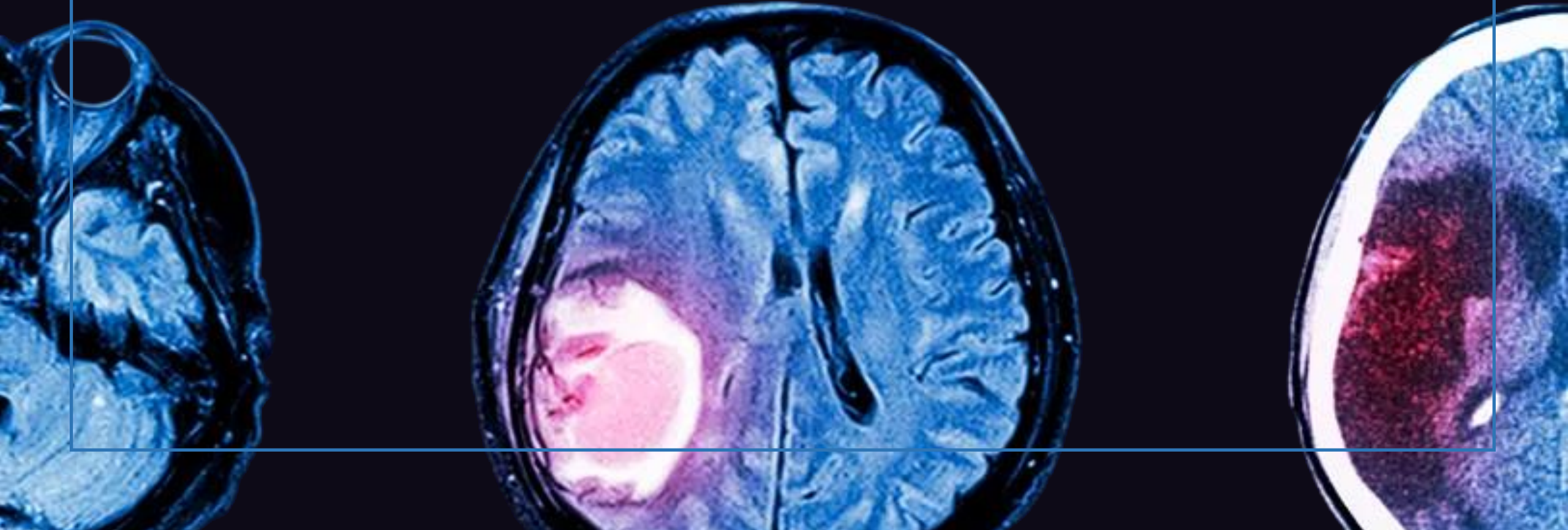


Micaela Floridia - Vanessa Galeano - Leandro Villanueva

CLASIFICACIÓN DE GLIOMAS CON TÉCNICAS DE APRENDIZAJE AUTOMÁTICO

INTRODUCCIÓN AL APRENDIZAJE AUTOMÁTICO – UNSAM 15-11-2024



A microscopic view of cells, showing various structures and processes, with a large, dark, irregular cell in the foreground and several smaller, more spherical cells in the background.

01

INTRODUCCIÓN

A close-up, slightly blurred image of a person's face, focusing on the eyes and nose, with a dark, textured background.

02

PRESENTACIÓN DEL DATASET Y EDA

A microscopic view of cells, showing various structures and processes, with a large, dark, irregular cell in the foreground and several smaller, more spherical cells in the background.

03

ENTRENAMIENTO Y EVALUACIÓN DE MODELOS

A photograph of a large, modern hospital building with a glass facade and a sign that reads "Hospital Provincial P". The building is surrounded by a paved area and some greenery.

04

COMPARACIÓN DE MODELOS

A photograph of a tall, multi-story hospital building with a sign that reads "HOSPITAL DE CLINICAS". The building is surrounded by a paved area and some greenery.

05

CONCLUSIÓN

01

INTRODUCCIÓN



LAS CÉLULAS GLIALES

SON CÉLULAS PERTENECIENTES AL SISTEMA NERVIOSO,
FUNDAMENTALES PARA EL FUNCIONAMIENTO DEL CEREBRO

SUMINISTRAN NUTRIENTES A LAS NEURONAS

CONTROLAN EL FLUJO SANGUÍNEO EN EL CEREBRO

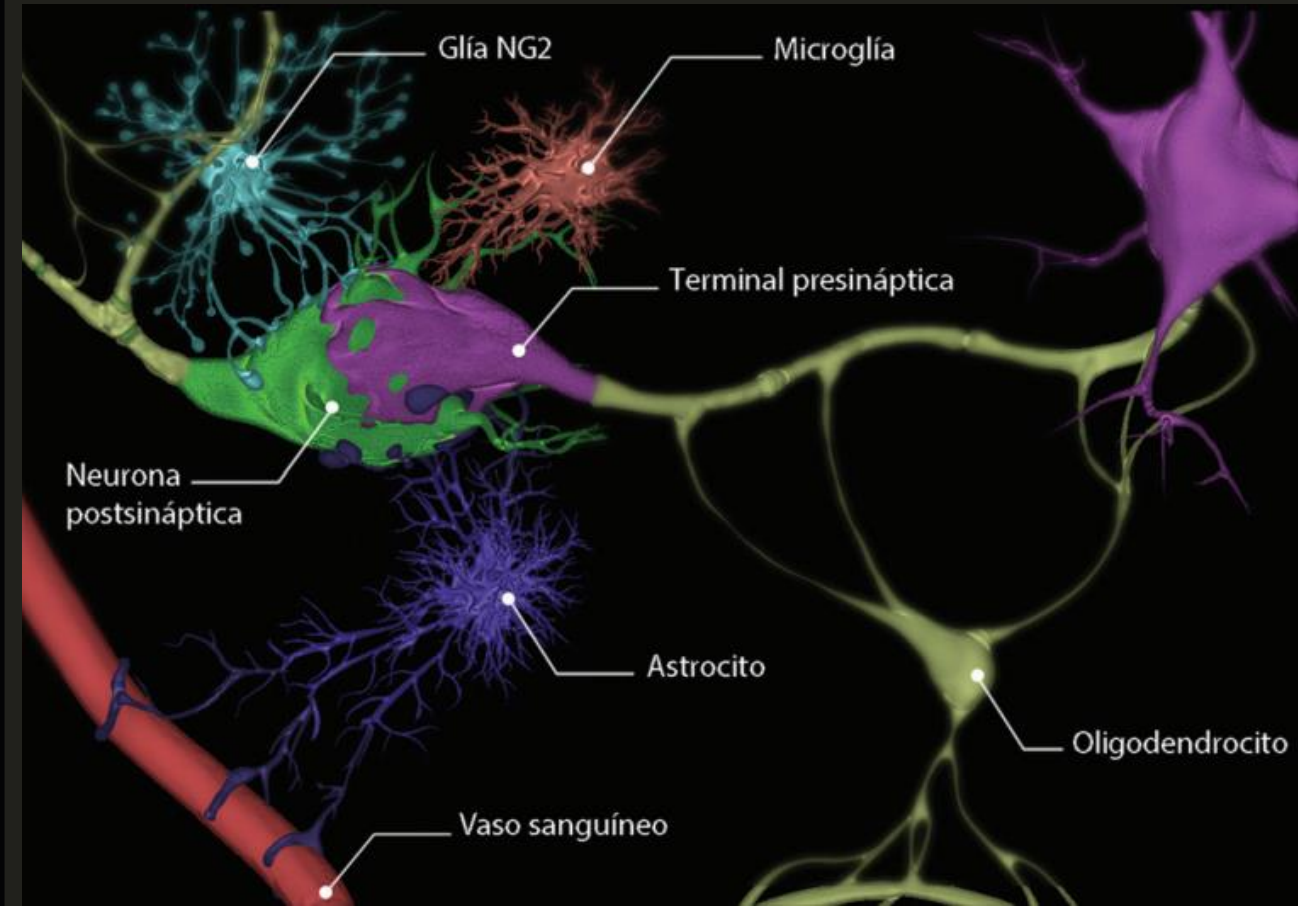
PRODUCEN MIELINA PARA AISLAR Y PROTEGER LOS AXONES

