

2025

Estudio de la relación morfología-densidad

Isabel Maldonado



**Universidad Autónoma de  
Querétaro**

**Facultad de Ingeniería**

**Estudio de la relación morfología-densidad  
en cúmulos de galaxias con rayos X**

**Tesis**

Que como parte de los requisitos para obtener el  
Grado de

Licenciada en

**Ingeniería Física**

Presenta

**Isabel Yenifer Maldonado García**

Dirigido por:

**Dr. Josué de Jesús Trejo Alonso**

Querétaro, Qro. a mayo de 2025

**La presente obra está bajo la licencia:**  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



**CC BY-NC-ND 4.0 DEED**

**Atribución-NoComercial-SinDerivadas 4.0 Internacional**

**Usted es libre de:**

**Compartir** — copiar y redistribuir el material en cualquier medio o formato

La licenciatario no puede revocar estas libertades en tanto usted siga los términos de la licencia

**Bajo los siguientes términos:**

 **Atribución** — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciatario.

 **NoComercial** — Usted no puede hacer uso del material con [propósitos comerciales](#).

 **SinDerivadas** — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

**No hay restricciones adicionales** — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

**Avisos:**

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.



**Universidad Autónoma de Querétaro  
Facultad de Ingeniería  
Licenciatura en Ingeniería Física**

**Estudio de la relación morfología-densidad en cúmulos de galaxias con rayos X**

**Tesis**

Que como parte de los requisitos para obtener el Grado de  
**Licenciada en Ingeniería Física**

Presenta

**Isabel Yenifer Maldonado García**

Dirigido por

**Dr. Josué de Jesús Trejo Alonso**

Sínodo

Dr. Josué de Jesús Trejo Alonso  
**Presidente**

\_\_\_\_\_

Firma

Dr. René Alberto Ortega Minakata  
**Secretario**

\_\_\_\_\_

Firma

Dr. José Alonso López Miranda  
**Vocal**

\_\_\_\_\_

Firma

Dr. Aldrin Cervantes Contreras  
**Suplente**

\_\_\_\_\_

Firma

Dr. Alberto Hernández Almada  
**Suplente**

\_\_\_\_\_

Firma

Campus Aeropuerto, Querétaro, Qro. México  
Mayo de 2025

**A quien creyó en mí incondicionalmente.**

---

## AGRADECIMIENTOS

---

Han pasado poco más de dos años desde que comenzó este viaje llamado *tesis*. Quiero agradecer profundamente a mi asesor, el Dr. Trejo, por su constancia, comprensión y, sobre todo, paciencia, que me ayudaron a seguir adelante incluso en los momentos más complicados. A mi familia, en especial a mis padres, por estar siempre ahí, animándome a continuar. A mis amigos, que hicieron estos años de universidad más llevaderos y memorables. A mi compañero de vida, Spin, por confiar en mí y en mi trabajo, incluso cuando yo misma dudaba.

Siento una gratitud inmensa hacia todas las personas con las que tuve la fortuna de coincidir en este camino.

Por último, quiero agradecer a la autora de este escrito, quien finalmente concluyó aquello que alguna vez pareció inalcanzable. Por su esfuerzo, resistencia, perseverancia y dedicación. Sé que en más de una ocasión pensó en rendirse, dejarlo todo. Sin embargo, continuó, a pesar de las dudas y los pensamientos negativos. Gracias por todo, Isa.

---

## ÍNDICE GENERAL

---

<b>Índice de figuras</b>	<b>5</b>
<b>Índice de tablas</b>	<b>7</b>
<b>Abreviaturas y siglas</b>	<b>7</b>
<b>Resumen</b>	<b>9</b>
<b>Abstract</b>	<b>10</b>
<b>1. Introducción</b>	<b>12</b>
<b>2. Antecedentes</b>	<b>14</b>
2.1. Morfología de las galaxias . . . . .	14
2.1.1. Clasificación de Hubble . . . . .	14
2.1.2. Galaxias elípticas . . . . .	15
2.1.3. Galaxias S0 o lenticulares . . . . .	15
2.1.4. Galaxias espirales . . . . .	15
2.1.5. Galaxias irregulares . . . . .	16
2.2. Cúmulos de galaxias . . . . .	16
2.2.1. Cúmulos de Abell . . . . .	17
2.3. Emisión de rayos X en cúmulos de galaxias . . . . .	17
2.4. Evolución de galaxias en cúmulos . . . . .	19
2.4.1. Procesos ambientales de transformación morfológica . . . . .	19
2.4.2. El rol de la masa en la evolución morfológica . . . . .	20
2.5. Relación morfología-densidad en cúmulos de galaxias . . . . .	20
2.6. Predicción de la morfología de las galaxias con CNNs . . . . .	22
2.6.1. Fundamentos y arquitectura . . . . .	23
2.6.2. Entrenamiento y optimización . . . . .	26
2.6.3. Aplicaciones en Astronomía . . . . .	27

<b>3. Metodología</b>	<b>28</b>
3.1. Selección de la muestra . . . . .	29
3.2. Estudio de membresía . . . . .	30
3.3. Base de datos con Galaxy Zoo . . . . .	31
3.4. Construcción de la CNN . . . . .	33
3.5. Clasificación de galaxias miembro faltantes . . . . .	37
3.6. Diagramas morfología-densidad . . . . .	38
<b>4. Resultados y discusión</b>	<b>41</b>
4.1. Selección de la muestra . . . . .	41
4.2. Estudio de membresía . . . . .	42
4.3. Base de datos de GZ . . . . .	45
4.4. Construcción de la CNN . . . . .	45
4.5. Clasificación de galaxias miembro faltantes . . . . .	48
4.6. Diagramas morfología-densidad . . . . .	49
4.7. Comparación con clasificaciones visuales . . . . .	58
<b>5. Conclusiones</b>	<b>60</b>
<b>Bibliografía</b>	<b>61</b>
<b>Apéndices</b>	<b>65</b>

---

## ÍNDICE DE FIGURAS

---

2.1.	Diagrama de la secuencia propuesta por Hubble que representa la morfología de las galaxias (Mo, Van den Bosch y White, 2010). . . . .	14
2.2.	Observaciones del Telescopio Espacial Hubble que muestran galaxias de distintos tipos morfológicos: E, S0, S, Irr (ordenadas de izquierda a derecha) (NASA Science, 2024). . . . .	16
2.3.	El cúmulo Abell 85 visto en longitud de onda óptica (izquierda) y en rayos X (derecha) (Max Planck Society, 2019; Chandra X-ray Observatory, 2008). . . . .	18
2.4.	Fracción de galaxias E, SO y S+I en función del logaritmo de la densidad proyectada (Dressler, 1980). . . . .	21
2.5.	Operación de convolución con un <i>stride</i> de 3 verticalmente y 2 horizontalmente (Zhang et al., 2023). . . . .	24
2.6.	Operación de convolución con <i>padding</i> (Zhang et al., 2023). . . . .	24
2.7.	Arquitectura de una CNN utilizada para identificar imágenes de perros (Alzubaidi et al., 2021). . . . .	25
3.1.	Diagrama representativo de las etapas seguidas a lo largo del estudio. . . . .	28
3.2.	Algunas columnas del catálogo original de Galaxy Zoo 2. . . . .	31
3.3.	Catálogo modificado de Galaxy Zoo 2 con las columnas de interés. . . . .	32
3.4.	Galaxias (a) elíptica y (b) espiral obtenidas de SDSS. . . . .	33
3.5.	Diagrama de flujo que representa el procesamiento de los datos a través de la CNN. . . . .	34
3.6.	Mapas de características de galaxias (a) elíptica y (b) espiral pertenecientes al conjunto de entrenamiento. . . . .	36
3.7.	Curvas de aprendizaje de GalaxyClassifier con 30 épocas. La precisión de validación final es 77.50%. . . . .	37
3.8.	Visión general de las fases seguidas para clasificar las galaxias restantes con la CNN entrenada. . . . .	37
3.9.	Proyección cartesiana del cúmulo A0085. . . . .	38
4.1.	Algunas de las columnas de los 11 cúmulos seleccionados para la muestra. . . . .	42

4.2. Proyección de los cúmulos en coordenadas ecuatoriales (ascensión recta y declinación) . . . . .	42
4.3. Datos de las galaxias de A0085 al aplicar el método de cáusticas. . . . .	43
4.4. Gráfico obtenido del análisis de cáusticas del cúmulo A0085. . . . .	43
4.5. Gráfico obtenido después del análisis de cáusticas para los cúmulos A0168 y A1139. . . . .	44
4.6. Gráfico resultante del análisis de cáusticas para los cúmulos A1750 y A2197. . . . .	44
4.7. Organización de las imágenes utilizadas para entrenar la CNN. . . . .	45
4.8. Curvas de aprendizaje del modelo M-a. . . . .	46
4.9. Curvas de aprendizaje del modelo M-b. . . . .	47
4.10. Curvas de aprendizaje del modelo M-c. . . . .	47
4.11. Diagrama para A0085, considerado como un cúmulo que sigue la tendencia esperada. . . . .	50
4.12. Diagramas de la relación morfología-densidad para los cúmulos A0168 y A1139, que van con la tendencia esperada. . . . .	51
4.13. Diagramas morfología-densidad de los cúmulos alineados con la tendencia esperada: A1767, A2197. . . . .	52
4.14. Diagramas de cúmulos que siguen la tendencia esperada: A2255, A2670.	53
4.15. Diagramas de dos cúmulos en los que las galaxias espirales predominan: A1750, A2244. . . . .	55
4.16. Diagramas de cúmulos que evidencian una tendencia inusual: A1205, A1650. . . . .	57
4.17. Diagramas de la relación morfología-densidad de A0085 y A0168 obtenidos mediante la clasificación de la CNN y visual. . . . .	59

---

## ÍNDICE DE TABLAS

---

2.1. Fracción de galaxias observadas por Dressler en un entorno de alta densidad, es decir, al centro del cúmulo (Spinrad, 2005). . . . .	21
2.2. Fracción de galaxias observadas por Dressler en un entorno de baja densidad o a las afueras del cúmulo (Spinrad, 2005). . . . .	21
3.1. Resumen de la arquitectura del modelo <code>GalaxyClassifier</code> . . . . .	35
4.1. Resumen de los entrenamientos realizados para el modelo de la CNN. Las pruebas reflejan las clasificaciones que se hicieron correctamente respecto al total. . . . .	46
4.2. 5 imágenes de galaxias categorizadas con la CNN entrenada. . . . .	48

---

## ABREVIATURAS Y SIGLAS

---

- **ANN.** Del inglés *Artificial Neural Network* o red neuronal artificial.
- **CNN.** Del inglés *Convolutional Neural Network* o red neuronal convolucional.
- **DM.** Del inglés *Dark Matter* o materia oscura.
- **E.** Notación para las galaxias elípticas.
- **ETG.** Del inglés *Early Type Galaxies* o galaxias de morfología temprana (elípticas y lenticulares).
- **GZ.** Abreviación de *Galaxy Zoo*.
- **ICM.** Del inglés *Intraccluster Medium* o medio intracumular.
- **Irr.** Notación para las galaxias irregulares.
- **LTG.** Del inglés *Late Type Galaxies* o galaxias de morfología tardía (espirales e irregulares).
- **$M_{\odot}$ .** Símbolo de masa solar que equivale a  $1.9891 \times 10^{30}$  kg (Carroll, 2014).
- **Pop04.** Abreviación para el nombre del artículo Popesso et al. (2004).
- **RASS.** Sondeo *ROSAT All-Sky Survey* hecho en el rango de los rayos X.
- **RPS.** Del inglés *Ram Pressure Stripping* o presión de arrastre ejercida sobre un cuerpo que se mueve a través de un medio fluido.
- **S.** Notación para las galaxias espirales.
- **S0.** Notación para las galaxias lenticulares.
- **SDSS.** Sondeo *Sloan Digital Sky Survey* hecho en el rango del espectro visible.

- **z** o **redshift**. Es el desplazamiento de la luz de objetos como estrellas o galaxias hacia longitudes de onda más largas (hacia el rojo en el espectro electromagnético) por la expansión del Universo (Carroll, 2014).

---

## RESUMEN

---

Este estudio investiga la relación entre la morfología de las galaxias y el entorno donde se encuentran, en una muestra de 11 cúmulos Abell en el rango  $0.02 \leq z \leq 0.095$ , a partir del catálogo de Popesso et al. (2004). Para clasificar las imágenes de dichas galaxias, desarrollamos una red neuronal convolucional (CNN) entrenada con imágenes del SDSS basadas en las etiquetas asignadas en el proyecto Galaxy Zoo 2, alcanzando una precisión del 77.50 %. Nuestros resultados revelan tres tendencias principales en comparación con el trabajo de Dressler (1980): (1) esperados, (2) con un dominio de espirales y (3) inesperados. Concluimos que, para cúmulos cercanos, factores como el equilibrio dinámico, el tiempo de formación, la presencia de subestructuras, así como las interacciones gravitacionales e hidrodinámicas, influyen en cómo se distribuyen las poblaciones de galaxias elípticas y espirales.

**Palabras clave:** cúmulos, galaxias, morfología-densidad, redes neuronales convolucionales, clasificación de imágenes.

---

## ABSTRACT

---

This study investigates the relationship between the morphology of galaxies and their environment in a sample of 11 Abell clusters in the range  $0.02 \leq z \leq 0.095$ , from the catalog of Popesso et al. (2004). To classify the images of such galaxies, a convolutional neural network (CNN) was designed and trained using SDSS images based on the labels assigned in the Galaxy Zoo 2 project, achieving an accuracy of 77.50 %. Our results reveal three main trends compared to the work of Dressler (1980): (1) expected, (2) spiral-dominated, and (3) unexpected. We conclude that, for nearby clusters, factors such as dynamical equilibrium, formation time, the presence of substructures, as well as gravitational and hydrodynamic interactions, influence how populations of elliptical and spiral galaxies are distributed.

**Keywords:** clusters, galaxies, morphology-density, convolutional neural networks, image classification.

# CAPÍTULO 1

---

## INTRODUCCIÓN

---

En el Universo, los cúmulos de galaxias son las estructuras más grandes ligadas gravitacionalmente. Estos representan un entorno denso en el que las galaxias experimentan interacciones que pueden afectar su desarrollo morfológico. Durante décadas, se ha observado que la morfología de la galaxia está relacionada con la densidad del entorno. Esto indica que los procesos como fusiones, interacciones entre galaxias y efectos del *medio intracumular*<sup>1</sup> (ICM, en inglés *intracluster medium*) juegan un papel importante en la transformación de las galaxias.

