

Artificial Intelligence for the Cosmos: Morphological Classification of Galaxies Using Neural Networks

F. Ortiz, B. Diaz, J. Aldama

Pontificia Universidad Católica del Perú, Facultad de Ciencias e Ingeniería

INTRODUCCIÓN

Este proyecto automatiza la clasificación de galaxias, superando las limitaciones de los métodos tradicionales y la inviabilidad de la clasificación manual. Se enfoca en cuatro tipos: **elípticas, espirales, irregulares y esféricas**, enfrentando desafíos como el ruido en las imágenes y la dependencia de datos etiquetados. Esto mejora la precisión y facilita el estudio de la evolución y distribución galáctica.

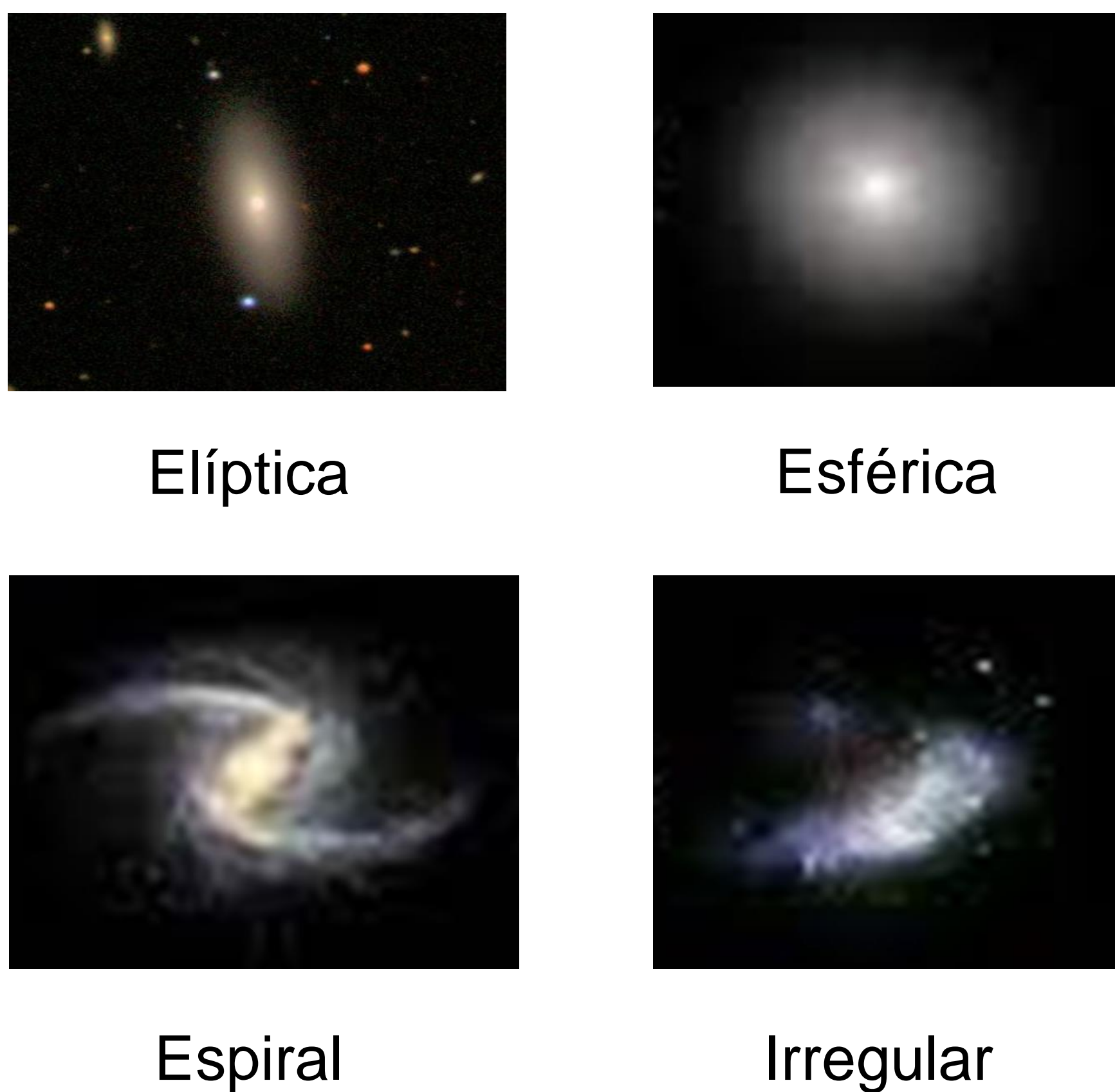


Fig.1: Clasificación morfológica de galaxias

OBJETIVOS

- **Desarrollar una red neuronal convolucional** para clasificar galaxias por morfología.
- **Depurar un conjunto de datos astronómicos**, eliminando valores atípicos y duplicados.
- **Nivelar la cantidad de imágenes** con dada augmentation.
- **Mejorar las técnicas de procesamiento de datos** en observaciones astronómicas.

METODOLOGÍA

Términos utilizados:

1. Red Neuronal Convolucional (CNN)
2. Softmax
3. Dropeo
4. Clasificación de Galaxias
5. Clases de Imágenes de Galaxias – Elípticas, Espirales, Irregulares, Esféricas

Construcción del modelo de Red Neuronal Convolucional:

Arquitectura de la CNN:

- 2 capas convolucionales
- 4 capas ocultas
- 2 capas densas penúltimas
- 1 capa de salida

Hiperparámetros:

- Tamaño de filtro: 3×3
- Tamaño de imagen: 128×128 píxeles
- Tamaño de lote: 1
- Épocas: 25
- Timestamps: 10 por época

Galaxy Types	Total Images	Training Set	Validation Set	Testing Set
Elliptical	17,295	12,106	2,594	2,595
Spiral	17,295	12,106	2,594	2,595
Irregular	17,295	12,106	2,594	2,595
Sphere	17,295	12,106	2,594	2,595

Tabla1: Diferentes clases con el número de imágenes de training, validation y testing

