

Data augmentation para imágenes médicas



Melissa Robles
Johan Rodríguez
Daniel Aguirre

Contexto

GENERATIVE ARTIFICIAL INTELLIGENCE FOR MEDICAL IMAGING

Alexandre Bône

Guerbet Research, Villepinte, France

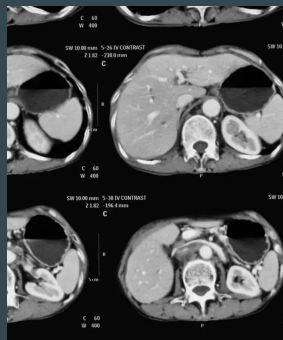
Tipos de imágenes médicas



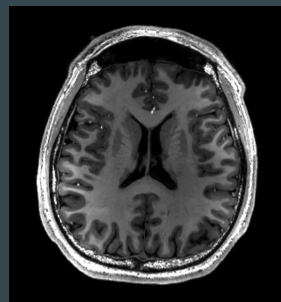
Rayos X



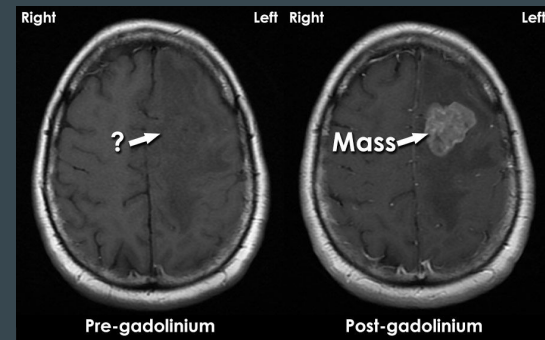
Ultrasonido



Tomografía



Resonancia
magnética



Medio de
contraste

Contexto

Rayos X

Radiación electromagnética descubierta en 1895 por el físico alemán Wilhelm Conrad Röntgen.

Las estructuras que son densas (como los huesos) bloquean la mayoría de las ondas de rayos x y aparecen de color blanco.



Procedimiento invasivo:
radiación



Procedimiento no
invasivo

Ultrasonido

Se compone de ondas mecánicas con frecuencia mayor a la que puede percibir un humano.

No tiene propiedades distintas al sonido más allá de que los humanos no podemos percibir las.

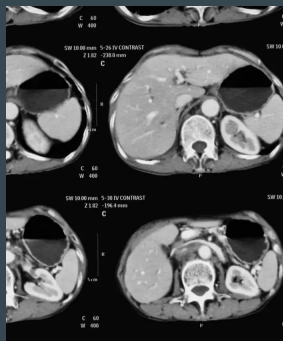


Contexto

Tomografía (TAC)

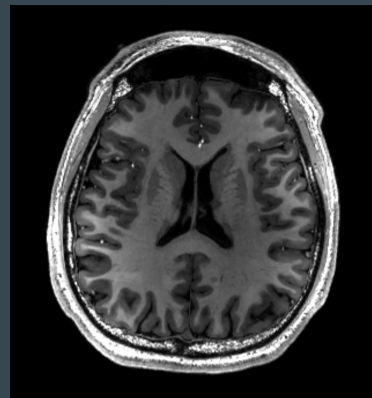
Se basa en el mismo principio que los rayos X, pero la máquina que emite los rayos X gira alrededor de una zona específica.

Genera imágenes 3D o por cortes 2D.



Procedimiento invasivo (radiación)

Procedimiento no invasivo pero altamente costoso.



Resonancia magnética (MRI)

Se crea un campo magnético y se aplica un impulso de radiofrecuencia.

Cuando se detiene el impulso, los núcleos de hidrógeno regresan a su alineación original con el campo magnético principal, liberando la energía.

Contexto

Medio de contraste - Costoso e invasivo

Sustancias que se utilizan para mejorar las imágenes diagnósticas: ayudan a resaltar órganos específicos y/o a resaltar zonas afectadas.

Tipos:

- Vía oral
- Vía rectal
- Intravenosos
- Inyectados en zonas específicas del cuerpo



Gadolinium
(inyectado)



Sulfato de bario
(vía oral)

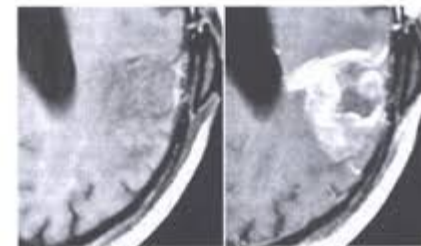
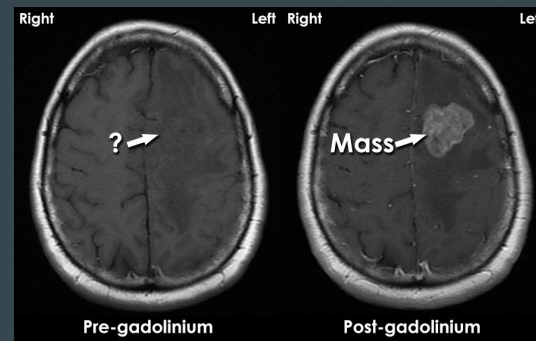


Fig. Lesión tumoral que capta contraste.



Efectos secundarios

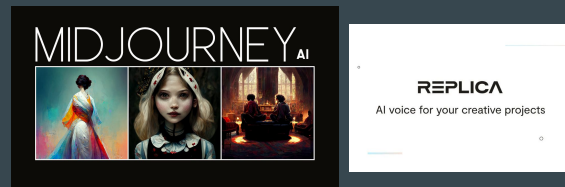
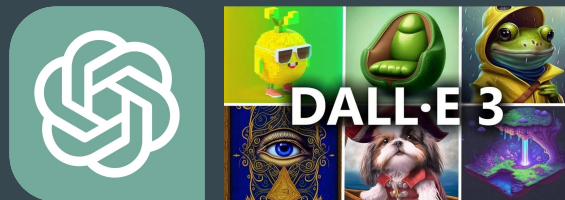
Leves - tos, vómito
Graves - anafilaxis, lesión renal,
fibrosis sistémica

Contexto

Usos de modelos generativos para imágenes médicas:

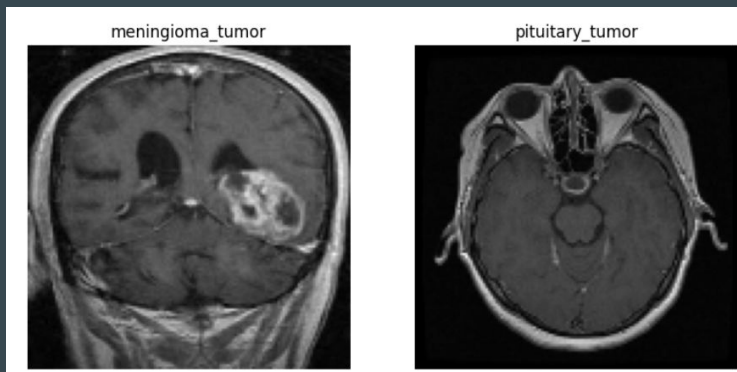
- Se utilizan tanto en investigación como en diagnóstico.

Inteligencia artificial generativa

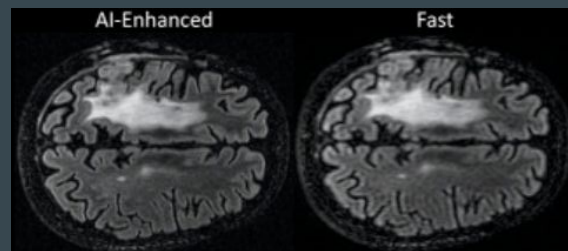


Diagnóstico

Clasificación



Reconstrucción de MRI



Rudie JD, Gleason T, Barkovich MJ, Wilson DM, Shankaranarayanan A, Zhang T, Wang L, Gong E, Zaharchuk G, Villanueva-Meyer JE. Clinical Assessment of Deep Learning-based Super-Resolution for 3D Volumetric Brain MRI. Radiol Artif Intell. 2022 Jan 12;4(2):e210059. doi: 10.1148/ryai.210059. PMID: 35391765; PMCID: PMC8980882.

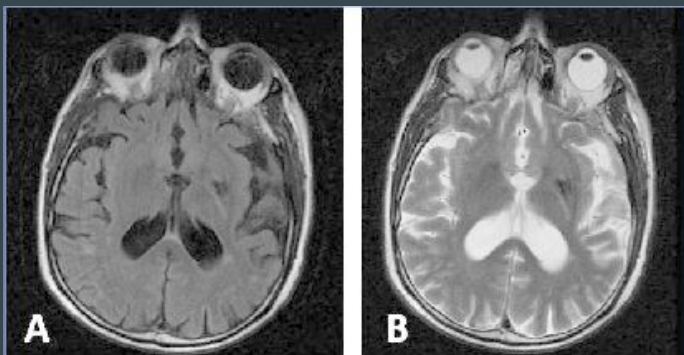
Contexto

Usos de modelos generativos para imágenes médicas:

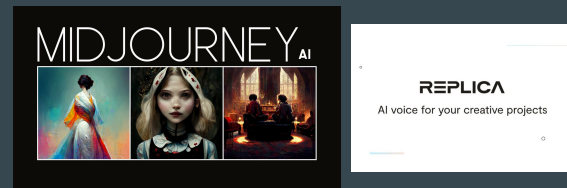
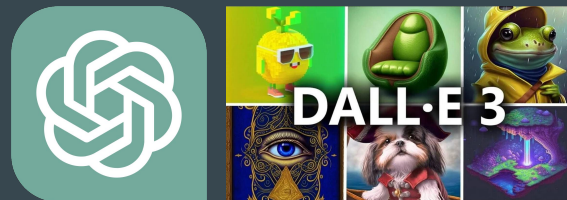
- Se utilizan tanto en investigación como en diagnóstico.