Uno de los primeros estudios que estableció esta relación fue el de Dressler (1980), quien analizó 55 cúmulos de galaxias y encontró que en regiones de alta densidad las galaxias *elípticas* y *lenticulares* eran más comunes, mientras que en regiones menos densas, la fracción de galaxias *espirales* aumentaba significativamente. Desde entonces, muchos estudios han confirmado esta tendencia en una variedad de cúmulos (Whitmore et al., 1993; Goto et al., 2003; van der Wel, 2008). Recientemente, se ha estudiado cómo la proporción de galaxias elípticas y lenticulares varía con diferentes *redshifts* (Sazonova et al., 2020; Mei et al., 2023). Además, el desarrollo de redes neuronales ha permitido una clasificación morfológica más precisa, minimizando la subjetividad humana y mejorando la identificación de la estructura de las galaxias (Pfeffer et al., 2023). No obstante, persisten diferencias en los valores de las fracciones morfológicas en ciertos contextos, lo que indica la existencia de otros elementos que influyen en la dinámica y evolución de los cúmulos de galaxias.

---

<sup>1</sup>El pozo de potencial gravitacional y la región que separa las galaxias en un cúmulo consiste en un medio compuesto por una población dispersa e irregular de estrellas, junto con un gas caliente que se distribuye de manera aproximadamente homogénea (Carroll, 2014).

Aunque se ha investigado ampliamente la relación morfología-densidad, todavía hay incertidumbre con respecto a la solidez y la universalidad de esta relación. En específico, se plantea que puede fluctuar dependiendo de las características particulares de cada cúmulo, como la regularidad estructural y el nivel de evolución (Fasano et al., 2015; Cava et al., 2017; Sazonova et al., 2020). Los resultados logrados en este trabajo indican que esta tendencia no se cumple completamente, lo que contribuye a las ideas anteriores.

El objetivo de esta investigación es analizar la relación entre morfología y densidad en 11 cúmulos cercanos ( $z < 0.1$ ) mediante una clasificación automatizada basada en redes neuronales convolucionales (CNN, por sus siglas en inglés). El **Capítulo 2** proporciona un marco teórico y una revisión de trabajos anteriores sobre este tema. El **Capítulo 3** detalla la metodología utilizada, incluida la selección de datos y el desarrollo de la CNN. Los resultados se presentan en el **Capítulo 4** y describen las tendencias principales identificadas. Finalmente, el **Capítulo 5** integra las conclusiones de la investigación y plantea preguntas para futuras investigaciones.

En este trabajo se adopta una cosmología plana del modelo  $\Lambda$ CDM, con los siguientes parámetros cosmológicos (Wright, 2006):

- $H_0 = 69.6 \text{ km s}^{-1}\text{Mpc}^{-1}$
- $\Omega_m = 0.286$
- $\Omega_\Lambda = 0.714$
- $h = 0.696$

# CAPÍTULO 2

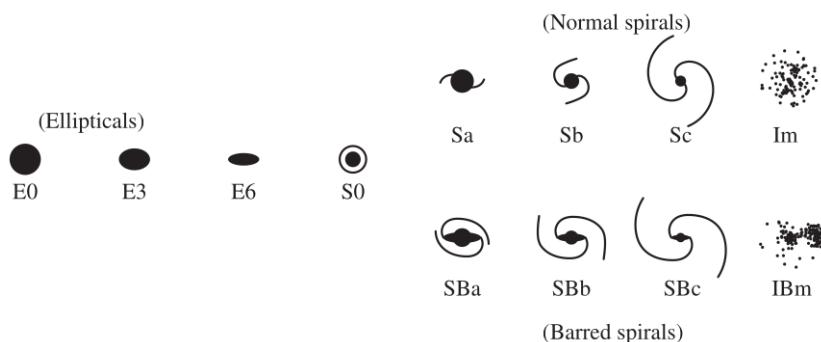
## ANTECEDENTES

### SECCIÓN 2.1

#### Morfología de las galaxias

##### 2.1.1. Clasificación de Hubble

En 1926, Edwin Hubble fue el primero en proponer una agrupación de las galaxias en función de su apariencia. Esta clasificación morfológica lleva el nombre de *diagrama Hubble* que divide las galaxias en elípticas (E's), lenticulares (S0), espirales (S's) e irregulares (Irr's), como se observa en la Figura 2.1. Al principio, Hubble pensó que el diagrama podía interpretarse como una secuencia evolutiva de las galaxias. Por lo tanto, denominó a las galaxias situadas a la izquierda del diagrama **tipos tempranos** (ETG, del inglés *Early Type Galaxies*) y a las situadas a la derecha **tipos tardíos** (LTG, del inglés *Late Type Galaxies*). Actualmente sabemos que la evolución se da de otra manera pero, por razones históricas, se sigue usando este nombre (Carroll, 2014).



**Figura 2.1:** Diagrama de la secuencia propuesta por Hubble que representa la morfología de las galaxias (Mo, Van den Bosch y White, 2010).

## 2.1.2. Galaxias elípticas

Las galaxias elípticas tienen distribuciones de luz suaves, y aparecen como elipses en las imágenes. Se designan con la letra  $E$ , seguida de un número entero  $n$  que representa su grado de *elipticidad aparente* en el cielo. Por convención,  $n$  es diez veces la elipticidad de la galaxia, redondeada al entero más próximo; se define como

$$e = 1 - \frac{b}{a} \quad (2.1)$$

Los parámetros  $a$  y  $b$  representan los ejes mayor y menor aparentes de la elipse, respectivamente. La elipticidad incrementa conforme se avanza de izquierda a derecha en el diagrama de Hubble, desde la casi circular  $E0$  hasta la más aplanada  $E7$  (Figura 2.2). Dado que en las elípticas predominan las estrellas viejas y están relativamente libres de polvo, su aspecto es muy similar a distintas longitudes de onda (Buta, 2013; Mo, Van den Bosch y White, 2010).

## 2.1.3. Galaxias S0 o lenticulares

En el centro del diagrama de Hubble, hay una clase intermedia de galaxias conocidas como lenticulares cuyo símbolo es  $S0$ . Estas galaxias están formadas por un bulbo central brillante, de aspecto similar al de una galaxia elíptica, rodeada de una estructura extendida en forma de disco (Figura 2.2). Las galaxias lenticulares no tienen una estructura espiral visible y no forman estrellas en cantidad significativa. El bulbo es a menudo la fuente dominante de luz en una galaxia lenticular (Buta, 2013; Mo, Van den Bosch y White, 2010).

## 2.1.4. Galaxias espirales

Según Buta (2013) y Mo, Van den Bosch y White (2010), las galaxias espirales están formadas por un disco aplanado, con estrellas que forman una estructura espiral (normalmente de dos brazos) y una concentración central de estrellas conocida como bulbo. También se observa en algunas espirales una estructura en forma de barra que se extiende desde el bulbo central, en cuyos extremos comienzan los brazos espirales (Figura 2.2). En el diagrama de Hubble, las espirales regulares ocupan la rama superior y se designan con la letra  $S$ , mientras que la rama inferior contiene las espirales barradas, cuyo símbolo es  $SB$ . Ambos tipos se dividen a su vez en tres clases:

- **Sa(SBa):** brazos lisos y compactos, con un bulbo central grande y brillante.
- **Sb(SBb):** brazos espirales más definidos que  $Sa(SBa)$ , con un bulbo más débil.
- **Sc(SBc):** brazos abiertos, claramente resueltos en cúmulos estelares y nebulosas individuales, con un bulbo más pequeño y tenue.

### 2.1.5. Galaxias irregulares

Las galaxias que no tienen una estructura regular (en forma de disco o elipsoidal), se denominan galaxias irregulares. Hubble definió dos clases:

- **Irr I:** tienen perfiles asimétricos y carecen de bulbo central o de una estructura espiral evidente; en su lugar, contienen muchos cúmulos individuales de estrellas jóvenes.
- **Irr II:** tienen un aspecto más suave y asimétrico, y no se resuelven claramente en estrellas individuales o cúmulos estelares.

Las irregulares se parecen a las espirales en que tienen estrellas jóvenes y viejas, así como polvo, gas atómico, molecular e ionizado, pero carecen de la estructura espiral que podría desencadenar la formación estelar (Buta, 2013; Mo, Van den Bosch y White, 2010).



**Figura 2.2:** Observaciones del Telescopio Espacial Hubble que muestran galaxias de distintos tipos morfológicos: E, S0, S, Irr (ordenadas de izquierda a derecha) (NASA Science, 2024).

#### SECCIÓN 2.2

### Cúmulos de galaxias

Las galaxias no se distribuyen aleatoriamente en el Universo, más bien, siguen una estructura a gran escala compuesta por filamentos y vacíos, moldeada por el potencial gravitacional. En las regiones donde convergen varios filamentos, se forman *grupos* y *cúmulos* (Bond, Kofman y Pogosyan, 1996).

Además de galaxias, los cúmulos se componen de gas caliente y materia oscura (DM, por sus siglas en inglés *Dark Matter*). Estos pueden contener de aproximadamente 50 galaxias (cúmulo **pobre**) a cientos de galaxias (cúmulo **rico**), en una región de espacio hasta  $6h^{-1}$  Mpc de diámetro. La dispersión de velocidades característica de un cúmulo es de  $800 \text{ km s}^{-1}$ .

Una masa típica de un cúmulo **virializado** (o en equilibrio dinámico) es de  $10^{14}\text{--}10^{15} M_{\odot}$  (Bower, 2013; Carroll, 2014). Este equilibrio está descrito por el **teorema del virial**, el cual establece que en un sistema ligado y estable bajo su propia gravedad, se cumple la relación

$$2\langle T \rangle + \langle U \rangle = 0 \quad (2.2)$$

donde  $\langle T \rangle$  es la energía cinética total promedio del sistema, asociada al movimiento de las galaxias dentro del cúmulo y  $\langle U \rangle$  corresponde a la energía potencial gravitacional promedio debida a la interacción entre galaxias, ICM y DM.

### 2.2.1. Cúmulos de Abell

En 1958, George Abell sistematizó por primera vez la detección de cúmulos ricos en galaxias como forma de investigar la distribución a gran escala de la materia. Seleccionó 1,682 cúmulos de galaxias del *Palomar Sky Survey* y fueron catalogados en base a dos criterios:

1. **Criterio de riqueza:** cada cúmulo debe contener al menos 50 galaxias miembros con *magnitudes aparentes*<sup>2</sup>  $m < m_3 + 2$ , donde  $m_3$  es la magnitud aparente del tercer miembro más brillante. La riqueza de un cúmulo se identifica como el número de galaxias miembros con magnitudes aparentes comprendidas entre  $m_3$  y  $m_3 + 2$ . Los cúmulos Abell ricos son aquellos con una riqueza superior a 50, aunque también registró cúmulos pobres con una riqueza entre 30 y 50.
2. **Criterio de compacidad:** únicamente se seleccionan como miembros las galaxias con distancias al centro del cúmulo inferiores a  $1.5h^{-1}$  Mpc (radio de Abell). Dado el criterio de riqueza, el criterio de compacidad actúa como un criterio de densidad.

Los cúmulos se consideran como **regulares** si su distribución de galaxias es más o menos circularmente simétrica y concentrada, en caso contrario como **irregulares** (Abell, 1958; Spinrad, 2005; Mo, Van den Bosch y White, 2010).

SECCIÓN 2.3

## Emisión de rayos X en cúmulos de galaxias

Aunque generalmente se consideran a los cúmulos como conjuntos de galaxias, en realidad, estos no son el elemento bariónico principal. Una porción significativamente mayor de los bariones dentro del sistema está vinculada a un gas que se encuentra atrapado en el *pozo de potencial* del cúmulo, lo que lleva a que se comprima y, en consecuencia, se caliente hasta conseguir temperaturas de  $10^7 - 10^8$  K, mientras que la densidad en

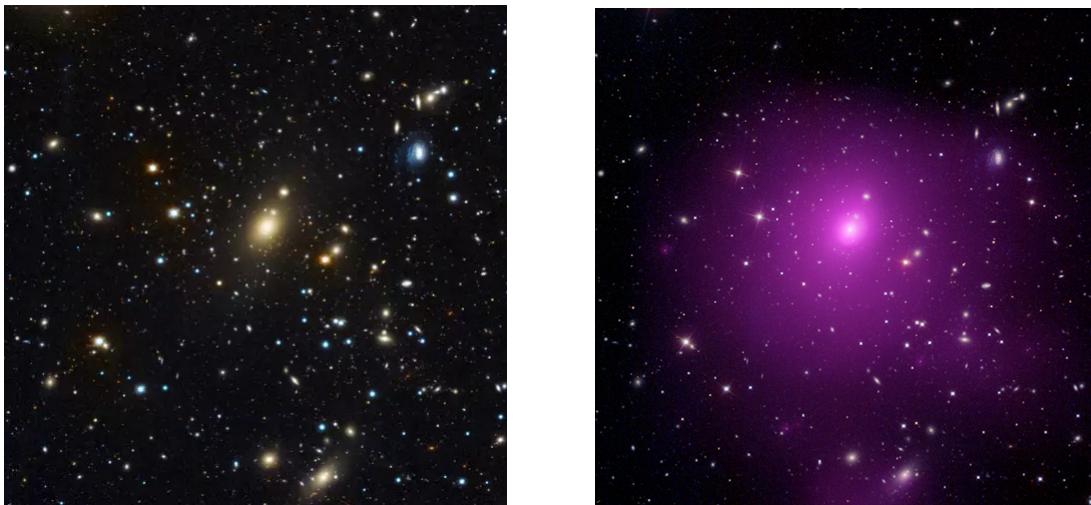
<sup>2</sup>Es la medida del brillo de una estrella observado desde la Tierra. Se mide en escala logarítmica donde los valores más bajos corresponden a objetos más brillantes. Históricamente, las estrellas más luminosas tenían  $m = 1$ , mientras que las más débiles  $m = 6$  (Carroll, 2014).

el núcleo alcanza entre 0.01 y 0.1 átomos por  $\text{cm}^{-3}$ . A esas temperaturas, el gas se encuentra altamente ionizado, lo que provoca la emisión de **rayos X** a través del proceso de *bremsstrahlung* (también conocido como emisión libre o radiación de frenado). El espectro de rayos X nos proporciona información sobre la profundidad del pozo potencial de un cúmulo, y las intensidades de las líneas de emisión revelan las abundancias de elementos como hierro, oxígeno y silicio en el ICM (Voit, 2005; Spinrad, 2005; Bower, 2013).

