



Modelos de predicción para diagnóstico de tumores cerebrales

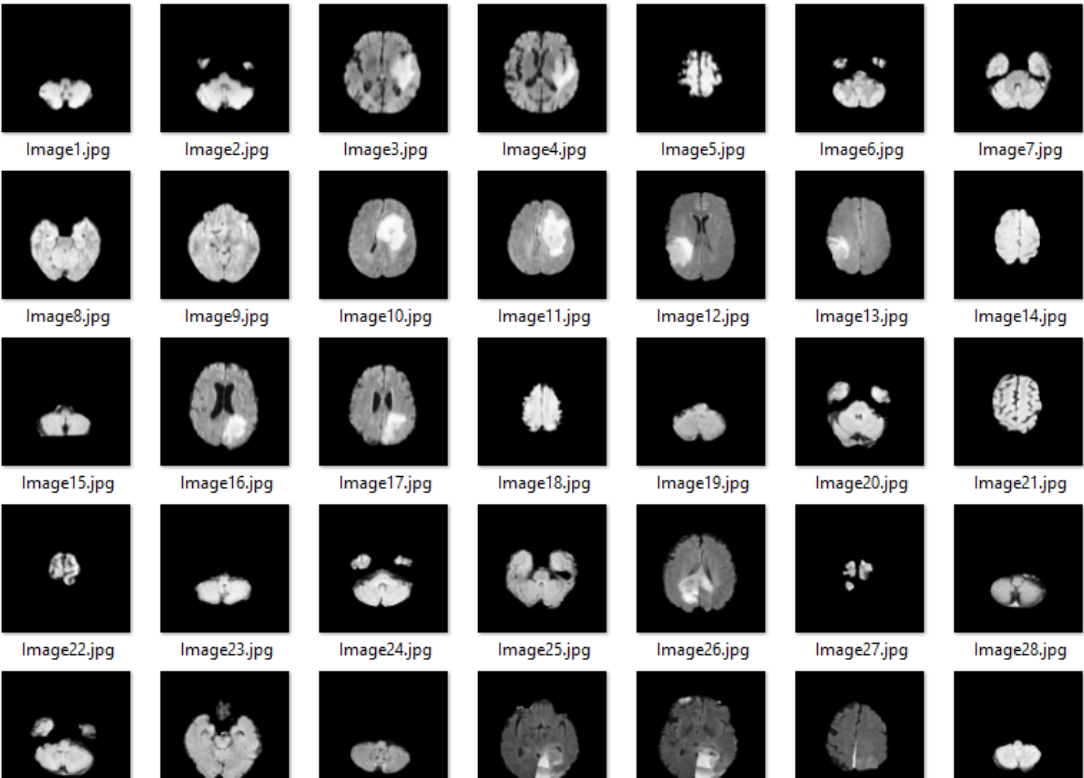
Arturo Moreno Puga

Bootcamp en Data Science & Machine Learning

ID Bootcamps

