

Detección Cancer De Pecho

David Barrera - 20002878
Jairo Garcia - 20002878
Pablo Rodríguez - 20002862



Introducción

Breast Cancer Detection o Detección de Cancer de Pecho siendo el tema principal del proyecto. Se solicito crear una inteligencia artificial la cual lograra detectar si habia cancer y Se nos brindaron un dataset que incluye cierta inforación que debias analizar para comprender con lo que estabamos trabajando, luego de analizar y entender el dataset nos percatamos que se tiene dos clases o tipos para clasificar el tejido del pecho siendo maligno o benigno.

Otro punto que tuvimos en cuenta fue que no hay imamgenes en el data set unicamente datos numeros, por lo que tambien nos da una idea que tenemos que utilizar un modelo de redes neuranales para empezar a crear nuestra inteligencia artificial. Partimos de lo que sabemos y entendemos para iniciar tenemos que tener en mente la estructura de nuestro modelo, que en este caso comenzamos con la importacion de librerria, creacion de funciones que tenemos en mente utilizar, preparar el dataset para que la AI pueda utilizarla, pre-procesar la data y por ultimo creamos el modelo donde estaremos haciendo uso de toda la data modificada para entrenar y, evaluar el modelo para observar si los resultados obtenidos son los deseados.

Descripción Del Dataset

Descripción del dataset:
1)Numero de ID
2)Diagnosis (M = maligno, B = benigno)
Distribucion de clase: 357 benigno, 212 maligno
Se calculan diez características de valor real para cada núcleo celular:
a)Radio
b)Textura
c)Perimetro
d)Area
e)Suavidad
f)compacidad
g)Concavidad
h)Putnos Cóncavos
i)Simetría
j)Dimensión Fractal
Se tuvo tomar en cuenta es que los numeros son medias y en 4 cifras significativas.

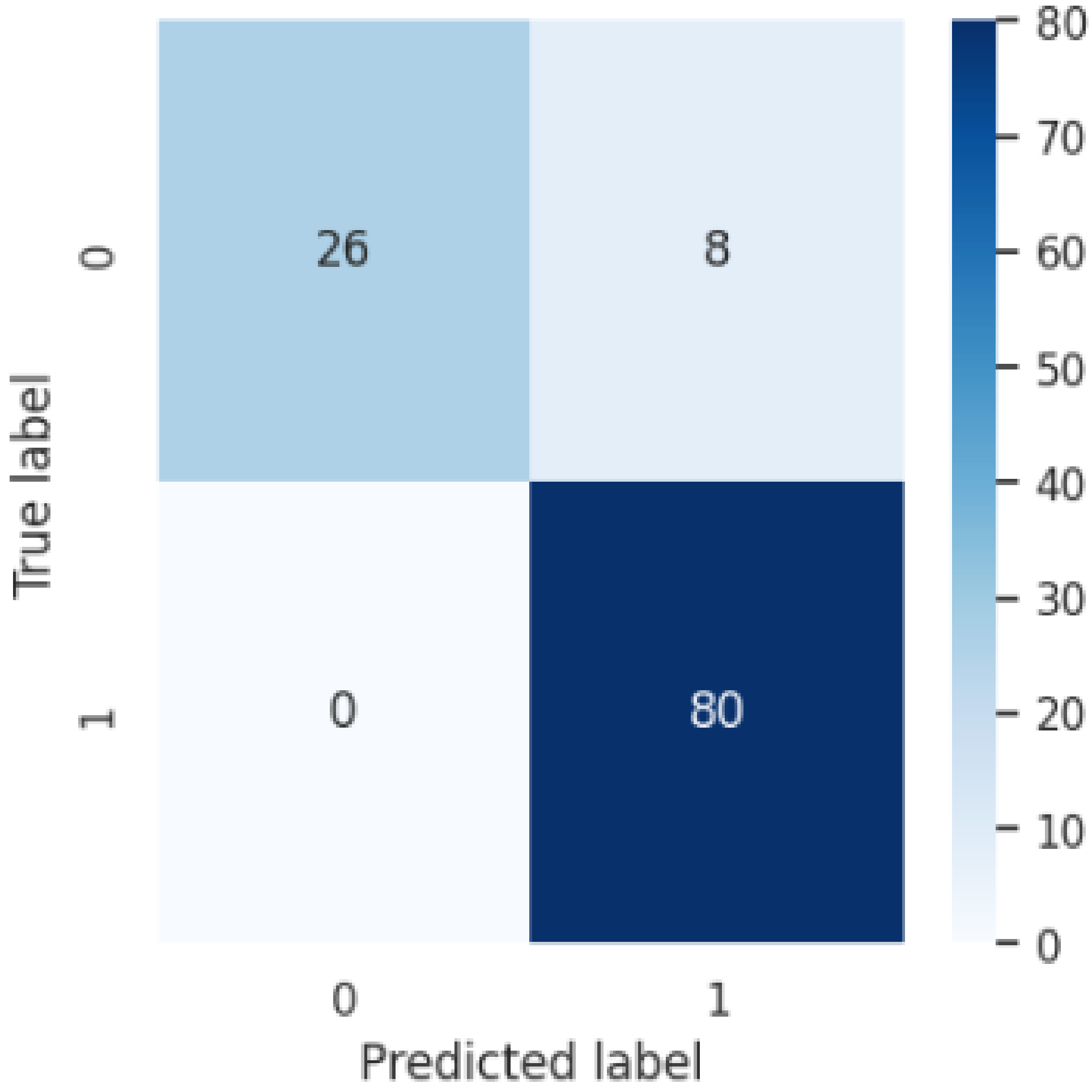
Metodología

Para este problema se decidio usar una red neuronal debido a que este es un problema con datos estructurados y de clasificación. Como primer punto, se normalizaron los datos para que estos estuvieran dentro de un rango de 0 a 1, permitiendo asi mejor manipulación de los datos. Luego, mediante la matriz de correlacion se comparaban los campos que no contribuian mucho en el diagnóstico y se eliminaban. Por ultimo, se trabajo con dos modelos, uno en el cual se trabajaba con la distribución de datos normal y otro donde se hizo un balanceo de clases.