Diagnóstico

Mejorar calidad de las imágenes obtenidas - posibilidad de reducir los medios de contraste



Inteligencia artificial generativa



Ammari S, Bône A, Balleyguier C, Moulton E, Chouzenoux É, Volk A, Menu Y, Bidault F, Nicolas F, Robert P, Rohé MM, Lassau N. Can Deep Learning Replace Gadolinium in Neuro-Oncology?: A Reader Study. Invest Radiol. 2022 Feb 1;57(2):99-107. doi: 10.1097/RLI.0000000000000811. PMID: 34324463.

Contexto

Usos de modelos generativos para imágenes médicas:

- Se utilizan tanto en investigación como en diagnóstico.

Investigación

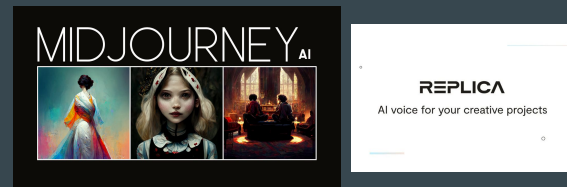
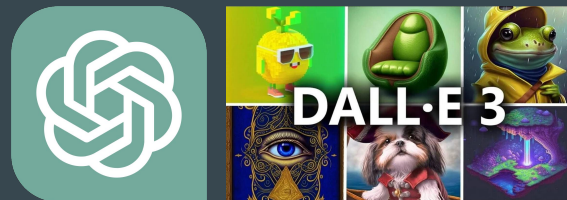
Problema:

- Muy pocas y muy costosas imágenes médicas.
- Problemas éticos (ej. datos de Single Cell de cerebro)

Solución:

- Crear imágenes “parecidas” a las imágenes reales.

Inteligencia artificial generativa

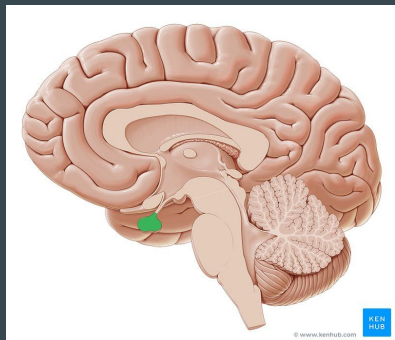


Motivación

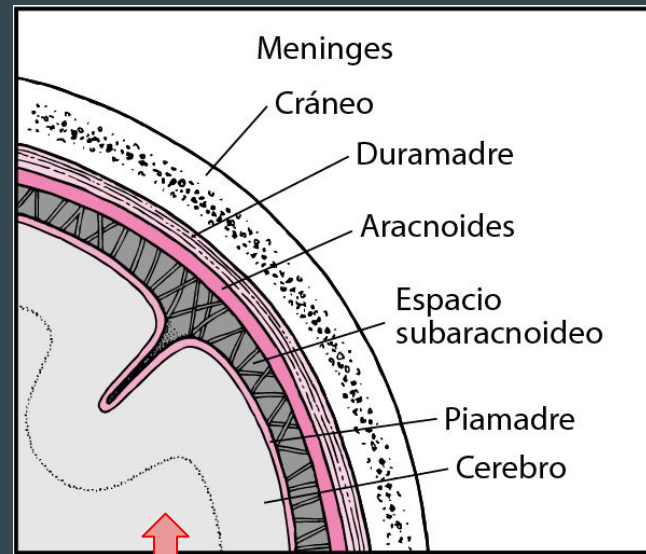
Un tumor cerebral es considerado una de las enfermedades más agresivas que un humano pueda tener a lo largo de su vida.

Hay muchos tipos de tumores cerebrales. Entre ellos están

- Meningioma: Surge de las meninges
- Glioma: Células gliales. Uno de los más comunes.
- Pituitario: Surge en la glándula pituitaria. Afecta las hormonas.



Glándula pituitaria



Células gliales

Datos

SARTAJ · UPDATED 3 YEARS AGO

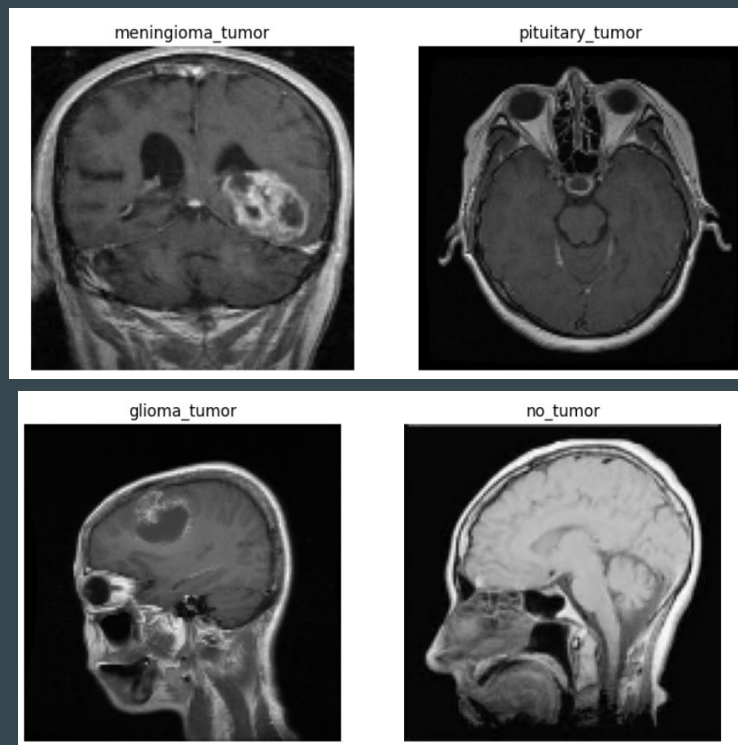
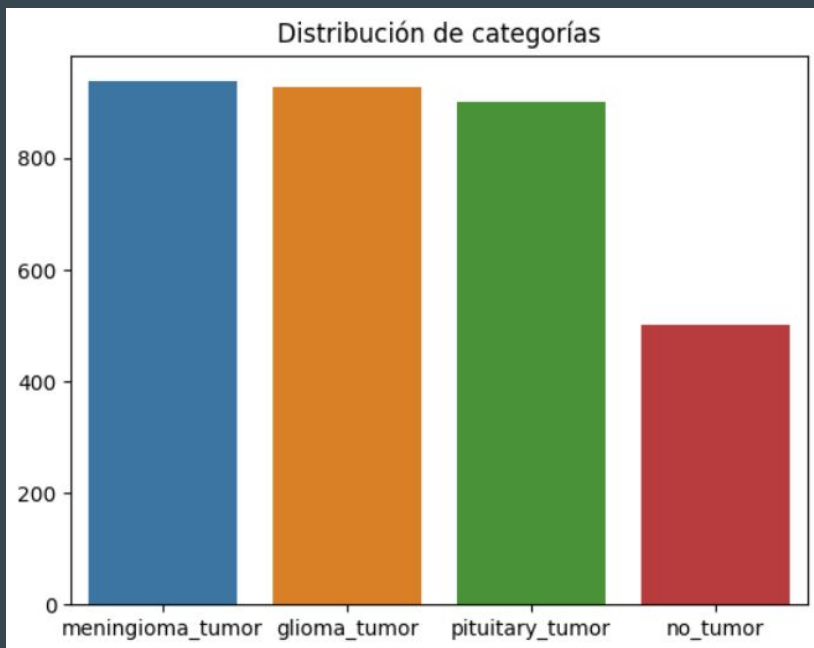

479

New Notebook

Download (91 MB)

Brain Tumor Classification (MRI)

Classify MRI images into four classes



Objetivo

Modelo baseline de clasificación: Transfer learning

split	accuracy_score	recall_score__micro	recall_score__macro	precision_score__micro	precision_score__macro	f1_score__micro	f1_score__macro
train	0.993531	0.993531	0.994334	0.993531	0.994046	0.993531	0.994170
test	0.963303	0.963303	0.968498	0.963303	0.965427	0.963303	0.966452

