PROYECTO FINAL

DEEP LEARNING

MANOLO IÑIGUEZ

INTRODUCCIÓN

Problema a resolver:
Comparar el rendimiento de diferentes modelos de deep learning.

Objetivos de la comparación de modelos:
Evaluar el rendimiento en clasificación de imágenes, entender los modelos, análisis de métricas.

Técnicas utilizadas para resolver el problema:
Red Neuronal Convolucional (CNN).
Transformer de Visión (ViT)

TÉCNICAS

MATERIALES Y MÉTODOS

Red Neuronal Convolucional (CNN).

Funciona con una operación fundamental que es la convolución.
Consiste en deslizar un filtro sobre la imagen de entrada.
El filtro es una pequeña matriz numérica que se utiliza para extraer características de la imagen.
Mientras el filtro se desliza se multiplican elementos del filtro y esa región de la imagen y se repite en diferentes ubicaciones de la imagen para extraer diversas características.

Se tiene:
Capas de convolución: Filtros
Capas de agrupación: Reducción de dimensionalidad
Capas completamente conectadas: Combinación de características extraídas y clasificación
Funciones de activación: En cada capa ReLu, introduce no linealidad
Capa de salida: Depende de la tarea especifica, clasificación o regresión
Entrenamiento: Ajuste de valores de filtros, pesos, etc.

TÉCNICAS

Transformer de Visión (ViT)

ViT toma una imagen y la divide en parches que son aplanados y transformados en una secuencia de vectores de características lineales.
A cada vector se le añade una incrustación (embedding) posicional para mantener la información espacial.
La clave de ViT es el mecanismo de atención, especialmente la atención multi-cabeza. Esto permite al modelo enfocarse en diferentes partes de la imagen para extraer características relevantes, dividiendo la consulta, la clave y el valor en múltiples "cabezas".
Esto significa que el modelo realiza el proceso de atención varias veces en paralelo, cada vez con una parte diferente de los vectores.

Se tiene:
Redes de capas de atención: múltiples capas de mecanismos de atención
Redes neuronales Feed-Forward: La información se mueve en una sola dirección.
- *Capas de Entrada: La red recibe datos de entrada*
- *Capas Ocultas: Procesan las entradas con pesos y función de activación.*
- *Capa de Salida: Resultado de la red*

METODOLOGÍA

DATASETS

Lenguaje de señas 1

Contiene imágenes de señas del lenguaje de señas representadas en un formato tabular.

- ***Etiqueta (Label):*** *Cada fila comienza con una etiqueta que es un número entero. Esta etiqueta representa la clase de la seña.*
- ***Píxeles:*** *Cada imagen está representada por una serie de valores de píxeles. Hay 784 columnas de píxeles, lo que sugiere que cada imagen tiene un tamaño de 28x28 píxeles.*
- **Shape:** *27455 x 785*

Lenguaje de señas 2

- **Shape:** *7172 x 785*

RESULTADOS

CNN Dataset Grande

All Metrics:

	fold	epoch	learning_rate	accuracy	auc	precision	recall	f1_score	mcc
0	1	100	0.0001	0.9854	0.9999	0.9924	0.9923	0.9922	0.9919
1	2	100	0.0001	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000
2	3	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	4	100	0.0001	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000
4	5	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	6	100	0.0001	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000
6	7	100	0.0001	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000
7	8	100	0.0001	0.9995	1.0000	1.0000	1.0000	1.0000	1.0000
8	9	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	10	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

CNN Dataset Pequeño

All Metrics:

	fold	epoch	learning_rate	accuracy	auc	precision	recall	f1_score	mcc
0	1	100	0.0001	0.9761	0.9999	0.9794	0.9767	0.9766	0.9772
1	2	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
2	3	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	4	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	5	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	6	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	7	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	8	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	9	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	10	100	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

ViT Dataset Grande

All Metrics:

	fold	epoch	learning_rate	accuracy	auc	precision	recall	f1_score	mcc
0	1	50	0.0001	1.0000	0.9997	1.0000	1.0000	1.0000	1.0000
1	2	50	0.0001	0.9995	0.9999	1.0000	1.0000	1.0000	1.0000
2	3	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	4	50	0.0001	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000
4	5	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	6	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	7	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	8	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	9	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	10	50	0.0001	0.9995	0.9432	0.6146	0.6084	0.5942	0.5830

