

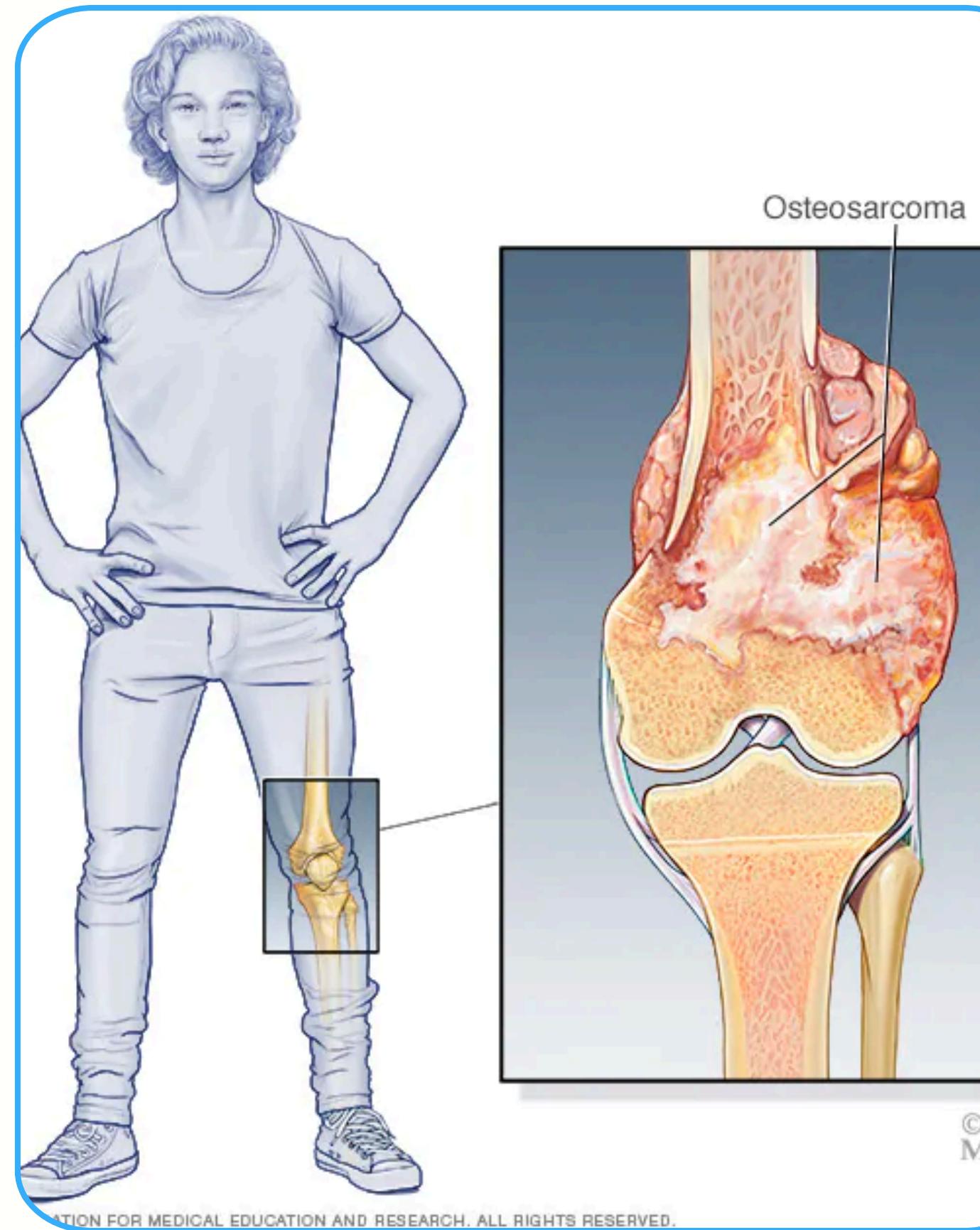


## ANÁLISIS EXPLORATORIO

# Identificación de tumores de osteosarcoma

By: Jose Villalta - Daniel Collaguazo ➤

01





# Descripción del dataset

El osteosarcoma es uno de los tipos más comunes de cáncer óseo primario, caracterizado por su rápida progresión y la tendencia a propagarse a otras partes del cuerpo. Este tipo de cáncer afecta principalmente a los huesos largos y puede generar dolor persistente, inflamación, y en algunos casos, la formación de masas visibles.

El dataset cuenta con la siguiente información:

**Nombre:** Osteosarcoma-Tumor-Assessment

**Propietario:** TCIA – UT Southwestern/UT Dallas

**Nº de Variables:** 68

**Numero de Pacientes:** 1144

**Variable de salida:** no tumor (0), viable, (1) tumor no viable (2), y tumor viable (3).

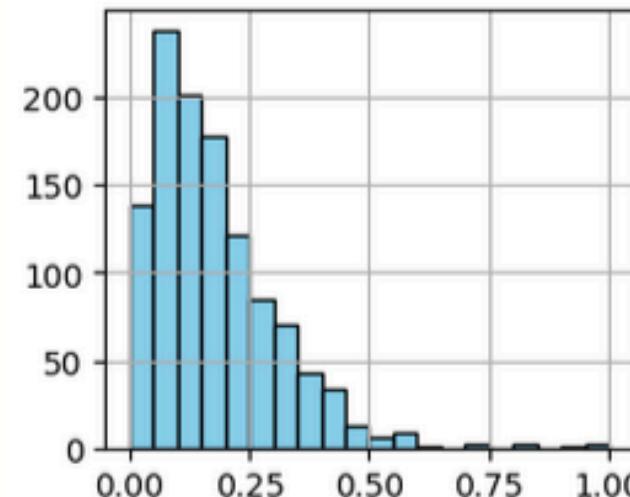
**Imagenes:** 1357 imágenes

**link:**

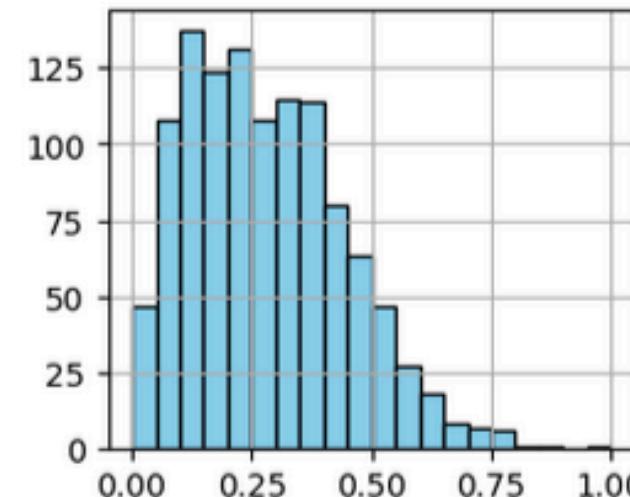
<https://www.cancerimagingarchive.net/collection/osteosarcoma-tumor-assessment/>