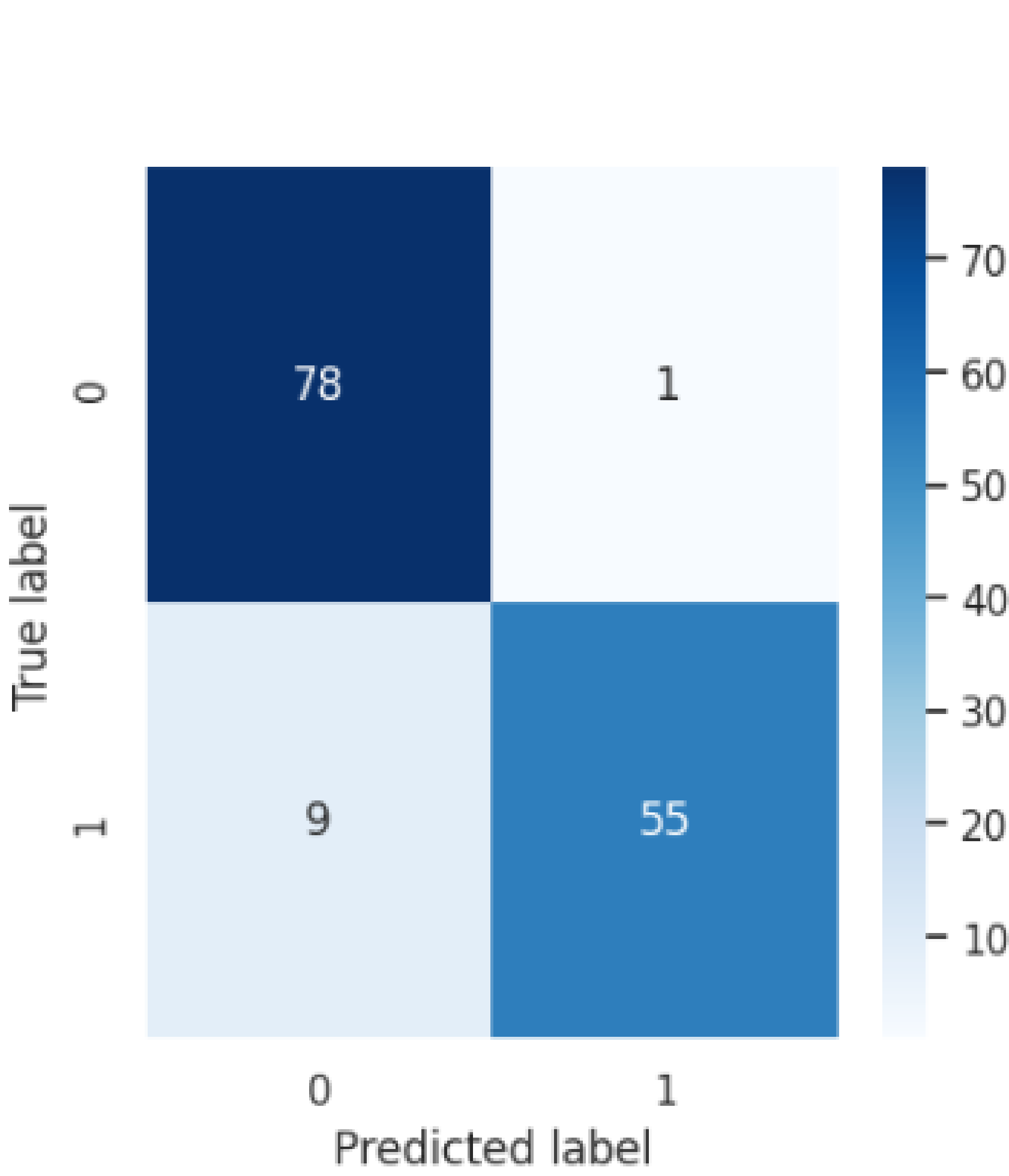
Modelo 1	Modelo 2
Utilizando una cantidad de 569 datos en total.	Utilizando una cantidad de 714 datos en total.
Una precisión del 92.98%	Una precisión del 93%
Mayor cantidad de falsos positivos (8)	Menor cantidad de falsos positivos (1)
Menor cantidad de falsos negativos (0)	Mayor cantidad de falsos negativos (9)

Resultados

Tumor Maligno = 0
Tumor Benigno = 1
Modelo 1 tiene una precision del 92.98%



Tumor Maligno = 0
Tumor Benigno = 1
Modelo 2 tiene una precision del 93%



Mejoras a Futuro

- Añadir mas campos relevantes al set de datos
- Experimentar con una mayor cantidad de datos
- Tener un dataset para el entrenamiento del modelo y otro para la evaluacion
- Implementacion de “computer vision” para poder clasificar los tumores utilizando imagenes

Conclusiones

- Obtuvimos resultados satisfactorios
- Se pudo clasificar de una buena manera sin obtener falsos negativos
- Nuestro modelo propuesto se adaptó a las necesidades del problema
- Aún hay puntos de mejora en nuestro problema

QR Code