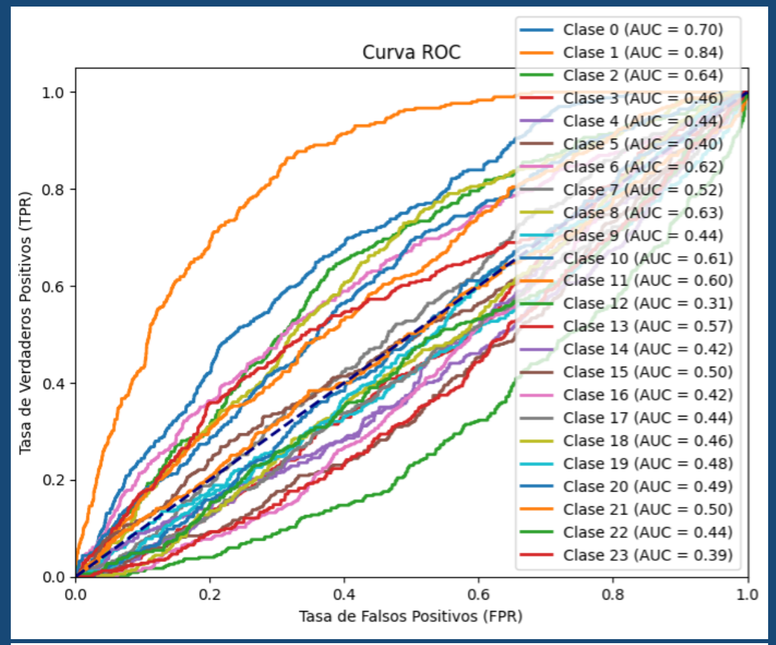
ViT Dataset Pequeño

All Metrics:

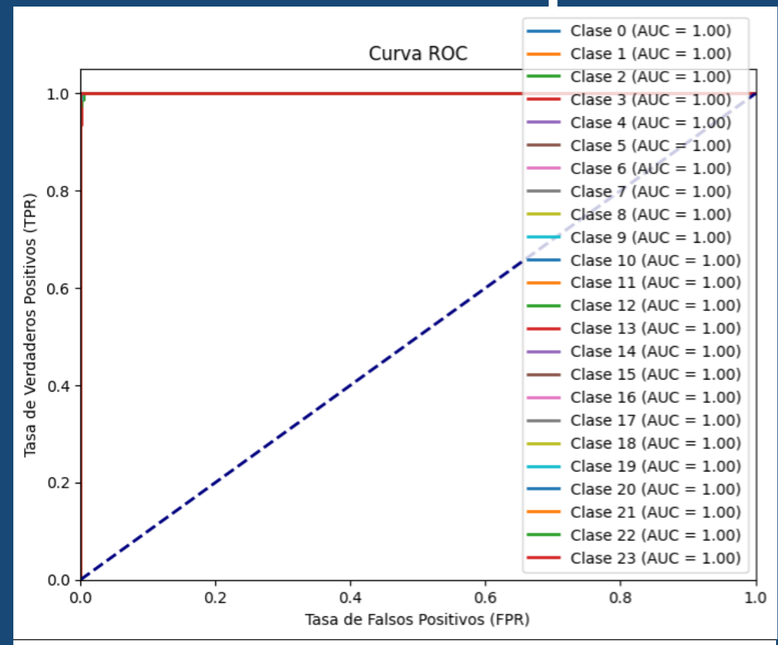
	fold	epoch	learning_rate	accuracy	auc	precision	recall	f1_score	mcc
0	1	50	0.0001	0.9900	0.9994	0.9943	0.9912	0.9925	0.9917
1	2	50	0.0001	1.0000	0.9997	1.0000	1.0000	1.0000	1.0000
2	3	50	0.0001	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000
3	4	50	0.0001	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000
4	5	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	6	50	0.0001	1.0000	0.9999	1.0000	1.0000	1.0000	1.0000
6	7	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	8	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	9	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	10	50	0.0001	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