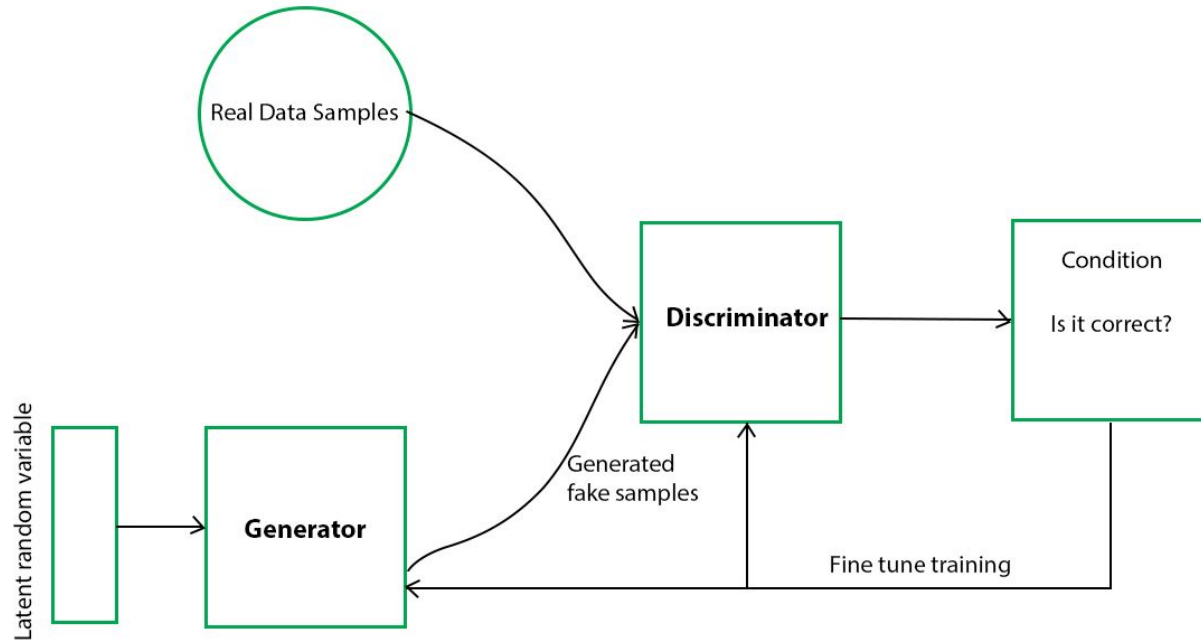
Probar tres modelos generativos distintos para mirar cómo puede impactar en la clasificación:

- Generative Adversarial Networks
- Conditional Variational Autoencoder
- Transfer learning

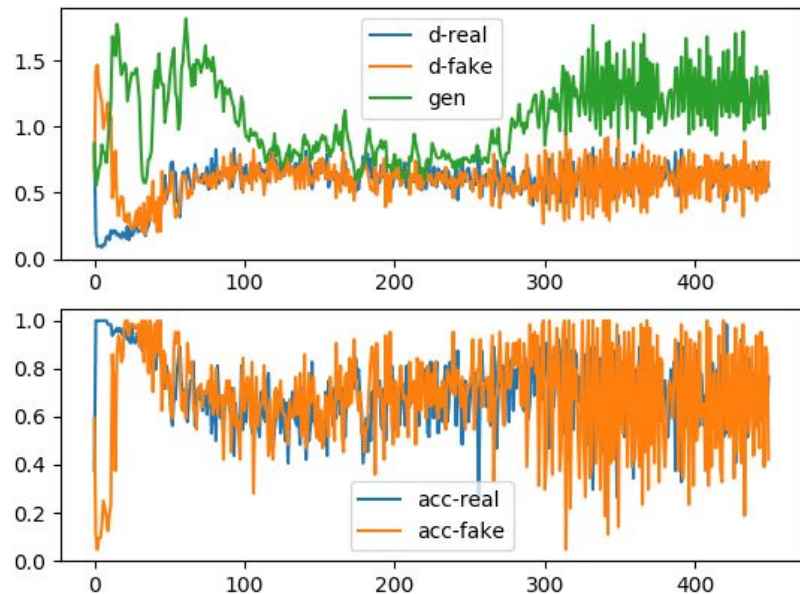
Se compara el modelo baseline con clasificadores entrenados con imágenes generadas.

Generative Adversarial Networks (GAN)

Arquitectura general de GAN

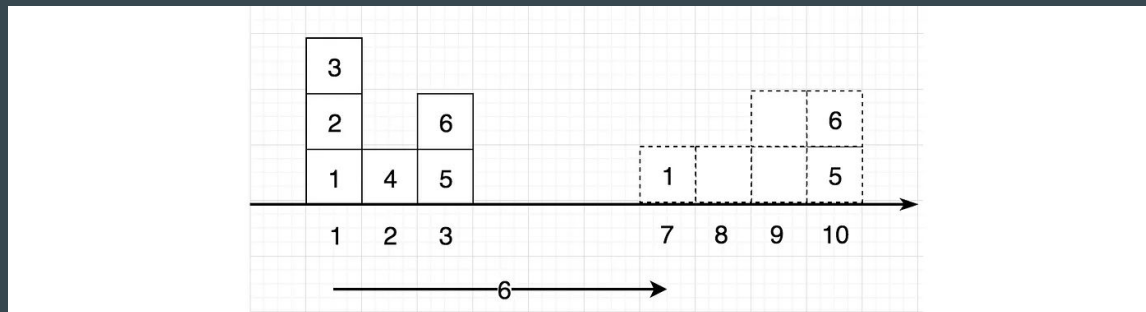


Problema de convergencia con las GAN



Wasserstein distance

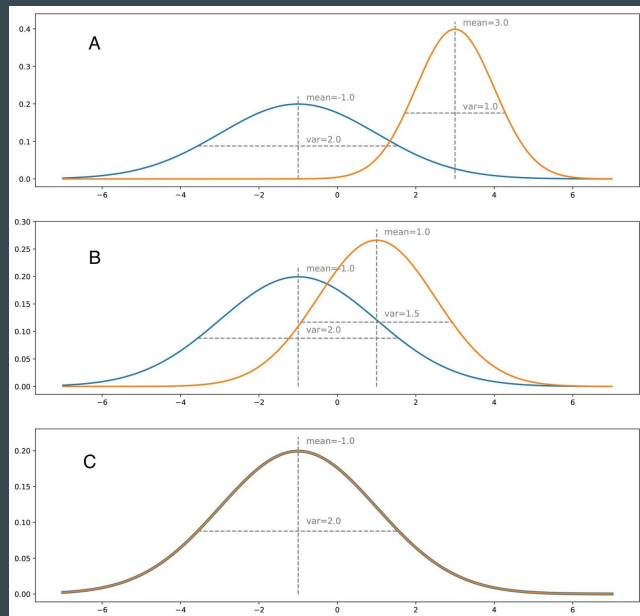
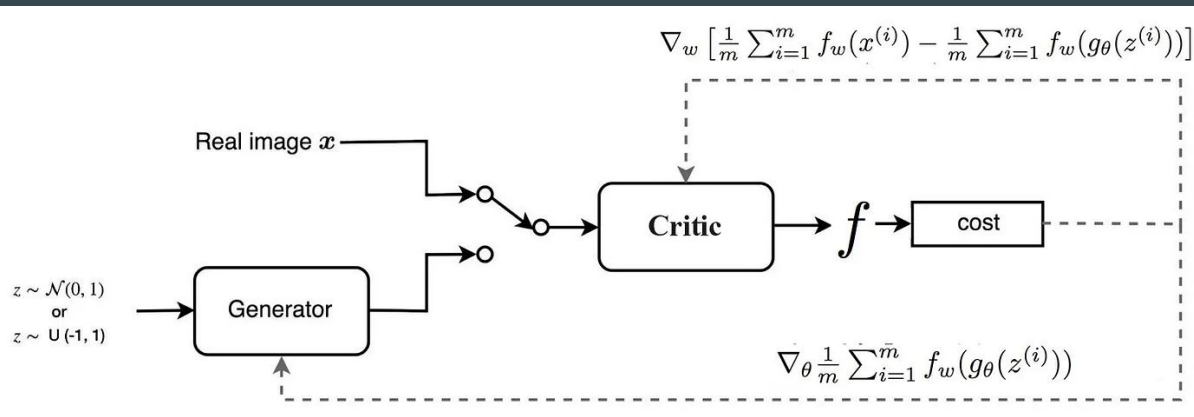
“Es el mínimo costo de mover las cajas de una configuración dada a otra”



El costo está expresado en términos del peso de la caja multiplicado la distancia para moverla.

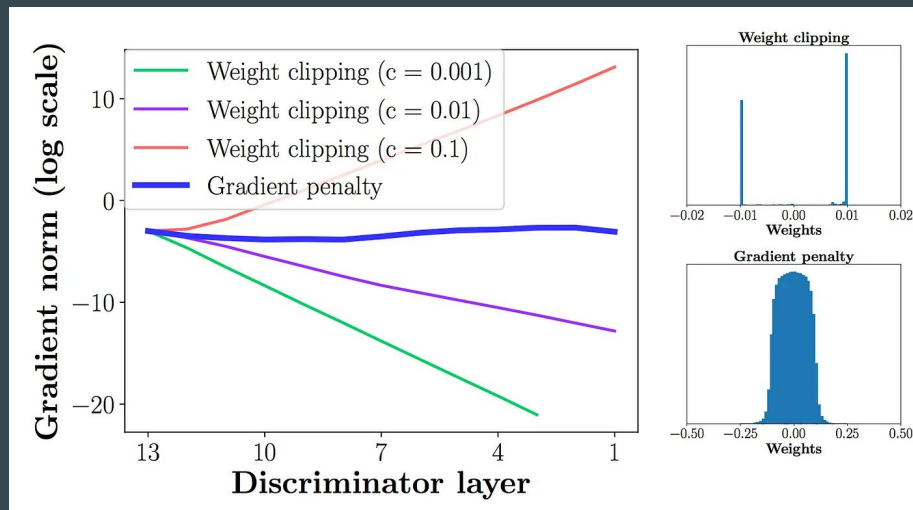
Wasserstein loss

- Asegura un gradiente más suave.
- Facilita la convergencia del modelo.