ARQUITECTURA DEL MODELO

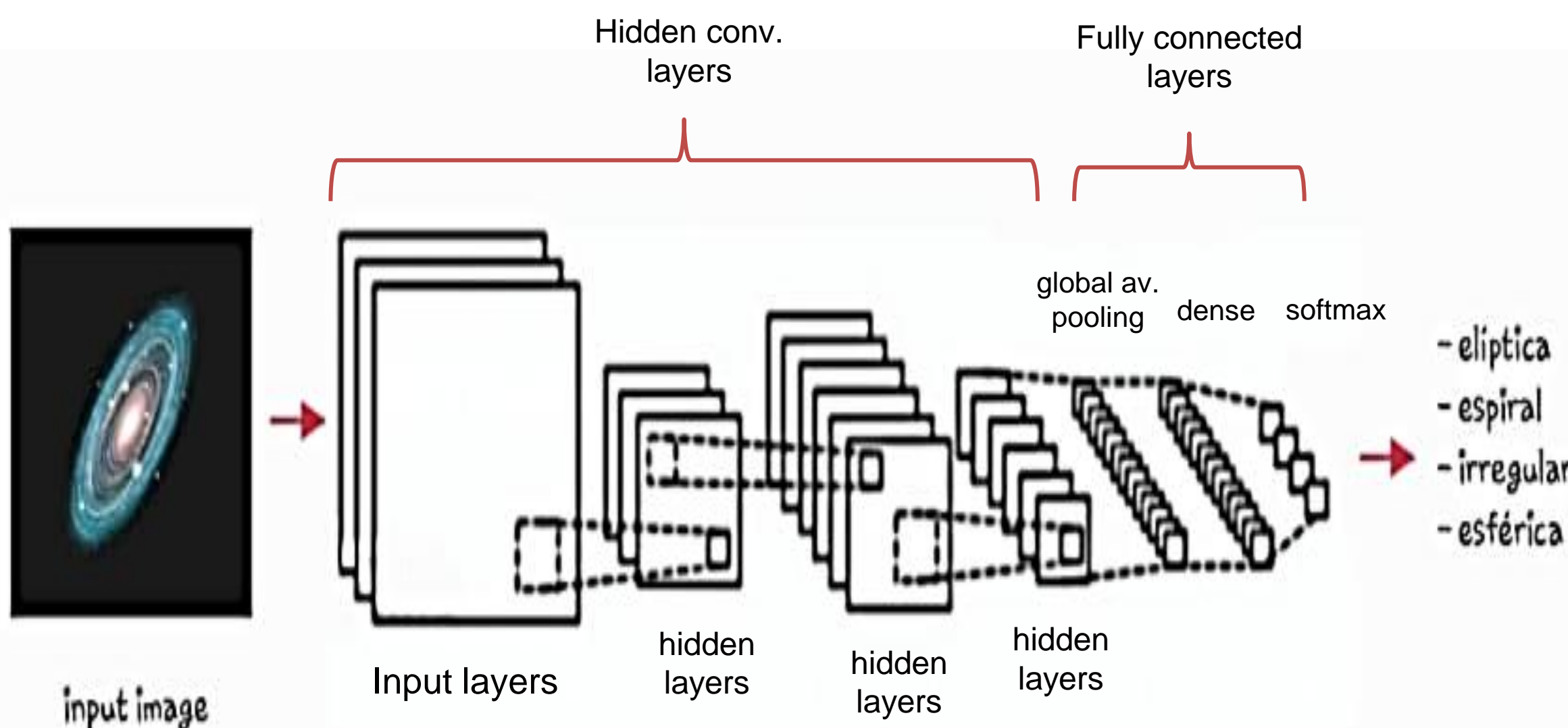


Fig.2: Arquitectura de la red neuronal convolucional para la clasificación de galaxias

RESULTADOS

Con respecto al accuracy y loss que presenta el modelo podemos decir que alcanzó **74%** de precisión en entrenamiento y **75%** en validación. La precisión en validación se estabiliza tras la época 15, indicando posible sobreajuste. La pérdida disminuyó de **1.14** a **0.63** en entrenamiento y **0.64** en validación, con fluctuaciones en las últimas épocas.