Curso 2022/2023

Deep Learning

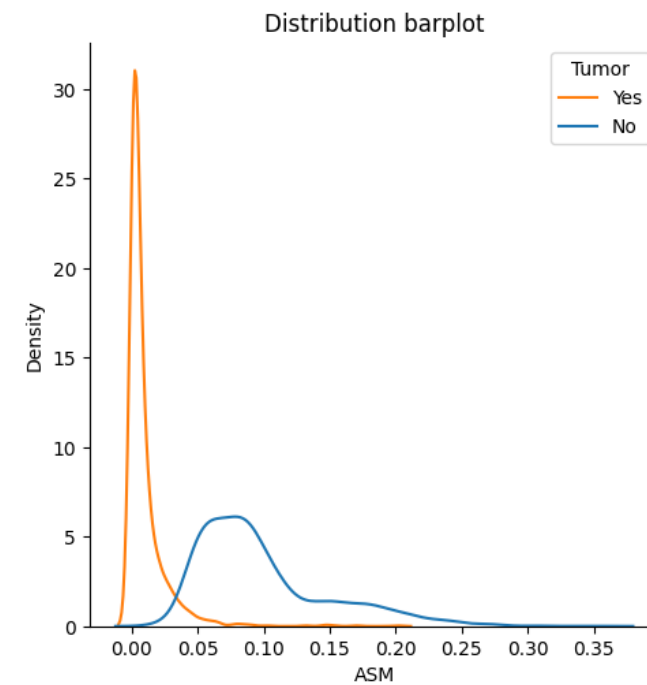
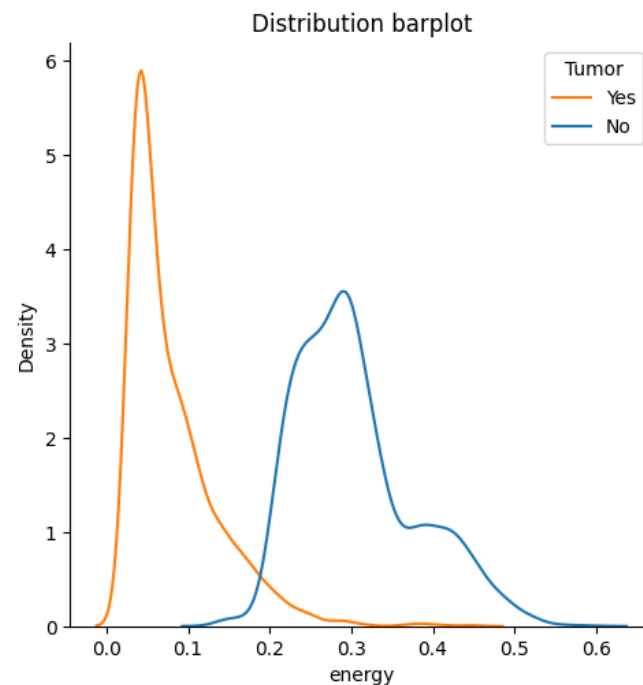
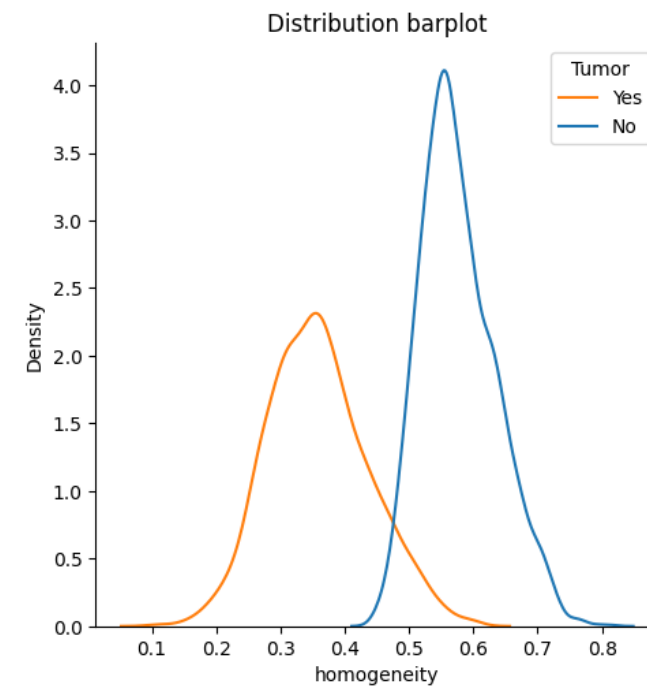
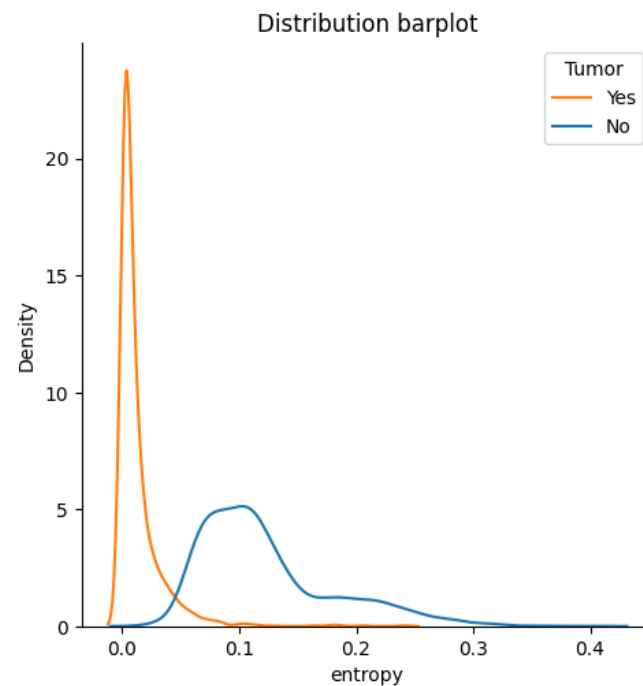