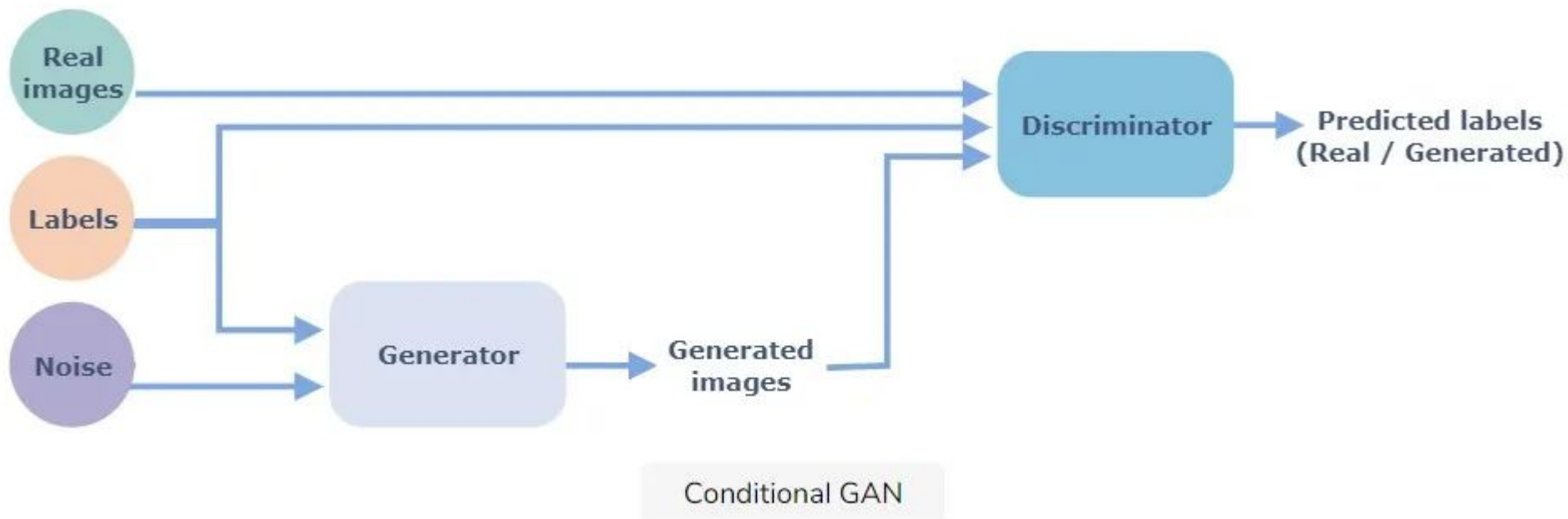


WGAN con Gradient Penalty

- En WGAN se hace clipping de los parámetros (restricción de 1-Lipschitz).
- WGAN puede no converger por desvanecimiento o explosión del gradiente.
- Gradient Penalty es asegurar que imágenes interpoladas entre reales y falsas generan un gradiente con norma 1 (restricción de 1-Lipschitz).

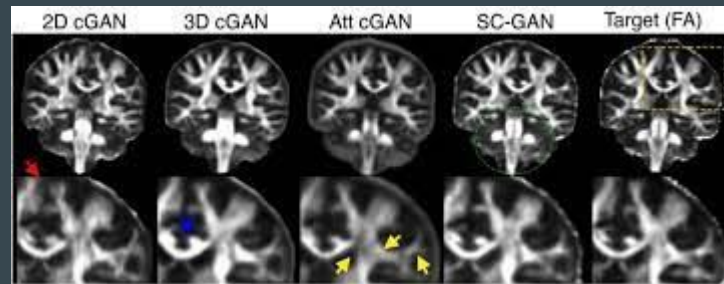
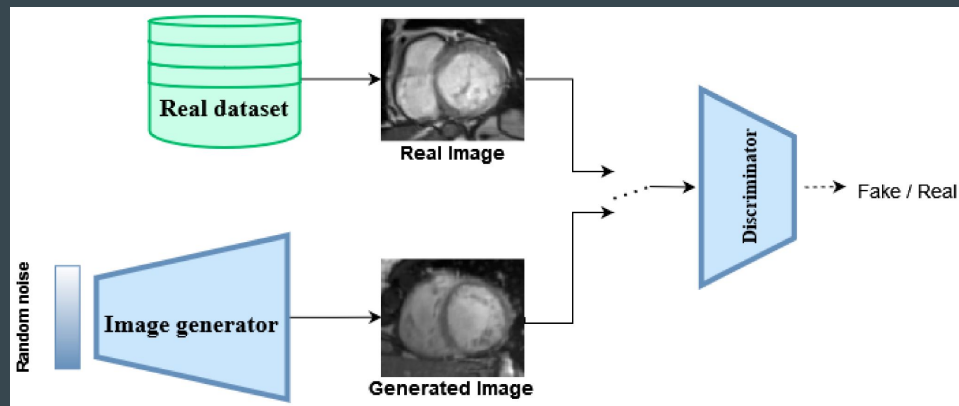


Arquitectura de Conditional GAN (cGAN)



Arquitectura general

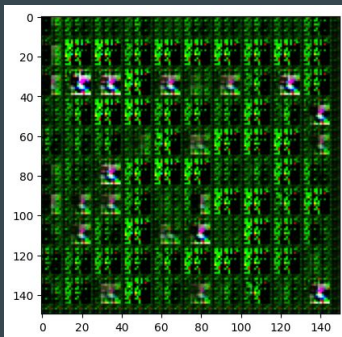
1. Exploración inicial de modelos GAN
2. Exploración de hiperparametros (WCGAN-GP)
3. Evaluación de métricas de clasificación y generación de imágenes



1. Exploración inicial de modelos GAN

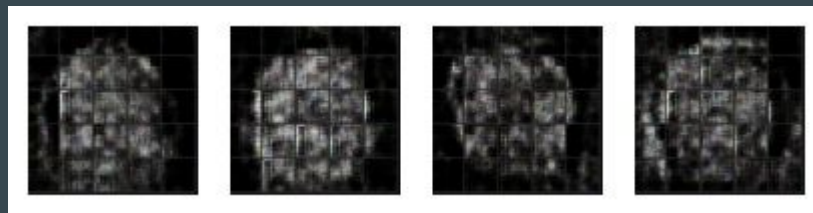
Conditional GAN (cGAN)

- Uso de probabilidad condicional en generador y discriminador.
- Fácil implementación de función de pérdida.
- Altamente inestable.



Wasserstein GAN (WGAN) with Gradient Penalty (GP) (WGAN-GP)

- Extensión de GAN con distancia de Wasserstein en la función de pérdida.
- Se penaliza el gradiente del discriminador para tener norma L2 cerca a 1, en vez de weight clipping.

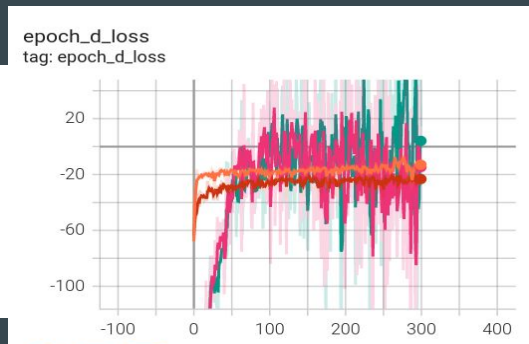


2. Exploración de hiperparámetros

Discriminador	
1	Conv2D + LeakyReLU + Dropout
2	Conv2D + Batch Normalization + LeakyReLU

Dimensión latente	
1	64
2	128
3	256

Generador	
1	Conv2DTranspose + BatchNormalization + LeakyReLU
2	Upsampling + Conv2D + BatchNormalization



