

## PROGETTO DI RICERCA

**Titolo:** Machine learning ed Intelligenza artificiale per l'armonizzazione e l'analisi di dati multiomici del progetto HARMONY-PLUS.

**Title:** Machine learning and Artificial Intelligence for multiomic data for the HARMONY-PLUS project

Questa richiesta di assegno di ricerca è motivata dalla necessità di sviluppare metodiche di Intelligenza Artificiale e di Big data analytics per il progetto H2020-JTI-IMI2 HARMONY PLUS GA 945406, e in particolare per dare un impulso alle attività del WP3.

La raccolta di dati multi-omici sta diventando sempre di più una pratica standard, in particolare nello studio di patologie onco-ematologiche e per questo servono metodiche di analisi accurate e sviluppate appositamente.

Il progetto H2020-JTI-IMI2 HARMONY PLUS , è la continuazione di HARMONY, sta raccogliendo questi dati nell'ambito di una collaborazioni Europea, con più di 50 partecipanti.

Questa tipologia di dati sta ottenendo sempre più successi nello spingere il trattamento dei pazienti oncoematologici verso l'obiettivo della medicina personalizzata con l'obbiettivo di poter trattare ciascuna patologia oncoematologica 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 di grandi quantità di dati 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 dati multi-omici (incluse anche le immagini mediche) tramite intelligenza artificiale, richiede la combinazione di tecniche provenienti dal mondo del Deep Learning come le Convolutional Neural Networks (CNN), (estremamente funzionali nel campo della computer vision), e quelle provenienti dai metodi di analisi di Big Data (Variational Autoencoders, etc) , in quanto questi dati possono occupare volumi considerevoli (anche diversi GB) 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 HARMONY-PLUS è l'evoluzione di HARMONY e 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 dati multiomici, in collaborazione con gli altri partner che partecipano al WP3
- 2) L'integrazione delle informazioni dai dati multi-omici per sviluppare modelli previsionali di osservabili cliniche e di risposte terapeutiche.

- 3) Preparazione di pannelli informativi per i partner clinici, per la preparazione di Sistemi di supporto delle decisioni cliniche (Clinical Decision Support Systems, CDSS)
- 4) Metodologie di Big Data Analytics, incluse metodiche di Deep learning e Statistical learning

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 il consorzio HARMONY-PLUS, per la condivisione di informazioni, dati e software. Questa attività sarà centrale sia per uno sviluppo della carriera dell'assegnista in direzione di ricerca nello sviluppo di nuovi metodi che di integrazione con le realtà cliniche coinvolte nel progetto.

## 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 HARMONY-PLUS:

- 1) gestire ed analizzare dati di multi-omici nei vari formati e in particolare in OMOP
- 2) implementare sistemi di intelligenza artificiale innovativi per la segmentazione, la quantificazione e la caratterizzazione di immagini mediche e per l'analisi dei dati multi-omici, all'interno della piattaforma di HARMONY-PLUS (derivata da HARMONY)
- 3) gestire ed analizzare dati clinici eterogenei longitudinali, integrandoli con database pubblici
- 4) integrare diverse fonti di dati in un quadro clinico unico del paziente, tramite metodi scalabili ed automatizzati
- 5) Sviluppare pipeline automatizzate per il pre-processing e per la previsione di outcomes clinici

Durante questo percorso di ricerca particolare enfasi verrà data allo stabilire connessioni i partner coinvolti in HARMONY-PLUS, e in particolare con quelli coinvolti nel task 3.2 del WP3.

L'assegnista sarà coinvolto direttamente nelle diverse attività in corso e sarà coinvolto attivamente in tutte le fasi che vanno dalla raccolta dei dati, all'analisi e all'interpretazione clinica

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

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 congressi nazionali e internazionali per imparare nuove tecniche e presentare alla comunità scientifica i dati ottenuti.