03

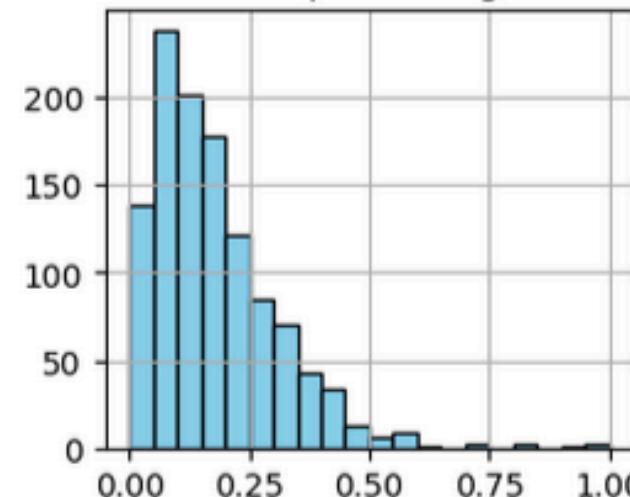
Blue.count



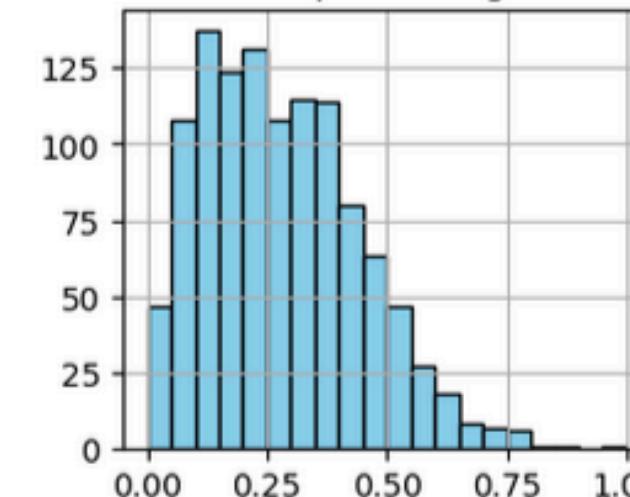
red.count



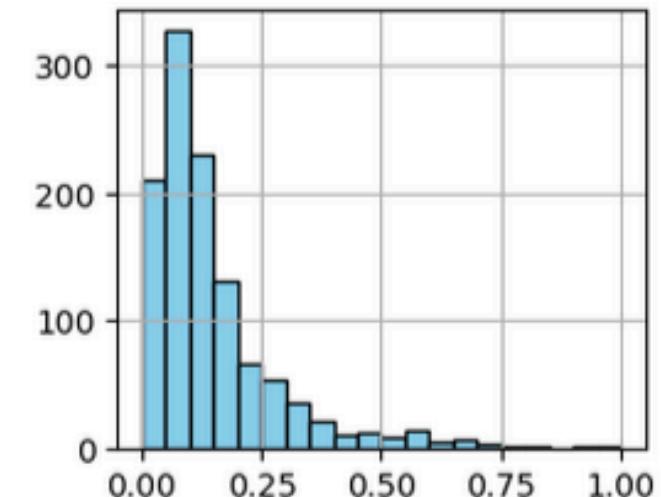
Blue.percentage



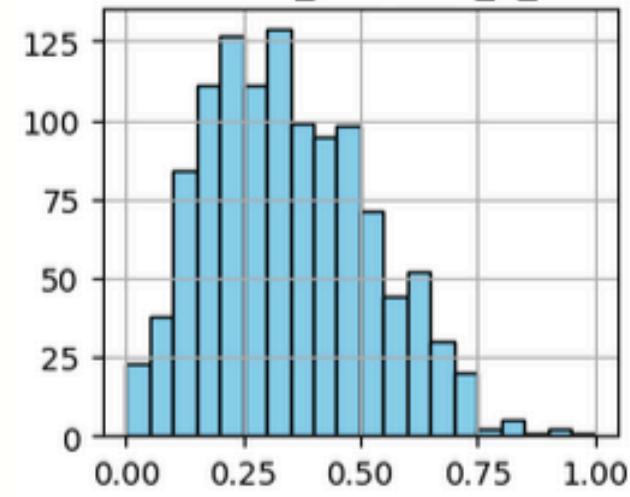
red.percentage



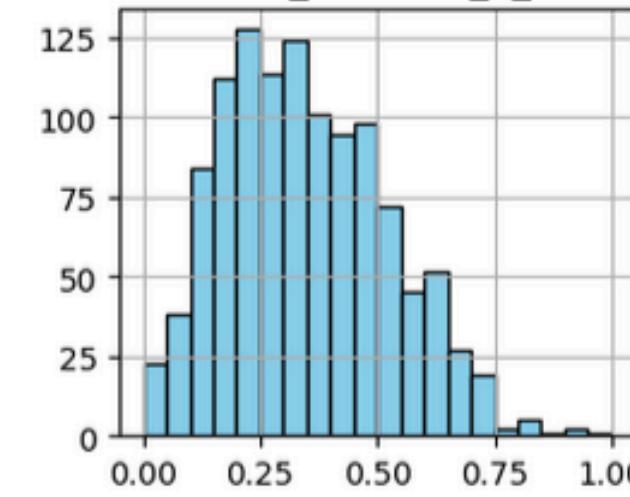
total.clusters



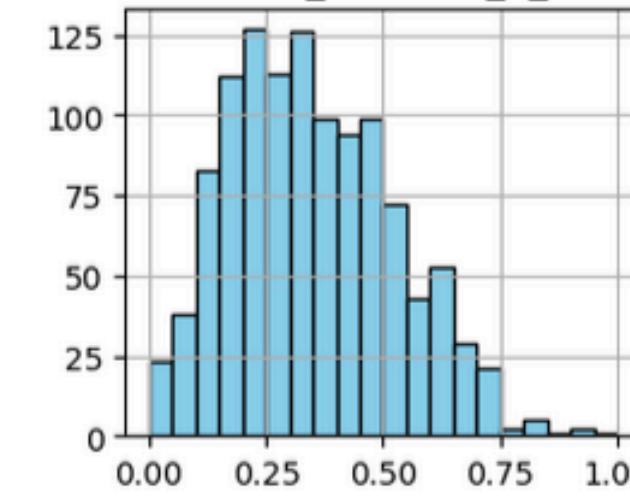
Texture\_Variance\_3\_0



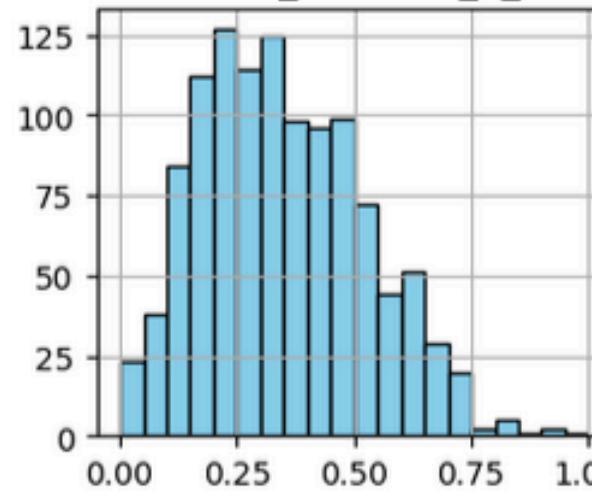
Texture\_Variance\_3\_135



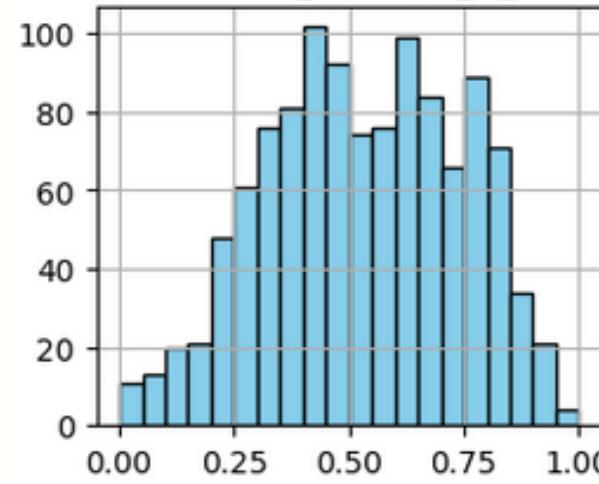
Texture\_Variance\_3\_45



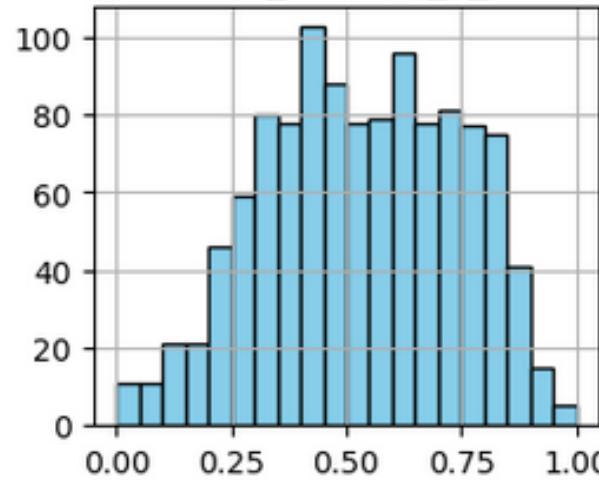
Texture\_Variance\_3\_90



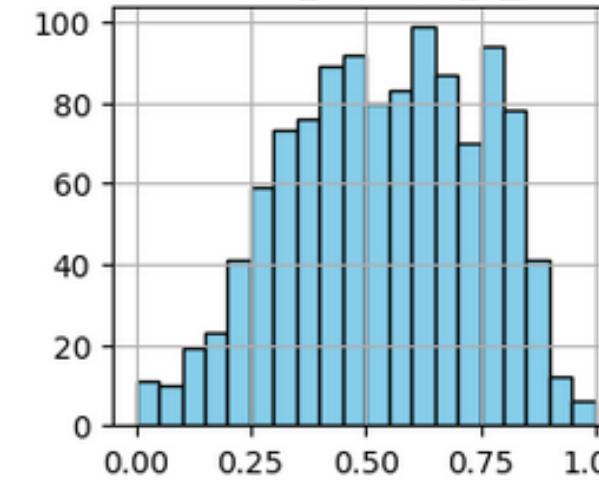
Texture\_Entropy\_3\_0



Texture\_Entropy\_3\_135

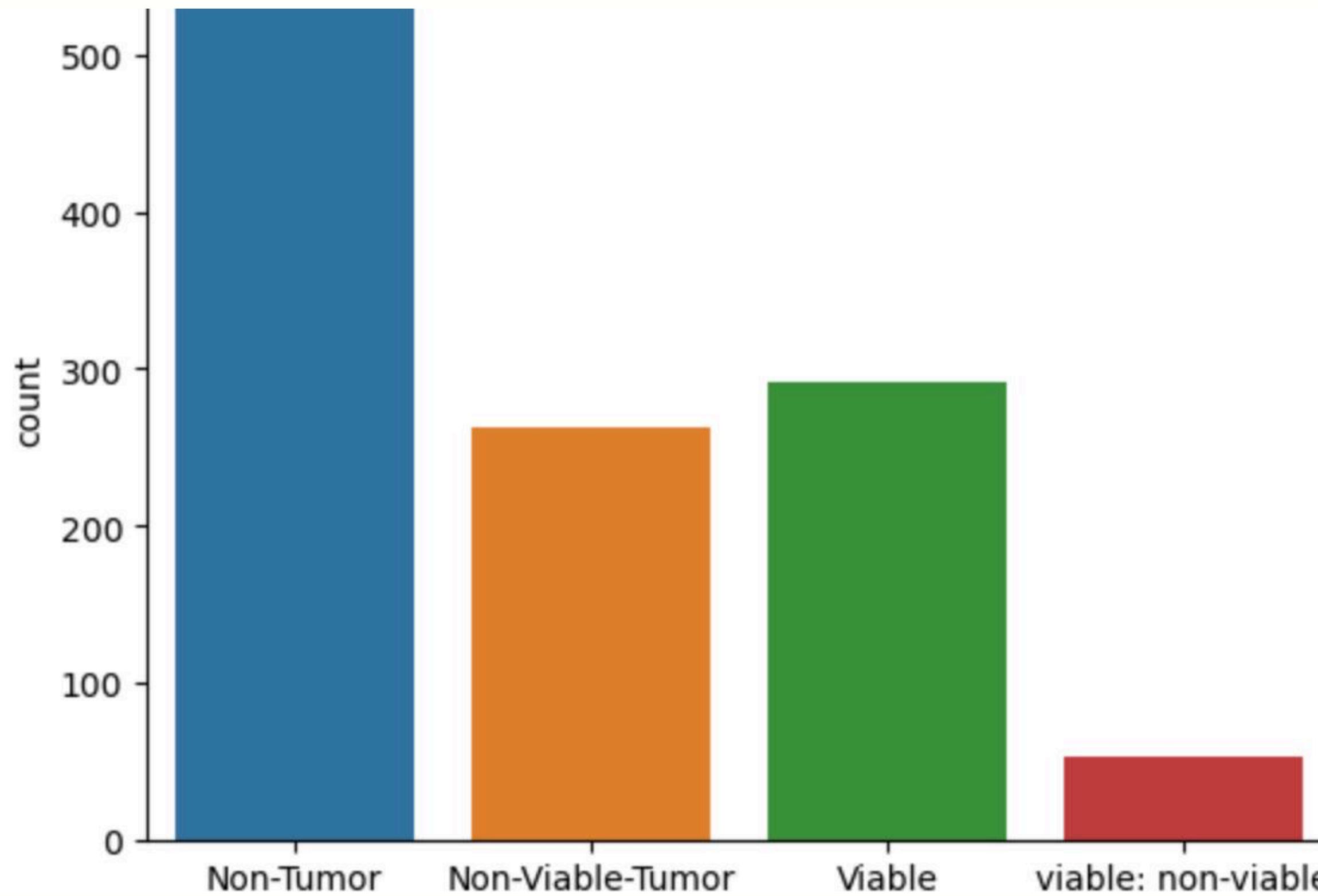


Texture\_Entropy\_3\_45

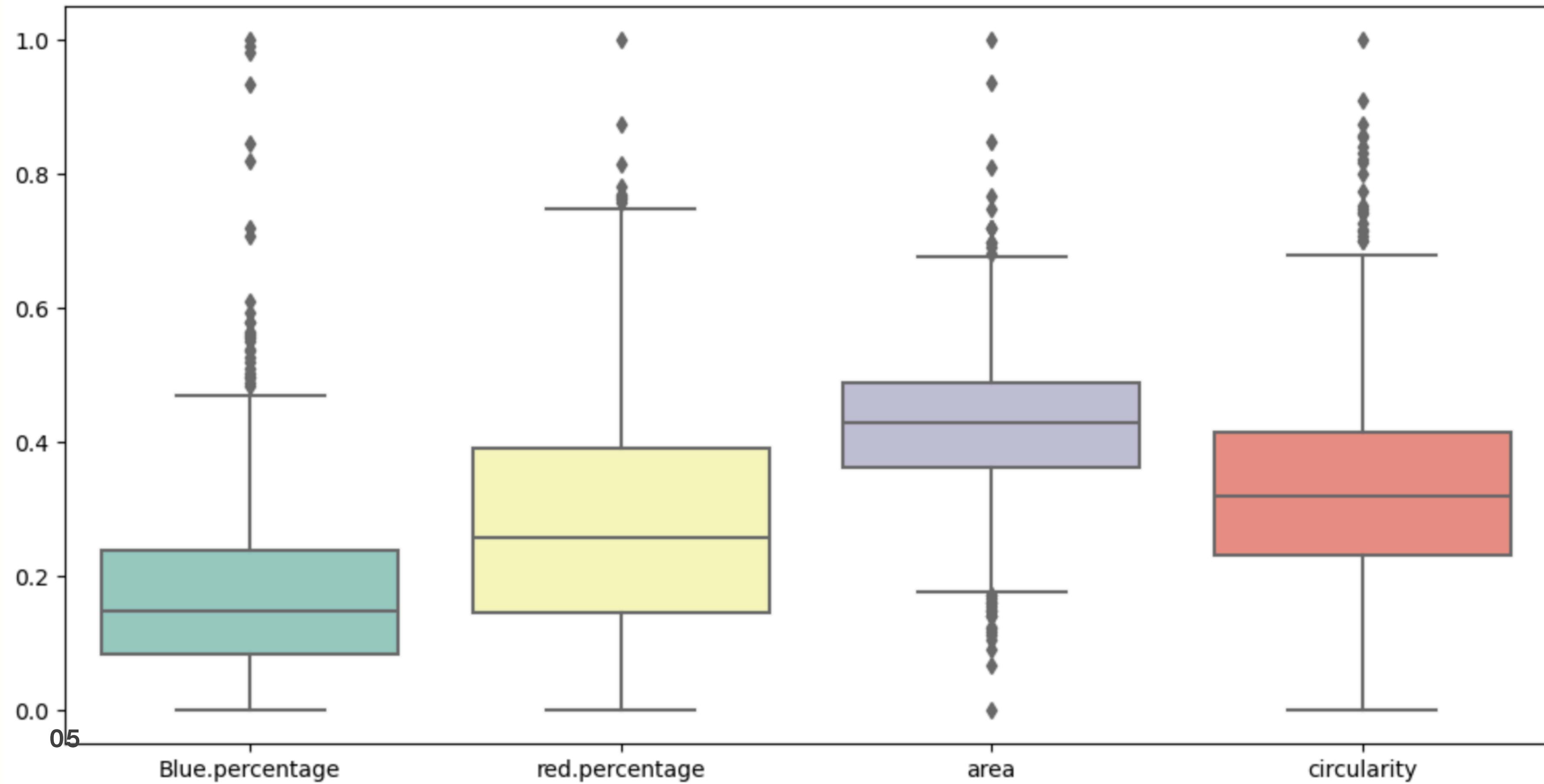


# Clasificación imágenes

04



Boxplot características clave