ELIMINAN DESHECHOS

COMBATEN INFECCIONES

REPARAN Y REGENERAN EL TEJIDO NERVIOSO

IMAGEN: ALGUNAS CÉLULAS GLIALES EN LAS TERMINALES DE 2 NEURONAS



LOS GLIOMAS

SON TUMORES ORIGINADOS EN LAS CÉLULAS GLIALES
EN PARTICULAR NOS CENTRAREMOS EN 2 TIPOS SEGÚN SU GRADO DE MALIGNIDAD

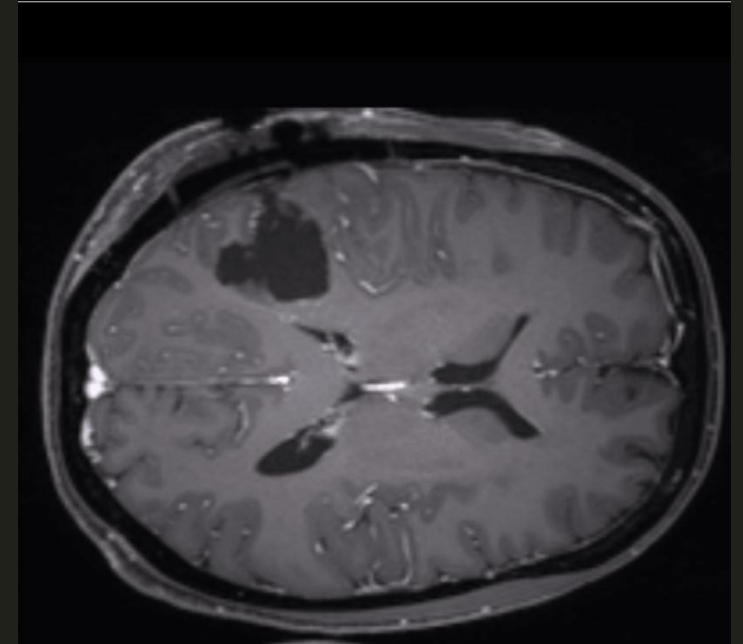
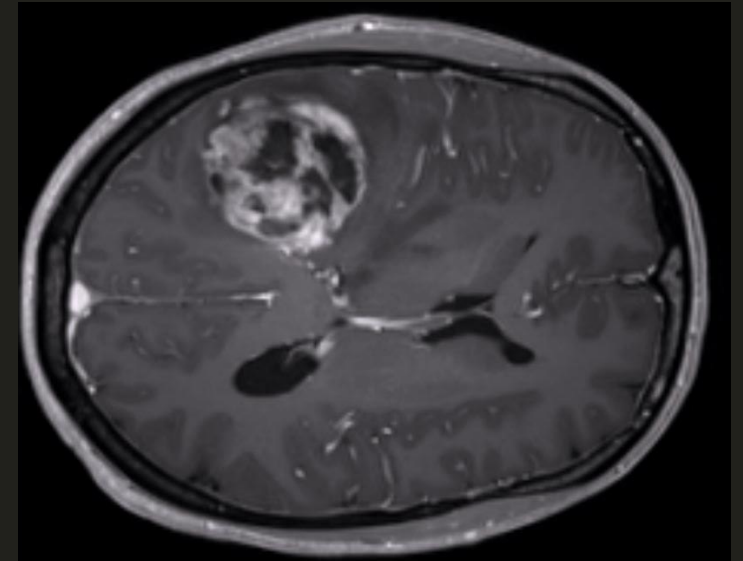
GLIOMAS DE BAJO GRADO (LGG)

MENOS AGRESIVO
CRECIMIENTO LENTO
MEJOR PRONÓSTICO

GLIOBLASTOMA MULTIFORME (GBM)