2. Exploración de hiperparametros

Modelo

Glioma

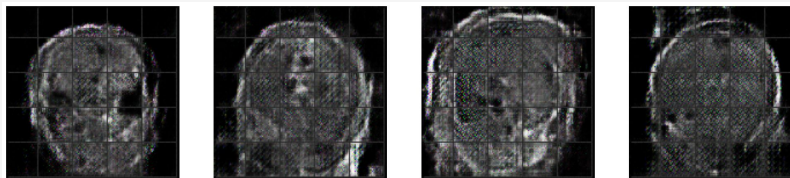
Normal

Meningioma

Pituitario

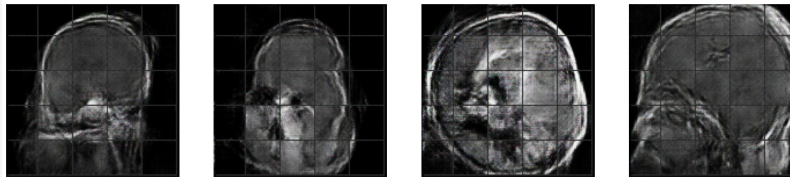
Loss - epoch
300

Generador1, discriminador1, 128 dim



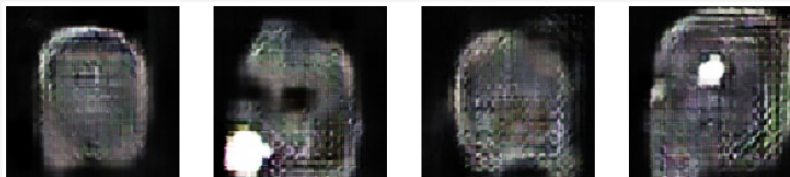
d_loss: -10.69
g_loss: -19.41

Generador1, discriminador2, 128 dim



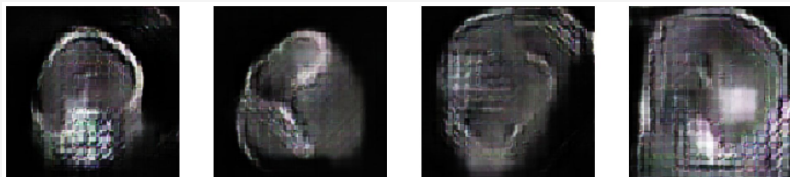
d_loss: -23.37
g_loss: 10.08

Generador2, discriminador1, 128 dim



d_loss: -23.9
g_loss: 1829

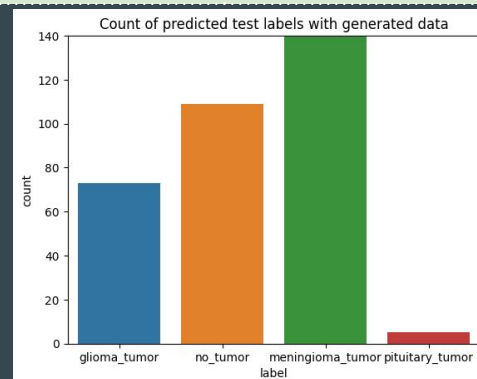
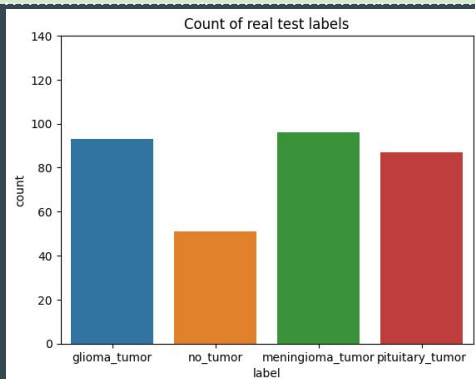
Generador2, discriminador2, 128 dim



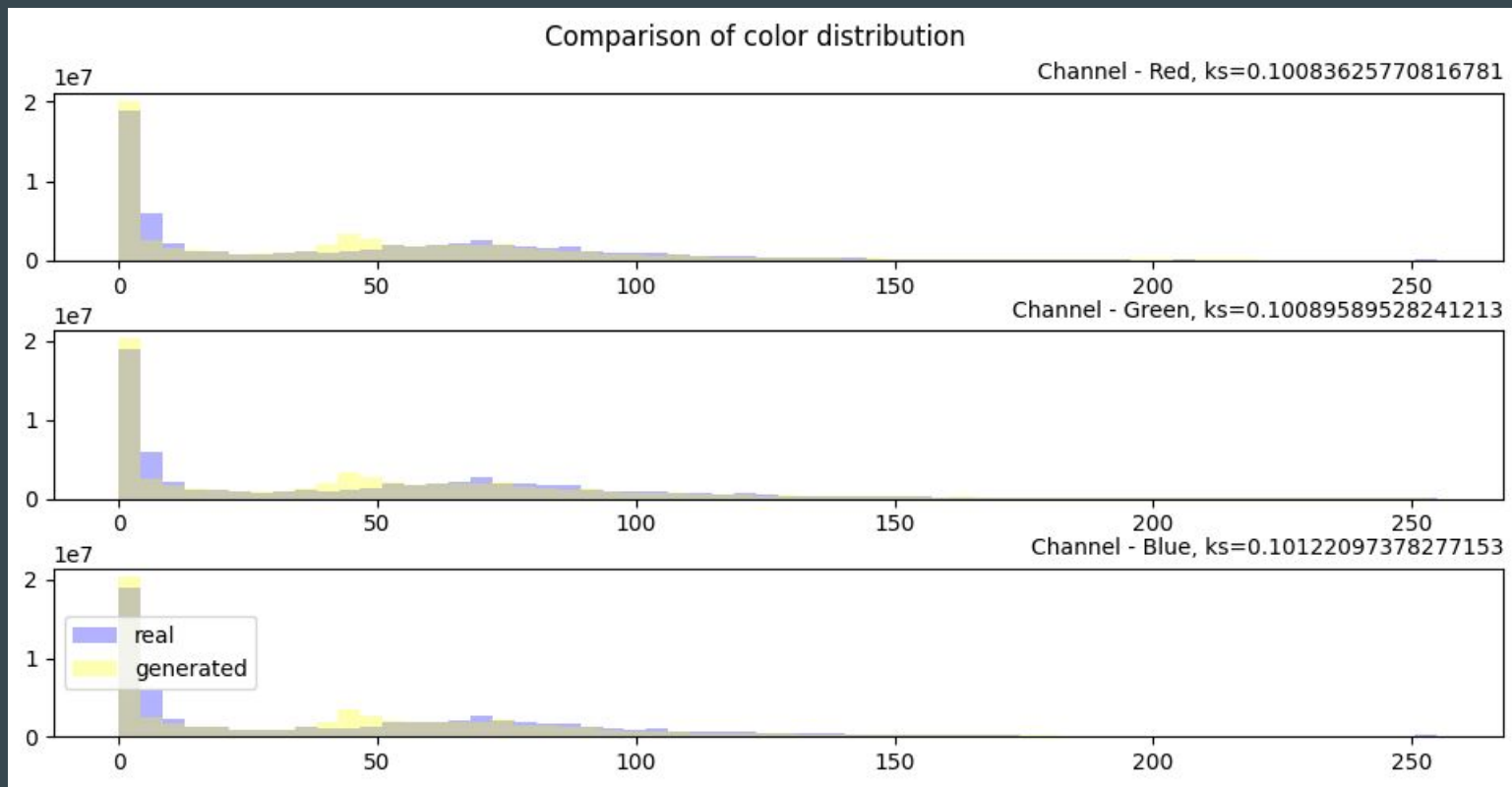
d_loss: 951
g_loss: -2400

3. Evaluación de métricas

data_type	split	accuracy_score	recall_score__micro	recall_score__macro	precision_score__micro	precision_score__macro	f1_score__micro	f1_score__macro
baseline	train	0.993531	0.993531	0.994334	0.993531	0.994046	0.993531	0.994170
	test	0.963303	0.963303	0.968498	0.963303	0.965427	0.963303	0.966452
generated	train	0.986721	0.986721	0.987813	0.986721	0.988114	0.986721	0.987826
	test	0.446483	0.446483	0.484022	0.446483	0.509142	0.446483	0.407214
mixed	train	0.979571	0.979571	0.979197	0.979571	0.981564	0.979571	0.98020
	test	0.935780	0.935780	0.940297	0.935780	0.942565	0.935780	0.93953



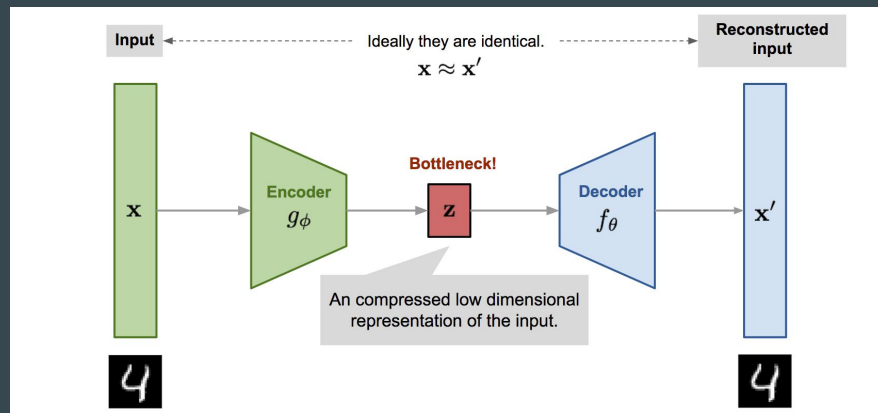
3. Evaluación de métricas (Kolmogorov - Smirnov)



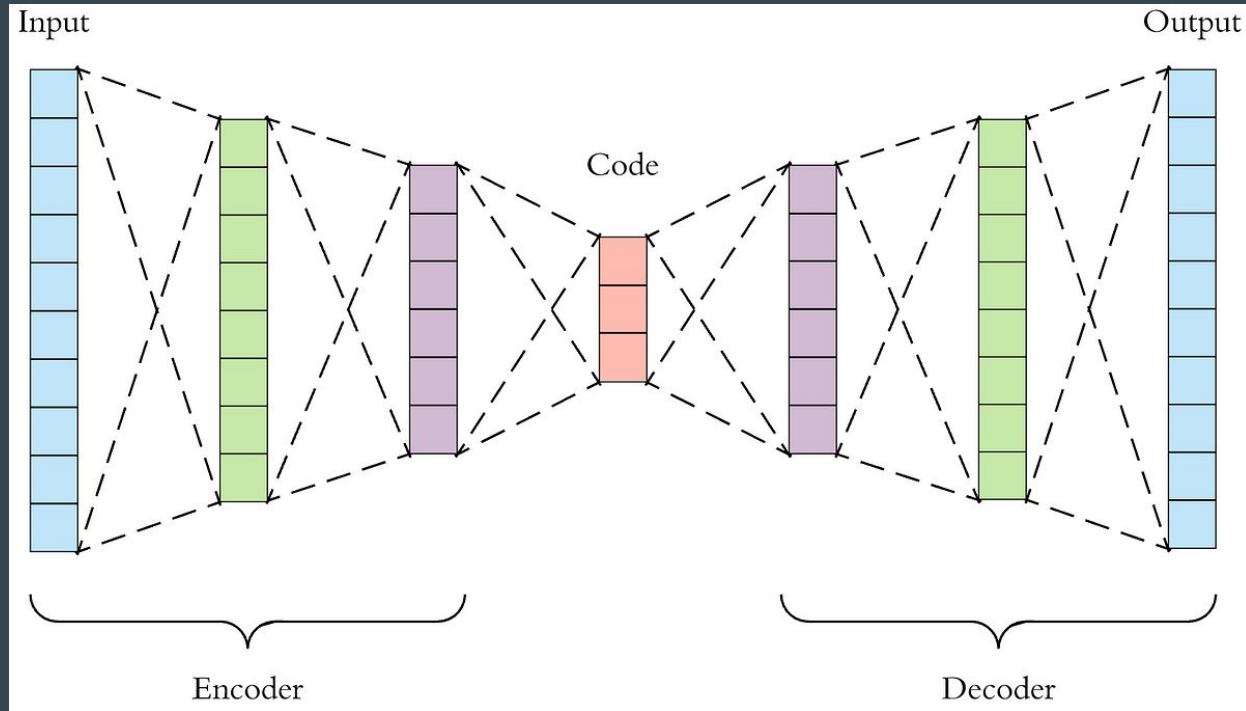
Conditional Variational Autoencoder

Procedimiento

1. Exploración inicial de modelos CVAE
2. Ajuste de código para imágenes
3. Exploración de hiperparámetros
4. Evaluación de métricas de clasificación y generación de imágenes
5. Pruebas con otra arquitectura agregando ruido