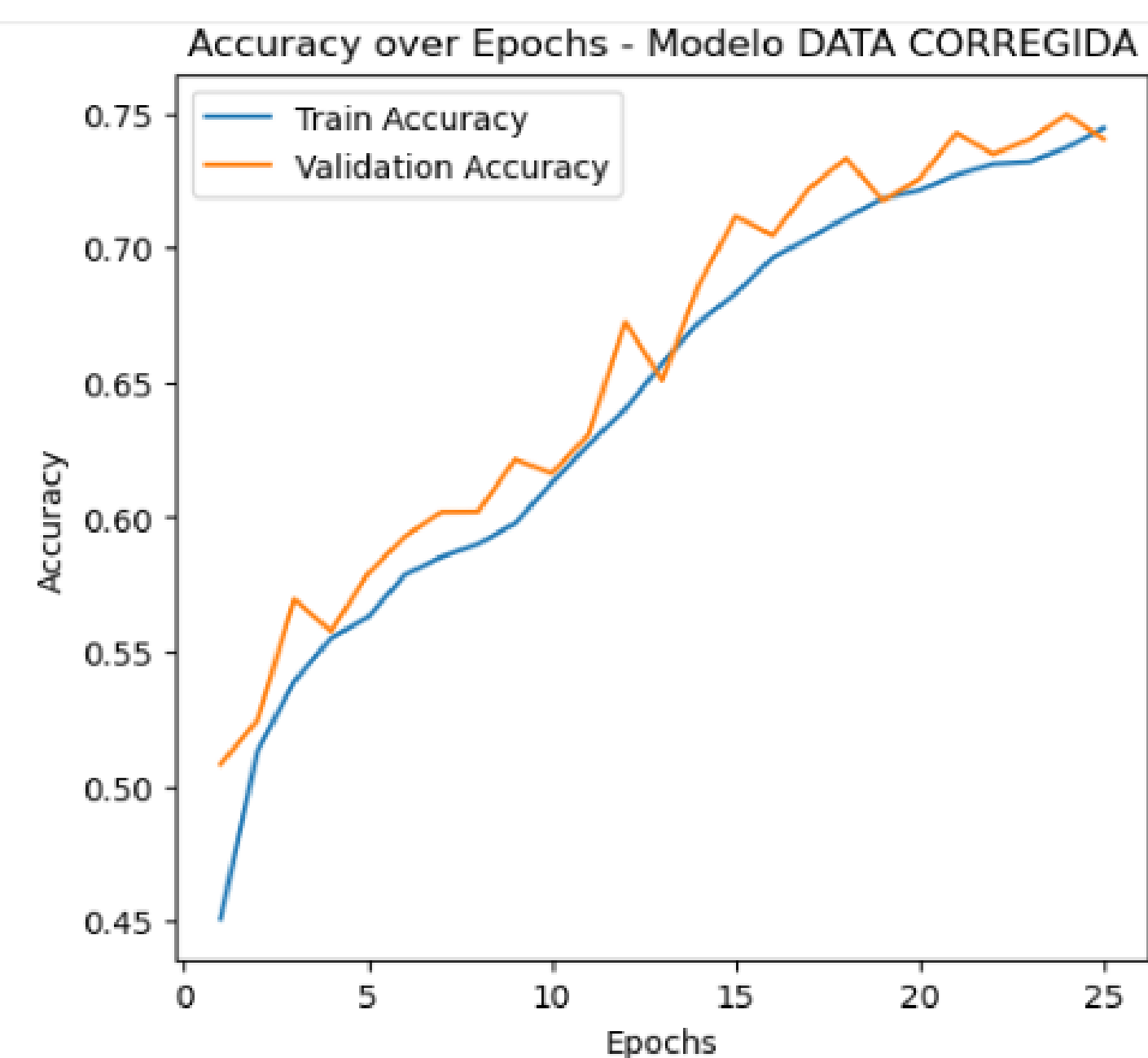