Como explica Bower (2013), conceptualmente, el *bremsstrahlung* surge cuando las trayectorias de los electrones son desviadas por los campos fuertes de los núcleos ionizados, incluyendo  $H^+$  (hidrógeno) y  $He^{++}$  (helio). La aceleración del electrón genera la emisión de un fotón. El cálculo del espectro real implica la integración sobre los parámetros de impacto y la distribución térmica de las velocidades de los electrones. Para un ion con carga  $Z$  y un plasma con temperatura de electrones  $T_e$  y densidades de electrones e iones  $n_e$  y  $n_i$ , respectivamente, la emisividad (por unidad de volumen por unidad de frecuencia) viene dada por:

$$\epsilon_Z(\nu) = A_Z n_e n_i T_e^{-1/2} e^{(-h\nu/k_B T_e)} \quad (2.3)$$

donde  $K_B$  es la constante de Boltzmann,  $h$  la constante de Plank y  $A_Z$  es una constante de normalización. Para más detalles, revisar el trabajo de Sarazin (1988). Así, la ecuación 2.3 indica que la emisión está influenciada por la densidad del ICM, donde se encuentra una mayor concentración de materia, generalmente en cúmulos ricos. Considerando que la temperatura del gas es del orden de  $10^7 - 10^8 \text{ K}$ , la emisión se encuentra en el rango de rayos X blandos ( $\sim 0.5 - 10 \text{ keV}$ ). Esto los hace sensibles a telescopios como Chandra y XMM-Newton (Böhringer y Werner, 2010), por ejemplo, el cúmulo de la Figura 2.3.



**Figura 2.3:** El cúmulo Abell 85 visto en longitud de onda óptica (izquierda) y en rayos X (derecha) (Max Planck Society, 2019; Chandra X-ray Observatory, 2008).

## Evolución de galaxias en cúmulos

### 2.4.1. Procesos ambientales de transformación morfológica

Existen múltiples mecanismos que se han sugerido para entender cómo evolucionan las galaxias en áreas de alta densidad. Estas explicaciones se dividen en dos grupos: las **perturbaciones gravitacionales**, que abarcan el acoso galáctico (*galaxy harassment*), interacciones entre galaxias y las interacciones entre cúmulos y galaxias donde intervienen gas, estrellas y DM; y los **procesos hidrodinámicos**, que incluyen la evaporación térmica y el *Ram Pressure Stripping* (RPS), que impactan de manera directa el gas que se encuentra en las galaxias y en el cúmulo (Lourenço et al., 2023). Entre los fenómenos físicos mencionados, se ha demostrado que el RPS es el que más influencia tiene en la evolución de las galaxias en los cúmulos (Boselli, Fossati y Sun, 2022), ya que el historial de formación estelar de una galaxia depende en gran medida del gas disponible: cualquier proceso que elimine, agregue o incluso perturbe el gas determina la evolución y el futuro de una galaxia, especialmente en lo referente a su contenido estelar (Pérez-Millán et al., 2023).

Es conocido que las galaxias están en un movimiento continuo dentro del ICM y son influenciadas por fuerzas hidrodinámicas. Específicamente, cuando una galaxia tiene un disco de gas frío, necesita desplazar el ICM para poder pasar a través de él. Si la atracción gravitatoria que mantiene el gas en el plano galáctico no es lo suficientemente fuerte, el gas se desprenderá. Esta condición, conocida como RPS, puede expresarse de la siguiente manera:

$$\rho_{ICM} V^2 > 2\pi G \Sigma_*(r) \Sigma_g(r) \quad (2.4)$$

donde  $\rho_{ICM}$  es la densidad del gas externo difuso en la velocidad orbital de la galaxia  $V$ ,  $\Sigma_*(r)$  es la densidad superficial estelar del disco de las galaxias en el radio  $r$ , y  $\Sigma_g(r)$  la densidad superficial del gas en este radio. Si se cumple la condición, el gas se desprenderá de este radio. Si la galaxia se mueve a una velocidad alta, todo el gas de la galaxia será eliminado. A menor velocidad únicamente se perderá el material de las partes exteriores. El mecanismo de RPS ofrece una posible explicación de la ausencia de galaxias espirales típicas en ciertos cúmulos. Sin embargo, el proceso se evidencia cuando la galaxia se mueve rápidamente y el ICM tiene una densidad adecuada (Bower, 2013).

## 2.4.2. El rol de la masa en la evolución morfológica

De igual manera, una característica física que parece tener un impacto importante en la evolución es la masa de las galaxias. Thomas (2011) analizó 3360 galaxias de tipo ETG, revisando sus edades y niveles de *metallicidad*<sup>3</sup> y la proporción de  $\alpha/Fe$ <sup>4</sup>, además de cómo estas características variaban al ser clasificadas por su masa y el ambiente que las rodea. Descubrió que las galaxias con mayor masa tienden a ser más antiguas y a exhibir mayor metalicidad, sin que el entorno juegue un papel importante, mientras que las galaxias de menor masa presentan una reciente actividad de formación estelar en áreas de baja densidad. Esto implica que la masa de las galaxias es el factor principal que influye en su evolución estelar, más que la densidad del entorno en el que se sitúan.

Estos hallazgos coinciden con investigaciones recientes que indican que la extinción de la formación estelar y otros procesos evolutivos están más vinculados a la masa que al entorno. Por ejemplo, Annunziatella et al. (2023) revisaron 323 galaxias masivas, comparando las edades de estas galaxias con su masa (a través de la luminosidad y modelos de evolución) y su entorno. Sus resultados revelan una relación bastante débil entre el entorno y las etapas iniciales de formación de las galaxias masivas que examinaron, concluyendo que la edad de las galaxias es más afectada por su masa que por el ambiente que las rodea.

Ambos estudios revelan que la evolución de las galaxias está principalmente determinada por su masa: cuanto más masivas son, más antiguas y ricas en metales, sin que el entorno sea un factor determinante.

### SECCIÓN 2.5

## Relación morfología-densidad en cúmulos de galaxias

En 1980, Dressler estudió 55 cúmulos en el Universo Local y estableció un vínculo entre la densidad del entorno y la morfología de las galaxias que los ocupan, dando origen así a la conocida *relación morfología-densidad*. Sus hallazgos indicaron que, a medida que la densidad del entorno aumenta, la fracción de galaxias elípticas crece al mismo tiempo que disminuye la de las espirales, un patrón que se ha corroborado en numerosos estudios posteriores. Esta tendencia se ilustra en la Tabla 2.1, Tabla 2.2 y en la Figura 2.4.

<sup>3</sup>Dado que el hidrógeno y el helio son los elementos primarios en la mayoría de las estrellas, los elementos más pesados como el carbono, el oxígeno y el hierro, se conocen como metales. Por lo tanto, la fracción de estos elementos se denomina metalicidad (Carroll, 2014).

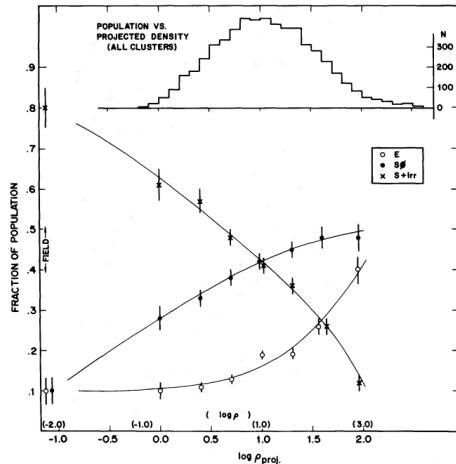
<sup>4</sup>Representa la razón entre la abundancia de los elementos alfa (N, O, Mg, Ca, Na, Ne, S, Si, Ti) y los elementos del pico del hierro (Cr, Mn, Fe, Co, Ni, Cu, Zn), que aporta información sobre la escala temporal de formación de las poblaciones estelares (Thomas et al., 2003).

Elípticas	41%
S0	48%
Espirales	11%
Irregulares	-

**Tabla 2.1:** Fracción de galaxias observadas por Dressler en un entorno de alta densidad, es decir, al centro del cúmulo (Spinrad, 2005).

Elípticas	13%
S0	21%
Espirales	61%
Irregulares	3%

**Tabla 2.2:** Fracción de galaxias observadas por Dressler en un entorno de baja densidad o a las afueras del cúmulo (Spinrad, 2005).



**Figura 2.4:** La fracción de galaxias E, SO y S+I en función del logaritmo de la densidad proyectada. El histograma superior muestra la distribución del número de galaxias en los intervalos de densidad proyectada (Dressler, 1980).

En el Universo Local ( $z < 0.1$ ), la relación morfología-densidad se hace clara al notar que en áreas más densas predominan las galaxias ETG, mientras que en zonas de menor densidad, las galaxias LTG forman la mayoría (Goto et al., 2003; Houghton, 2015). Otros estudios como Nuijten et al. (2005) y Pfeffer et al. (2023) han demostrado que existe una influencia del entorno en la evolución de espirales a elípticas. Además, esta relación ha sido detectada en altos *redshifts* ( $z > 0.5$ ), lo que sugiere que los procesos evolutivos de las galaxias han estado activas aún en cúmulos distantes (Desai et al., 2007; Mei et al., 2023).

Aunque se ha discutido si la relación morfología-densidad debe ser considerada un principio universal en la astronomía extragaláctica, es un hecho que existen excepciones fundamentadas en varias investigaciones. Se han notado variaciones en esta relación derivadas de factores tales como la evolución cósmica, interacciones entre galaxias, masa estelar, dinámica de los cúmulos y características intrínsecas de las galaxias. Por ejemplo, a *redshifts* intermedios ( $z \sim 0.5$ ) se manifiesta el **efecto Butcher-Oemler**, caracterizado por un exceso de galaxias LTG en comparación con lo observado a *redshifts* menores (Butcher y Oemler, 1978; Kodama y Bower, 2001). Asimismo, en cúmulos irregulares o con subestructuras, la correlación entre morfología y densidad no es evidente, lo que sugiere que solo se establece en cúmulos más evolucionados dinámicamente (Dressler et al., 1997; Fasano et al., 2015; Cava et al., 2017). Estudios en *redshifts* altos ( $z \geq 1$ ) revelan que en ciertos cúmulos la proporción de galaxias ETG es menor y no se observa la relación morfología-densidad, lo que indica que evoluciona con el tiempo (Postman et al., 2005; Sazonova et al., 2020).

En el contexto del modelo evolutivo de galaxias ampliamente aceptado ( $S \rightarrow S0 \rightarrow E$ ), se han propuesto mecanismos que pueden explicar el proceso, entre ellos, la relación morfología-densidad. Sin embargo, investigaciones como Park y Hwang (2009), Fasano et al. (2015) y Sazonova et al. (2020) señalan que, aunque el entorno juega un papel importante, no es el único determinante. Otros mecanismos, como las interacciones hidrodinámicas entre galaxias, las fuerzas gravitacionales dentro del cúmulo, la masa de la galaxia y la tasa de formación estelar también desempeñan un papel esencial en la evolución, independientemente del medio en el que se encuentren.

#### SECCIÓN 2.6

## Predicción de la morfología de las galaxias con CNNs

Las redes neuronales artificiales, también conocidas como ANN en inglés, son sistemas computacionales que se basan en la manera en que operan los sistemas nerviosos biológicos, como el cerebro de humanos o animales. Estas redes consisten en muchos nodos conectados entre sí, llamados **neuronas**, que colaboran para aprender de la información que reciben y mejorar sus salidas finales.

Para facilitar este proceso de aprendizaje, las ANN tienen una organización que se compone de diversas capas. Los datos, que suelen presentarse como un vector en múltiples dimensiones, ingresa por la *capa de entrada* y se transmiten hacia las *capas ocultas*. Estas capas se encargan de procesar la información determinando cómo diferentes variaciones pueden afectar el resultado final, lo cual es esencial para el aprendizaje de la red. Cuando una ANN posee múltiples capas ocultas apiladas, se le denomina **aprendizaje profundo**.

Según la cantidad de etiquetas en los datos utilizados para entrenar, las ANN pueden aplicar diferentes métodos de aprendizaje. En el **aprendizaje supervisado**, la red hace uso de ejemplos que ya están etiquetados, donde cada conjunto de datos de entrada tiene un resultado predefinido. La meta es reducir el error de clasificación modificando los parámetros del modelo. Por otro lado, en el **aprendizaje no supervisado**, la red opera con datos que carecen de etiquetas y se enfoca en detectar patrones o estructuras ocultas mediante la optimización de una función de coste (O'Shea y Nash, 2015).

### 2.6.1. Fundamentos y arquitectura

Las redes neuronales convolucionales conocidas como CNN se asemejan a las ANN convencionales ya que se componen de neuronas que mejoran su rendimiento a través del aprendizaje. Cada una de estas neuronas continúa recibiendo datos y realizando un proceso denominado **convolución** (O'Shea y Nash, 2015). En matemáticas, la convolución entre dos funciones  $f$  y  $g$ , está definida como:

$$(f * g)(x) = \int f(z)g(x - z)dz \quad (2.5)$$

Es decir, medimos el solapamiento entre  $f$  y  $g$  cuando una función se invierte y se desplaza por  $x$ . Siempre que tengamos objetos discretos, la integral se convierte en una suma:

$$(f * g)(i) = \sum_a f(a)g(i - a) \quad (2.6)$$

Para los tensores bidimensionales, la operación de convolución se expresa como una suma con índices  $(a, b)$  para  $f$  y  $(i - a, j - b)$  para  $g$ , respectivamente:

$$(f * g)(i, j) = \sum_a \sum_b f(a, b)g(i - a, j - b) \quad (2.7)$$

En terminología de las redes convolucionales, el primer parámetro de esta operación, identificado como la función  $f$ , es lo que se denomina **entrada**, mientras que el segundo parámetro,  $g$ , es conocido como **núcleo** o *kernel*. El resultado de la convolución proporciona una nueva forma de la entrada, que se llama **mapa de características**, donde se reconocen patrones locales en los datos (Goodfellow, Bengio y Courville, 2016; Zhang et al., 2023).

Dado que la convolución es fundamental en el procesamiento en una CNN, las arquitecturas típicas de estas redes incluyen capas alternas de convolución y agrupación (*pooling*), seguidas de una o más capas totalmente conectadas (*fully-connected*). Cada capa de una CNN recibe la entrada en un formato tridimensional  $m \times m \times r$ , donde  $m$  se refiere a la altura y el ancho de la imagen, y  $r$  indica la profundidad o número de canales. Por ejemplo, una imagen en formato RGB tiene una profundidad de  $r = 3$ . Los *kernel*s o filtros en una capa convolucional, representados como  $k$ , también poseen tres dimensiones  $(n \times n \times q)$ , con  $n$  menor que  $m$  de la entrada y  $q$  igual o menor que el número de canales  $r$ , actuando como pesos en el proceso de convolución (Alzubaidi et al., 2021).

