

# Fondamenti di Intelligenza Artificiale 2021 – 2022

## Corso di laurea Magistrale in *Ingegneria dei Sistemi Intelligenti*

### Competition

La competizione avrà luogo su un task di Radiomica.

*Per **Radiomica** si intende l'analisi delle immagini mediche volta ad ottenere, tramite opportuni metodi matematici e l'uso del calcolatore, informazioni di tipo quantitativo da queste non rilevabili tramite la loro semplice osservazione visiva da parte dell'operatore [Wikipedia].*

L'obiettivo del progetto consiste nell'analizzare dati che provengono da 79 pazienti anonimizzati dai quali sono state acquisite immagini TAC del distretto anatomico dell'addome.

A tutti i 79 pazienti è stata inizialmente diagnosticata una neoplasia maligna alla prostata e si sono sottoposti ad una prostatectomia radicale.

Successivamente, durante un esame di “controllo” nel quale sono state raccolte le TAC in analisi, in 45 dei pazienti è stata riscontrata una recidiva del tumore.

Si chiede di sviluppare un sistema in IA che, utilizzando il dato acquisito in questa visita di controllo, predica la presenza di recidiva del tumore.

**Il task consiste quindi nel predire l'etichetta binaria *Recidiva/Non\_Recidiva* per ogni paziente.**

I dati a disposizione consistono in un dataset tabulare composto da attributi clinici associati ai pazienti e da feature quantitative estratte direttamente dai volumi 3D del distretto anatomico dell'addome:

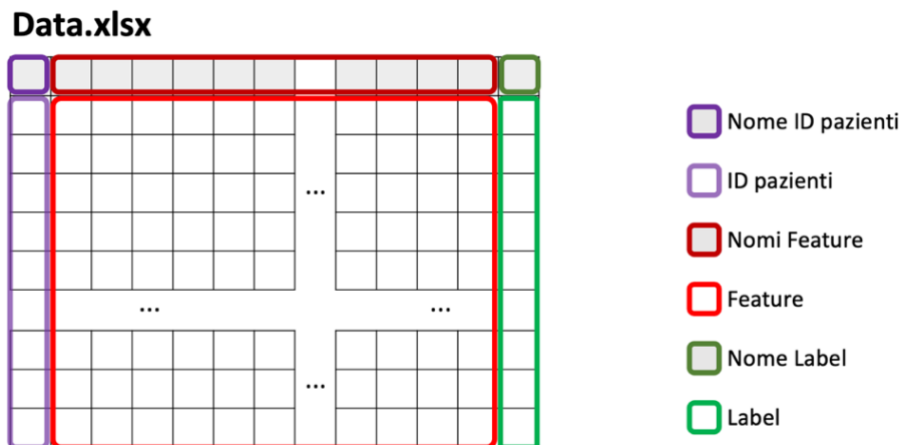
- cliniche raccolte dai medici (7):
  - *Epoca\_TC*: se la TAC è stata acquisita prima della prostatectomia radicale o dopo, durante la radioterapia;
  - *Area\_grasso\_periviscerale*: volume del grasso periviscerale;
  - *Area\_grasso\_sottocutaneo*: volume del grasso sottocutaneo;
  - *Istologia*: risultato istologico del tessuto analizzato;
  - *GS\_alla\_diagnosi*: Gleason Score alla diagnosi;
  - *TNM\_alla\_diagnosi*: Estensione del tumore;
  - *Eta\_alla\_RP*: Età del paziente alla prostatectomia radicale;
- feature estratte dai VOI:
  - 12 estratte dall'*istogramma del I ordine* dell'immagine;
  - 182 estratte dalle matrici *GLCM 3D* (istogrammi del II ordine), come 7 feature dalle matrici calcolate da 26 direzioni differenti;
  - 48 estratte dall'*istogramma del I ordine* delle trasformazioni *TOP-LBP*, come 12 feature da 4 differenti varianti della trasformazione.

L'ultima colonna del dataset infine rappresenta la classe binaria da predire *Recidiva/Non\_Recidiva* per ogni paziente.

Quindi verrà fornito un file “Data.xlsx”, in cui sono state estratte e riportate in tabella diverse feature cliniche e quelle quantitative direttamente dai volumi 3D.

- La tabella è composta da 79 righe e 251 colonne;

- La prima riga e la prima colonna sono quindi da ignorare ai fini della computazione;



### Modalità di svolgimento

*Il progetto non richiede il superamento di alcuna soglia prestabilita, ma permette agli studenti di applicare gli strumenti appresi durante il corso di Fondamenti di Intelligenza Artificiale. Sarà a discrezione dello studente l'adozione di operazioni di pre-processing, di feature selection, la scelta del modello, delle metriche di valutazione e del metodo di validazione. Ciò significa che lo studente avrà la libertà di scegliere il metodo di sua preferenza per risolvere il problema assegnato.*

### Ulteriori informazioni

*Si dovrà consegnare:*

- una tabella analitica delle prestazioni;
- il codice commentato (in Python);
- un file contenente le istruzioni per eseguire il codice;
- presentazione in PowerPoint del pitch di presentazione della soluzione adottata (durata del pitch 7 minuti + domande);