Machine Learning

	image	tumor	mean	variance	standard_deviation	entropy	skewness	kurtosis	contrast	energy	ASM	homogeneity	dissimilarity	correlation	coarseness
0	Image1	0	6.535339	619.587845	24.891522	0.109059	4.276477	18.900575	98.613971	0.293314	0.086033	0.530941	4.473346	0.981939	7.458341e-155
1	Image2	0	8.749969	805.957634	28.389393	0.266538	3.718116	14.464618	63.858816	0.475051	0.225674	0.651352	3.220072	0.988834	7.458341e-155
2	Image3	1	7.341095	1143.808219	33.820234	0.001467	5.061750	26.479563	81.867206	0.031917	0.001019	0.268275	5.981800	0.978014	7.458341e-155
3	Image4	1	5.958145	959.711985	30.979219	0.001477	5.677977	33.428845	151.229741	0.032024	0.001026	0.243851	7.700919	0.964189	7.458341e-155
4	Image5	0	7.315231	729.540579	27.010009	0.146761	4.283221	19.079108	174.988756	0.343849	0.118232	0.501140	6.834689	0.972789	7.458341e-155
...
3757	Image3758	0	21.234512	1208.850174	34.768523	0.063774	2.082079	4.647310	158.437600	0.220666	0.048693	0.487131	5.211739	0.950972	7.458341e-155
3758	Image3759	0	20.435349	1227.151440	35.030721	0.066763	2.144625	4.882034	161.158675	0.225931	0.051045	0.502712	5.083126	0.952749	7.458341e-155
3759	Image3760	0	18.011520	1151.582765	33.934978	0.068396	2.308349	5.579498	167.130118	0.228930	0.052409	0.492269	5.103700	0.952181	7.458341e-155
3760	Image3761	0	13.330429	945.732779	30.752769	0.087872	2.732822	7.757570	223.812932	0.261527	0.068397	0.480064	6.439784	0.940898	7.458341e-155
3761	Image3762	0	6.110138	480.884025	21.929068	0.118171	4.110669	17.538826	239.251388	0.306224	0.093773	0.494333	6.787329	0.938731	7.458341e-155
3762 rows x 15 columns															

Features más relevantes:

- ASM
- Energy
- Entropy
- Homogeneity



Optimizar recall

Mayor recall



Menor cantidad de
falsos negativos (FN)



Menor cantidad de pacientes
que hemos diagnosticado sin
tumor cuando realmente sí lo
tienen

		Predicción	
		No hay tumor	Sí hay tumor
Realidad	No hay tumor	TN	FP
	Sí hay tumor	FN	TP

$$recall = \frac{TP}{TP + FN}$$

Machine Learning

KNN (k-Nearest Neighbors)

- `n_neighbors = 1`
- `weights = uniform`
- `leaf_size = 10`

Recall	Precision	Accuracy	F1-Score
0.9760	0.9939	0.9867	0.9849

Decision Tree

- `criterion = gini`
- `max_depth = 150`

Recall	Precision	Accuracy	F1-Score
0.9671	0.9788	0.9761	0.9729

Deep Learning

Modelo **CNN** (red neuronal convolucional)

- Nº de épocas = 100
- Regularizador L1 → reduce overfitting
- Optimizador Adam
- Función de pérdida binary_crossentropy

→ Conv2D(filters = 16)
Conv2D(filters = 16)

→ Conv2D(filters = 32)
Conv2D(filters = 32)

→ Conv2D(filters = 64)
Conv2D(filters = 64)

→ Capas de agrupamiento
(MaxPooling2D)

→ Capas de aplanamiento
(Flatten)

→ Capas completamente
conectadas (Dense)