Los componentes principales de la arquitectura básica (Figura 2.7) de una CNN en el contexto de clasificación de imágenes son los siguientes (Khan et al., 2020; Alzubaidi et al., 2021):

- **Capa de convolución.** Está formada por un conjunto de filtros, conocidos como *kernels*, que actúan sobre la imagen de entrada que identifican patrones específicos al dividir la imagen en partes pequeñas, llamadas *campos receptivos*. Cuando un *kernel* se mueve a lo largo de la imagen, lleva a cabo una operación de convolución, multiplicando sus valores con los del campo receptivo correspondiente y sumando los resultados para producir un único valor en el mapa de características.

El movimiento del *kernel* por la imagen está determinado por un hiperparámetro llamado ***stride***, que especifica cuántos píxeles se desplaza el *kernel* en la imagen; un valor alto produce una salida con menor resolución y reduce la cantidad de datos procesados, mientras que un valor más bajo posibilita un análisis más preciso de la información espacial (Figura 2.5). Asimismo, para gestionar el tamaño de la salida y prevenir la pérdida de datos en los bordes, se introduce el concepto de ***padding*** que consiste en añadir valores (normalmente ceros) alrededor de la imagen de entrada, lo que permite que el *kernel* pueda trabajar con los píxeles en los bordes, contribuyendo a mantener el tamaño de la salida lo más parecido posible al de la entrada (Figura 2.6).

Input	Kernel	Output
$  \begin{bmatrix}  0 & 0 & 0 & 0 & 0 \\  0 & 0 & 1 & 2 & 0 \\  0 & 3 & 4 & 5 & 0 \\  0 & 6 & 7 & 8 & 0 \\  0 & 0 & 0 & 0 & 0  \end{bmatrix}  $	$*$	$  \begin{bmatrix}  0 & 1 \\  2 & 3  \end{bmatrix}  = \begin{bmatrix}  0 & 8 \\  6 & 8  \end{bmatrix}  $

**Figura 2.5:** Operación de convolución con un *stride* de 3 verticalmente y 2 horizontalmente (Zhang et al., 2023).

Input	Kernel	Output
$  \begin{bmatrix}  0 & 0 & 0 & 0 & 0 \\  0 & 0 & 1 & 2 & 0 \\  0 & 3 & 4 & 5 & 0 \\  0 & 6 & 7 & 8 & 0 \\  0 & 0 & 0 & 0 & 0  \end{bmatrix}  $	$*$	$  \begin{bmatrix}  0 & 1 \\  2 & 3  \end{bmatrix}  = \begin{bmatrix}  0 & 3 & 8 & 4 \\  9 & 19 & 25 & 10 \\  21 & 37 & 43 & 16 \\  6 & 7 & 8 & 0  \end{bmatrix}  $

**Figura 2.6:** Operación de convolución con *padding* (Zhang et al., 2023).

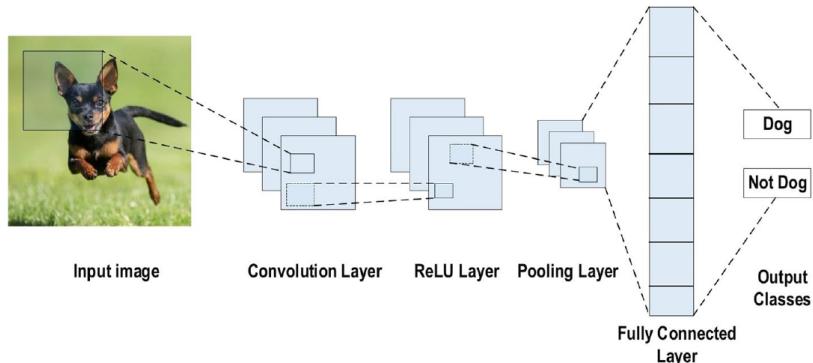
- **Capa de agrupación.** Es un proceso de submuestreo que disminuye las dimensiones de los mapas de características creados después de la convolución, manteniendo los datos más significativos. Utiliza un *kernel* y un *stride* definidos para realizar el submuestreo, lo que ayuda a optimizar la capacidad de generalización del modelo al conservar las características más importantes a pesar de ligeras variaciones en la información.
- **Función de activación.** Se trata de una función no lineal que se aplica luego de las capas que poseen parámetros ajustables (como las convolucionales) para insertar no linealidad en la transformación de entrada a salida. Facilita que la red reconozca patrones complejos y características clave. La selección de esta función impacta en la efectividad del aprendizaje y el rendimiento del modelo. En este contexto, las funciones más comunes son:
  - ReLU. Convierte los valores enteros de la entrada en números positivos. La principal ventaja de ReLU sobre los demás es el menor gasto computacional. Su representación matemática es:

$$f(x)_{ReLU} = \max(0, x) \quad (2.8)$$

- Sigmoide. La entrada de esta función de activación son números reales, mientras que la salida está restringida entre cero y uno. La curva tiene forma de S y puede representarse matemáticamente por:

$$f(x)_{sigm} = \frac{1}{1 + e^{-x}} \quad (2.9)$$

- **Capa totalmente conectada.** Es una capa en la que cada neurona está enlazada a todas las neuronas de la capa anterior. Generalmente, está situada al final de la CNN y actúa como un clasificador, fusionando de forma no lineal las características extraídas para producir la salida final del modelo.



**Figura 2.7:** Arquitectura de una CNN utilizada para identificar imágenes de perros (Alzubaidi et al., 2021).

## 2.6.2. Entrenamiento y optimización

En la capa de salida también se aplica una **función de pérdida** para evaluar la diferencia entre las predicciones del modelo y los valores reales. Esta función requiere dos entradas: la predicción generada por la red y la etiqueta correcta de los datos. Su propósito es medir el error y reducirlo durante el proceso de entrenamiento.

Como mencionan Alzubaidi et al. (2021), la elección de la función de pérdida depende del tipo de problema que se quiere abordar. En tareas de clasificación, una elección frecuente es la entropía cruzada (*cross-entropy*), que se utiliza junto con la función de activación *softmax*. Esta se encarga de convertir las salidas de la última capa en valores de probabilidad, garantizando que la suma total de todas las salidas sea igual a 1. La función *softmax* se expresa como

$$p_i = \frac{e^{a_i}}{\sum_{k=1}^N e^a_k} \quad (2.10)$$

donde  $e^{a_i}$  es la salida no normalizada de la red para la clase  $i$ , y  $N$  es el número total de clases en la capa de salida. La función de pérdida de entropía cruzada se expresa como:

$$H(p, y) = - \sum_i y_i \log(p_i), \quad i \in [1, N] \quad (2.11)$$

$y_i$  es el valor real (etiqueta) para la clase  $i$  y  $p_i$  es la probabilidad predicha para dicha clase.

Durante el proceso de entrenamiento, la CNN ajusta sus parámetros (incluyendo pesos y sesgos) para minimizar la función de pérdida. Para lograr esto, se emplean algoritmos de optimización que se basan en gradientes, los cuales ajustan los parámetros a lo largo de varias **épocas** (ciclos completos del conjunto de entrenamiento). La magnitud de cada modificación está determinada por la **tasa de aprendizaje**, que establece el tamaño del ajuste en cada iteración.

Uno de los algoritmos más utilizados es el **descenso de gradiente** (*Gradient Descent*), que actualiza los parámetros en la dirección contraria al gradiente de la función de pérdida. Este procedimiento se realiza a través de *retropropagación*, la cual transmite los errores desde la salida hacia las capas anteriores para lograr un ajuste eficiente de los pesos.

Entre las variantes del descenso de gradiente, una de las más populares es **Adam** (*Adaptive Moment Estimation*) que combina las ventajas de dos métodos (momentum y RMSprop) proporcionando una optimización efectiva con menor uso de memoria y un mejor desempeño en redes profundas.

### 2.6.3. Aplicaciones en Astronomía

Las CNNs han transformado la manera en que los astrónomos categorizan las galaxias, simplificando el estudio de grandes volúmenes de imágenes con una notable precisión. Por ejemplo, el trabajo de Barchi et al. (2020) evidenció que estos sistemas pueden reconocer galaxias elípticas y espirales con una exactitud del 99 %, superando las técnicas tradicionales. De forma similar, Cheng et al. (2021) crearon un catálogo que contiene más de 20 millones de galaxias clasificadas, utilizando información del *Dark Energy Survey*, con resultados igualmente precisos para las galaxias brillantes.

Sin embargo, la capacidad de estas redes va más allá de una simple clasificación binaria (elípticas o espirales). Estudios como en el de Domínguez Sánchez et al. (2018), desarrollaron una robusta CNN que clasifica galaxias de acuerdo con la secuencia de Hubble y el esquema *Galaxy Zoo 2*, llegando a una precisión mayor al 97 %. Más recientemente, Urechiatu y Frincu (2024) propusieron una arquitectura personalizada capaz de clasificar galaxias en cinco tipos distintos, superando modelos populares como *ResNet50* y *EfficientNet*, con un 96.83 % de precisión.

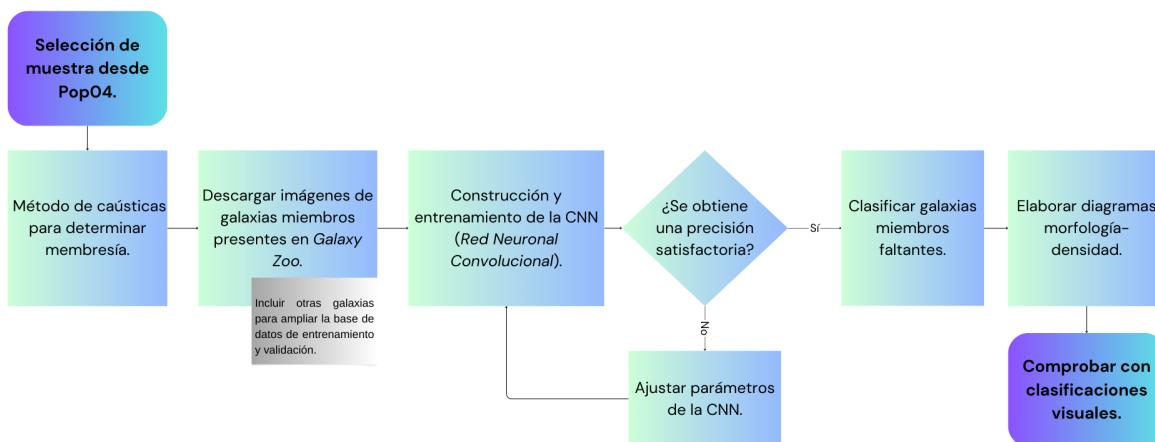
La principal ventaja de las CNNs es su habilidad para realizar el trabajo pesado de la clasificación de grandes conjuntos, disminuyendo notablemente el esfuerzo de la clasificación visual. A pesar de su considerable efectividad, estas redes aún enfrentan ciertos obstáculos: dependen de datos etiquetados por humanos, lo que puede traducirse a sesgos que afectan su precisión, especialmente en galaxias más débiles o distantes (Cheng et al., 2021). Asimismo, el desbalance en los conjuntos de datos sigue siendo un reto, especialmente al intentar clasificar estructuras menos frecuentes, como fusiones o tipos intermedios (Reza, 2021).

# CAPÍTULO 3

## METODOLOGÍA

Después de la revisión del estado del arte, esta investigación utiliza un método cuantitativo y computacional, fundamentado en el estudio de un conjunto de 11 cúmulos, elegidos del catálogo de Popesso et al. (2004) (en adelante Pop04), datos observacionales y técnicas de aprendizaje automático para identificar galaxias mediante CNNs, lo que permitió obtener información significativa para comprobar si hay o no una relación entre la forma de las galaxias y la densidad del entorno en el que se encuentran.

Se sugiere que las galaxias ubicadas en regiones de alta densidad son un 90% más susceptibles a poseer una forma de tipo ETG (elíptica), en comparación con aquellas que se hallan en áreas de baja densidad, donde predomina la forma LTG (espiral), tal como Dressler (1980). El proceso utilizado para confirmar esta hipótesis se ilustra en la Figura 3.1.



**Figura 3.1:** Diagrama representativo de las etapas seguidas a lo largo del estudio.

Esta investigación abarca una serie de etapas diseñadas para verificar la relación morfología-densidad en los cúmulos seleccionados. En este sentido, el procedimiento incluye tanto el desarrollo de un clasificador de galaxias basado en CNNs como un análisis de los diagramas morfología-densidad con el fin estudiar la distribución morfológica según la densidad local de cada cúmulo.

Para abordar el análisis propuesto, es fundamental detallar claramente las etapas que conforman el estudio. Comenzamos con la *selección de la muestra*, donde se presenta el catálogo de Pop04, sus hallazgos y los criterios relevantes para dicha selección. Posteriormente, se describe la implementación del *estudio de membresía* y su utilidad en el contexto del presente análisis. Luego, se revisan y descargan las imágenes de las galaxias ya clasificadas en *Galaxy Zoo 2* (GZ) cuyo propósito principal es formar la base de datos utilizada en el siguiente paso. En la *construcción de la CNN*, se explica su estructura y cómo se adapta a nuestros fines. Seguidamente, se presenta la clasificación de *galaxias miembro faltantes* a través de la CNN entrenada. Finalmente, se presentan los *diagramas morfología-densidad* y se lleva a cabo una comprobación mediante *clasificaciones visuales* para verificar la precisión de nuestros resultados.

### SECCIÓN 3.1

## Selección de la muestra

Pop04 integra en su catálogo dos de los más significativos sondeos disponibles: el **ROSAT All-Sky Survey** (Voges et al. (1999), RASS) y el **Sloan Digital Sky Survey** (York et al. (2000), SDSS). Este conjunto de información es fundamental para investigar las características de los cúmulos de galaxias, ya que relaciona las observaciones ópticas con el ICM en rayos X.

El descubrimiento más importante es la conexión entre la luminosidad en rayos X y la temperatura del ICM con la luminosidad óptica, en particular en las bandas *i* y *z*. Esta conexión es crucial para estimar la masa de un cúmulo a partir de sus características en rayos X, siendo este un enfoque más sólido en comparación con la identificación puramente óptica. Así, empleando las correlaciones obtenidas de la banda *z*, demostraron que, a partir de las propiedades ópticas de un cúmulo, es posible predecir la luminosidad en rayos X y la temperatura con un margen de precisión del 60%, y viceversa. De este modo, la información disponible en este catálogo incluye tanto observaciones ópticas como de rayos X, ofreciendo una perspectiva más integral de las propiedades de los cúmulos, lo que resulta esencial para analizar la morfología y la densidad.