AUC-ROC

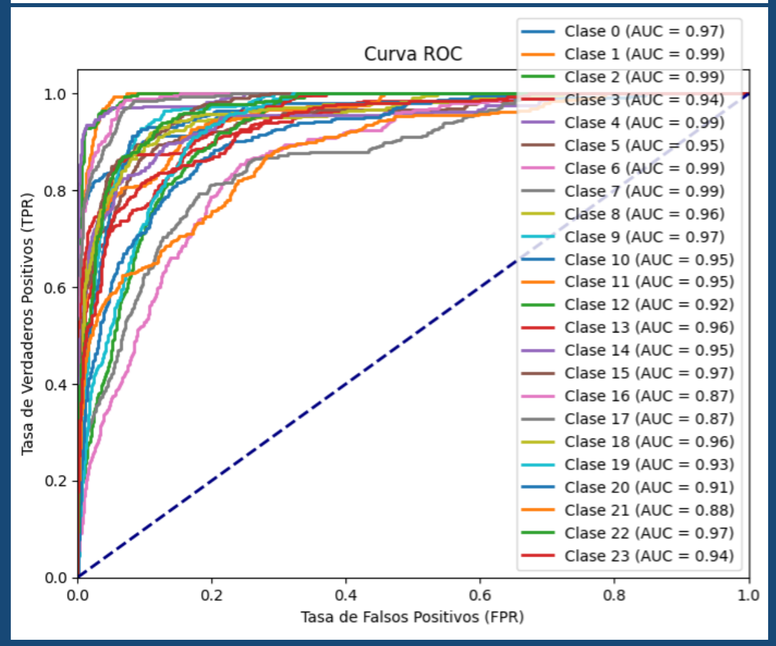
CNN Dataset Grande



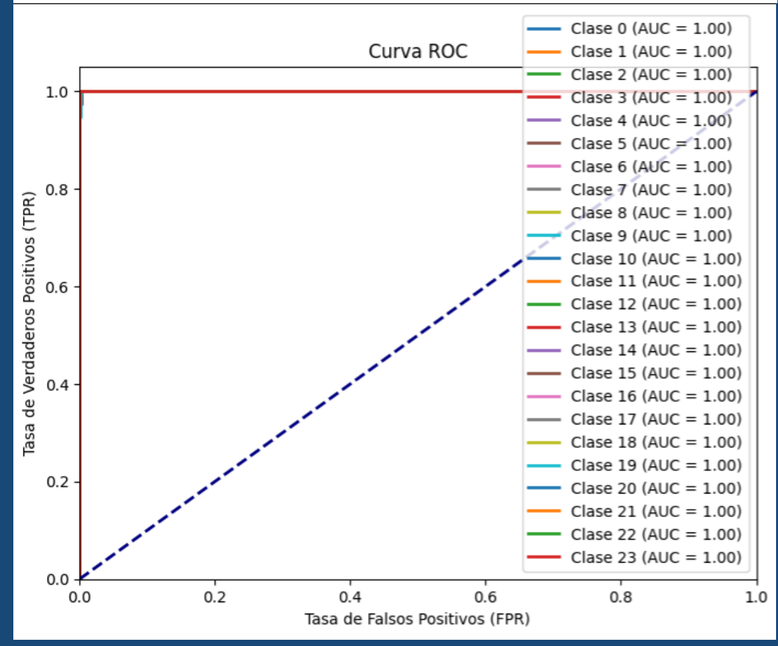
CNN Dataset Pequeño



ViT Dataset Grande

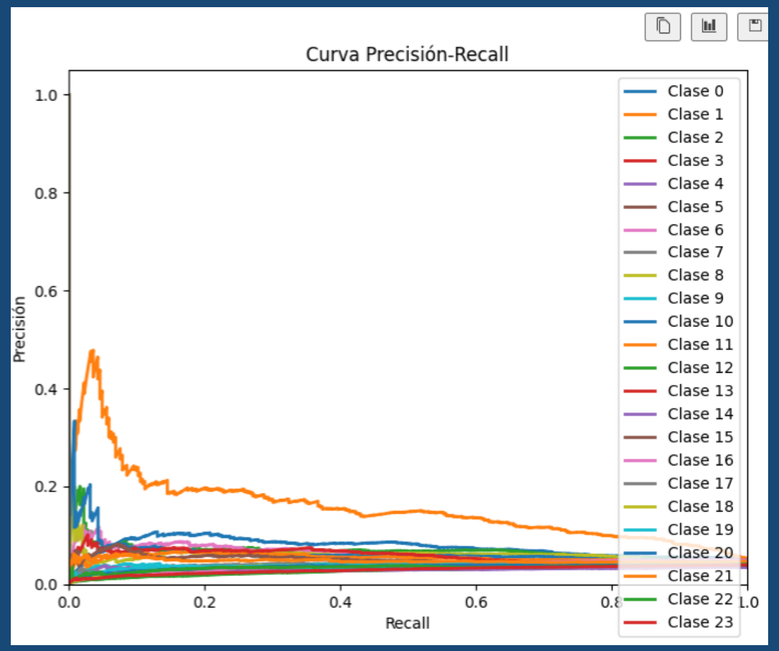


ViT Dataset Pequeño