- 256 neuronas, ReLU
- 128 neuronas, ReLU
- 1 neurona, sigmoid

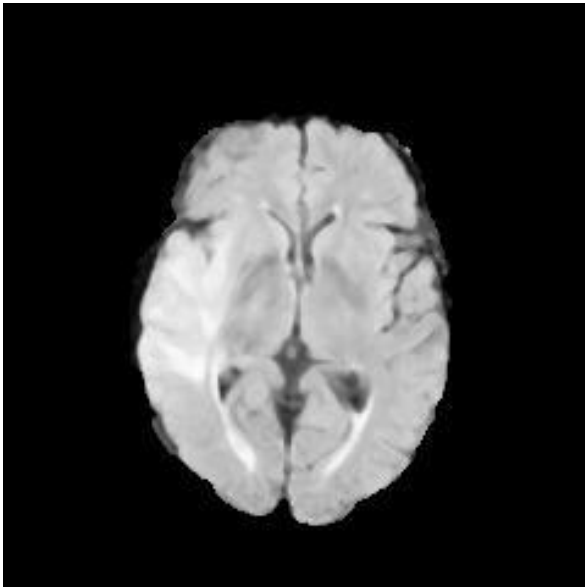
- **Épocas:** Cantidad de veces que se pasa por todo el conjunto de entrenamiento.
- **Filters:** Cantidad de filtros en cada capa. Cada filtro es una matriz de pesos que extrae características.

$$n = \underbrace{\left(\text{pesos} \right)}_{\text{filtro}} n-1$$

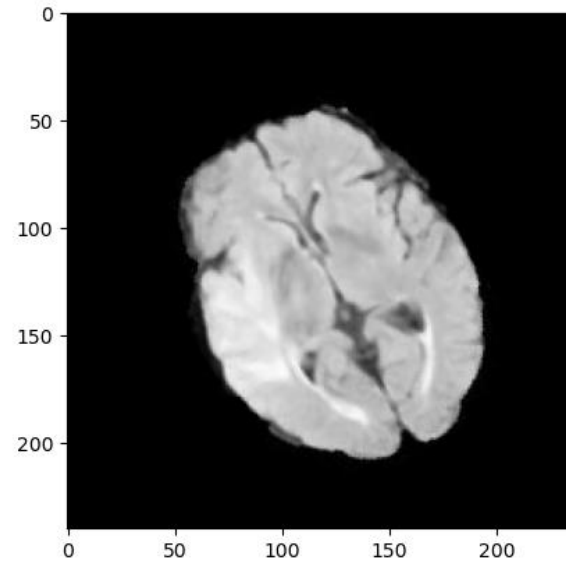
- En cada época los pesos se ajustan iterativamente.

Deep Learning

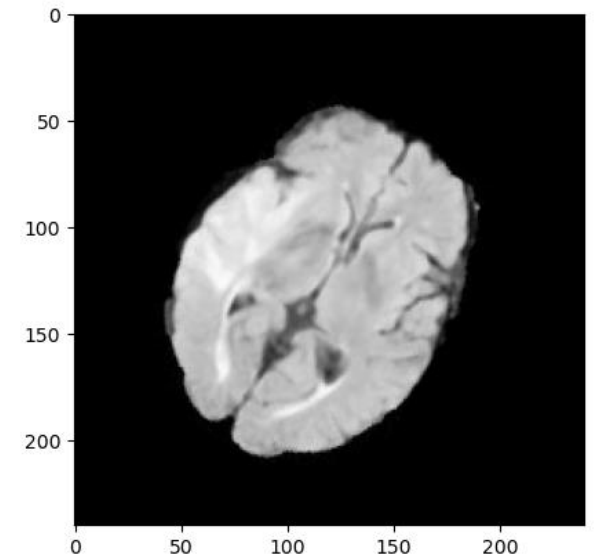
Rotaciones de 30° y -30° \longrightarrow Aumenta conjunto de entrenamiento