# Correlación de variables con presencia de tumor

classification_weighted	<b>1.000000</b>	Texture_DifferenceVariance_3_135	<b>0.421818</b>
Texture_DifferenceEntropy_3_0	<b>0.609410</b>	Texture_DifferenceVariance_3_0	<b>0.406402</b>
Texture_DifferenceEntropy_3_135	<b>0.608757</b>	Texture_DifferenceVariance_3_45	<b>0.403474</b>
Texture_SumEntropy_3_135	<b>0.603431</b>	Texture_InfoMeas1_3_0	<b>0.400888</b>
Texture_SumEntropy_3_45	<b>0.602175</b>	Texture_Contrast_3_90	<b>0.397180</b>
Texture_DifferenceEntropy_3_45	<b>0.601817</b>	Texture_InfoMeas1_3_135	<b>0.394811</b>
Texture_SumEntropy_3_0	<b>0.601213</b>	Texture_InfoMeas1_3_45	<b>0.361402</b>
Texture_SumEntropy_3_90	<b>0.597034</b>	Texture_DifferenceVariance_3_90	<b>0.360444</b>
Texture_Entropy_3_0	<b>0.585934</b>	Blue.percentage	<b>0.349015</b>
Texture_Entropy_3_135	<b>0.585706</b>	Blue.count	<b>0.349015</b>
Texture_Entropy_3_45	<b>0.581157</b>	Texture_InfoMeas1_3_90	<b>0.324697</b>
Texture_DifferenceEntropy_3_90	<b>0.569198</b>	Texture_SumAverage_3_0	<b>0.259088</b>
Texture_Entropy_3_90	<b>0.568891</b>	Texture_SumAverage_3_90	<b>0.259059</b>
total.clusters	<b>0.466746</b>	Texture_SumAverage_3_135	<b>0.258977</b>
average.clusters.32	<b>0.466746</b>	Texture_SumAverage_3_45	<b>0.258976</b>
Texture_Contrast_3_135	<b>0.456215</b>	Texture_Gabor_3	<b>0.239860</b>
Texture_Contrast_3_45	<b>0.447556</b>	Texture_Variance_3_135	<b>0.229298</b>
Texture_Contrast_3_0	<b>0.439193</b>	Texture_Variance_3_90	<b>0.229285</b>
Count_Nuclei	<b>0.428861</b>	Texture_Variance_3_0	<b>0.229232</b>
		Texture_Variance_3_45	<b>0.229193</b>
		red.percentage	<b>0.223055</b>
		red.count	<b>0.223055</b>