Se seleccionaron 11 cúmulos de galaxias del catálogo de Pop04, incluidos también en el catálogo de Abell (1958), teniendo en cuenta los siguientes parámetros:

- *Redshift* en el rango  $0.02 \leq z \leq 0.095$
- Radio mínimo de búsqueda  $r_{min} = 2.5$  Mpc

El valor de  $z$  se eligió para trabajar con galaxias cercanas, lo cual facilita la observación de su morfología en las imágenes que se emplearán posteriormente. Por otro lado,  $r_{min}$  corresponde al radio desde el centro del cúmulo para la búsqueda de galaxias. Los cúmulos generalmente tienen radios entre 2 y 3 Mpc (Bahcall, 1999), entonces, con este radio aproximado se asegura que no se pierdan galaxias antes del estudio de membresía.

SECCIÓN 3.2

## Estudio de membresía

Para identificar las galaxias pertenecientes a cada cúmulo, se utilizó el **método de cáusticas**, que permite calcular la masa de los cúmulos a través de la evaluación de la velocidad de escape de las galaxias que los componen (Diaferio, 1999; Serra et al., 2011; Gifford et al., 2013).

El método consiste en identificar el límite o **cáustica** donde las galaxias están unidas por gravedad, lo cual se determina mediante la conexión entre la velocidad de escape  $v_{esc}(r)$  y el potencial gravitacional  $\Phi(r)$  del cúmulo, la cual se define de la siguiente manera:

$$v_{esc}^2(r) = -2\Phi(r) \quad (3.1)$$

Esta ecuación implica que las galaxias unidas al pozo de potencial del cúmulo existen dentro de una región claramente delimitada del **espacio fase  $\mathbf{r} - \mathbf{v}$** , donde  $r$  es el radio físico o proyectado desde el centro del cúmulo y  $v$  es la velocidad tridimensional o proyectada, respecto al movimiento del cúmulo.

El algoritmo empleado para llevar a cabo el estudio de membresía fue desarrollado por Gifford et al. (2013)<sup>5</sup>. El gráfico de la cáustica se construye a partir de las posiciones (`dproj`) y velocidades (`vlos`) proyectadas de cada galaxia. También, se usa  $r_{200}$  un radio en el que la densidad media del cúmulo equivale a 200 veces la densidad crítica del Universo, que se define como

$$\rho_c = \frac{3H_0^2}{8\pi G} \quad (3.2)$$

De esta manera, se delimita la región donde se estima que las galaxias están ligadas gravitacionalmente. Así mismo, se calcula la densidad del espacio fase a través de un kernel gaussiano 2D. Posteriormente, se ajusta una curva suave utilizando `smooth.spline()` para generar la superficie de escape de la cáustica.

Adicionalmente, encontramos un ajuste del *perfil NFW* (*Navarro-Frenk-White*) utilizada en los cálculos de la masa del cúmulo. Para nuestros fines, la información relevante del método son los miembros o `members` resultantes dentro de la cáustica (como se aprecia en las Figuras 4.4-4.6).

<sup>5</sup>Consultar código escrito en Python: <https://github.com/giffordw/CausticMass>

## Base de datos con Galaxy Zoo

Para este trabajo, se utilizó una base de datos construida a partir de imágenes provenientes del SDSS y, a su vez, de sus clasificaciones proporcionadas en el proyecto Galaxy Zoo 2 (GZ)<sup>6</sup>, que está orientado a la clasificación morfológica de galaxias, donde las imágenes son evaluadas por voluntarios de diversas partes del mundo, quienes asignan categorías siguiendo un árbol de decisiones jerárquico. Según Willett et al., 2013, las clasificaciones permiten distinguir principalmente entre elípticas y espirales, de acuerdo a la respuesta más votada.

Las galaxias clasificadas como elípticas tienen etiquetas *E*, seguidas de una letra que representa su forma aparente: *r* (completamente redonda), *i* (intermedia) y *c* (alargada como cigarro). Las galaxias con estructuras en disco tienen etiquetas *S*; si contienen una barra, la letra siguiente es una *B* mayúscula. La prominencia del bulbo se codifica con letras: *d* (sin bulbo), *c* (débil), *b* (marcado), *a* (dominante). En presencia de brazos espirales, se incluye un número que indica cuántos hay (1, 2, 3, 4, +, ?), seguido por una letra que describe su grado de apertura: *t* (cerrados), *m* (intermedios), *l* (abiertos).

En el contexto de esta investigación, las clasificaciones realizadas en GZ son una fuente clave para entrenar la CNN ya que se trata de un conjunto de datos etiquetados de forma precisa y confiable cuya clase puede identificarse claramente. Aunque las clasificaciones visuales pueden contener ciertas variaciones subjetivas, en GZ se han implementado métodos de consenso para tener una buena precisión.

El catálogo originalmente consiste de una tabla csv de **239,695** filas por **231** columnas (véase Figura 3.2) que brindan datos sobre las clasificaciones de los voluntarios en las clases propuestas, además de algunas características como el número de brazos espirales, la existencia de bulbo o barra, la redondez y suavidad de la elipse, entre otras.

dr7objid	ra	dec	rastring	decstring	sample	gz2_class	total_classifications	total_votes	t01_smooth_or_features_a01_smooth_count
1	587732591714893851	179.04298	60.52252	11:56:10.32	+60:31:21.1	original	Sc*t	45	342
2	588009360545984617	135.0844	52.49424	09:00:20.26	+52:29:39.3	original	Sb*t	42	332
3	587732484359913515	183.37198	50.74151	12:13:29.27	+50:44:29.4	original	Ei	36	125
4	587741723357282317	186.35195	28.5586	12:25:00.47	+28:33:31.0	original	Sc*t	28	218
5	587738410866965777	161.0864	14.08447	10:44:20.73	+14:05:04.1	original	Er	43	151

**Figura 3.2:** Algunas columnas del catálogo original de Galaxy Zoo 2. Vista de TOPCAT.

<sup>6</sup>Página oficial: <https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/>

Teniendo en cuenta la cantidad de estos datos, se utilizó la terminal de Linux, programas en `bash` y `TOPCAT` para limpiar y organizar la información relevante de las galaxias de cada cúmulo. También se filtró la columna `g2_class`, a partir de las cadenas de caracteres, para tener únicamente las galaxias elípticas y espirales. El catálogo depurado con los datos necesarios (`id, ra, dec, g2_class`) tomó la siguiente forma:

	dr7objid	ra	dec	gz2_class
1	587732591714893851	179.04298	60.52252	Sc+t
2	588009368945984617	135.0844	52.49424	Sb+t
3	587732484359913515	183.37198	50.74151	Ei
4	587741723357282317	186.25195	28.5586	Sc+t
5	587738410866966577	161.0864	14.08447	Er

**Figura 3.3:** Catálogo modificado de *Galaxy Zoo 2* con las columnas de interés. Vista de *TOPCAT*.

Esto facilitó la ejecución del siguiente código escrito en `bash` que permitió una conexión con el servidor público de SDSS *Data Release 7* que nos proporcionó una cantidad adecuada de imágenes para la base de datos <sup>7</sup>.

```

1 #Obtener imágenes de SDSS
2 #Modificar los parámetros necesarios
3 #(ancho, alto, escala)
4 file=lis.dat
5 while read -r id ra dec class;
6 do
7
8 echo $id $ra $dec $class
9 wget "http://skyservice.pha.jhu.edu/DR7/ImgCutout/getjpeg.aspx?ra=$ra&
  dec=$dec&scale=0.15&width=150&height=150&opt=&query=" ; mv getjpeg.
  aspx\?ra\=$ra\&dec\=$dec\&scale\=0.15\&width\=150\&height\=150\&opt
  \=\&query\= $class$id.jpg
10
11 done < "$file"
```

El código obtiene imágenes en formato *jpg* desde el servidor SDSS (ver ejemplos en Figura 3.4). Procesa una lista llamada `lis.dat` que incluye las coordenadas (`ra, dec`) y la categoría (`class`) de las galaxias. En este caso, se trata de galaxias de la muestra que coinciden con las clasificadas en GZ. Los parámetros ajustables son los siguientes:

- **scale**: determina el nivel de ampliación o reducción de la imagen.
- **width, height**: especifica el tamaño de la imagen en píxeles.

<sup>7</sup>Ver <https://classic.sdss.org/dr7/> para más información.



(a) *Ec587725041172086986.*



(b) *Sb2m5877324843537531.*

**Figura 3.4:** Galaxias (a) elíptica y (b) espiral obtenidas de SDSS. Tienen un tamaño de **150 × 150** pixeles a una escala de **0.15**. El nombre del archivo jpg se compone de su clase y su id, respectivamente.

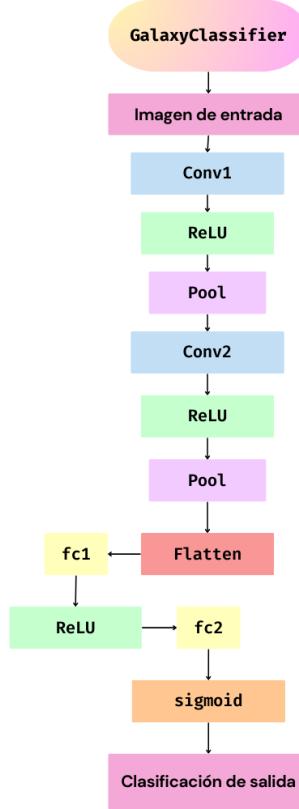
Particularmente, de las 1624 galaxias miembro presentes en la muestra de los 11 cúmulos de Abell, 727 galaxias ya habían sido clasificadas en GZ. A partir de esto, se añadieron 873 galaxias más del catálogo con el fin de ampliar el conjunto de datos. Posteriormente, el 80% de estos datos se destinó para entrenamiento, mientras que el 20% restante a la validación. En total, se tomaron **919** galaxias elípticas y **681** galaxias espirales para el *dataset*.

#### SECCIÓN 3.4

## Construcción de la CNN

Para la tarea de clasificar imágenes de galaxias, se optó por una red neuronal de tipo *convolucional* utilizando PyTorch en *Google Colab*, gracias a su efectividad en el manejo de datos visuales. Estas redes están diseñadas para reconocer patrones sencillos que evolucionan hacia patrones más elaborados a medida que las capas profundas analizan la información. La configuración de esta red integra capas de convolución para la detección de características, capas de agrupamiento para disminuir la dimensionalidad y capas completamente conectadas para realizar las clasificaciones.

El modelo tiene una arquitectura básica de clasificación de imágenes en dos clases, compuesta por una serie de capas operativas que analizan, filtran y organizan las imágenes que se ingresan hasta llegar a una clasificación final (ver Figura 3.5).



**Figura 3.5:** Diagrama de flujo que representa el procesamiento de los datos a través de la CNN. Para más detalles, consultar Apéndice 1.

El funcionamiento de las capas puede definirse de la siguiente forma:

- **Capas convolucionales (nn.Conv2d).** Aplican filtros 2D a la entrada para extraer características como bordes y texturas.
- **Capas de activación (nn.ReLU).** Introducen no linealidad en la red, permitiendo que aprenda funciones más complejas.
- **Capas de agrupamiento (nn.MaxPool2d).** Reducen la dimensionalidad de las características, manteniendo las más importantes.
- **Capas de aplanamiento (nn.Flatten).** Convierte la matriz multidimensional de las características extraídas a un vector unidimensional para ser la entrada de las siguientes capas.
- **Capas completamente conectadas (nn.Linear).** Mapean las características extraídas a las clases de salida.

Capa	Tipo	Entrada	Salida	Descripción
1	Conv2d	(3, 150, 150)	(64, 150, 150)	64 filtros, stride=1, padding=1
2	ReLU	(64, 150, 150)	(64, 150, 150)	Función de activación
3	MaxPool2d	(64, 150, 150)	(64, 74, 74)	kernel=3, stride=2, padding=0
4	Conv2d	(64, 74, 74)	(32, 74, 74)	32 filtros, stride=1, padding=1
5	ReLU	(32, 74, 74)	(32, 74, 74)	Función de activación
6	MaxPool2d	(32, 74, 74)	(32, 36, 36)	kernel=3, stride=2, padding=0
7	Flattten	(32, 36, 36)	(32 * 36 * 36)	Aplana a un vector
8	Linear	(32 * 36 * 36)	(96)	Totalmente conectada, 96 neuronas
9	Linear	(96)	(2)	Totalmente conectada, 2 neuronas
10	Sigmoid	(2)	(2)	Función de activación de probabilidad

**Tabla 3.1:** Resumen de la arquitectura del modelo *GalaxyClassifier*.

Dentro de la definición del modelo, podemos encontrar la **función forward** que básicamente define el flujo de los datos a través de la red neuronal:

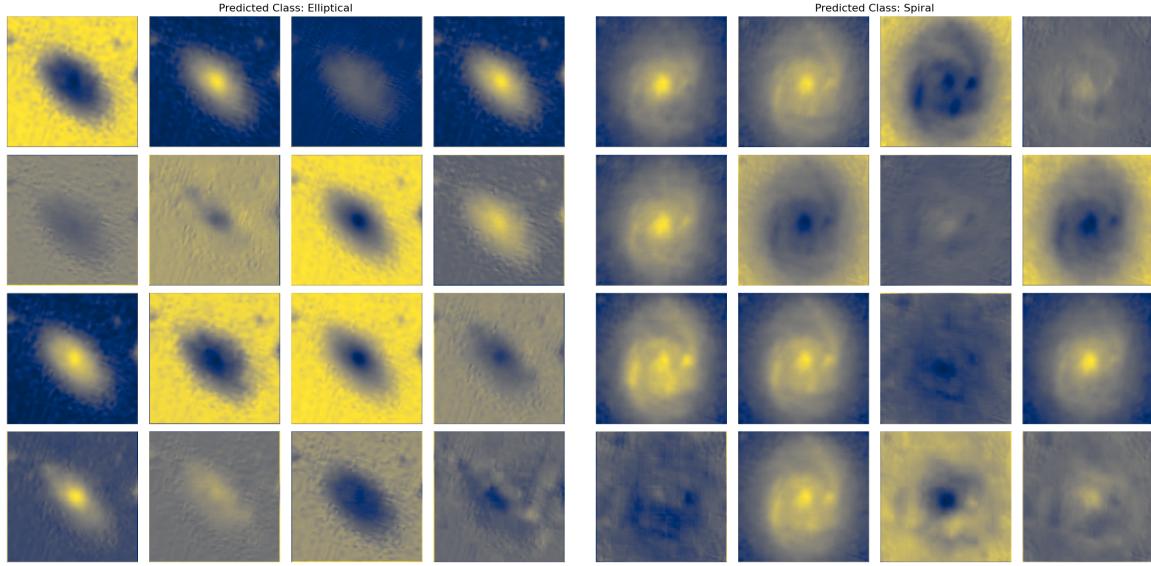
```

1 def forward(self, x):
2     #Extracción de características
3     x = F.relu(self.conv1(x))
4     x = self.pool(x)
5     x = F.relu(self.conv2(x))
6     x = self.pool(x)
7     #Aplanar la salida
8     x = self.flatten(x)
9     #Clasificación
10    x = F.relu(self.fc1(x))
11    x = torch.sigmoid(self.fc2(x))
12    return x

```

En pocas palabras, la imagen de entrada pasa por la capa `conv1`, se aplica la función de activación `ReLU` y se reduce con `pool`. Posteriormente, hace lo mismo pero para la capa `conv2`. La matriz resultante, se aplaná con `flatten` y va a la primera capa conectada `fc1`. Finalmente, pasa a `fc2` y se aplica la función `sigmoid` para obtener la probabilidad de pertenecer a una de las dos clases.