MÁS AGRESIVO
CRECIMIENTO RÁPIDO E INVASIVO
PEOR PRONÓSTICO

IMAGEN: GBM PRE Y POSOPERATORIO



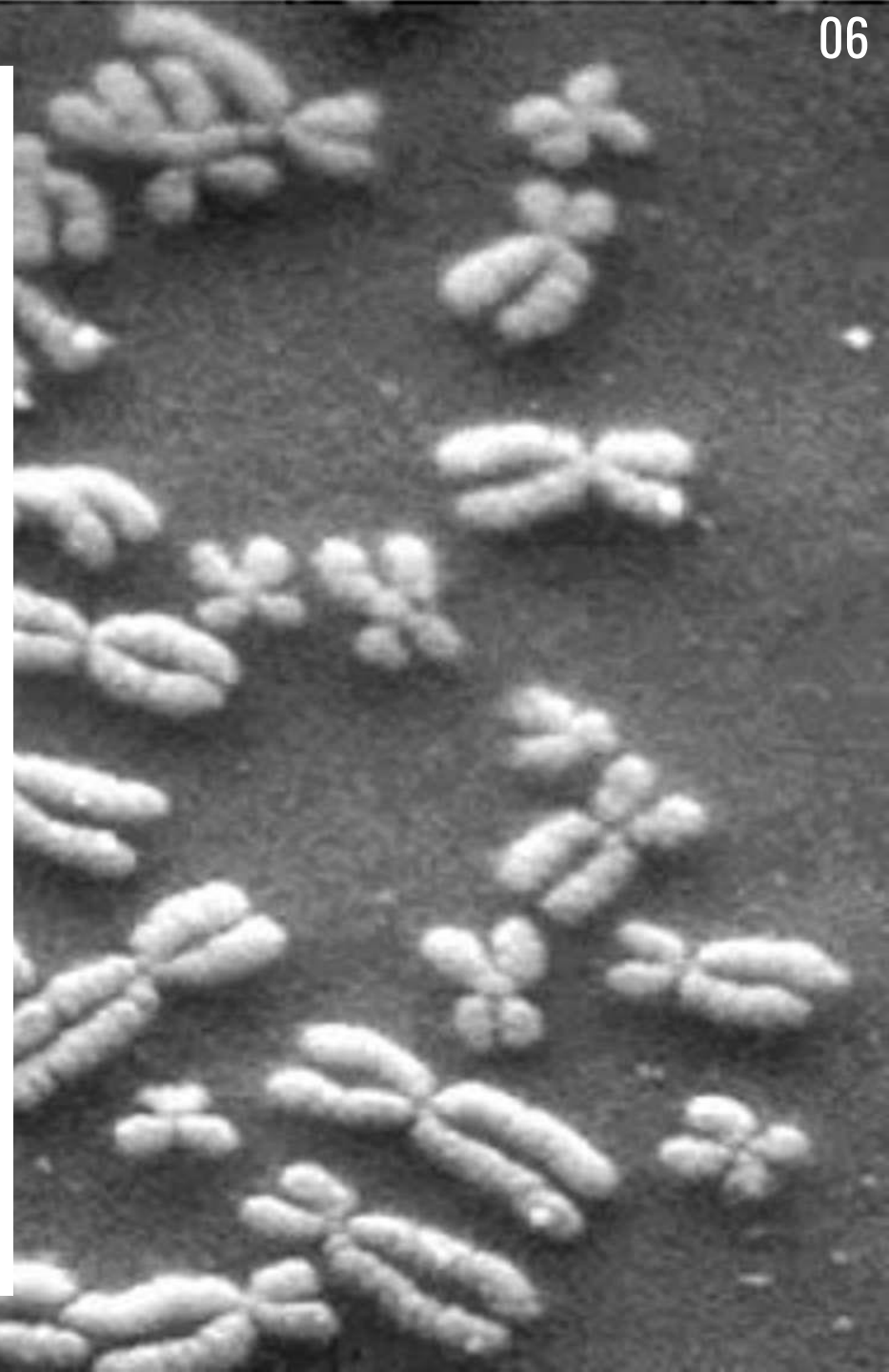
LOS GENES

SON UNIDADES DE INFORMACIÓN GENÉTICA. SECCIONES EN EL ADN CON INSTRUCCIONES PARA DIVERSAS FUNCIONES

→ PRODUCCIÓN DE PROTEÍNAS Y ARN → REGULACIÓN Y HERENCIA GENÉTICA

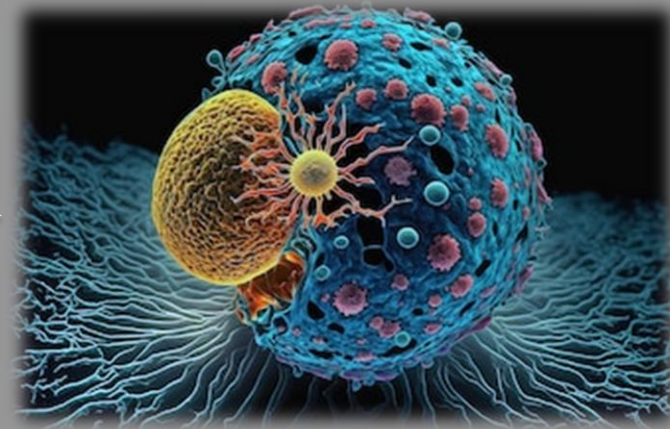
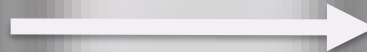
LAS MUTACIONES SON CAMBIOS EN LOS GENES QUE PUEDEN ALTERAR LA FUNCION NORMAL DE UNA CÉLULA E INCIDIR EN EL CANCER

→ OBTENCION DE ENERGÍA [IDH1]	→ FRENO AL CRECIMIENTO Y PROLIFERACIÓN CELULAR [PTEN]
→ ACTIVACIÓN DELCRECIMIENTO Y PROLIFERACIÓN CELULAR [EGFR]	→ CONTROL DE LA PRODUCCIÓN CELULAR [CIC]
→ MANTENIMIENTO / REPARACIÓN DEL ADN [ATRX]	→ AFECCIÓN DE LA APOPTOSIS [TP53]



OBJETIVOS

- OPTIMIZAR LA IDENTIFICACIÓN DEL TIPO DE GLIOMA UTILIZANDO TÉCNICAS DE APRENDIZAJE AUTOMÁTICO
- ANALIZAR LAS MUTACIONES GENÉTICAS ASOCIADAS PARA MEJORAR EL DIAGNÓSTICO
- FACILITAR UN TRATAMIENTO MÁS PRECISO Y EFECTIVO PARA CADA PACIENTE