# Imágenes:

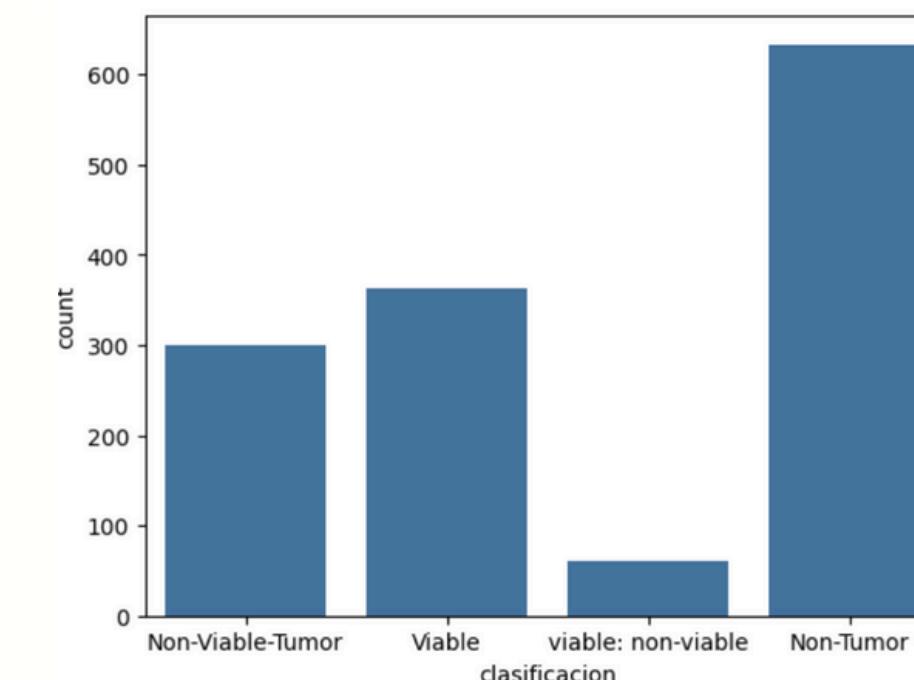
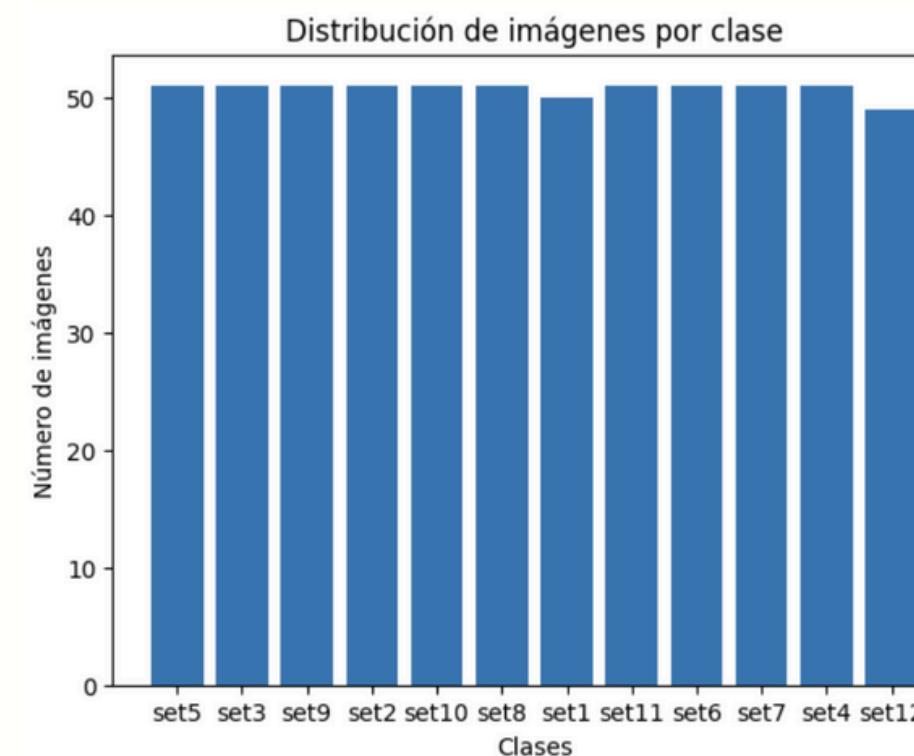
Las imágenes estaban representadas en carpetas para su organización, mas no tenían relación con la clasificación del estudio, dentro de cada carpeta existe un archivo de información que clasifica cada imagen.

## Training-Set-1

Categoría: set5, Número de imágenes: 51  
Categoría: set3, Número de imágenes: 51  
Categoría: set9, Número de imágenes: 51  
Categoría: set2, Número de imágenes: 51  
Categoría: set10, Número de imágenes: 59  
Categoría: set8, Número de imágenes: 51  
Categoría: set1, Número de imágenes: 27  
Categoría: set11, Número de imágenes: 49  
Categoría: set6, Número de imágenes: 51  
Categoría: set7, Número de imágenes: 51  
Categoría: set4, Número de imágenes: 51

## Training-Set-2

Categoría: set5, Número de imágenes: 51  
Categoría: set3, Número de imágenes: 51  
Categoría: set9, Número de imágenes: 51  
Categoría: set2, Número de imágenes: 51  
Categoría: set10, Número de imágenes: 51  
Categoría: set8, Número de imágenes: 51  
Categoría: set1, Número de imágenes: 50  
Categoría: set11, Número de imágenes: 51  
Categoría: set6, Número de imágenes: 51  
Categoría: set7, Número de imágenes: 51  
Categoría: set4, Número de imágenes: 51  
Categoría: set12, Número de imágenes: 49