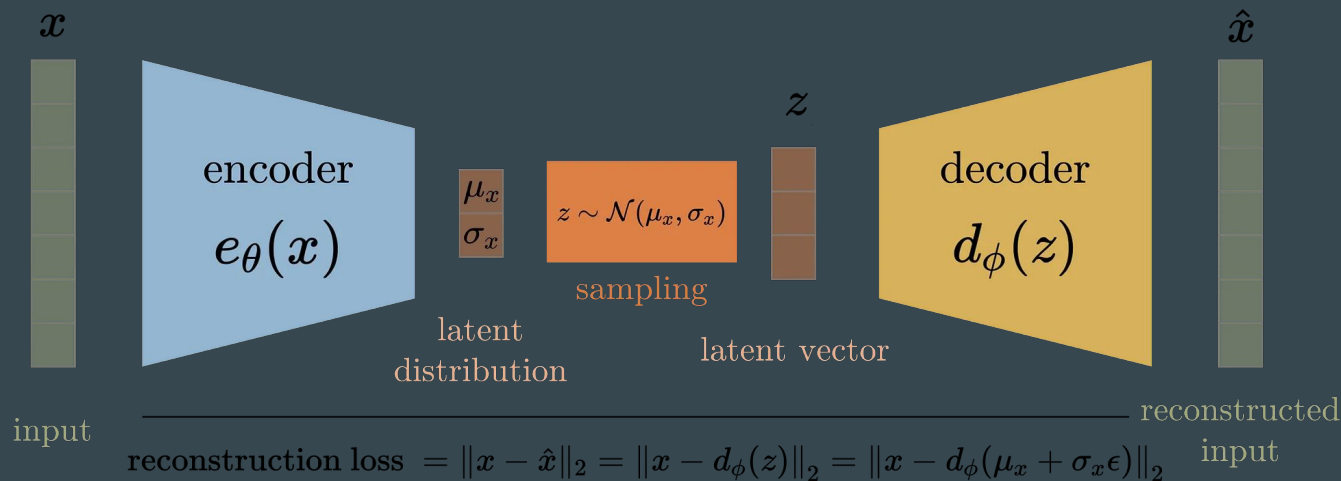


Autoencoder



<https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798>

Variational autoencoder

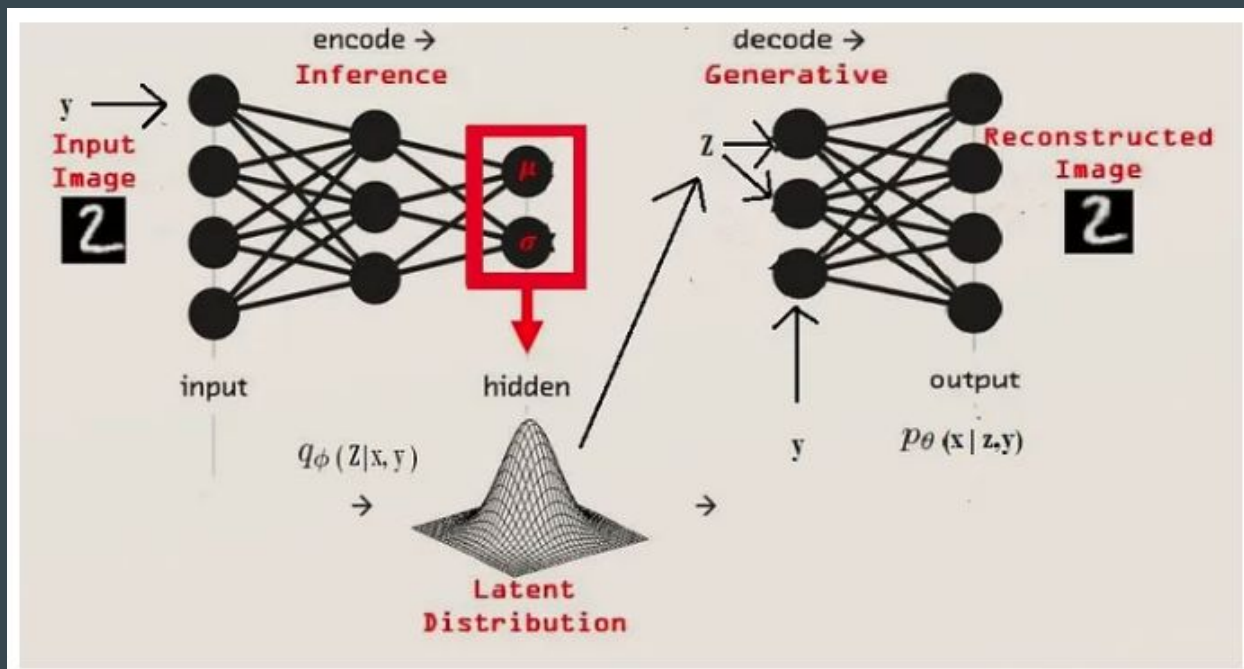


$$\mu_x, \sigma_x = e_{\theta}(x), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\text{similarity loss} = KL \text{ Divergence} = D_{KL}(\mathcal{N}(\mu_x, \sigma_x) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

$$\text{loss} = \text{reconstruction loss} + \text{similarity loss}$$

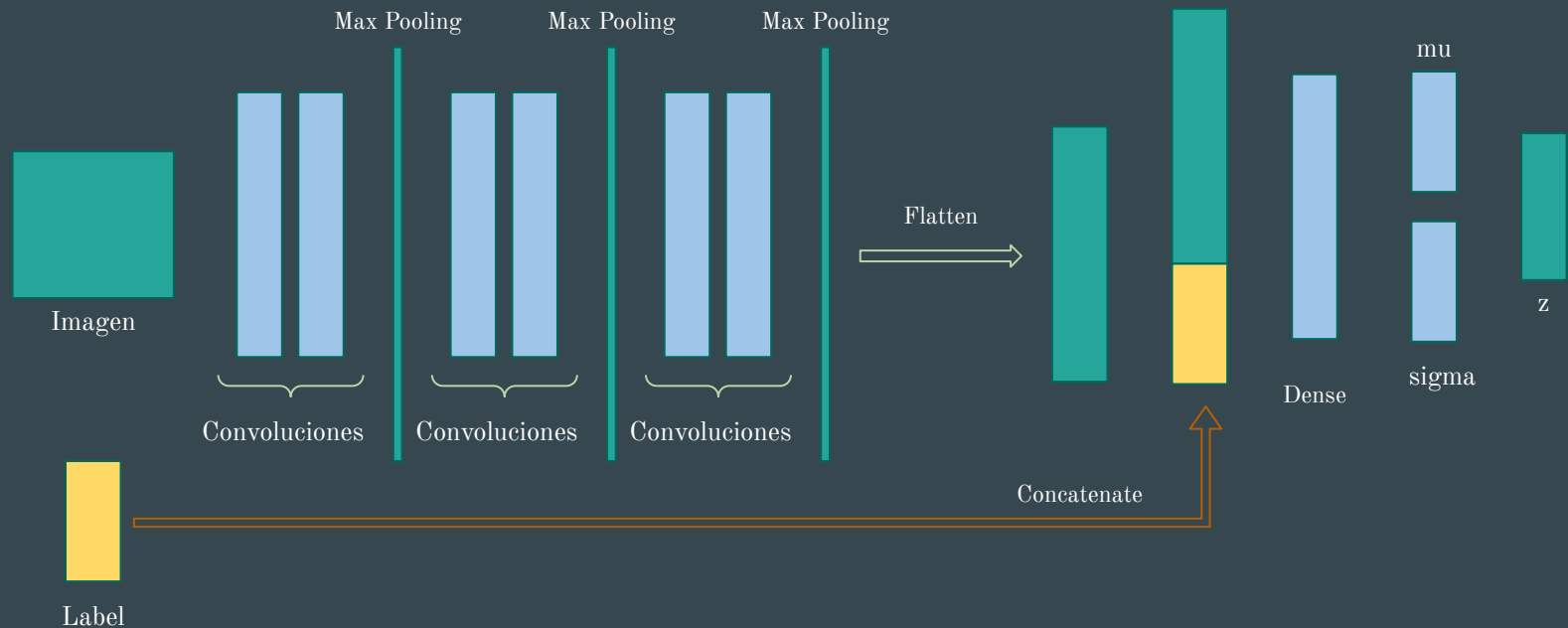
Conditional variational autoencoder



Arquitectura utilizada

Se modificó la arquitectura para que, en vez de hacer un Flatten en la imagen de entrada, se utilizaran capas convolucionales sobre la imagen.