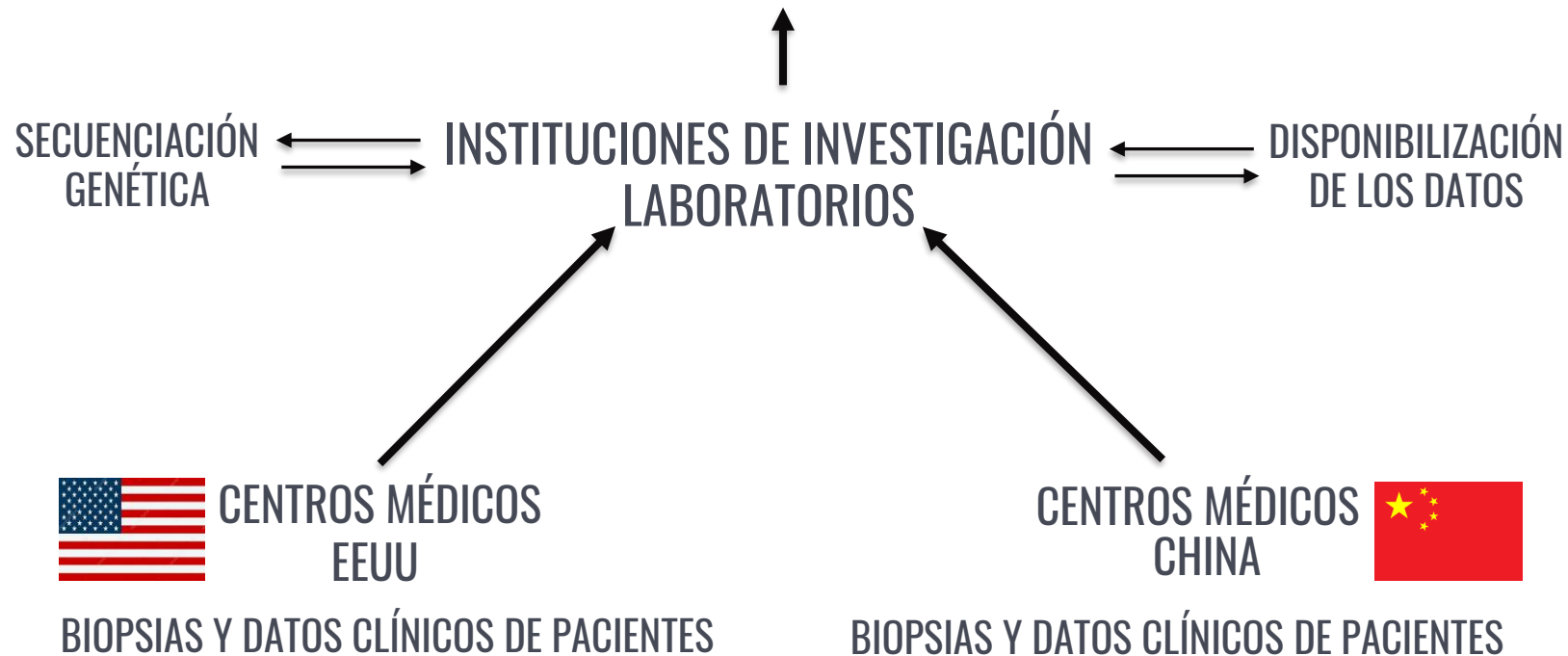
02

PRESENTACIÓN DEL DATASET Y EDA



ORIGEN DEL DATASET

NATIONAL CANCER INSTITUTE [EEUU] DIVISION RADIOTERAPIA



ESTE DATASET CONTIENE INFORMACIÓN GENÉTICA Y CLÍNICA DE PACIENTES CON GLIOMAS

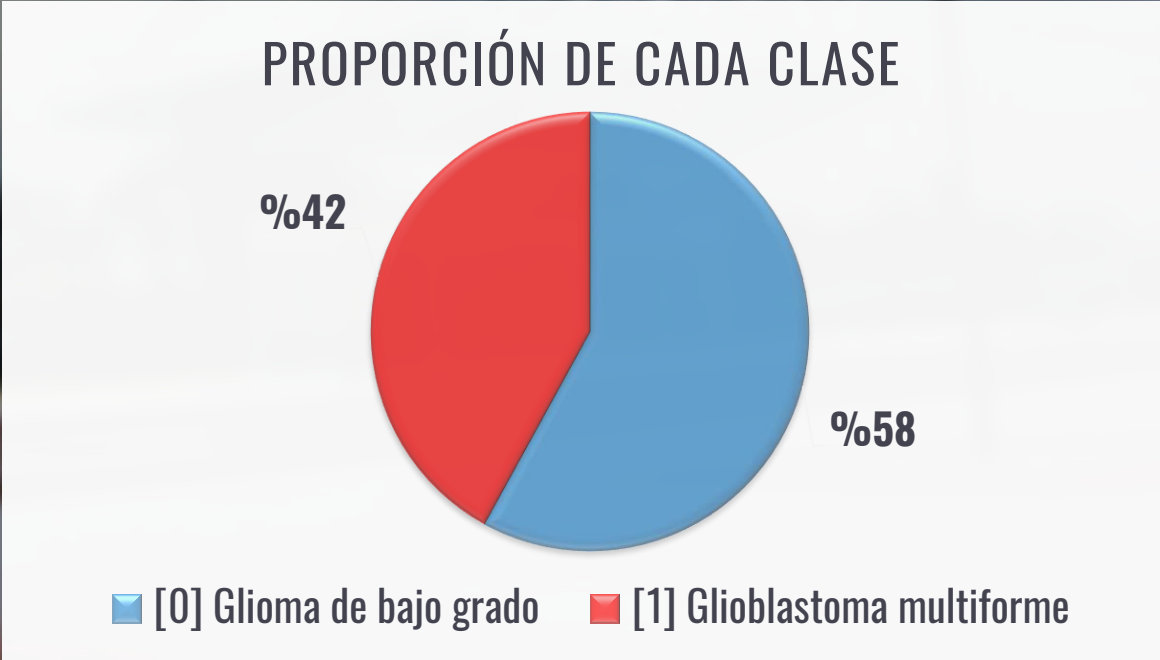
COLUMNAS PRINCIPALES

GRADE	INT	[0] GLIOMA DE BAJO GRADO [1] GLIOBLASTOMA MULTIFORME
GENDER	INT	[0] FEMENINO – [1] MASCULINO
AGE_AT_DIAGNOSIS	FLOAT	14~89
RACE	INT	[0] BLANCO – [1] ASIATICO [2] AFROAMERICANO [3] NO REPORTADO
MUTACIONES GENÉTICAS IDH1-TP53-ATRX-PTEN-EGFR- CIC-MUC16-PIK3CA-NF1- PIK3R1-FUBP1-RB1-NOTCH1- BCOR-CSMD3-SMARCA4- GRIN2A-IDH2-FAT4-PDGFRA	INT	[0] NO MUTADO - [1] MUTADO

COMPOSICIÓN DEL DATASET

COLUMNAS	24
FILAS	839

SIN NA's



DESCRIPCIÓN DEL PROBLEMA

CLASIFICACIÓN BINARIA

TARGET: 'GRADE'

LGG:
GLIOMA DE BAJO GRADO

CLASE 0:
NEGATIVA

GBM:
GLIOBLASTOMA MULTIFORME

CLASE 1:
POSITIVA

MÉTRICAS

EXACTITUD
PRECISIÓN

EXHAUSTIVIDAD (RECALL)

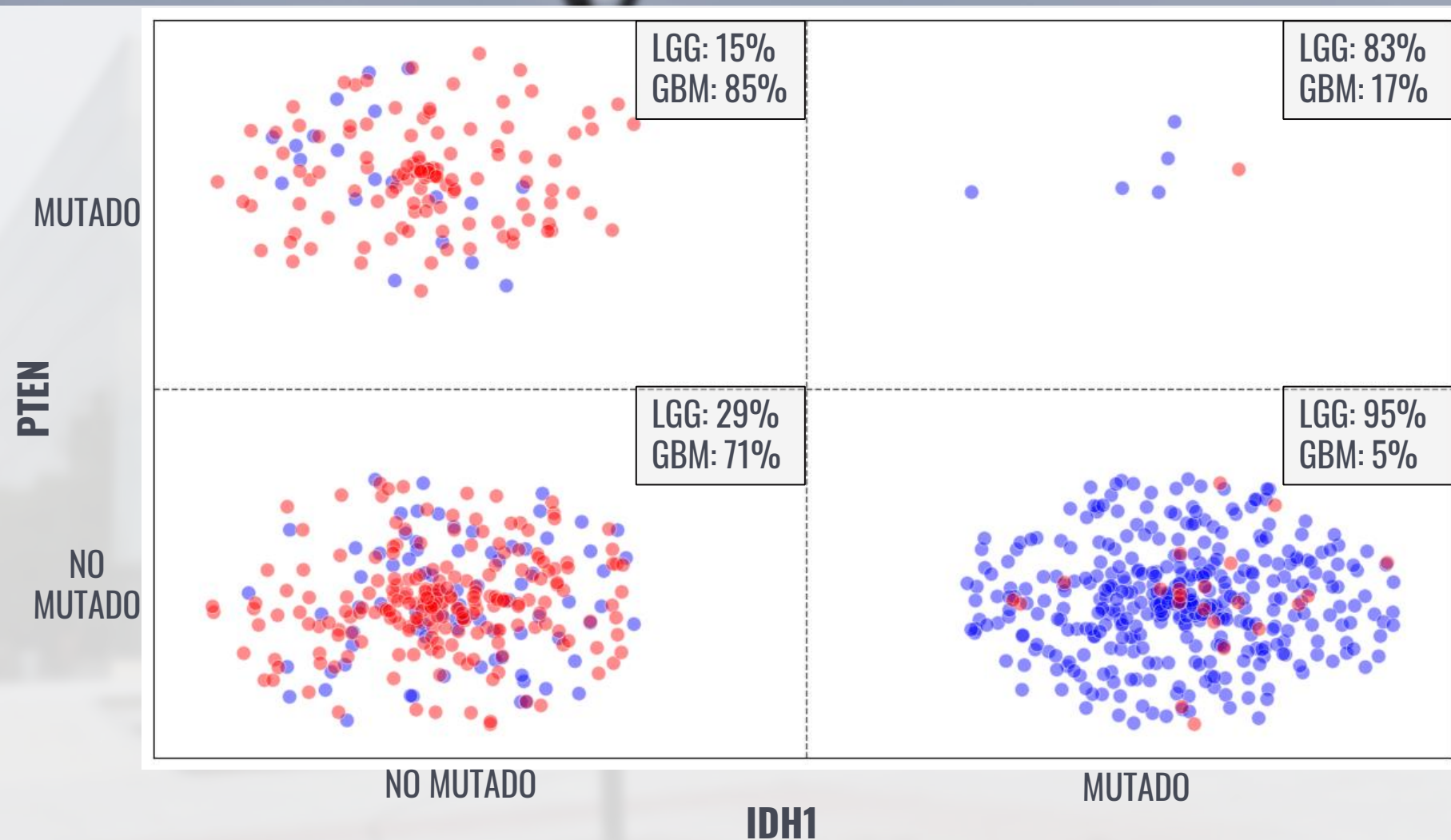
IDENTIFICACIÓN CORRECTA DE LOS GBM

MINIMIZACIÓN DE FALSOS NEGATIVOS

MATRIZ DE CORRELACIÓN

	Grade	Age	IDH1	PTEN	ATRX	CIC	EGFR
Grade	1	0.53	-0.71	0.37	-0.31	-0.30	0.24
Age	0.53	1	-0.57	0.26	-0.42	-0.12	0.23
IDH1	-0.71	-0.57	1	-0.39	0.45	0.32	-0.36
PTEN	0.37	0.26	-0.39	1	-0.19	-0.14	0.18
ATRX	-0.31	-0.42	0.45	-0.19	1	-0.17	-0.16
CIC	-0.30	-0.12	0.32	-0.14	-0.17	1	-0.12
EGFR	0.24	0.23	-0.36	0.18	-0.16	-0.12	1