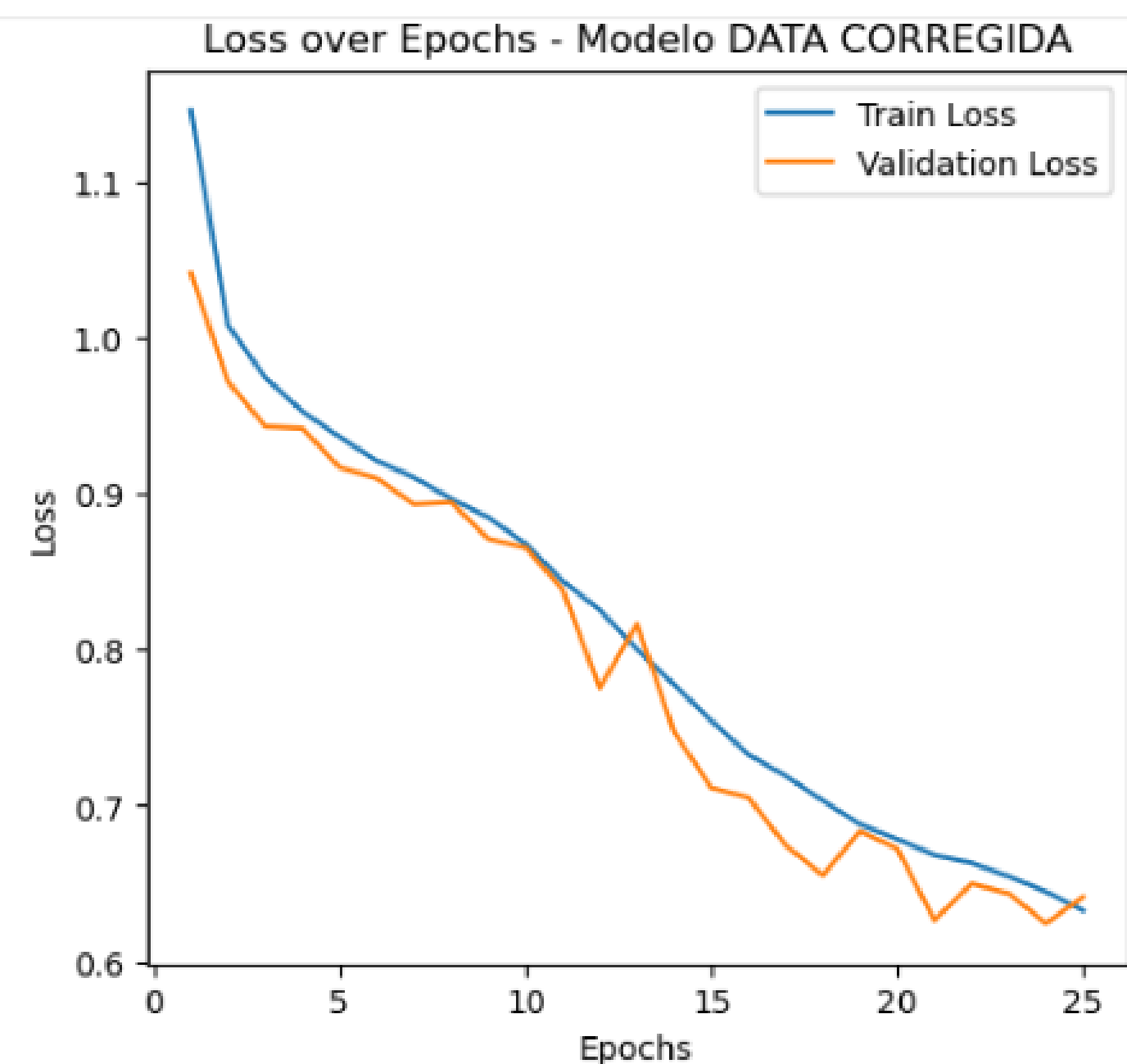


