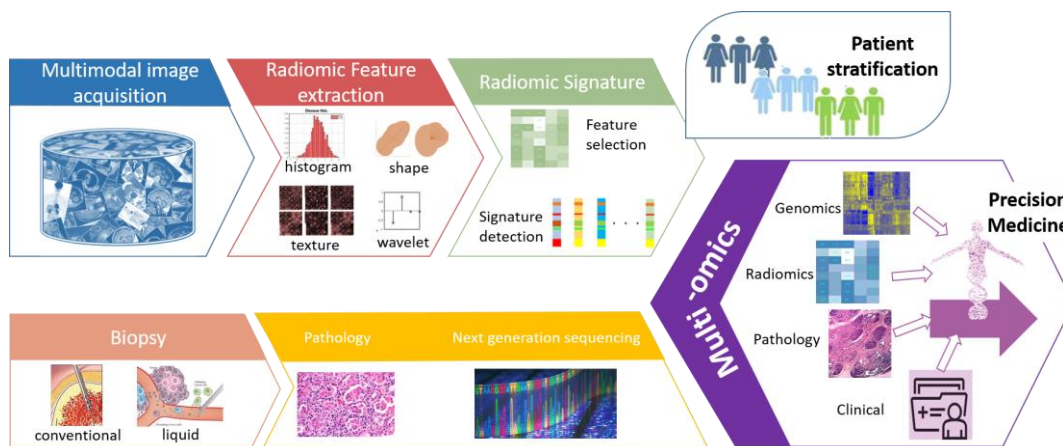


Figura 1. La pipeline della radiomica verso la medicina di precisione

individuare pattern fenotipici da collegare a end-point biologici o clinici. Ciò si traduce nell'individuazione di una *signature* radiomica oppure di *biomarker* quantitativi da usare a scopi prognostici o terapeutici.

L'analisi radiomica prevede un flusso di operazioni, come riportato in Figura 1. Per ciascuna delle modalità di *imaging* (MRI, TC, PET, US), la prima fase prevede una segmentazione semiautomatica o automatica delle regioni pertinenti; questa è seguita dall'estrazione automatica di un gran numero di *features* (descrittive dell'istogramma, della forma, della trama e della texture). Queste caratteristiche vengono quindi elaborate con tecniche di *machine learning*, come tecniche di selezione delle caratteristiche e di analisi comparativa, per selezionare quelle più stabili che si correlano ai biomarcatori di ground truth, usati nella pratica clinica in riferimento al quesito diagnostico, prognostico o terapeutico che si sta affrontando. I dati radiomici e gli eventuali biomarker da questi estratti si possono combinare con altri dati omici, costituendo una fonte di rilevanza notevole per lo sviluppo di modelli predittivi di medicina di precisione.

2 Le attività in corso

Con l'idea di combinare le varie competenze rilevanti, quali la fisica medica, la biologia, l'oncologia, l'ingegneria biomedica, la matematica e l'informatica, un gruppo di ricercatori di due istituti del CNR e di medici di diverse aziende ospedaliere e centri universitari stanno lavorando alla definizione di varie soluzioni basate su radiomica e tecniche induttive di analisi dei dati. In particolare, le attività si possono raggruppare in tre grandi filoni:

- la raccolta e la conservazione per diverse neoplasie solide di grandi quantità di dati di *imaging* multimodale, corredati da altri dati omici per costituire una **Bio banca di imaging**. La bio banca è concepita per agevolare le attività di ricerca e validazione di biomarker di imaging, anche attraverso piattaforme di ricerca collaborative (*virtual research infrastructure*) in ottica *open science*
- lo studio di **signature radiomiche** per diversi quesiti prognostici (e.g., il *grading* del tumore alla prostata e del tumore al polmone [Barucci et al., 2018]), valutando anche l'impiego di **metodi induttivi di representation learning**
- lo sviluppo di applicazioni di supporto ai processi decisionali clinici per l'intero iter oncologico (e.g., per preven-

zione, diagnosi e prognosi di alcuni tumori solidi, come quello al seno).

In particolare, nell'ambito del secondo filone di attività, è stato condotto uno studio per l'identificazione della *signature* radiomica del tumore alla prostata in grado di correlare/sostituire il Gleason score calcolato da biopsia solida. Utilizzando una coorte retrospettiva di oltre 100 pazienti sottoposti ad MRI, sono state valutate le caratteristiche radiomiche (circa 800) sulle aree tumorali segmentate dai radiologi. Un metodo *feed-forward* di selezione delle feature di tipo *wrapper* è stato utilizzato per selezionare le quattro *feature* più rilevanti. Queste sono state usate per addestrare (*10-fold cross-validation*) un classificatore neurale *narrow* in grado di predire il Gleason score. L'accuratezza ottenuta è maggiore dei lavori presenti in letteratura (>93%). Un paper descrittivo è in fase di sottomissione a BIBE2019 e all'RSNA2020, mentre è in corso un lavoro di ulteriore raccolta dati.

Riferimenti bibliografici

- [Barucci et al, 2018] A. Barucci, P. Bastiani, R. Carpi, S. Fondelli, A. Giannetti, M. Olmastroni, R. Pini, F. Ratto, M. Rucco, G. Zatelli, M. Esposito, *Prostate cancer Radiomics using multiparametric MR imaging: an exploratory study*, Physica Medica: European Journal of Medical Physics, Vol. 56 , 246, 2018
- [ESR 2015] European Society of Radiology (ESR). *Medical imaging in personalised medicine: a white paper of the research committee of the European Society of Radiology*. Insights Imaging (2015) 6:141–155
- [Gigin Lin et al, 2017] Gigin Lin, Kayvan R. Keshari, and Jae Mo Park. *Cancer Metabolism and Tumor Heterogeneity: Imaging Perspectives Using MR Imaging and Spectroscopy*. Contrast Media Mol Imaging 2017
- [Trattnig 2017] Siegfried Trattnig. *The Shift in Paradigm to Precision Medicine in Imaging: International Initiatives for the Promotion of Imaging Biomarkers*. Imaging Biomarkers, 2017