SCATTERPLOT IDH1 y PTEN



- CLASE 0 LGG (BAJO GRADO)
- CLASE 1 GBM (AGRESIVO)

03

ENTRENAMIENTO Y EVALUACIÓN DE MODELOS



MODELOS UTILIZADOS

- BENCHMARK
- ÁRBOL DE DECISIÓN CON UN ÚNICO ATRIBUTO: IDH1
- ÁRBOLES DE DECISIÓN CON 2 ATRIBUTOS

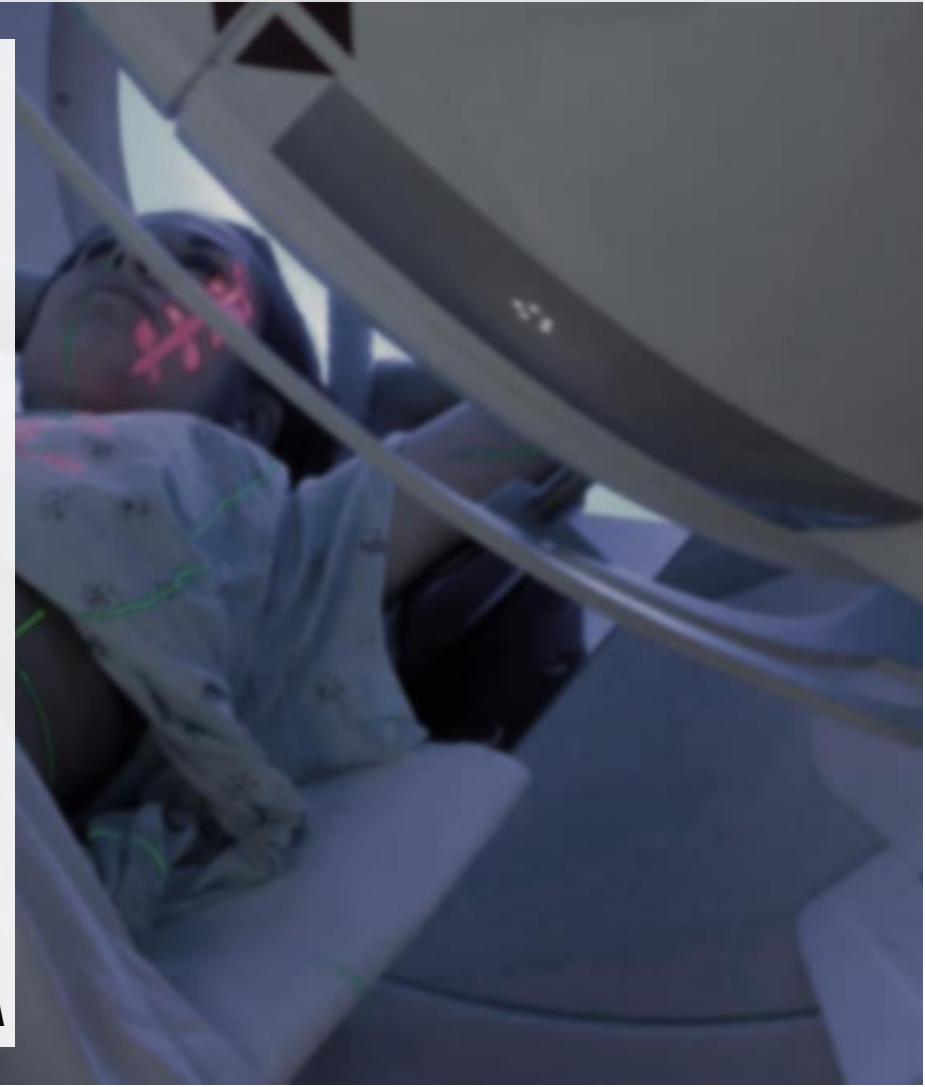
ATRIBUTO 1	ATRIBUTO 2
IDH1	PTEN
	ATRX
	EGFR
	CIC

TRAIN TEST

- ÁRBOL DE DECISIÓN CON CINCO ATRIBUTOS: IDH1, PTEN, ATRX, EGFR, CIC.

- RANDOM FOREST CON TODOS LOS GENES COMO ATRIBUTOS →

TRAIN TEST
VALIDACIÓN CRUZADA



BENCHMARK

DUMMY CLASSIFIER (STRATEGY = 'STRATIFIED')

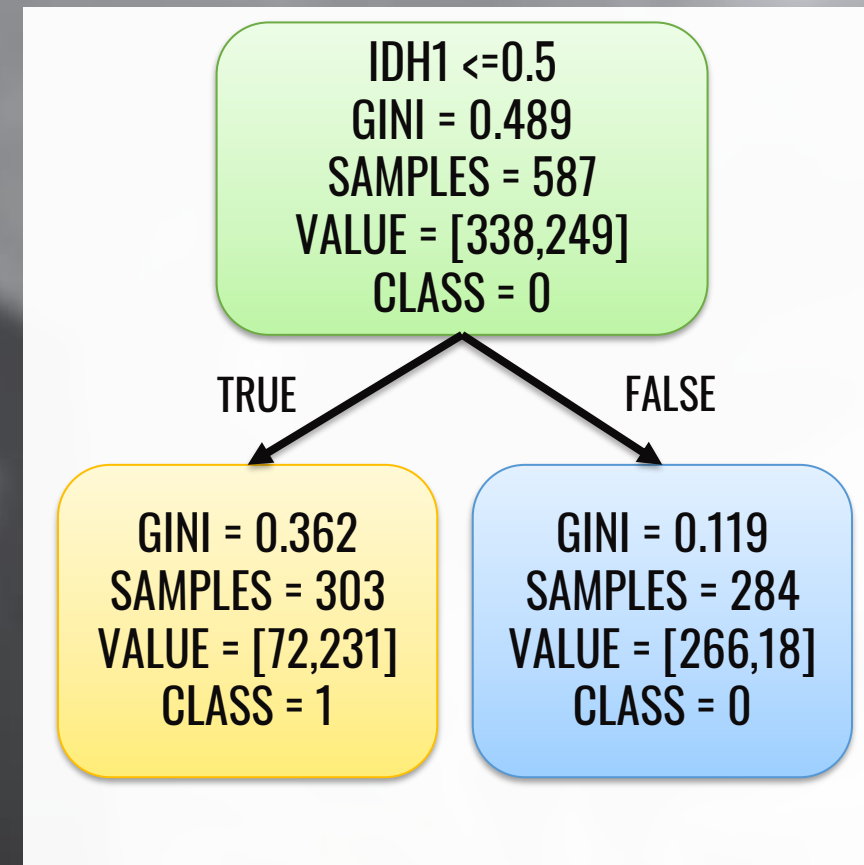
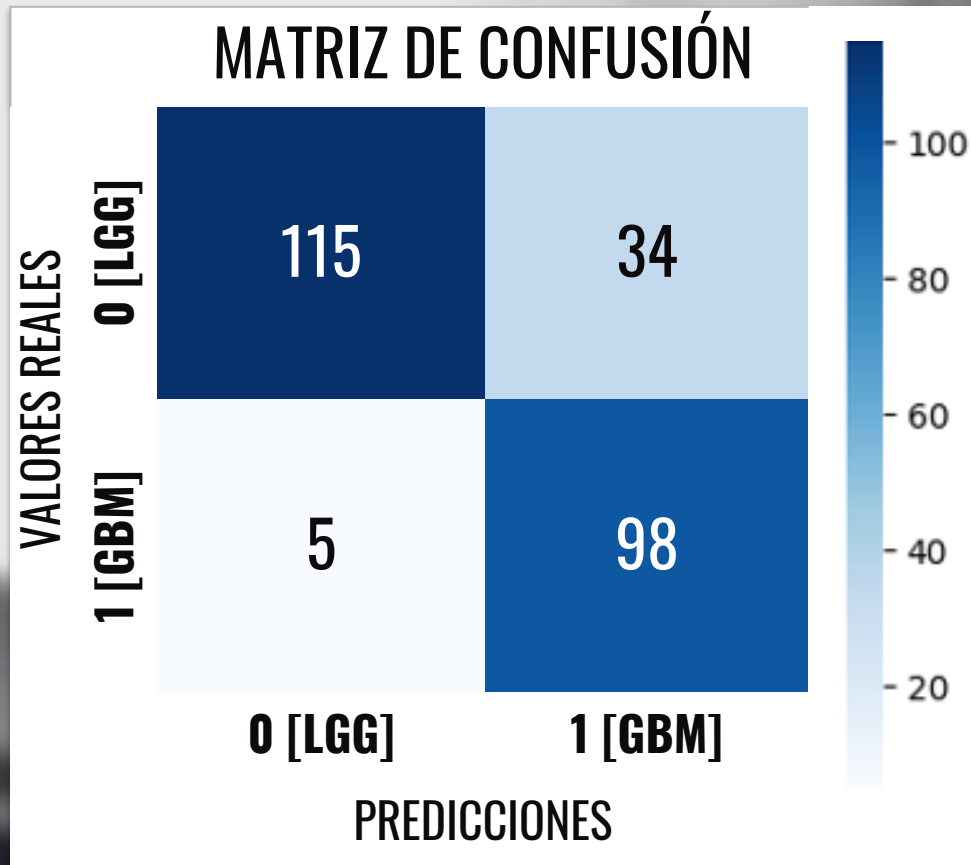
GENERA PREDICCIONES MANTENIENDO LA DISTRIBUCIÓN DE CLASES DEL CONJUNTO DE ENTRENAMIENTO