Con el fin de ilustrar el flujo de la información dentro de la CNN, empleamos las galaxias de la Figura 3.4 para graficar los mapas de características haciendo uso de la librería `matplotlib`. Los resultados se muestran en la Figura 3.6.



(a) *Ec587725041172086986*.

(b) *Sb2m5877324843537531*.

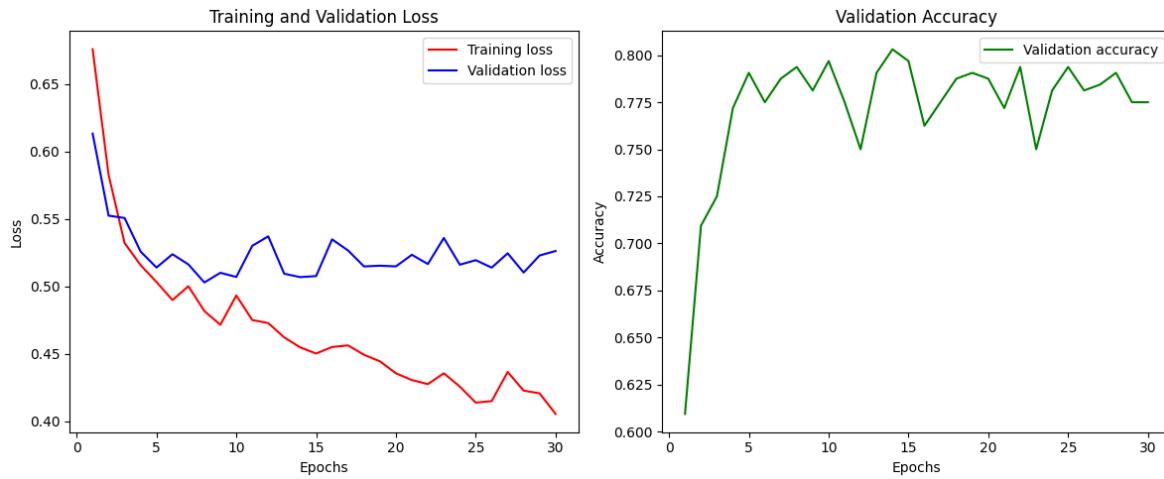
**Figura 3.6:** Mapas de características de galaxias (a) elíptica y (b) espiral pertenecientes al conjunto de entrenamiento. Las partes más brillantes de cada imagen representan patrones o detalles aprendidos por la red después de pasar por los filtros de las capas.

A pesar de que todas las imágenes fueron descargadas con un tamaño de  $150 \times 150$  pixeles, se normalizó el conjunto con las medias y desviaciones estándar usadas comúnmente en modelos de aprendizaje profundo como *ResNet50* (Torchvision Contributors, 2024):

```
1   transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])
```

En el caso del entrenamiento, se usa la función de pérdida `nn.CrossEntropyLoss()` para calcular la discrepancia entre las predicciones y las etiquetas verdaderas. Mientras, el optimizador `optim.Adam()`, ajusta los parámetros del modelo en base al gradiente (Kingma y Ba, 2014).

Por lo que vemos en la Figura 3.7, el modelo demuestra un aprendizaje adecuado en cuanto a las pérdidas durante el entrenamiento y la validación; es decir, conforme se incrementan las épocas, se ajusta mejor a los datos. En términos generales, logra una precisión satisfactoria, lo que sugiere un rendimiento aceptable para la tarea principal de clasificar galaxias.



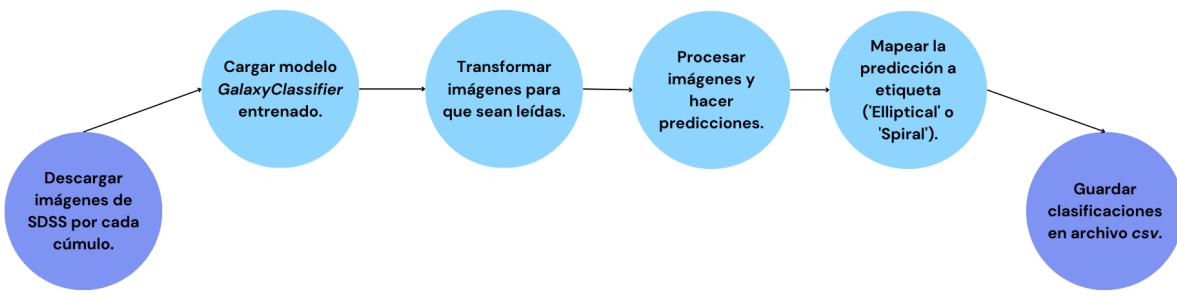
**Figura 3.7:** Curvas de aprendizaje de *GalaxyClassifier* con 30 épocas. La precisión de validación final es 77.50%.

### SECCIÓN 3.5

## Clasificación de galaxias miembro faltantes

Hasta este momento, la CNN estaba entrenada para cumplir con su objetivo. Por tanto, el siguiente paso fue descargar las imágenes de la muestra que aún no contaban con clasificación. Para ello, se volvió a emplear el código en `bash` 3.3 con una ligera modificación. En este caso, el archivo `lis.dat` contenía las coordenadas de las galaxias que, precisamente, no se encontraron en el catálogo de GZ, lo cual involucró un total de **897** galaxias.

A continuación, las imágenes fueron procesadas con PyTorch, siguiendo un flujo de transformación, carga, predicción y guardado de datos para cada cúmulo. Para una mejor comprensión del proceso realizado, se puede recurrir a la Figura 3.8.



**Figura 3.8:** Este diagrama da una visión general de las fases que se siguieron con el propósito de clasificar las galaxias restantes a través de la CNN entrenada.

## Diagramas morfología-densidad

Una vez que se completó la clasificación de todas las galaxias pertenecientes a los cúmulos, se procedió al estudio de morfología-densidad conforme a Dressler (1980). Los diagramas morfología-densidad nos muestran la relación que existe entre la fracción de galaxias y la densidad local del entorno. Justamente, el análisis brindado por estos diagramas es importante para definir si, además de las interacciones galácticas, el ambiente en el que se encuentran también influye en las propiedades morfológicas que presentan.

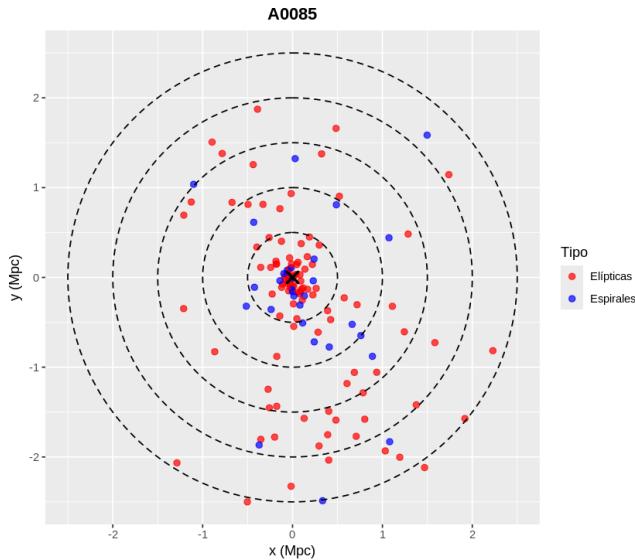
Para este trabajo, se utilizó una ecuación de densidad con modificaciones que consideraron la naturaleza de nuestros datos. Sabemos que

$$\rho = \frac{m}{V} \quad (3.3)$$

y podemos proyectar a dos dimensiones

$$\rho = \frac{f_g}{A} = \frac{f_g}{\pi(r_{ex}^2 - r_{in}^2)} \quad (3.4)$$

donde definimos  $f_g$  como la fracción de galaxias de cada tipo morfológico en un anillo y  $A$  como el área de una diferencia de radios que determinan el anillo de interés.



**Figura 3.9:** Proyección cartesiana del cúmulo A0085. Los círculos concéntricos ayudan a dividir el cúmulo en anillos de 0.5 Mpc que se utilizan para calcular la densidad proyectada. Se calcula del centro hacia afuera.

La Figura 3.9 muestra la proyección bidimensional del sistema. Cada anillo está caracterizado por un radio interior y un radio exterior que lo delimitan. A partir del área ocupada por cada anillo, se contabiliza la cantidad de galaxias de cada tipo y se calcula su proporción respecto al número total de miembros del cúmulo.

Así, se escribió un código en R que implementa la ecuación 3.4 con las operaciones e iteraciones necesarias para generar los diagramas, teniendo como entrada un archivo csv con la información del cúmulo y de sus respectivas galaxias. Su funcionamiento se comprende más claramente al dividirlo por secciones. Tomemos como ejemplo el cúmulo A0085:

1. **Carga de datos y parámetros iniciales.** Desde un archivo csv, se incluyen las coordenadas de las galaxias junto con sus distancias al centro ( $d_{cc}$ ).

```
1 data <- read.csv("A0085_end.csv", header = TRUE)
2 centro_ra <- 10.4587
3 centro_dec <- -9.3019
```

2. **Definición de anillos concéntricos.** Se crean radios en un rango de 0.5 a 2.5 Mpc en pasos de 0.5. Cada radio delimita un anillo o región dentro del cúmulo.

```
1 radios <- seq(0.5, 2.5, by = 0.5)
```

3. **Cálculo de galaxias por radio.** Se seleccionan las galaxias dentro de cada anillo y se hace un conteo del número de elípticas y espirales sin tomar en cuenta los anillos anteriores.

```
1 for (i in seq_along(radios)) {
2   galaxias_en_radio <- which(distancias <= radios[i])
3   if (i > 1) {
4     galaxias_en_radio <- setdiff(galaxias_en_radio, which(
5       distancias <= radios[i-1]))
6   }
7   E[i] <- sum(substr(data$gz2_class[galaxias_en_radio], 1, 1)
8   == "E")
8   S[i] <- sum(substr(data$gz2_class[galaxias_en_radio], 1, 1)
9   == "S")
10 }
```

4. **Cálculo de densidades proyectadas.** Se obtienen al dividir el conteo de galaxias entre el área de anillo dada por  $area = \pi(r_{exterior}^2 - r_{interior}^2)$ . Definimos rho\_E y rho\_S para cada tipo.

```
1 area <- pi * (radios[i]^2 - (radios[i] - 0.5)^2)
2 data_densidades[i, c("rho_E", "rho_S")] <- sapply(c(elipticas[i],
3   espirales[i]), function(x) x / area)
```

5. **Generación de gráfico.** Se utiliza ggplot para graficar en el eje  $y$  las fracciones y el logaritmo la densidad proyectada en el eje  $x$  unidas a través de líneas.

```
1 plot <- ggplot(data_densidades) +
2   geom_line(aes(x = log10(rho_E), y = E/total_galaxias,
3     linetype = "E")) +
4   geom_line(aes(x = log10(rho_S), y = S/total_galaxias,
5     linetype = "S")) +
6   xlab(expression(log[10]*rho)) +
7   ylab("Fraction of Galaxies") +
8   theme_gray()
```

En resumen, los diagramas morfología-densidad se construyeron siguiendo tres pasos principales: en primer lugar, se estimó la fracción de galaxias ( $f_g$ ) de cada tipo utilizando el número de clasificaciones realizadas por la CNN respecto a la totalidad de galaxias dentro del cúmulo. Posteriormente, se calculó la densidad proyectada ( $\rho$ ) para cada un anillo concéntrico en torno al centro del cúmulo y, para finalizar, se graficaron los valores obtenidos mostrando  $f_g$  en función de  $\log_{10}(\rho)$ .

# CAPÍTULO 4

---

## RESULTADOS Y DISCUSIÓN

---

Los resultados que se obtuvieron durante el desarrollo de esta investigación se presentan en este capítulo, evaluando tanto los procedimientos como los hallazgos más relevantes. Cada sección está dedicada a un aspecto específico del estudio. Se comienza con una visualización de la muestra utilizada, seguida de gráficas del estudio de membresía mediante el método de cáusticas. Luego, se detalla el entrenamiento del modelo de la CNN así como los parámetros ajustados para llegar al `GalaxyClassifier` final. Por último, se analizan las relaciones morfología-densidad a través los diagramas y se comparan con clasificaciones visuales de dos cúmulos en particular.

SECCIÓN 4.1

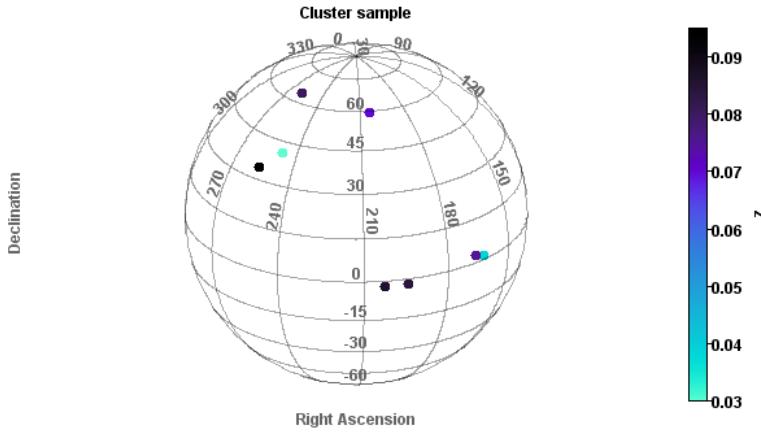
### Selección de la muestra

La información derivada del catálogo Pop04 constituyó el núcleo de nuestros cálculos; las coordenadas y el *redshift* tanto de los cúmulos como de las galaxias analizadas, se emplearon a lo largo de toda la investigación. En las Figura 4.1 se presenta la disposición de los datos obtenidos directamente de dicho catálogo. Dado que los programas fueron diseñados para seleccionar únicamente los elementos esenciales, el orden de las columnas no tuvo impacto significativo.

