

INFORME DEL PROYECTO

OncoLens

Clasificación de tumores a partir de la expresión genica

Este documento es una plantilla provisional. Se ha generado para dejar operativos los enlaces de descarga de la web mientras se termina la versión definitiva. No contiene todavía el contenido final del proyecto.

1 · Resumen

OncoLens es una aplicación web de aprendizaje automático que, a partir de un perfil de expresión genica obtenido por RNA-Seq, determina si una muestra de tejido es tumoral y, en caso afirmativo, de cual de cinco órganos procede: mama, riñón, pulmón, próstata o colon.

2 · Datos

1.680 muestras del proyecto TCGA, con 20.531 genes por muestra. 1.344 son tumorales y 336 corresponden a tejido normal adyacente. 320 pacientes aportaron ambas.

Cohorte	Órgano	Tumor	Sano	Total
BRCA	Mama	448	112	560
KIRC	Riñón	288	72	360
LUAD	Pulmón	236	59	295
PRAD	Próstata	208	52	260
COAD	Colon	164	41	205
Total		1.344	336	1.680

3 · Metodología

Partición agrupada por paciente (GroupShuffleSplit y StratifiedGroupKFold) para evitar fuga de información. Selección de genes por ANOVA F-test validada con pruebas de permutación. Regresión logística con class_weight balanceado.

4 · Resultados

Modelo 1 (tumoral / sano): F1-macro 0,974. Modelo 2 (tipo de tumor): F1-macro 1,00 sobre el conjunto retenido.

5 · Limitaciones

No sustituye al diagnostico histopatologico. Solo reconoce cinco tipos de tumor. Los datos proceden de poblacion mayoritariamente estadounidense. El 2% de tumores clasificados como sanos no se recupera en la segunda etapa.

Pendiente de completar: metricas detalladas por clase, matrices de confusion, analisis de interpretabilidad y bibliografia.