PROGETTO DI RICERCA

Titolo: Machine learning ed Intelligenza artificiale per l’armonizzazione e l’analisi di dati multiomici in Oncoematologia (Progetto GenoMed4All).

Title: Machine learning and Artificial Intelligence for harmonization and analysis of multiomic data in Oncohematology (Project GenoMed4All).

ITA

Nella pratica clinica corrente la raccolta di dati di imaging e dati multi-omici sta diventando sempre di più una pratica standard, in particolare nello studio di patologie onco-ematologiche. Questa tipologia di dati sta ottenendo sempre più successi nello spingere il trattamento dei pazienti verso l’obiettivo della medicina personalizzata, l’idea di poter trattare ciascuna patologia in base alle sue specifiche caratteristiche ed alla loro interazione con il paziente stesso. Al momento il limite principale dell’uso di queste informazioni consiste nell’interpretazione che il medico deve farne per poterlo applicare alla pratica clinica. Questo è particolarmente evidente nel caso dei dati multi-omici (genetica, proteomica, etc…) e della loro integrazione con dati di imaging che è ancora molto limitata, e l’analisi stessa è spesso effettuata in modo quasi completamente qualitativo.

L’analisi di immagini mediche tramite intelligenza artificiale, in particolare per le immagini istopatologiche, richiede la combinazione di tecniche provenienti dal mondo del Deep Learning come le Convolutional Neural Networks (CNN), che si sono dimostrate estremamente funzionali nel campo della computer vision in generale, e quelle provenienti dai metodi di analisi di Big Data, in quanto queste immagini possono occupare volumi considerevoli (anche diversi GB per immagine) sia per via della loro risoluzione che della precisione necessaria per mantenere tutta l’informazione acquisita. Un’immagine Whole Slide (WSI) può infatti avere una risoluzione oltre i 40'000 x 40'000 pixel, più di 10'000 volte la dimensione di una classica immagine radiologica (tipicamente con una risoluzione per slide di 512x512 pixels). Simili considerazioni possono essere fatte per i dati di Next Generation Sequencing (NGS), che possono contenere centinaia di milioni di reads per paziente, con dimensioni di decine o centinaia di GB.

Il progetto GenoMed4All.. si propone di sviluppare metodiche di Artificial Intelligence per l’analisi di Big Data nell’ambito delle patologie ematologiche maligne.

L’obiettivo di questo progetto di ricerca è quindi lo sviluppo di metodi per:

1) L’analisi automatica tramite metodi di computer vision ed intelligenza artificiale di immagini istopatologiche, per supportare l’anatomopatologo in modo simile a quanto viene correntemente fatto per le immagini radiologiche (radiomica)

2) L’integrazione delle informazioni raccolte dalle immagini istopatologiche (sia in campo chiaro, in fluorescenza ed immunoistochimica) con informazioni omiche, in parallelo a quanto fatto nella ricerca radiogenomica per le immagini radiologiche

3) Preparazione di pannelli informativi per il medico curante, per la preparazione di Sistemi di supporto delle decisioni cliniche (Clinical Decision Support Systems, CDSS)

4) Integrazione dei dati di imaging PET e dati genomici

Il ricercatore dovrà quindi avere confidenza sia con le tecniche di intelligenza artificiale e deep learning, analisi di dati multi-omici e processamento di immagini (ed in particolare immagini istopatologiche), di

sviluppo su piattaforme ad alta performance, con lo sviluppo di codice ad alta performance con linguaggi multipli appropriati ai diversi campi di applicazione (come Python e C++)

Nel corso di questo progetto, l’assegnista formerà un collegamento fra l’università e l’azienda ospedaliera, per la condivisione di informazioni e dati utili che possano essere utilizzati nell’ambito di progetti di ricerca e trial clinici. Questo network sarà centrale sia per uno sviluppo della carriera dell’assegnista in direzione di ricerca nello sviluppo di nuovi metodi che di integrazione nelle aziende sanitarie o delle imprese regionali che forniscono servizi in ambito biomedico e sanitario.