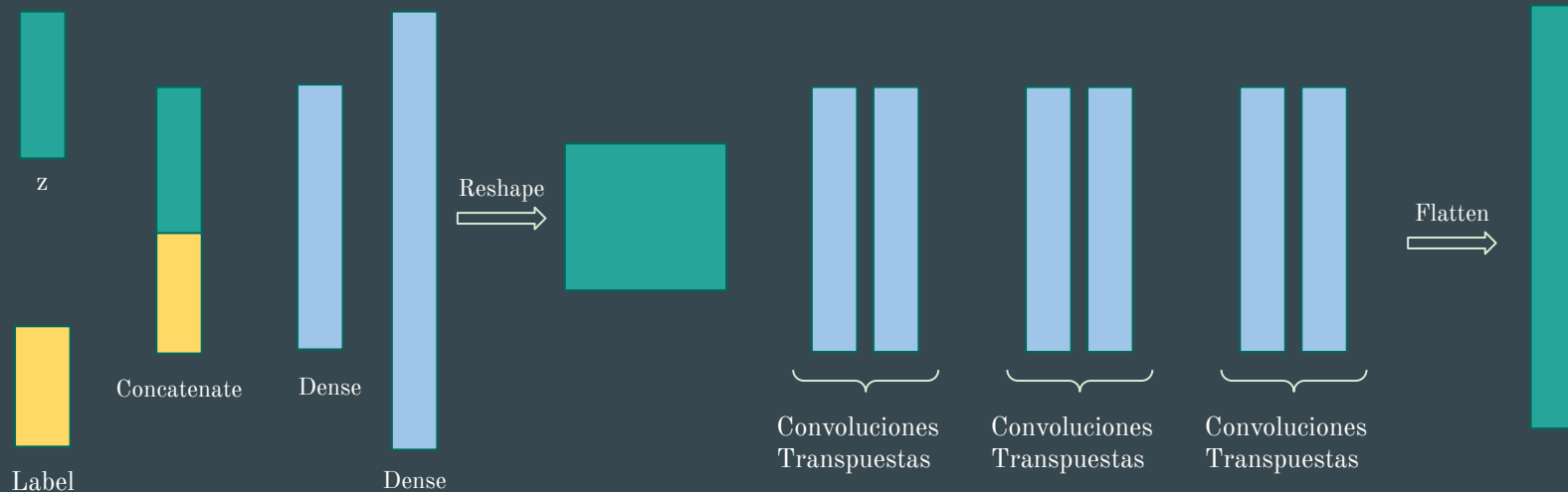
Encoder



Arquitectura utilizada

Se modificó la arquitectura para que, en vez de hacer un Flatten en la imagen de entrada, se utilizaran capas convolucionales sobre la imagen.

Decoder



Exploración de hiperparámetros

Encoder	
1	3 Conv2D + Dense
2	6 Conv2D + MaxPooling + Dense

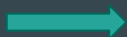
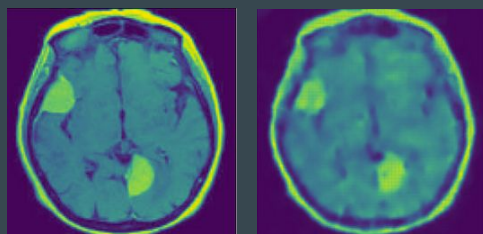
Decoder	
1	Dense + 3 Conv2DTranspose
2	Dense + 6 Conv2DTranspose

Dimensión latente	
2	256
3	512

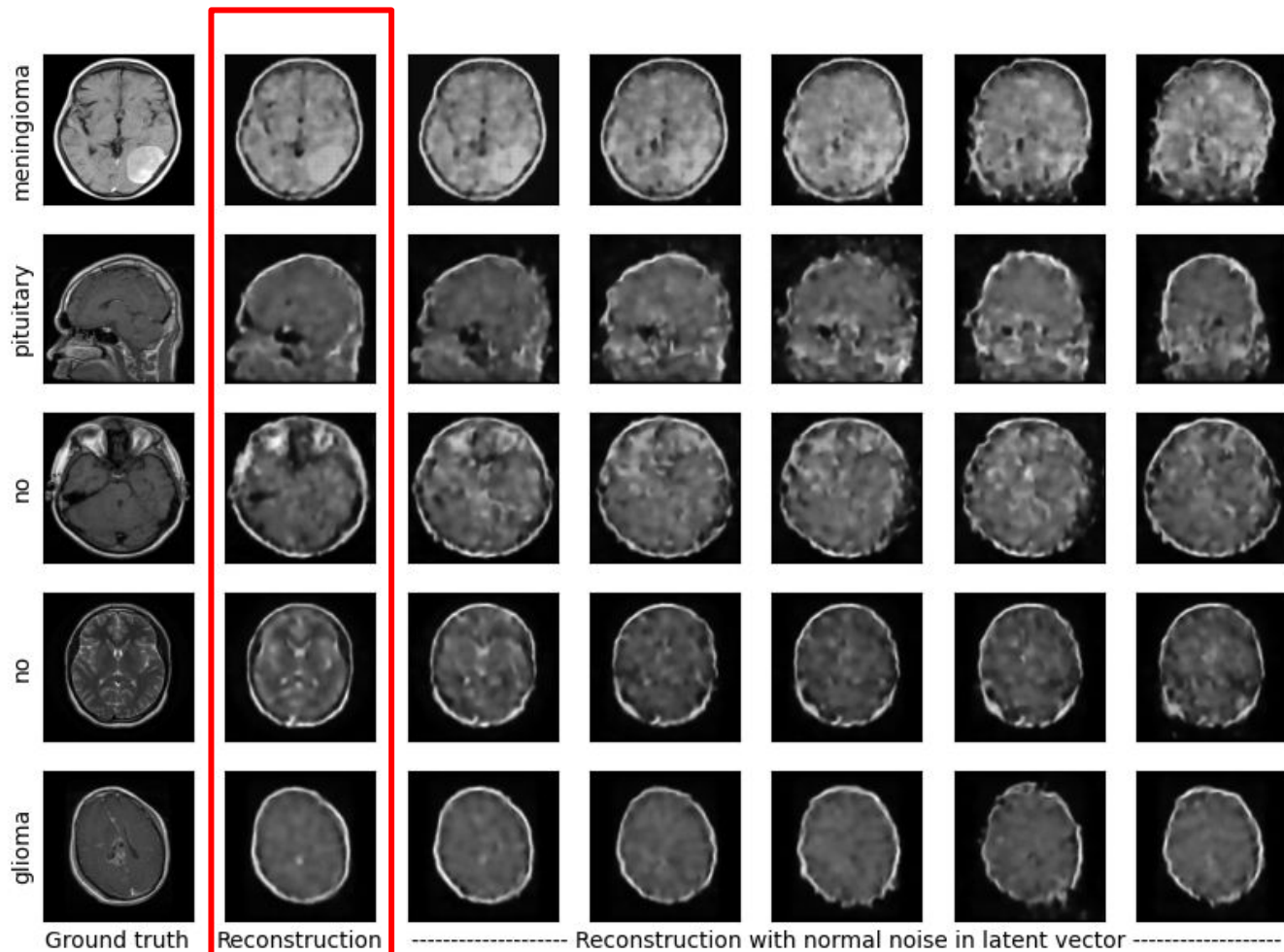
Resultados

(400 épocas)

	Test	Validation
KL Loss	59.63	57.63
Recon Loss	5990.6	6035.97



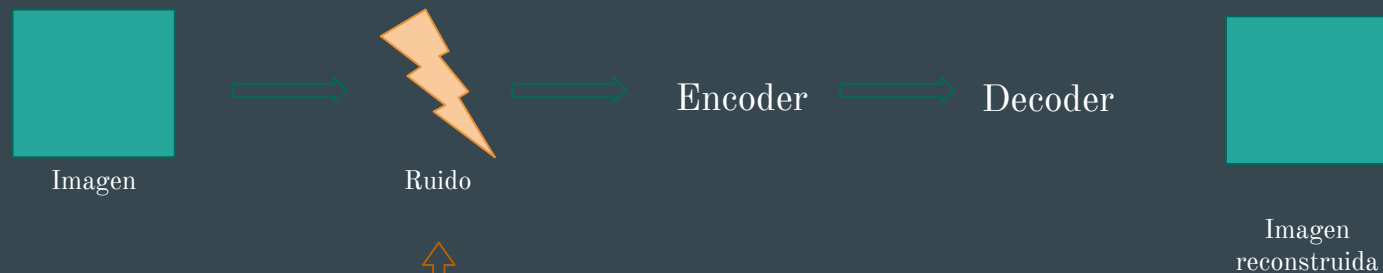
Reconstructions of conditional convolutional variational autoencoder