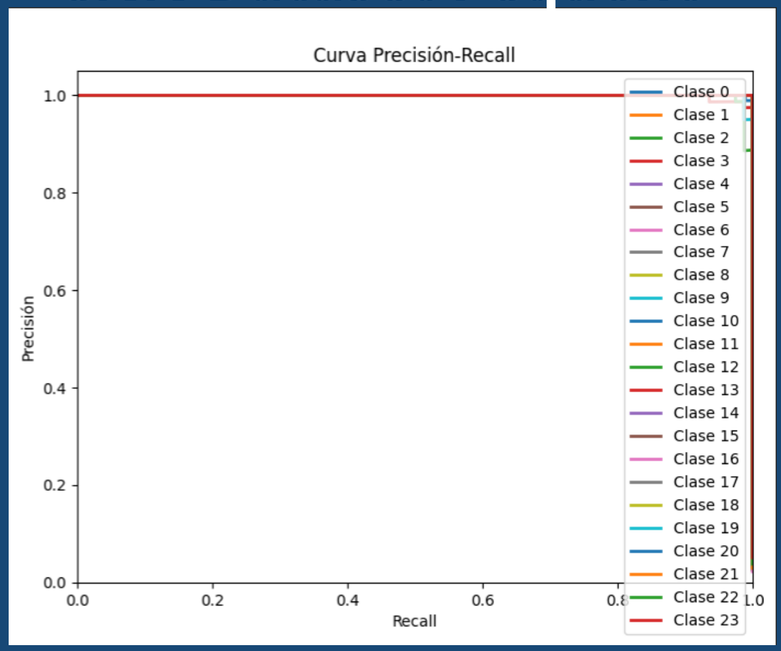


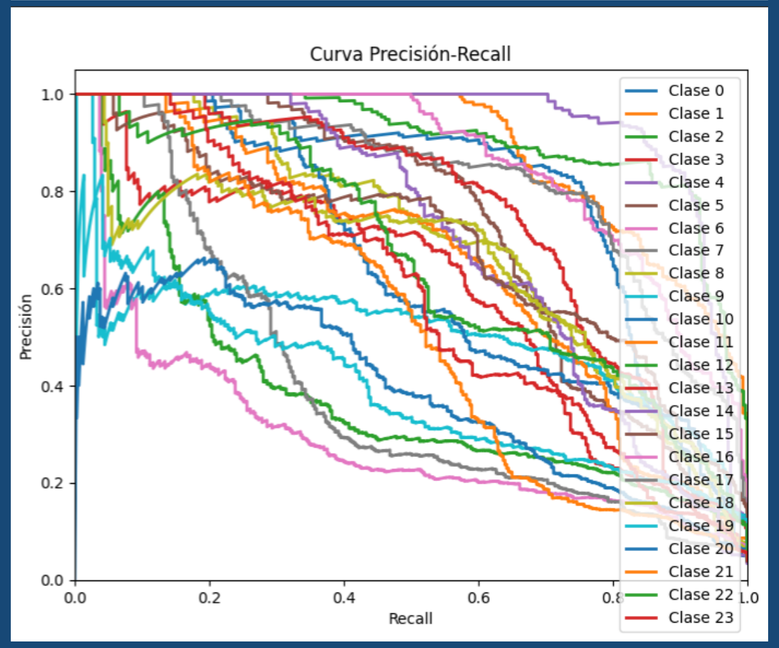
Precisión-Recall

CNN Dataset Grande

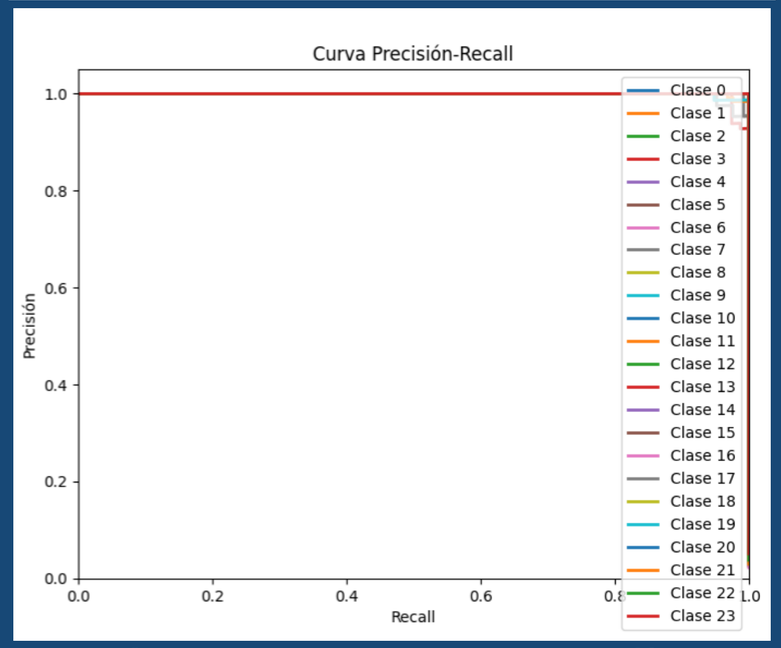


CNN Dataset Pequeño





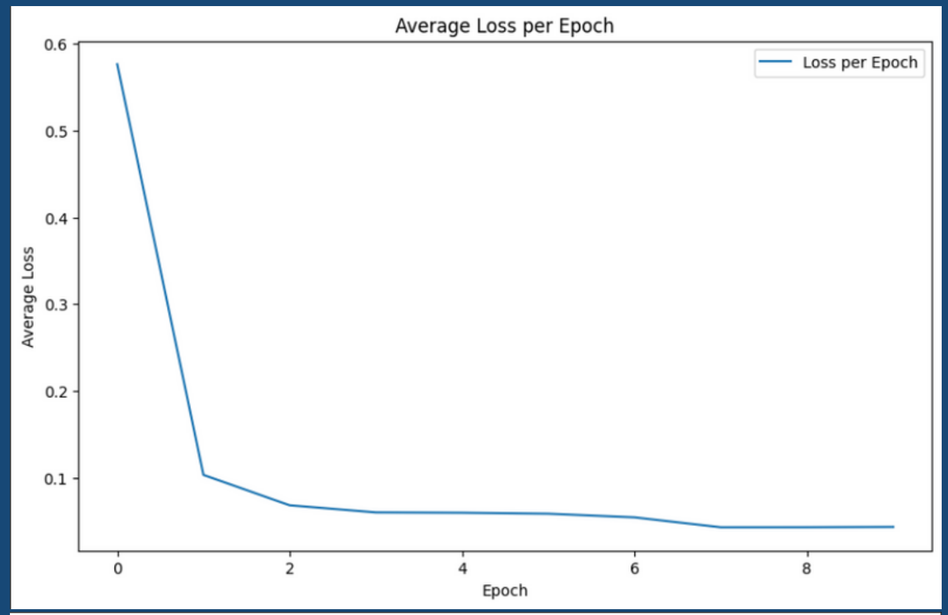
ViT Dataset Grande