**ENG**

In current clinical practice, the collection of imaging data and multi-omics data is increasingly becoming a standard practice, particularly in the study of onco-haematological diseases. This type of data is obtaining more and more success in pushing the treatment of patients towards the goal of personalized medicine, the idea of ​​being able to treat each pathology based on its specific characteristics and their interaction with the patient himself. The main limitation of the use of this information consists in the interpretation that the doctor must make of it in order to apply it to clinical practice. This is particularly evident in the case of multi-omic data (genetics, proteomics, etc ...) and their integration with imaging data which is still very limited, and the analysis itself is often carried out in an almost completely qualitative way.

The analysis of medical images using artificial intelligence, in particular for histopathological images, requires the combination of Deep Learning techniques such as Convolutional Neural Networks (CNN) which have proved extremely functional in the field of computer vision in general, and those coming from Big Data analytics, as these images can occupy considerable volumes (even several GB per image) both because of their resolution and the precision necessary to keep all the information acquired. A Whole Slide (WSI) image can in fact have a resolution of over 40'000 x 40'000 pixels, more than 10'000 times the size of a classic radiological image (typically with a resolution per slide of 512x512 pixels). Similar considerations can be made for Next Generation Sequencing (NGS) data, which can contain hundreds of millions of reads per patient, with sizes of tens or hundreds of GB.

The GenoMed4All project aims to develop Artificial Intelligence methods for Big Data analysis in the field of malignant haematological diseases.

The goal of this research project is therefore the development of methods for:

1) The automatic analysis of histopathological images using computer vision and artificial intelligence methods, to support the pathologist in a similar way to what is currently done for radiological images (radiomics)

2) The integration of information collected from histopathological images (both in bright field, in fluorescence and immunohistochemistry) with omics information, in parallel to what has been done in radiogenomic research for radiological images

3) Preparation of information panels for the attending physician, for the preparation of Clinical Decision Support Systems (CDSS)

4) Integration of PET imaging data and genomic data

The researcher must therefore be familiar with both artificial intelligence and deep learning techniques, multi-omic data analysis and image processing (and in particular histopathological images) with the development of high performance code with multiple languages ​​appropriate to different fields of application (such as Python and C ++)

During this project, the research fellow will form a link between the university and the hospital, for the sharing of information and useful data that can be used in research projects and clinical trials. This network will be central both for the development of the research fellow's career in the direction of research in the development of new methods and for integration in healthcare companies or regional companies that provide services in the biomedical and healthcare fields.

**PIANO DI ATTIVITÀ**

Il progetto si propone di formare un nuovo tipo di ricercatore con competenze computazionali che sappia gestire ed analizzare i dati provenienti da GenoMed4All:

1) gestire ed analizzare dati di imaging e radiomici in vari formati

2) implementare sistemi di intelligenza artificiale innovativi per la segmentazione, la quantificazione e la caratterizzazione di immagini mediche

3) gestire ed analizzare dati clinici eterogenei longitudinali, integrandoli con database pubblici ed interni alle strutture ospedaliere

4) integrare diverse fonti di dati in un quadro clinico unico del paziente, tramite metodi scalabili ed automatizzati

5) analizzare i dati multi-omici provenienti da GenoMed4All.

6) interfacciarsi con il personale medico e clinico per determinare direzioni comuni di collaborazione e ricerca.

Durante questo percorso di ricerca particolare enfasi verrà data allo stabilire connessioni con le aziende ospedaliere del territorio, in particolar modo con il policlinico Sant’Orsola.

Il ricercatore sarà coinvolto direttamente nei diversi progetti in corso e sarà coinvolto attivamente in tutte le fasi che vanno dalla raccolta dei dati, all’analisi e valutazione della diagnosi e prognosi del paziente.

Il ricercatore sarà invitato a elaborare i dati ottenuti sotto forma di report e articoli scientifici in lingua inglese.

Il ricercatore parteciperà inoltre nella produzione di software di analisi statistica e di intelligenza artificiale che verrà distribuito in modo open source tramite piattaforme collaborative di ampio uso.

Inoltre, il ricercatore sarà invitato a partecipare a lab meeting, congressi nazionali e internazionali per imparare nuove tecniche e presentare alla comunità scientifica i dati ottenuti.