EXACTITUD [ACCURACY] : 0.52

EXHAUSTIVIDAD [RECALL] : 0.43

PRECISION : 0.41

ARBOL DE DECISIÓN CON 1 ATRIBUTO: IDH1

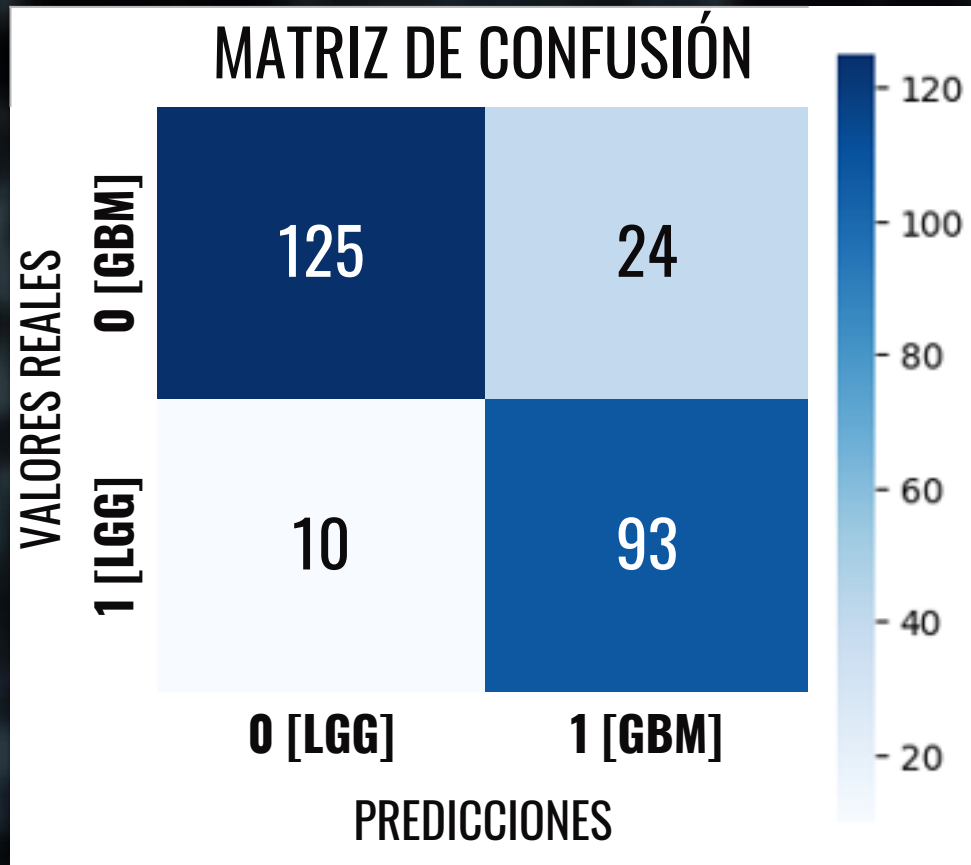


EXACTITUD [ACCURACY] : 0.85

EXHAUSTIVIDAD [RECALL] : 0.95

PRECISION : 0.74

ARBOL DE DECISIÓN CON 5 ATRIBUTOS: IDH1/PTEN/ATRX/EGFR/CIC



↑ ACCURACY
↑ PRECISION
↓ RECALL