Fig.3: Gráfica del accuracy del modelo

Fig.4: Gráfica del loss del modelo



Además, podemos observar en las matrices de confusión la alta precisión en galaxias elípticas y esféricas, mientras que las espirales e irregulares presentan más confusiones. Esto sugiere que el modelo tiene dificultades en diferenciar características sutiles entre estas clases.

Fig.5: Matriz de confusión de la data de validación

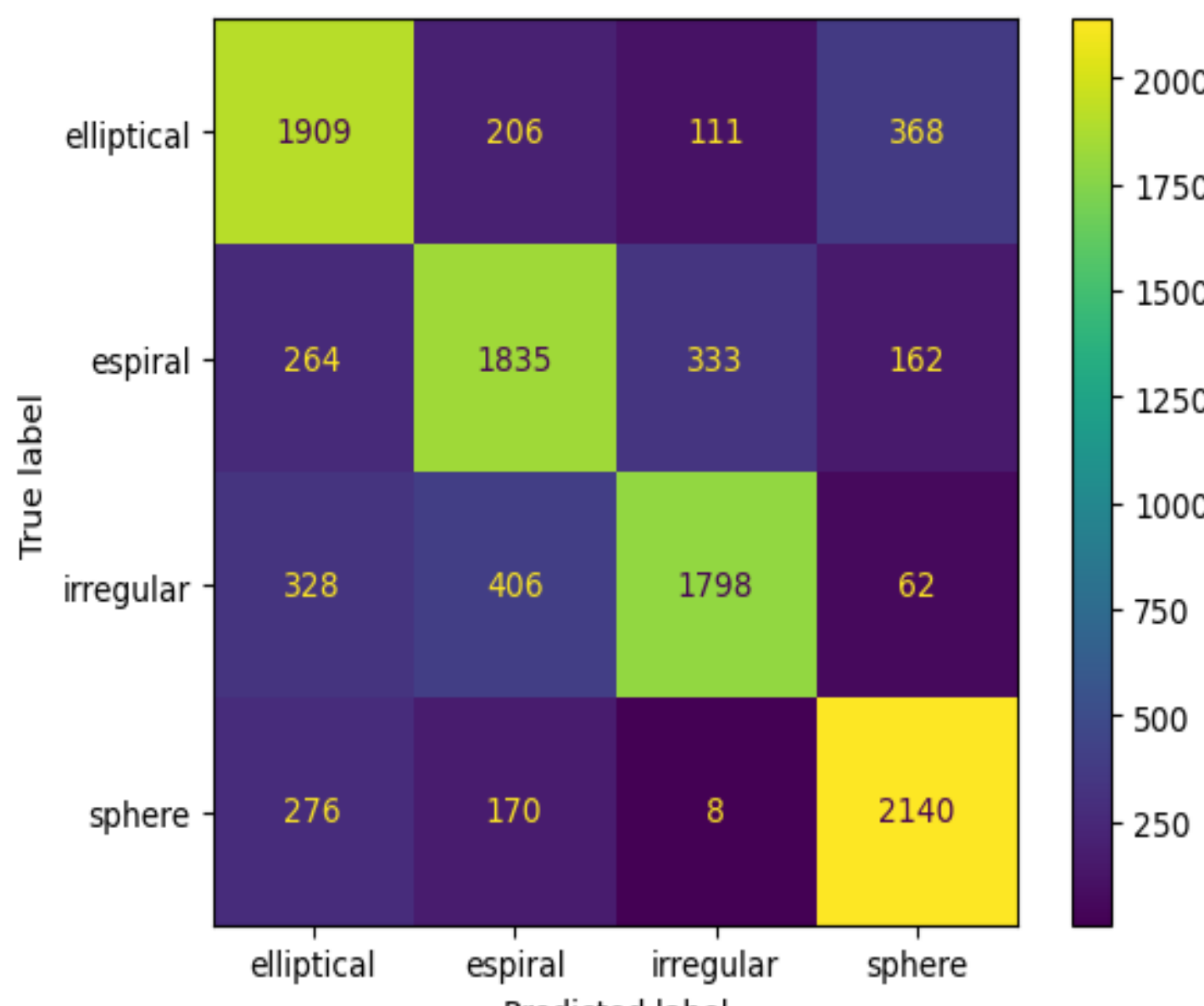
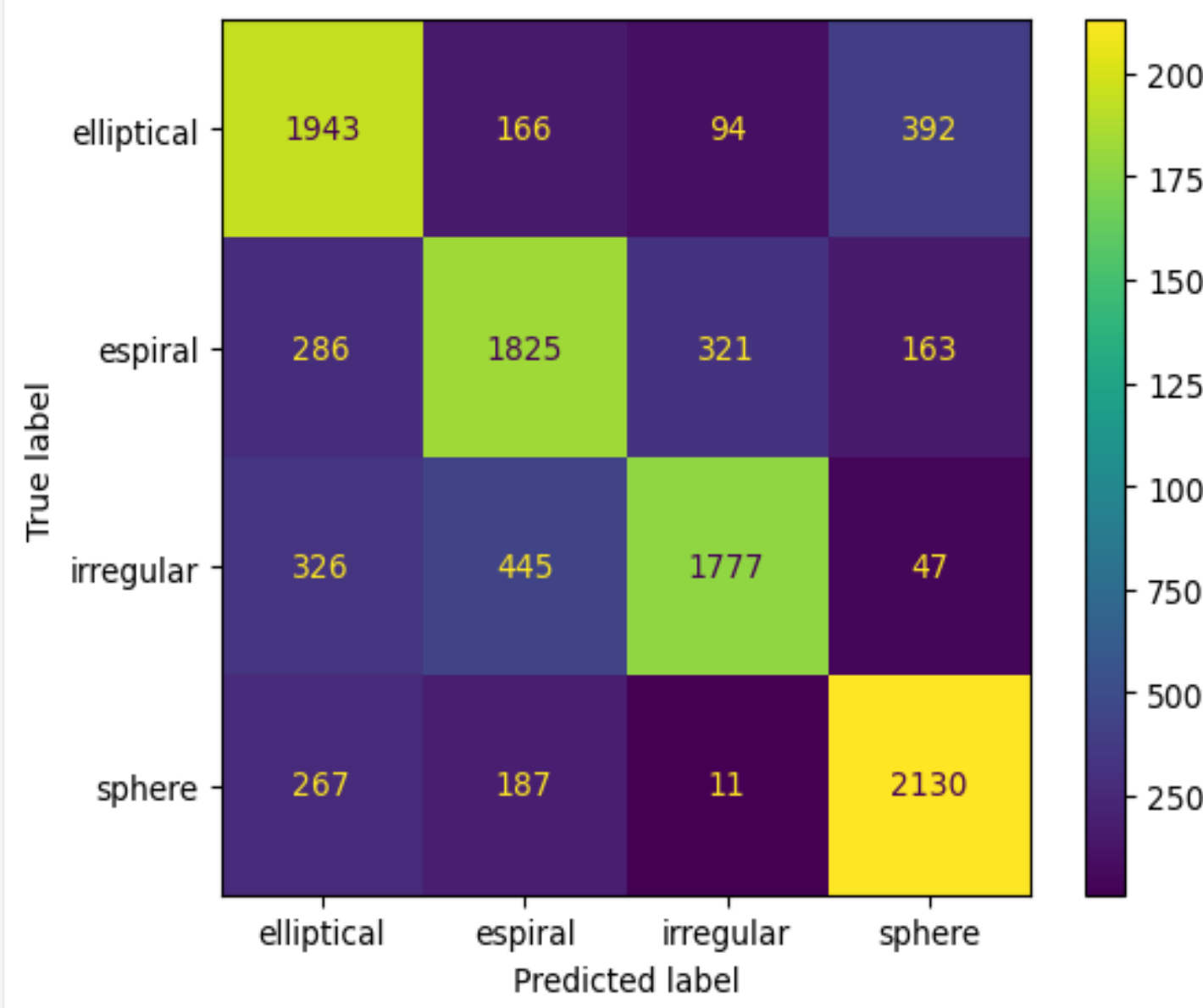


Fig.6: Matriz de confusión de la data de testeo



CONCLUSIÓN

El modelo presenta un buen desempeño en la clasificación de galaxias, pero aún puede mejorarse. Para optimizar su rendimiento, se recomienda ajustar la arquitectura, refinar los hiperparámetros y aplicar técnicas de aumento de datos. Estas mejoras permitirán reducir errores y mejorar la capacidad de generalización en la identificación de distintas clases de galaxias.

ALCANCE

Este proyecto tiene aplicaciones en astronomía e investigación, además de servir como herramienta educativa. Puede emplearse en la divulgación científica mediante un juego didáctico que ayude a estudiantes a clasificar galaxias, fomentando el interés en la ciencia desde edades tempranas.

REFERENCIAS

Do, T., Boscoe, B., Jones, E., Li, Y. Q., & Alfaro, K. (2024). *GalaxiesML: A dataset of galaxy images, photometry, redshifts, and structural parameters for machine learning*. arXiv. <https://arxiv.org/abs/2410.00271>