0°



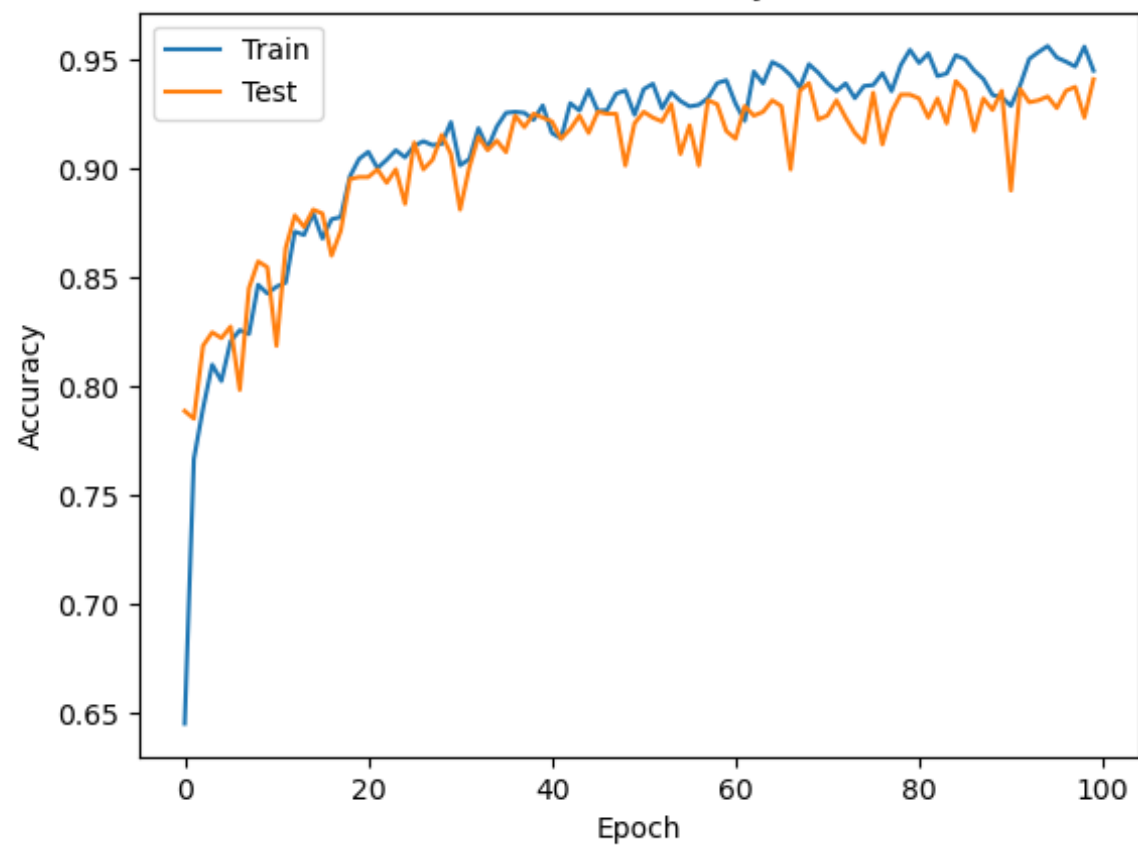
30°



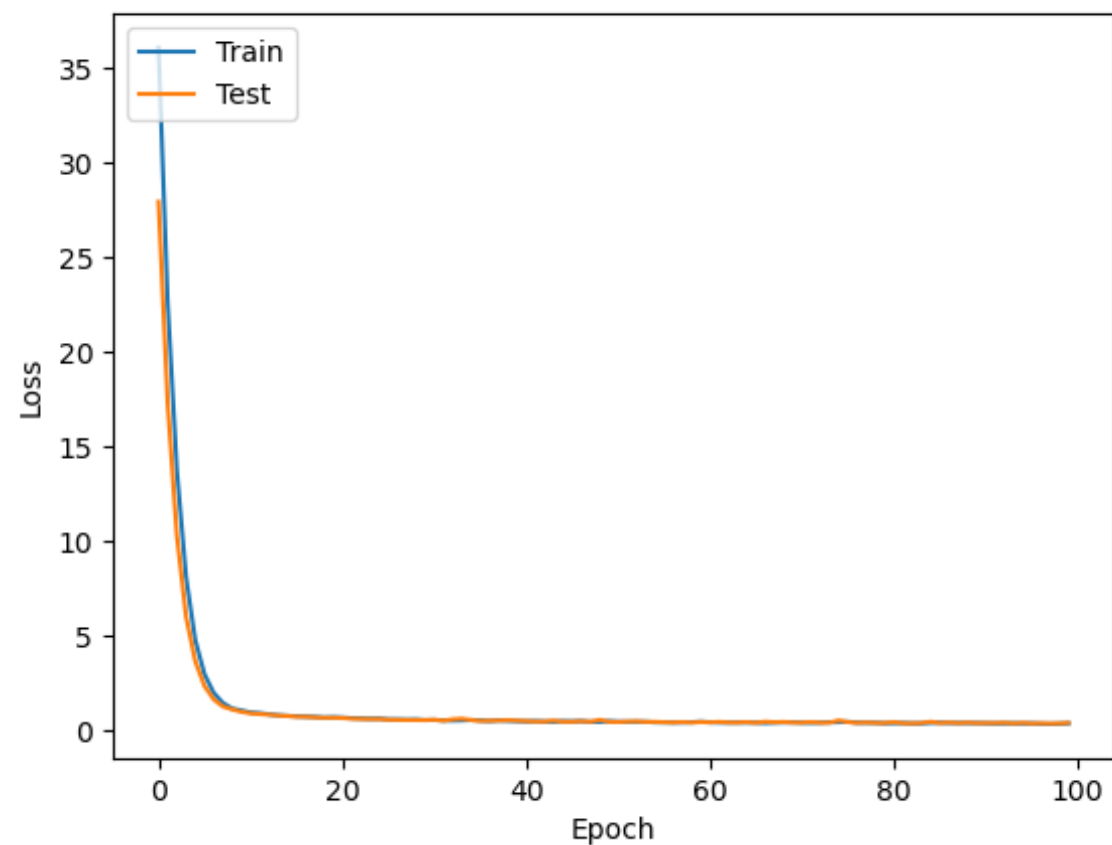
-30°

Deep Learning

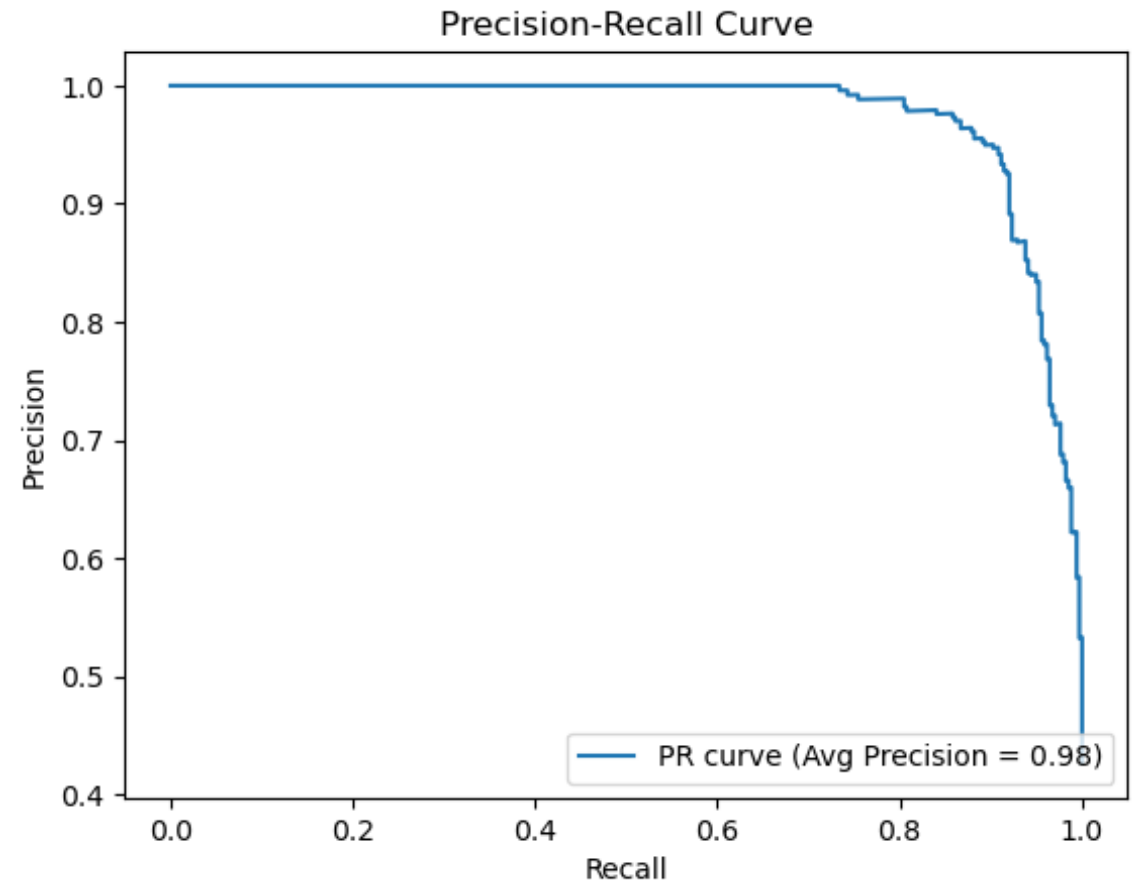
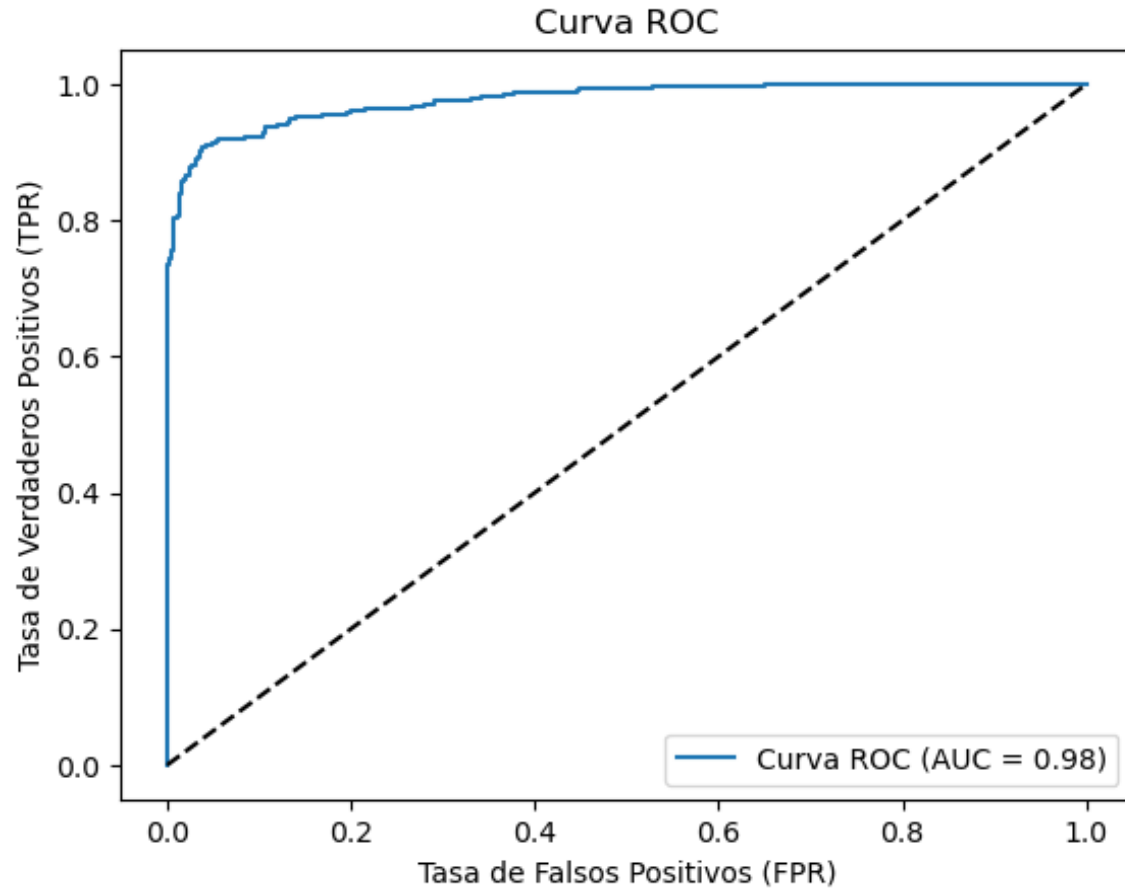
Model accuracy



Model loss



Deep Learning



Resultados finales

	Recall	Precision	Accuracy	F1-Score
KNN	0.9760	0.9939	0.9867	0.9849
CNN	0.9201	0.9120	0.9281	0.9160

- Métricas para KNN mayores que para CNN.
- KNN requiere extracción de features previa, CNN la realiza internamente.
- Para ambos modelos, métricas mayores a 0.9.

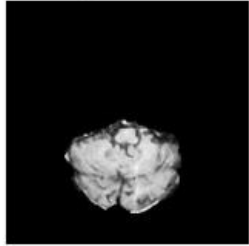
Conclusiones

- Aunque KNN presenta mejores métricas, CNN tiene una aplicación más sencilla.
- Ambos modelos muestran un buen rendimiento.
- Se requeriría un estudio posterior sobre un dataset desbalanceado para observar el comportamiento en un caso más real.