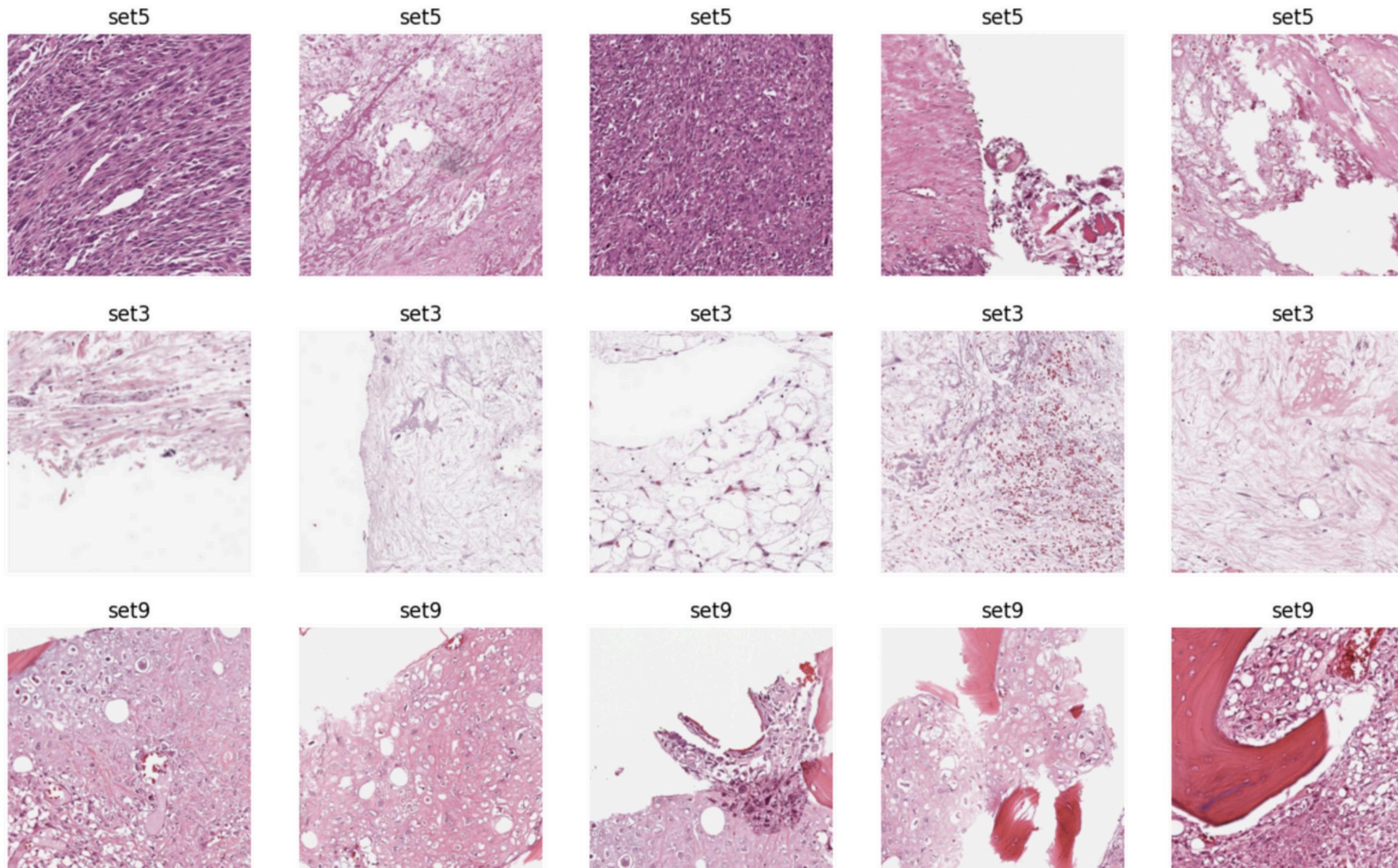


	ruta	clasificacion	carpeta
0	Case 4 C29-22504-33746.jpg	Non-Viable-Tumor	set5
1	Case 4 C34-17486-25030.jpg	Non-Viable-Tumor	set5
2	Case 4 C28-27602-26398.jpg	Non-Viable-Tumor	set5
3	Case 4 C29-16874-21863.jpg	Non-Viable-Tumor	set5
4	Case 4 C29-22107-10482.jpg	Viable	set5
...	...	...	...
1352	Case 4 C24-37864-31739.jpg	Non-Viable-Tumor	set4
1353	Case 4 C28-36370-18227.jpg	Viable	set4
1354	Case 4 C21-42043-16377.jpg	Viable	set4
1355	Case 4 C27-45200-21938.jpg	Non-Tumor	set4
1356	Case 4 C28-35755-28991.jpg	Non-Viable-Tumor	set4