RXC	Name	RAJ2000	DEJ2000	z	FX	e_FX	LX	rho	LF	e_LF	LC	e_LC	sigma0	sigmab	e_sigmab	
1	J0041.8+0918	A0085	10.4587	-9.3019	0.052	67.612	3.	7.87	2227	1.81	0.22	1.87	0.06	5731	2295	456
2	J0114.9+0022	A0168	18.735	0.3746	0.047	8.725	8.7	0.81	759	1.27	0.18	1.34	0.04	2281	4540	106
3	J1058.4+0134	A1139	164.6152	1.5682	0.039	4.175	13.6	0.2	53	0.65	0.14	0.67	0.46	2101	3524	193
4	J1113.3+0232	A1205	169.3446	2.5424	0.075	8.034	9.3	1.54	2778	1.45	0.19	1.6	0.08	4223	5110	135
5	J1258.6-0145	A1650	194.6712	-1.7569	0.084	20.64	6.4	6.28	1810	1.43	0.2	1.4	0.08	13691	3945	210
6	J1330.8-0152	A1750S	202.7081	-1.8728	0.086	6.373	12.	2.03	3900	4.04	1.	2.87	0.07	11468	4081	1372
7	J1336.1+5912	A1767	204.0272	59.2074	0.07	12.435	5.3	2.61	808	0.96	0.16	0.92	0.08	7928	4181	73
8	J1627.6+0455	A2197	246.9172	40.9208	0.03	4.244	11.	0.12	1742	2.14	0.19	2.19	0.05	2173	6097	108
9	J1702.7+3403	A2244	255.6776	34.0613	0.095	18.942	3.8	7.31	3441	2.07	0.24	2.05	0.08	19128	5340	79
10	J1712.7+6403	A2255	258.1984	64.0632	0.08	17.678	2.1	4.87	2816	3.36	0.34	3.21	0.08	11799	3934	170
11	J2354.2-1024	A2670	358.5557	-10.4129	0.076	10.083	9.7	2.5	2425	2.54	0.29	2.55	0.03	17421	5214	46

**Figura 4.1:** Algunas de las columnas de los 11 cúmulos seleccionados para la muestra. Vista de TOPCAT.

A su vez, la Figura 4.2 presenta una proyección en el cielo que permite observar la distribución espacial de los cúmulos.



**Figura 4.2:** Proyección de los cúmulos en coordenadas ecuatoriales (ascensión recta y declinación). Los puntos indican la posición de cada cúmulo y los colores corresponden a su redshift, donde tonos claros indican valores bajos y tonos oscuros valores altos.

#### SECCIÓN 4.2

## Estudio de membresía

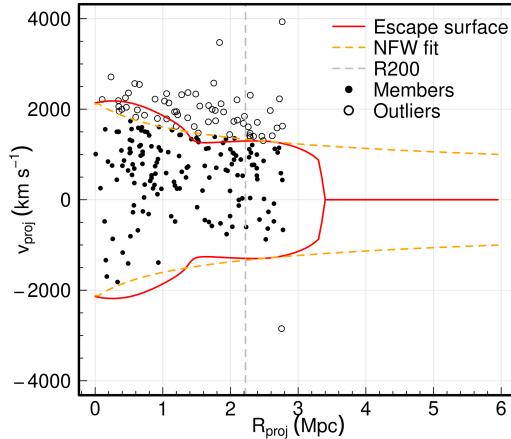
Con la muestra definida, se empleó el algoritmo basado en el **método de cáusticas** para determinar las galaxias pertenecientes al cúmulo. Este proceso genera como salida una gráfica y un archivo **csv** con las galaxias consideradas ligadas gravitacionalmente.

La Figura 4.3 muestra el contenido del archivo **csv**. Los datos iniciales se obtuvieron de la base de datos del SDSS, considerando un radio de búsqueda de 2.5 Mpc para identificar galaxias cercanas al centro del cúmulo, cuyas coordenadas fueron determinadas según el catálogo de Pop04. Las dos primeras columnas incluyen los cálculos realizados por el algoritmo de Gifford et al. (2013).

	dcid	dvel	ObjID	ra	dec	z	zErr	dered_g	dered_r	class	vel	evel
1	0.0081	1007.631	1.237653E18	10.46027	-9.30315	0.05536	1.569514E-5	18.98963	18.20053	GALAXY	16607.631	4.70854
2	0.07839	254.55	1.237653E18	10.45054	-9.28422	0.05265	1.440608E-5	16.10053	15.24489	GALAXY	15854.55	4.32182
3	0.13993	-478.839	1.237653E18	10.49078	-9.31533	0.0504	2.154027E-5	18.29741	17.48467	GALAXY	15121.161	6.46208
4	0.14881	1555.917	1.237653E18	10.48979	-9.282	0.05719	1.723868E-5	18.01665	17.20059	GALAXY	17155.917	5.1686
5	0.15756	-1335.963	1.237653E18	10.42211	-9.31588	0.04755	1.725678E-5	16.51344	15.93722	GALAXY	14264.037	5.17703
6	0.17222	-1697.031	1.237653E18	10.43663	-9.33819	0.04634	1.514009E-5	17.62876	16.83477	GALAXY	13902.969	4.54203
7	0.18947	843.201	1.237653E18	10.41986	-9.27547	0.05481	2.000571E-5	18.21876	17.38121	GALAXY	16443.201	6.00171
8	0.25063	754.752	1.237653E18	10.40088	-9.3251	0.05452	1.662786E-5	17.09507	16.22086	GALAXY	16354.752	4.98836

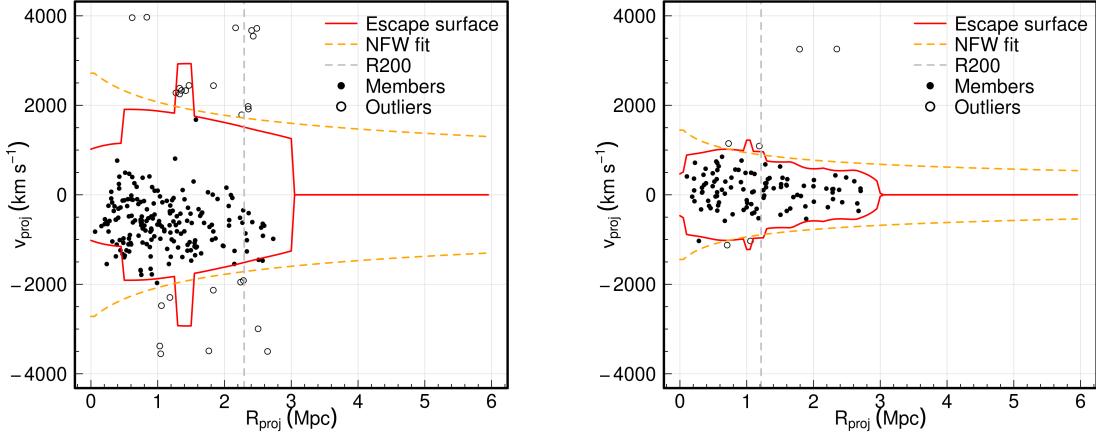
**Figura 4.3:** Datos de las galaxias de A0085 al aplicar el método de cáusticas. Vista de TOPCAT.

En la Figura 4.4 se presenta el diagrama de cáusticas para el cúmulo A0085, construido a partir del espacio de fase proyectado ( $R_{\text{proj}}-v_{\text{proj}}$ ). Las líneas rojas definen la *superficie de escape*, que delimita la región dinámica dominada por el potencial gravitacional del cúmulo. Las galaxias localizadas dentro de esta superficie (puntos negros) fueron clasificadas como *miembros* del cúmulo. La línea discontinua naranja representa un ajuste teórico de la velocidad de escape basado en un *perfil de masa NFW*, utilizado como referencia para validar la forma de las cáusticas.

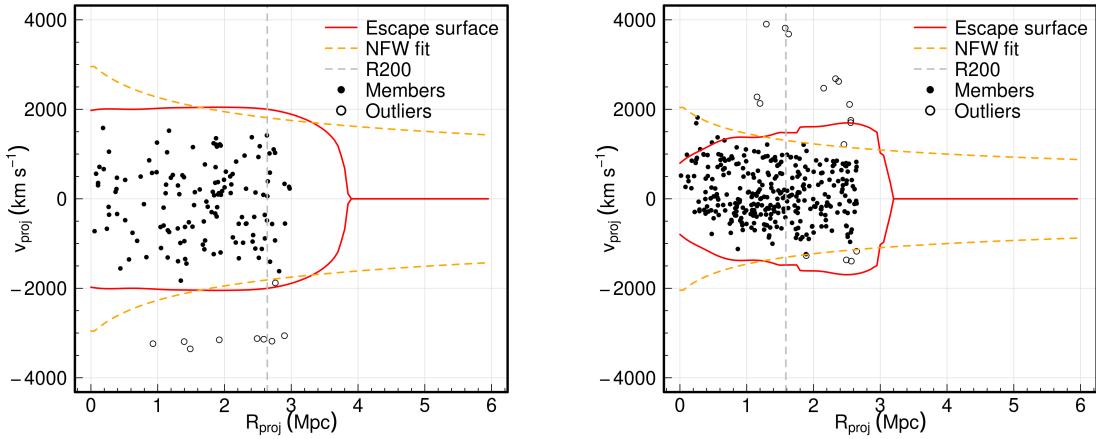


**Figura 4.4:** Gráfico de A0085 que muestra la velocidad proyectada ( $v_{\text{proj}}$ ) en función del radio proyectado ( $R_{\text{proj}}$ ) desde el centro del cúmulo. La curva roja separa las galaxias ligadas gravitacionalmente de las no ligadas.

Las Figuras 4.5 y 4.6 permiten observar cómo la superficie de escape varía entre cúmulos debido a diferencias en sus propiedades físicas y dinámicas. Por ejemplo, los cúmulos con mayor **masa** crea pozos de potencial más profundos, lo que resulta en velocidades de escape más altas (Diaferio, 1999). Por otro lado, el **equilibrio dinámico** también influye: cúmulos estables muestran superficies más definidas, mientras que aquellos en proceso de interacción o formación presentan irregularidades causadas por perturbaciones gravitacionales (Serra et al., 2011).



**Figura 4.5:** Gráfico obtenido después del análisis de cáusticas para los cúmulos **A0168** (izquierda) y **A1139** (derecha).



**Figura 4.6:** Gráfico obtenido tras el análisis de cáusticas para los cúmulos **A1750** (izquierda) y **A2197** (derecha).

Una de las principales ventajas del método de cáusticas para la determinación de membresía en cúmulos es que no depende de supuestos sobre el equilibrio dinámico o la simetría del sistema, ya que se basa en una propiedad física (la curva de escape del potencial gravitacional) permitiendo identificar las galaxias ligadas, de manera confiable.

## Base de datos de GZ

A partir del *estudio de membresía*, se verificó si las galaxias identificadas como miembros de los cúmulos estaban incluidas en el catálogo de GZ, con el propósito de utilizarlas como parte del conjunto de entrenamiento de la CNN. Siguiendo el proceso propuesto en la Sección 2.3, se constituyó un *dataset* con imágenes provenientes de SDSS (en el espectro visible).

Los datos se organizan en dos carpetas principales (ver Figura 4.7): una contiene imágenes de galaxias elípticas (919 archivos *jpg*), mientras que la segunda agrupa galaxias clasificadas como espirales (681 archivos *jpg*). Cada archivo se identifica por un nombre único basado en su clasificación y en el identificador del objeto.



**Figura 4.7:** Organización de las imágenes utilizadas para entrenar la CNN. Se guardaron en Google Drive para facilitar la conexión al notebook de Colab.

## Construcción de la CNN

Se exploraron diversas variaciones numéricas de los parámetros (*batch\_size*, *normalización* y *épocas*) en el entrenamiento de la red neuronal con la finalidad de determinar la configuración que se ajustaba mejor a las características de las imágenes del *dataset*. En la Tabla 4.1 se presentan las diferentes configuraciones utilizadas centrándose particularmente en la cantidad de épocas debido a que este parámetro resultó más influyente en el aprendizaje y por tanto, en la precisión final.

Cabe destacar que se hizo una comparación con dos grupos de galaxias aleatorias de “prueba” del catálogo de GZ:

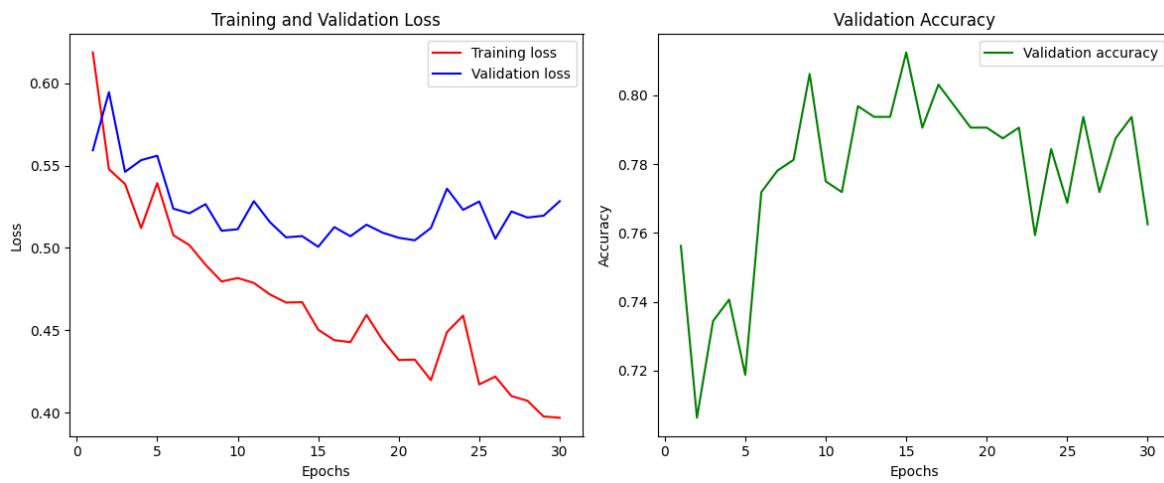
- **Prueba 1.** Filas 1000 - 1050
- **Prueba 2.** Filas 238868 - 238968

Esto con la intención de evaluar la generalización del modelo en datos no vistos.