Resultados

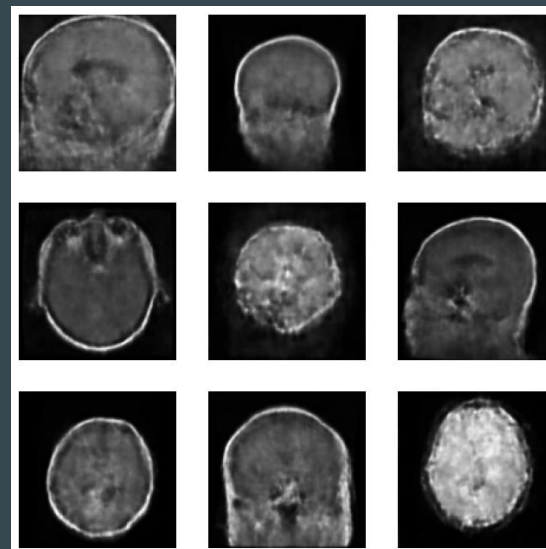
data_type	split	accuracy_score	recall_score__micro	recall_score__macro	precision_score__micro	precision_score__macro	f1_score__micro	f1_score__macro
baseline	train	0.993531	0.993531	0.994334	0.993531	0.994046	0.993531	0.994170
	test	0.963303	0.963303	0.968498	0.963303	0.965427	0.963303	0.966452
generated	train	0.979911	0.979911	0.981210	0.979911	0.980914	0.979911	0.981045
	test	0.522936	0.522936	0.5622936	0.522936	0.719677	0.522936	0.499308
mixed	train	0.989713	0.989713	0.990235	0.989713	0.990462	0.989713	0.990271
	test	0.949974	0.949974	0.950381	0.949974	0.951113	0.949974	0.950317

Cambio de arquitectura



Hiperparámetros:

- Tamaño de vector latente: 256, 512
- Número de capas convolucionales:
 - 13, 16, 19 bloques (inspirado en VGG16-19)
- Skip-connections.
- Residual blocks.



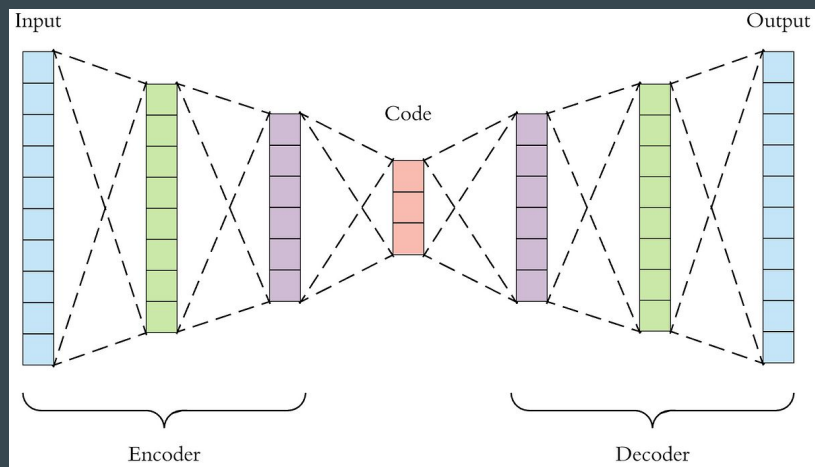
Cambio de arquitectura - resultado

data_type	split	accuracy_score	recall_score__micro	recall_score__macro	precision_score__micro	precision_score__macro	f1_score__micro	f1_score__macro
baseline	train	0.993531	0.993531	0.994334	0.993531	0.994046	0.993531	0.994170
	test	0.963303	0.963303	0.968498	0.963303	0.965427	0.963303	0.966452
generated	train	0.934968	0.934968	0.926880	0.934968	0.938648	0.934968	0.931982
	test	0.535168	0.535168	0.551517	0.535168	0.595428	0.535168	0.516590
mixed	train	0.930031	0.930031	0.931964	0.930031	0.929121	0.930031	0.930357
	test	0.920489	0.920489	0.924353	0.920489	0.933008	0.920489	0.926584

Transfer Learning

Metodología - Modelos pre-entrenados

Se utilizó un dataset de 7000 imágenes de MRI de cerebro para el entrenamiento de un variational autoencoder



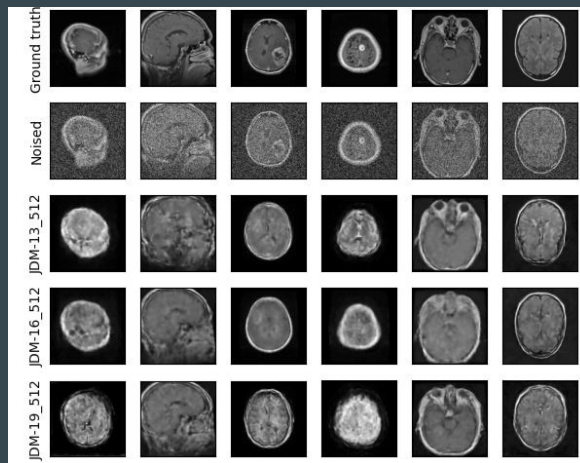
El modelo pre-entrenado se entrenó por 400 épocas. Se probaron distintas variables de modelo pre-entrenado.

- JDM13-256: modelo con 13 bloques convolucionales en el encoder y 13 en el decoder.
- JDM13-512, JDM16-256, JDM16-512, JDM19-256, JDM10-512.

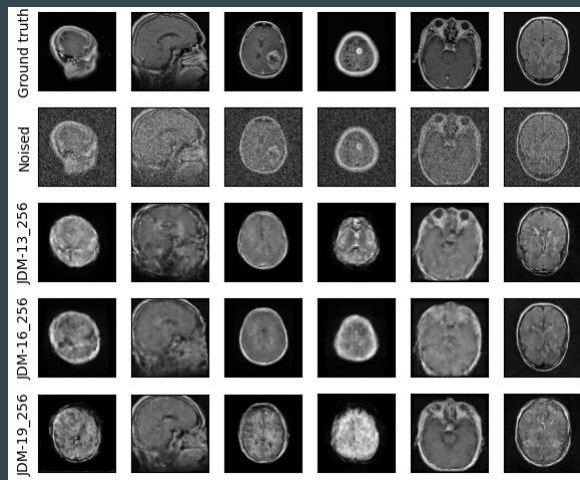
Resultados

Estos modelos preentrenados se probaron agregando ruido a la imagen y pasando la imagen con ruido al variational autoencoder. Encontramos resultados positivos para los 3 modelos distintos. En particular el vector latente no es un hiperparámetro relevante como lo es el número de capas convolucionales.

vector latente
512

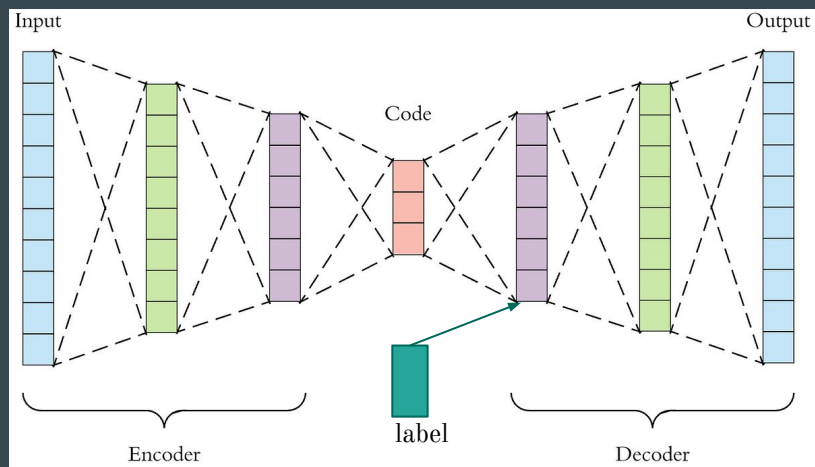


vector latente
256



Metodología - CVAE Transfer Learning

Se utilizó un dataset de 3000 imágenes de MRI de cerebro para el entrenamiento de un variational autoencoder



Se creó un VAE condicional que tuviera la misma arquitectura que el modelo preentrenado (con diferencia en los inputs y en las capas concatenadas por la etiqueta)

Se congelaron las capas convolucionales (no dependen de la etiqueta) y se entrenó el modelo por 30 épocas.

Se descongelaron las capas y se entrenó por 10 épocas.

Resultados

data_type	split	accuracy_score	recall_score__micro	recall_score__macro	precision_score__micro	precision_score__macro	f1_score__micro	f1_score__macro
baseline	train	0.993531	0.993531	0.994334	0.993531	0.994046	0.993531	0.994170
	test	0.963303	0.963303	0.968498	0.963303	0.965427	0.963303	0.966452
generated	train	0.805924	0.805924	0.797374	0.805924	0.856619	0.805924	0.808658
	test	0.437309	0.437309	0.479192	0.437309	0.604011	0.437309	0.384512
mixed	train	0.913006	0.913006	0.914103	0.913006	0.922535	0.913006	0.915647
	test	0.93578	0.93578	0.946286	0.93578	0.934789	0.93578	0.938577

Conclusiones

Conclusión

- Un Loss de generador cercano a cero no necesariamente implica que las imágenes son realistas, sino que siguen patrones de distribución de color similar a las imágenes originales.
- Los modelos generativos generalmente requieren millones de datos para obtener un resultado aceptable, eso puede influenciar el bajo rendimiento en test.
- Todos los modelos de clasificación entrenados con imágenes generadas muestran un overfitting.
- Aunque se ven imágenes que una persona reconocería como un cerebro, no se evidencian completamente los tumores.
- Para los VAE los hiperparámetros más relevantes corresponden a las capas convolucionales.
- Qué podría mejorar los resultados:
 - Más datos.
 - Otras arquitecturas más complejas.
- Con transfer-learning bastaron 40 épocas para tener resultados comparables con los demás modelos.