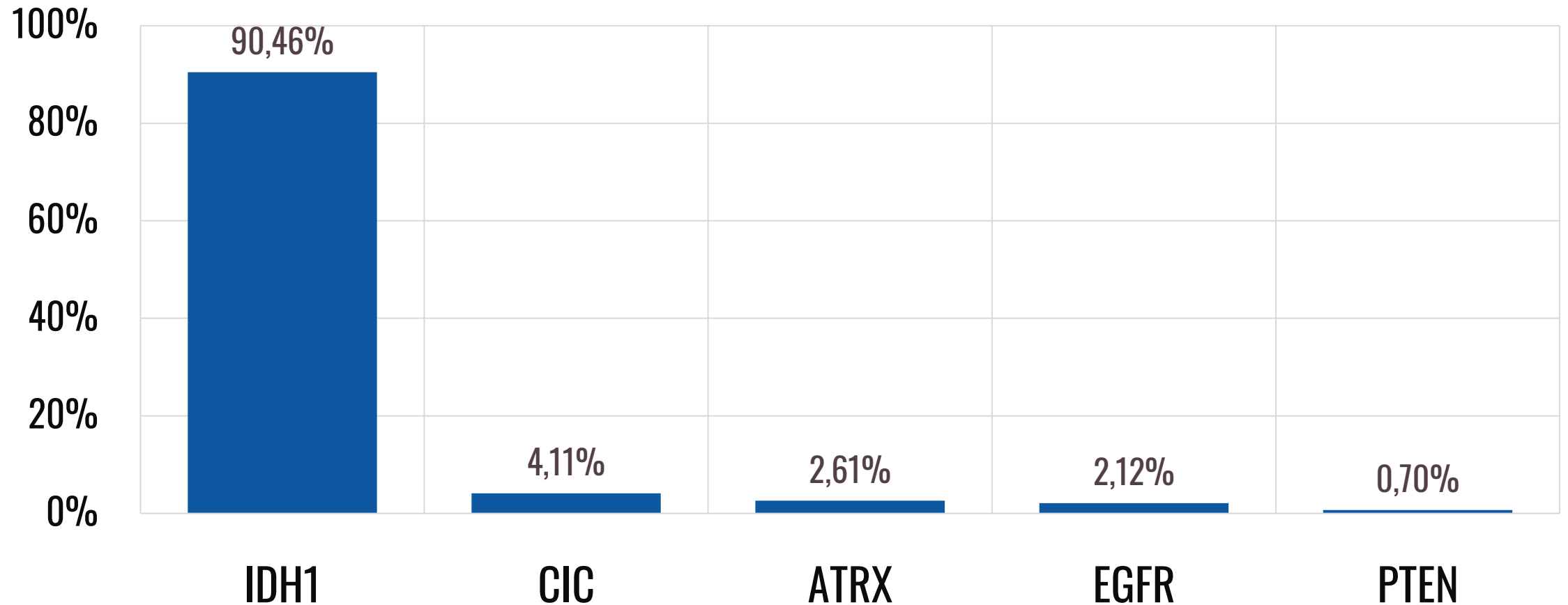
EXACTITUD [ACCURACY] : 0.87

EXHAUSTIVIDAD [RECALL] : 0.90

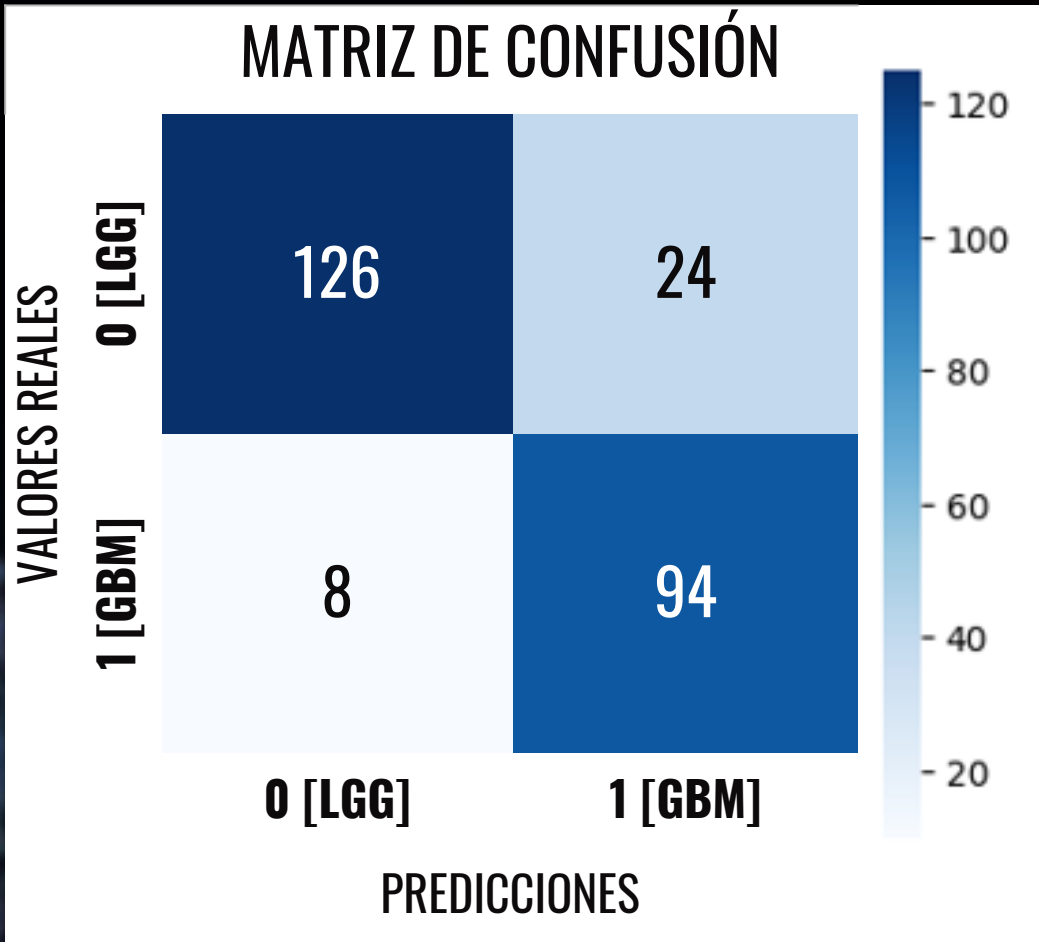
PRECISION : 0.79

ARBOL DE DECISIÓN CON 5 ATRIBUTOS: IDH1/PTEN/ATRX/EGFR/CIC

IMPORTANCIA DE LOS ATRIBUTOS [%]



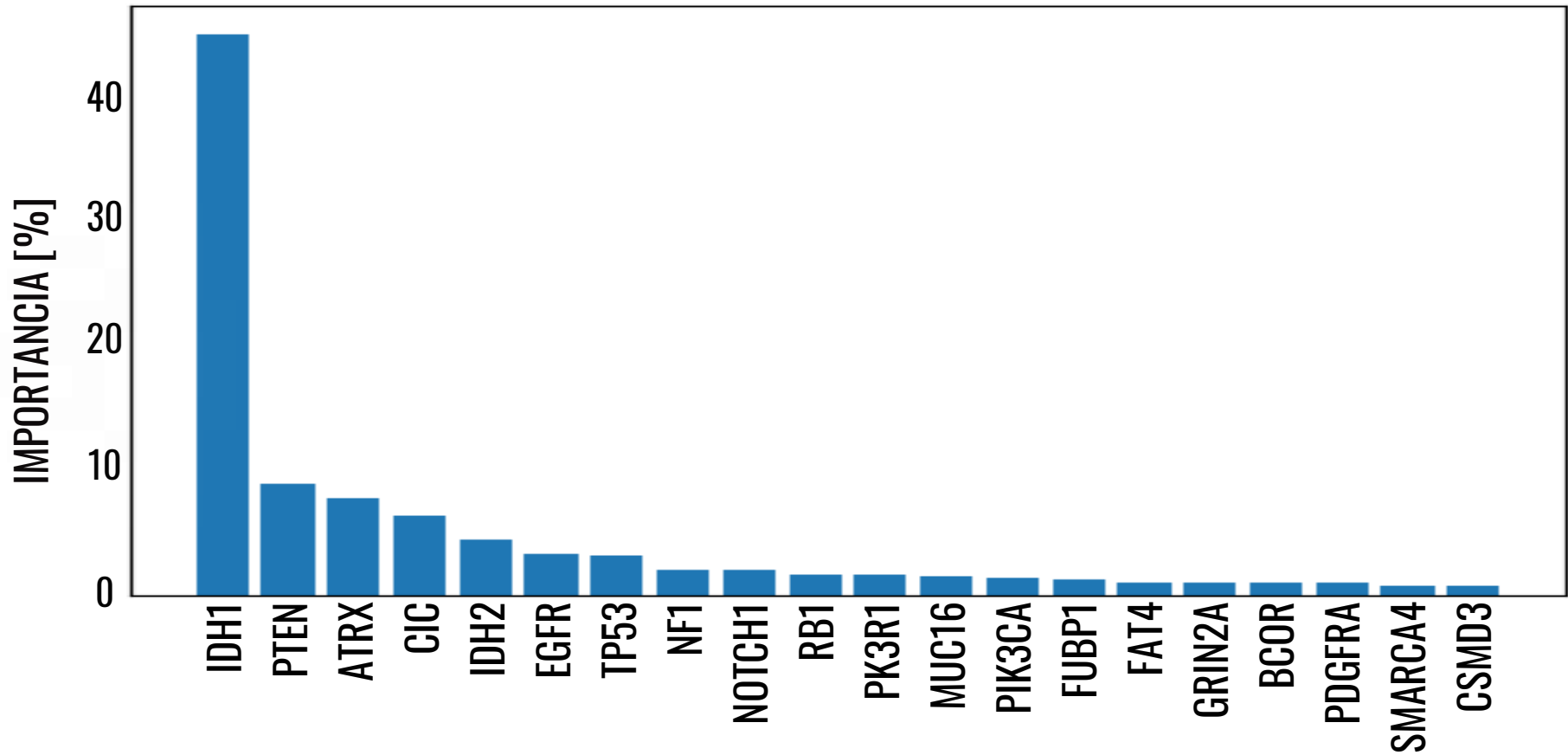
RANDOM FOREST CON TODOS LOS ATRIBUTOS [20]