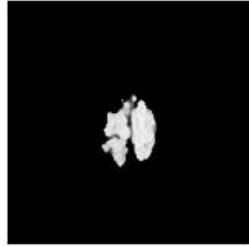


Predictions Examples

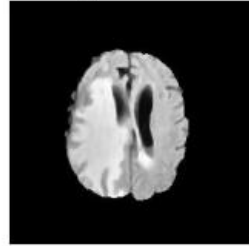
Tumor: No
Tumor prediction: No



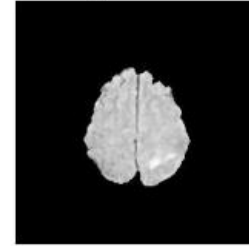
Tumor: No
Tumor prediction: No



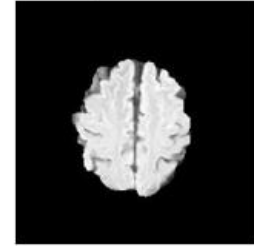
Tumor: Yes
Tumor prediction: Yes



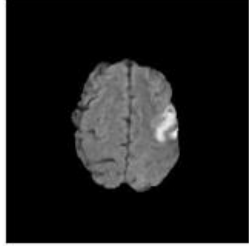
Tumor: Yes
Tumor prediction: No



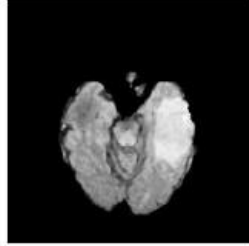
Tumor: No
Tumor prediction: No



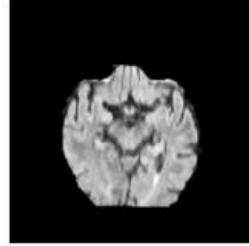
Tumor: Yes
Tumor prediction: Yes



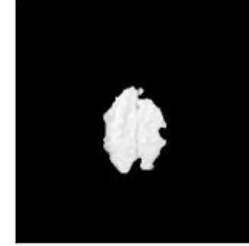
Tumor: Yes
Tumor prediction: Yes



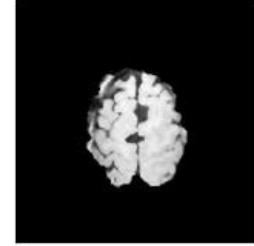
Tumor: No
Tumor prediction: Yes



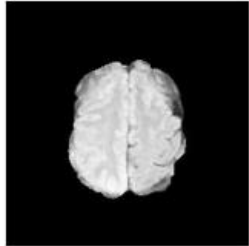
Tumor: No
Tumor prediction: No



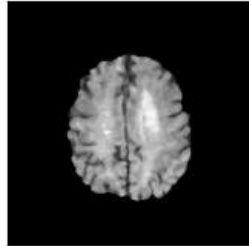
Tumor: No
Tumor prediction: No



Tumor: Yes
Tumor prediction: No



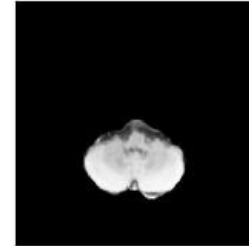
Tumor: Yes
Tumor prediction: Yes



Tumor: No
Tumor prediction: No



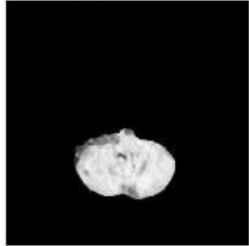
Tumor: No
Tumor prediction: No



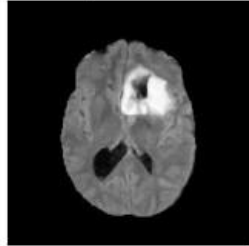
Tumor: No
Tumor prediction: No



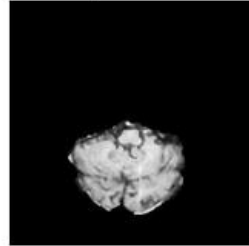
Tumor: No
Tumor prediction: No



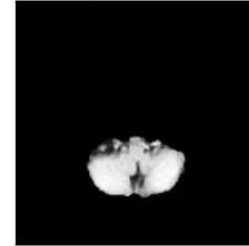
Tumor: Yes
Tumor prediction: Yes



Tumor: No
Tumor prediction: No



Tumor: No
Tumor prediction: No



Tumor: Yes
Tumor prediction: Yes