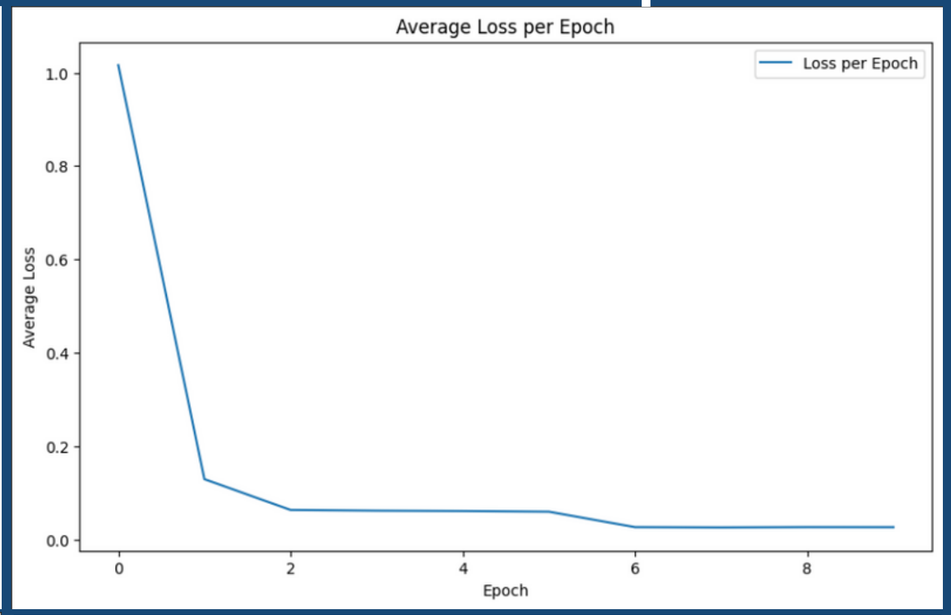
ViT Dataset Pequeño

Loss vs Epoch

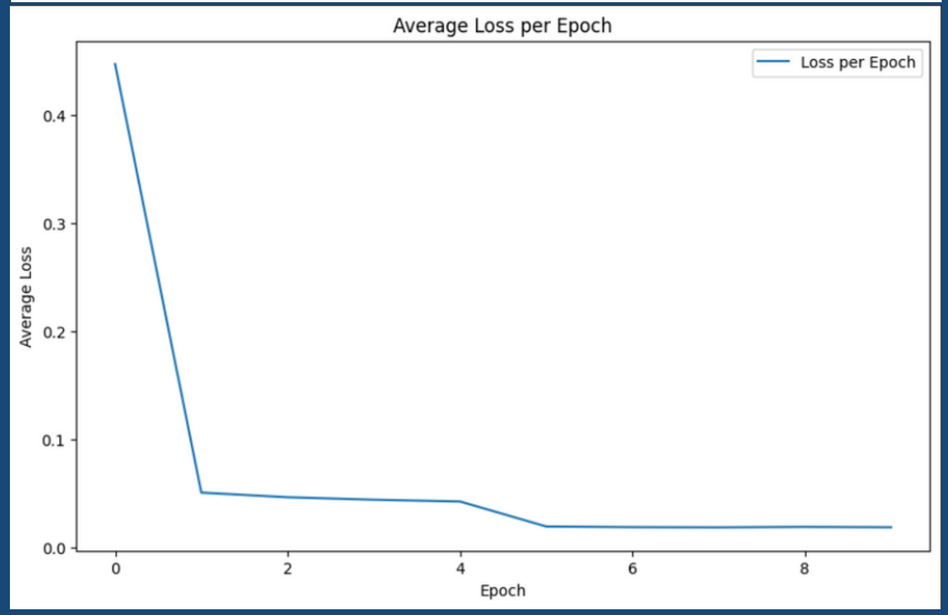
CNN Dataset Grande



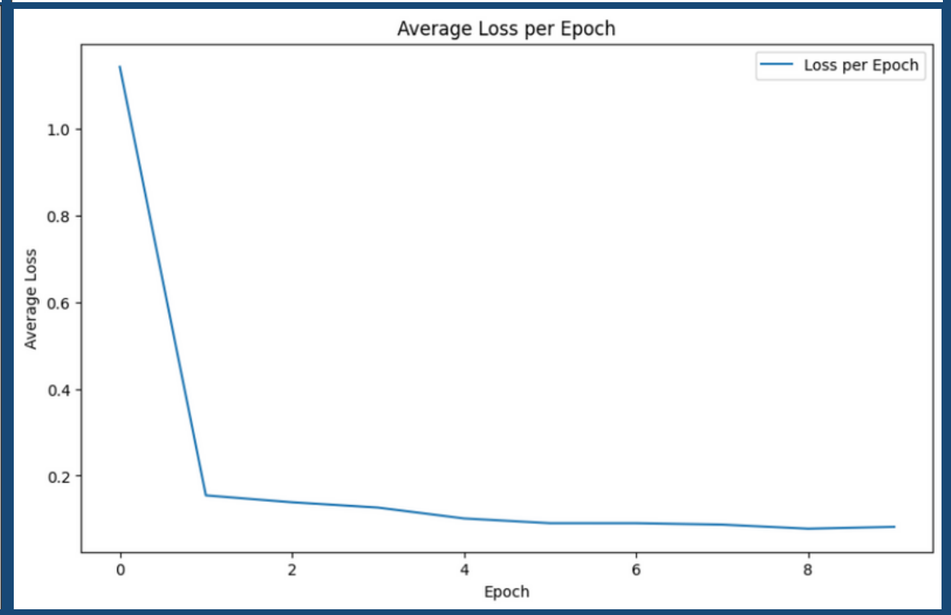
CNN Dataset Pequeño



ViT Dataset Grande