ACCURACY	RECALL	PRECISION
0.86	0.92	0.79
0.86	0.94	0.78
0.85	0.90	0.78
0.88	0.90	0.83
0.85	0.92	0.77
0.86 ± 0.01	0.92 ± 0.01	0.79 ± 0.02

RANDOM FOREST CON TODOS LOS ATRIBUTOS [20]

IMPORTANCIA DE LOS ATRIBUTOS



ATRIBUTO	IMPORTANCIA [%]
IDH1	46.75
PTEN	9.24
ATRX	8.11
CIC	6.59
IDH2	4.70
EGFR	3.45



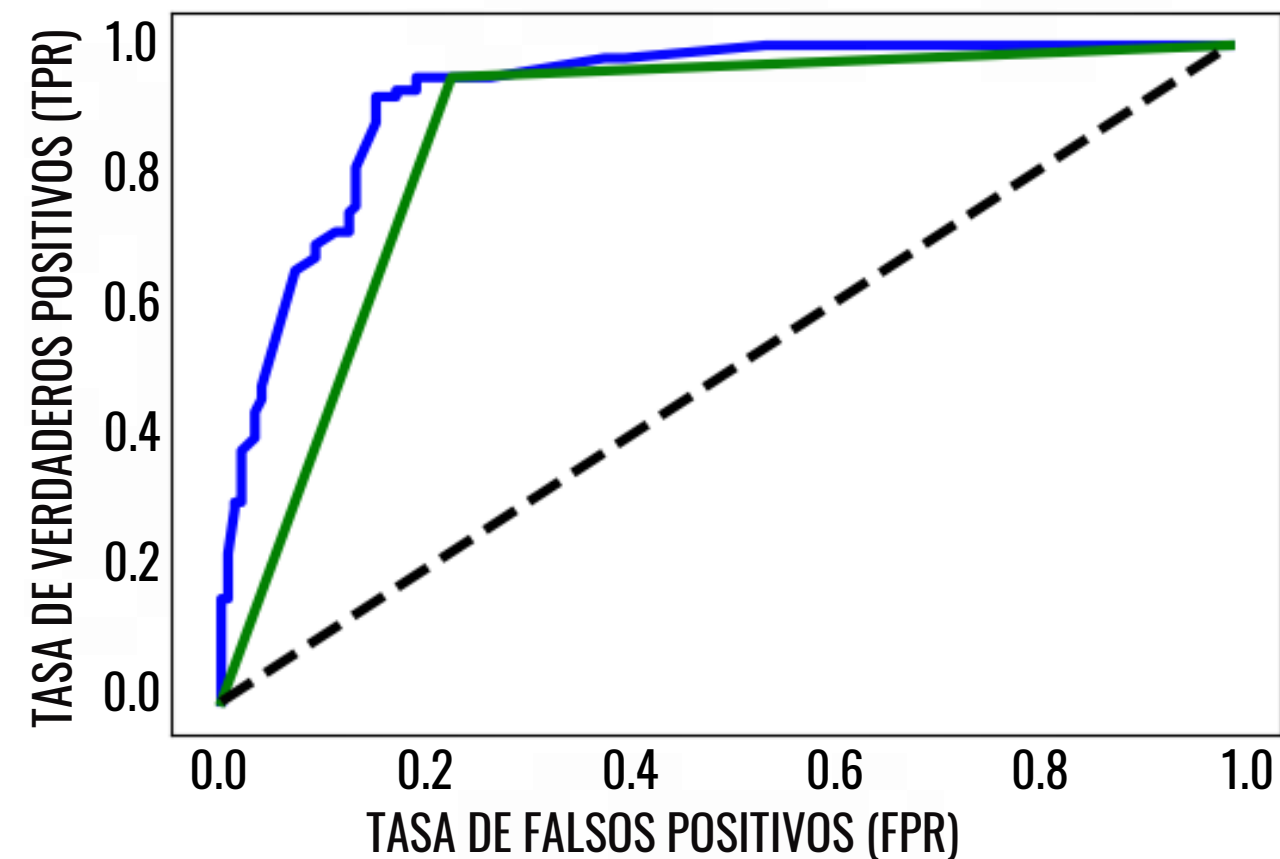
04

COMPARACIÓN DE MODELOS



COMPARATIVA DE MODELOS – AUC ROC

CURVA ROC COMPARATIVA



— RANDOM FOREST (AUC =0.93)

MAYOR QUE

— ÁRBOL DE DECISIÓN (AUC=0.86)

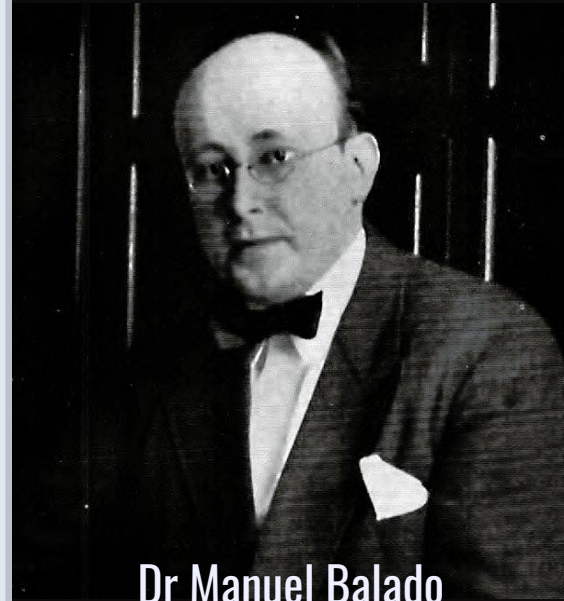
COMPARATIVA DE MÉTRICAS

	ACCURACY	RECALL	PRECISION
ARBOL DE DECISIÓN CON 1 ATRIBUTO: IDH1	0.85	0.95	0.74
ÁRBOL DE DECISIÓN CON 5 ATRIBUTOS: IDH1/PTEN/ATRX/EGFR/CIC	0.87	0.90	0.79
RANDOM FOREST CON TODOS LOS GENES COMO ATRIBUTOS [20]	0.86 ± 0.01	0.92 ± 0.01	0.79 ± 0.02

05 CONCLUSIÓN



- IDH1 MUTADO: LGG
 - IDH1 NO MUTADO: GBM
- } IDH1 BUEN PREDICTOR
- OTROS GENES: CONTRIBUCIÓN BAJA COMPARADA CON IDH1.
 - MODELO ELEGIDO: ÁRBOL DE DECISIÓN CON LA MUTACIÓN DE IDH1 COMO ÚNICO ATRIBUTO (PRIORIZANDO RECALL).
 - TRABAJO A FUTURO: REPETIR EL ANÁLISIS EXCLUYENDO IDH1, PARA EVALUAR SI OTROS GENES PREDICEN EL GRADO DEL GLIOMA.



Dr Manuel Balado

¡MUCHAS GRACIAS!

ESPACIO DE PREGUNTAS



LA CIENCIA NO ES UN LUJO,
ES UNA NECESIDAD PARA EL
PROGRESO DE LA HUMANIDAD

LUIS FEDERICO LELOIR

A microscopic image of a cell, likely a cancer cell, with a large, irregular, blue nucleus and several smaller, red, spherical structures attached to its surface. The cell is surrounded by a dense network of fine, blue, hair-like structures. The entire image is set against a dark background.

EXTRA



MODELO CON REGRESIÓN LOGÍSTICA

Atributo predictor: “Age_at_diagnosis”

Target: “Grade”

Usamos la edad de los pacientes para generar el modelo

ACCURACY

0.73

RECALL

0.64

PRECISIÓN

0.67

Las métricas no son tan buenas
como en los otros modelos.