# Imágenes:

Ejemplo de imágenes

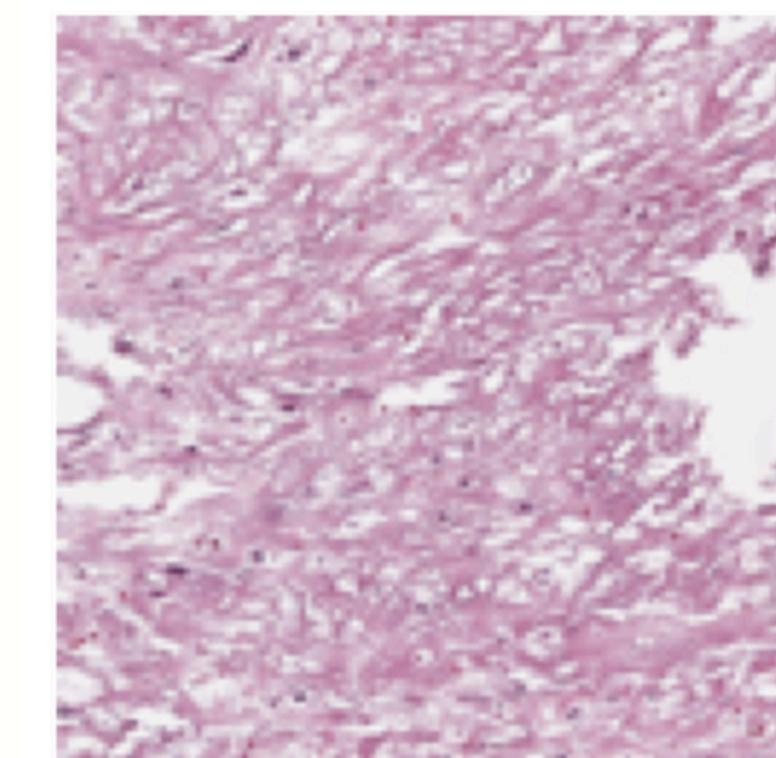
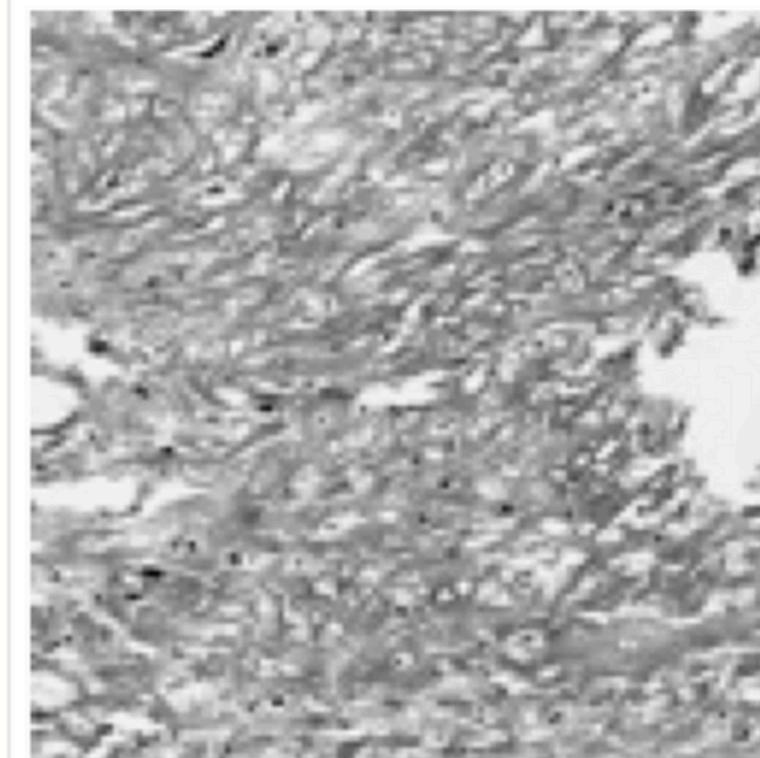
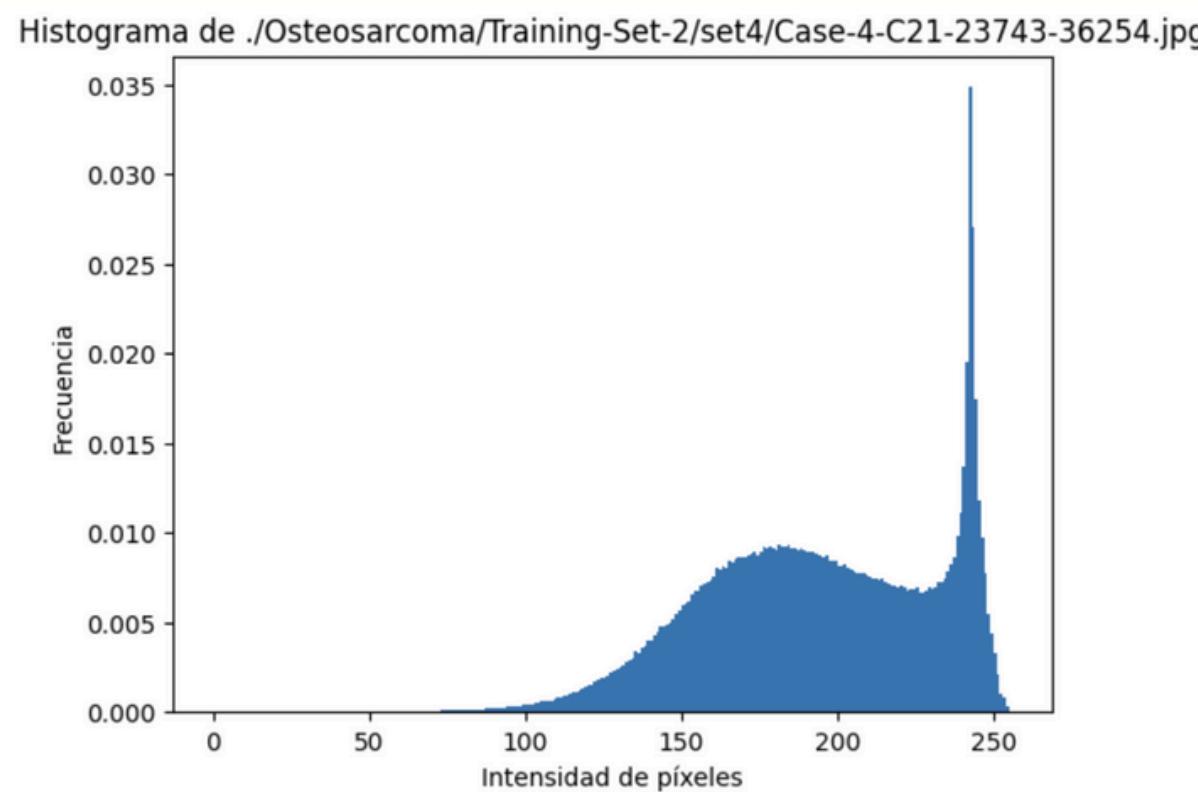




# Imágenes:

Características de las imágenes

Dimensionalidades de las imágenes:  $\{(1024, 1024, 3)\}$





# Conclusiones:

Mediante el análisis exploratorio realizado podemos concluir lo siguiente:

1. Los histogramas de las variables como Blue.count, red.count, Blue.percentage, y red.percentage indican que la mayoría de las muestras tienen concentraciones relativamente bajas de azul y rojo.
2. Los histogramas de las variables relacionadas con la textura (p. ej., Texture\_Variance\_3\_0, Texture\_Variance\_3\_135)
3. El boxplot muestra la dispersión de estas variables, destacando que hay muchos valores atípicos (outliers) en todas las características.
4. Diversidad de Clases: El dataset contiene imágenes clasificadas en cuatro categorías importantes para evaluar la viabilidad de tejidos tumorales
5. Características de Color y Textura: La extracción de estas características es relevante para distinguir entre tumores viables y no viables, así como tejido no tumoral, destacando la importancia de un análisis exhaustivo de las propiedades visuales de las imágenes.
6. Formato de las imágenes: Las imágenes cuentan con un formato de dimensiones y capas correctos para su análisis.
7. La clasificación de las imágenes permiten tener un mejor estudio y mayor organización de ellas, aunque estén separadas en conjuntos con un nombre no relacionado, la etiqueta de cada una permite clasificarlas.



# Gracias!!