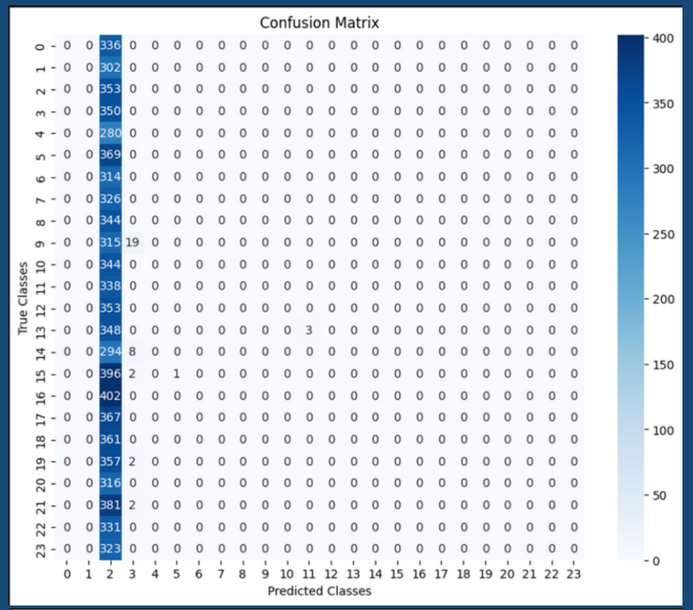


ViT Dataset Pequeño

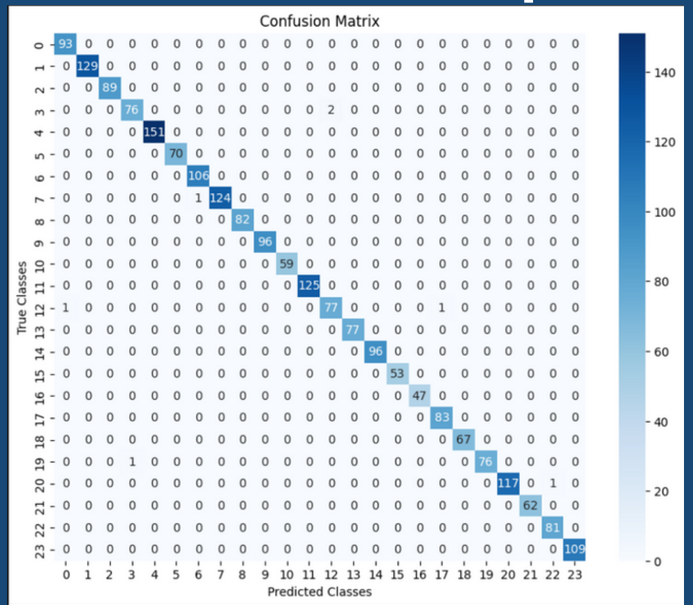


Matrices de confusión

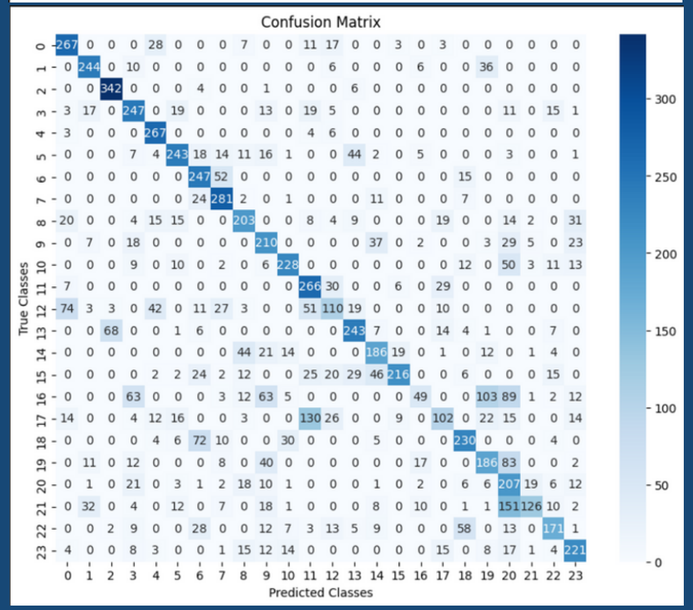
CNN Dataset Grande



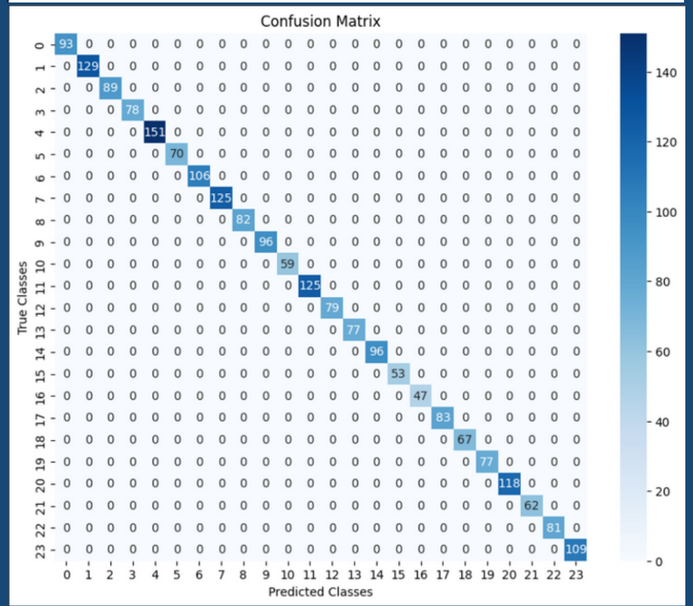
CNN Dataset Pequeño



ViT Dataset Grande



ViT Dataset Pequeño



CONCLUSIÓN

- *Los modelos CNN y ViT han demostrado tener métricas de rendimiento similares en el conjunto de datos con el que fueron evaluados.*
- *Ambos modelos son candidatos viables para tareas de clasificación de imágenes.*
- *Se puede ajustar los hiperparametros de los modelos aun mas.*
- *El modelo ViT tiende a superar al modelo CNN cuando se trabaja con conjuntos de datos más grandes.*

Probar modelos en vivo?

GRACIAS