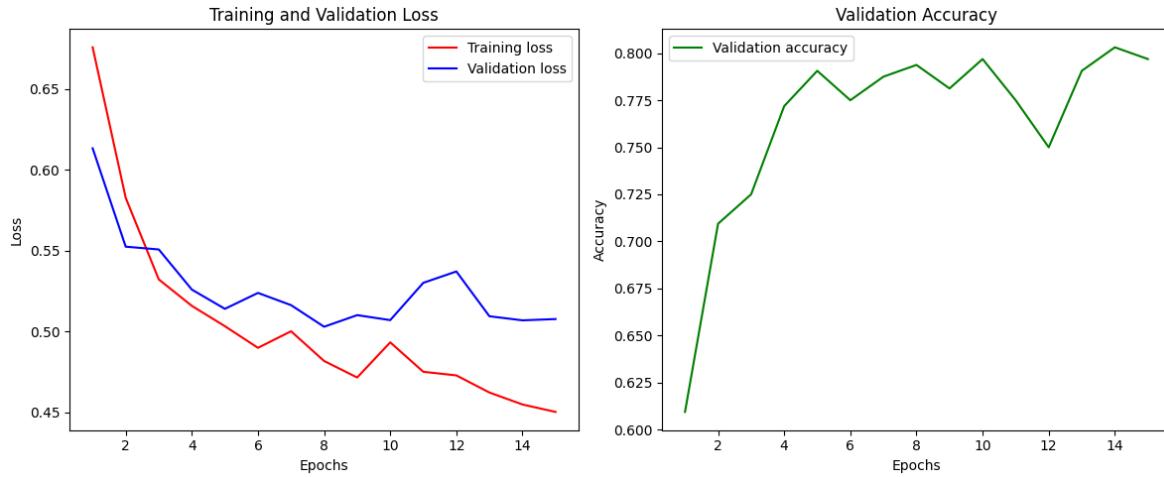
Modelo	Épocas	Tiempo de entrenamiento	Precisión final	Prueba 1	Prueba 2
M-a	30	59m 17s	78.44 %	41/50	79/100
M-b	15	28m 96s	79.69 %	39/50	79/100
M-c	40	78m 46s	76.88 %	42/50	67/100
M-d	30	58m 07s	77.50 %	40/50	85/100

**Tabla 4.1:** Resumen de los entrenamientos realizados para el modelo de la CNN. Las pruebas reflejan las clasificaciones que se hicieron correctamente respecto al total.

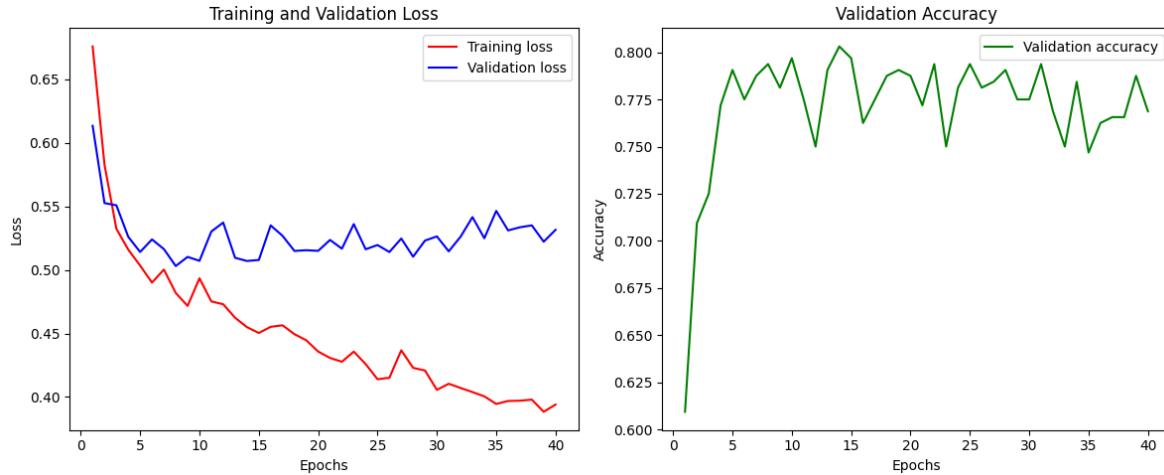
El desempeño del modelo al variar los parámetros puede ser evaluado mediante la pérdida y la precisión representados en las curvas de aprendizaje:



**Figura 4.8:** M-a. Notamos que la curva de pérdida de validación (azul) fluctúa y no decrece de manera consistente, lo que indica una tendencia al sobreajuste. La curva de precisión (verde) también presenta variaciones significativas.



**Figura 4.9:** *M-b.* La curva de pérdida de validación sigue la tendencia de la curva de entrenamiento (roja) y la precisión es menos fluctuante. Sin embargo, puede mejorar.

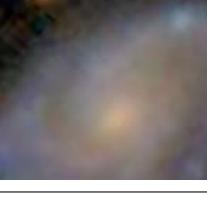


**Figura 4.10:** *M-c.* La curva de validación va decreciendo bien sin demasiadas fluctuaciones como en el modelo *M-a*. Por su parte, la precisión se mantiene en un rango entre 75% y 80%. Aún así, hay más de estabilidad en el modelo *M-d* (véase Figura 3.7).

Teniendo los resultados de las pruebas para los distintos modelos, tomamos como un indicador importante el número de clasificaciones correctas en cuanto a las definidas en GZ pues nos brindó una forma de confirmar que, para las galaxias sin categoría, habría buenas clasificaciones a pesar del  $\approx 20\%$  de error como consecuencia de la precisión final. Entonces, el modelo más viable fue el *M-d* que llamamos *GalaxyClassifier*.

## Clasificación de galaxias miembro faltantes

Con el seguimiento de las etapas propuestas en el diagrama de la Figura 3.8, se muestra la Tabla 4.2 que resumen las clasificaciones de galaxias faltantes en el conjunto de miembros. En ella se incluyen tanto las predicciones logradas por la red neuronal, como su información distintiva (*ObjID*, *z* y cúmulo). Incluso, se proporciona la imagen de cada galaxia para una mejor visualización de su correlación con la clasificación predicha.

Galaxia	ObjID	z	Cúmulo	galaxy_class
	1237654029251706880	0.03940599	A1139	Elliptical
	1237654030327021824	0.07843165	A1205	Spiral
	1237671939804627456	0.07763089	A2255	Elliptical
	1237652900210606336	0.07903229	A2670	Spiral
	1237659326566039552	0.02587877	A2197	Spiral

**Tabla 4.2:** Se exponen 5 imágenes de galaxias que se les ha asignado una categoría de acuerdo a la CNN entrenada.

## Diagramas morfología-densidad

Con la clasificación morfológica de las galaxias completada para los 11 cúmulos, el siguiente paso fue ejecutar el código descrito en la Sección 2.6. Para cada cúmulo, se ajustaron los parámetros de entrada, como los archivos de datos y las coordenadas del centro, organizando la información en carpetas independientes. El cálculo de la densidad local y la fracción de galaxias según su morfología se realizó de manera uniforme, garantizando consistencia en la metodología.

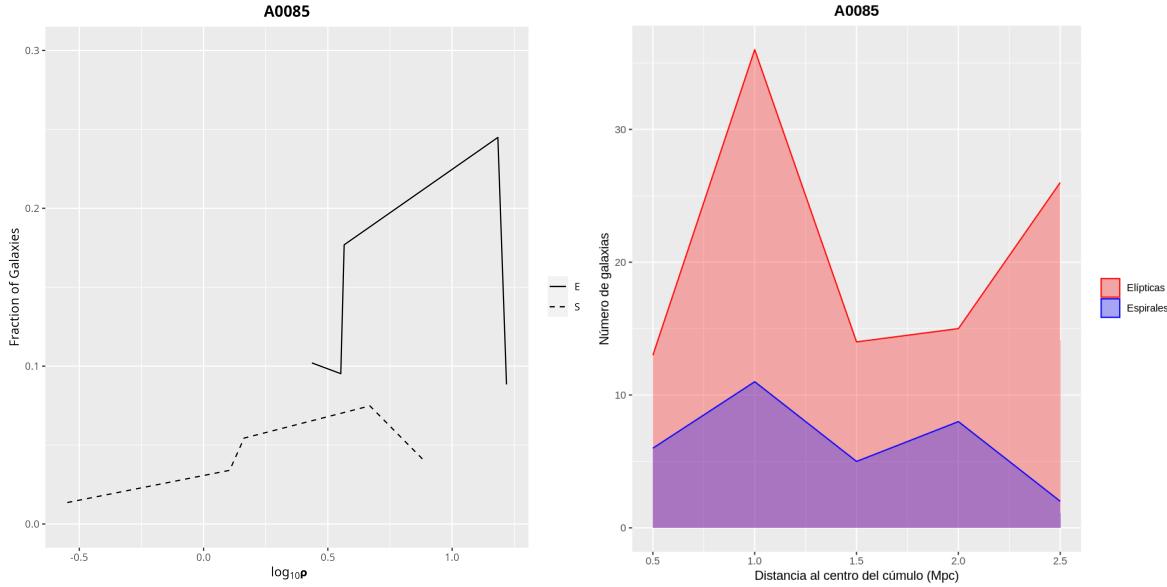
Los resultados se resumen en dos archivos principales: *data\_density.csv*, que registra los conteos morfológicos (clasificados mediante CNN y GZ) y las densidades proyectadas; y *Axxxx\_MD.png*, que representa gráficamente la fracción de galaxias como función del logaritmo de la densidad local.

A través del análisis de los diagramas morfología-densidad, buscamos comprender cómo el entorno influye en la forma de las galaxias que lo habitan. Por ello, hemos organizado los cúmulos en tres grupos, en virtud de algunas características que comparten sus diagramas.

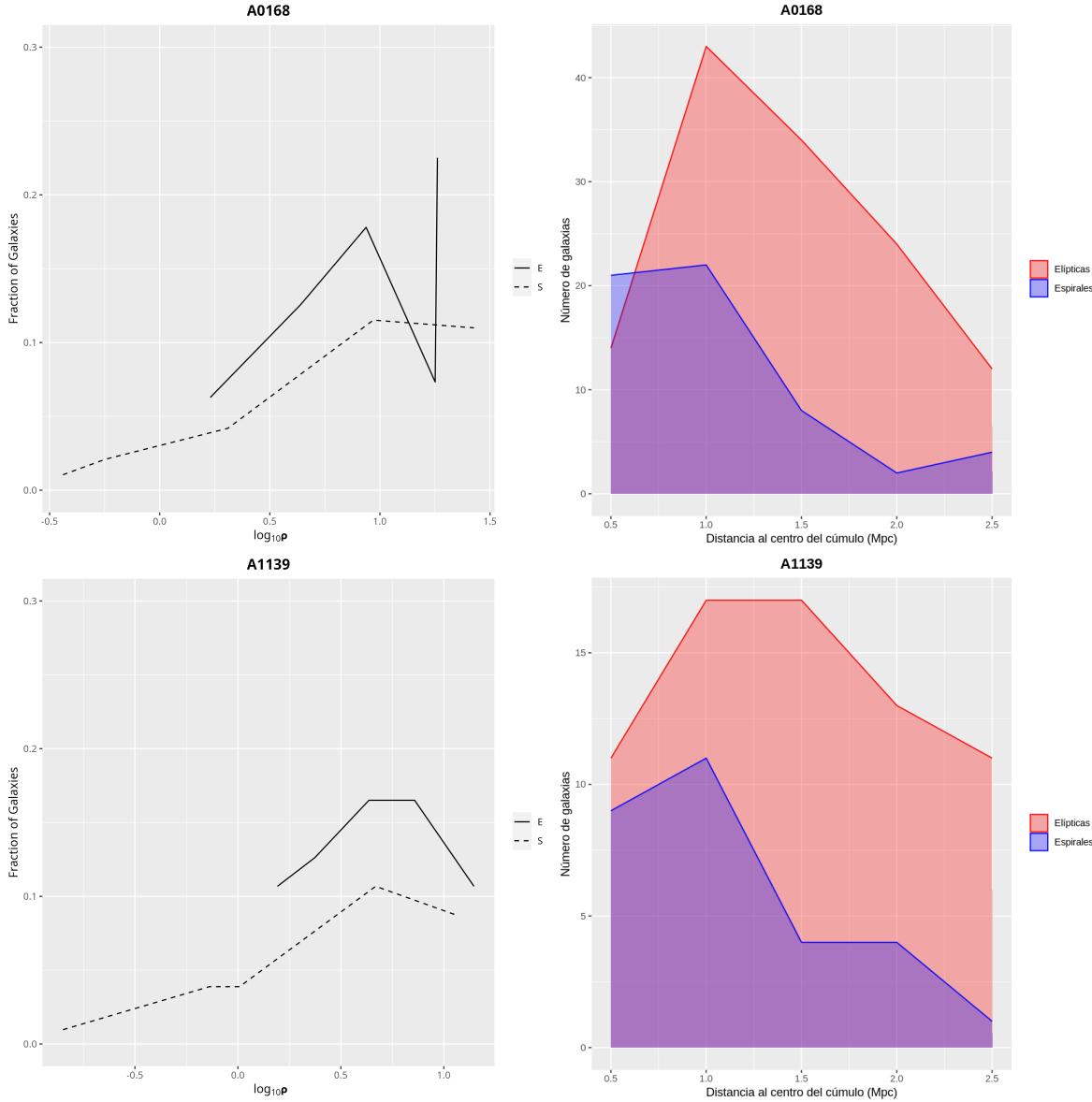
### 1. Tendencia esperada

Los cúmulos **A0085, A0168, A1139, A1767, A2197, A2255 y A2670** conforman el **63.6%** de la muestra total y fueron agrupados bajo la categoría de *tendencia esperada* en función del comportamiento de sus curvas morfología-densidad, especialmente en los extremos de densidad alta y baja. Al examinar los diagramas del panel izquierdo en las Figuras 4.11-4.14, se aprecia que en  $\log_{10} \rho \approx 1.0$  la fracción de galaxias elípticas supera consistentemente a la de espirales, mientras que en  $\log_{10} \rho \approx 0$  las espirales dominan la población, aunque no exclusivamente.

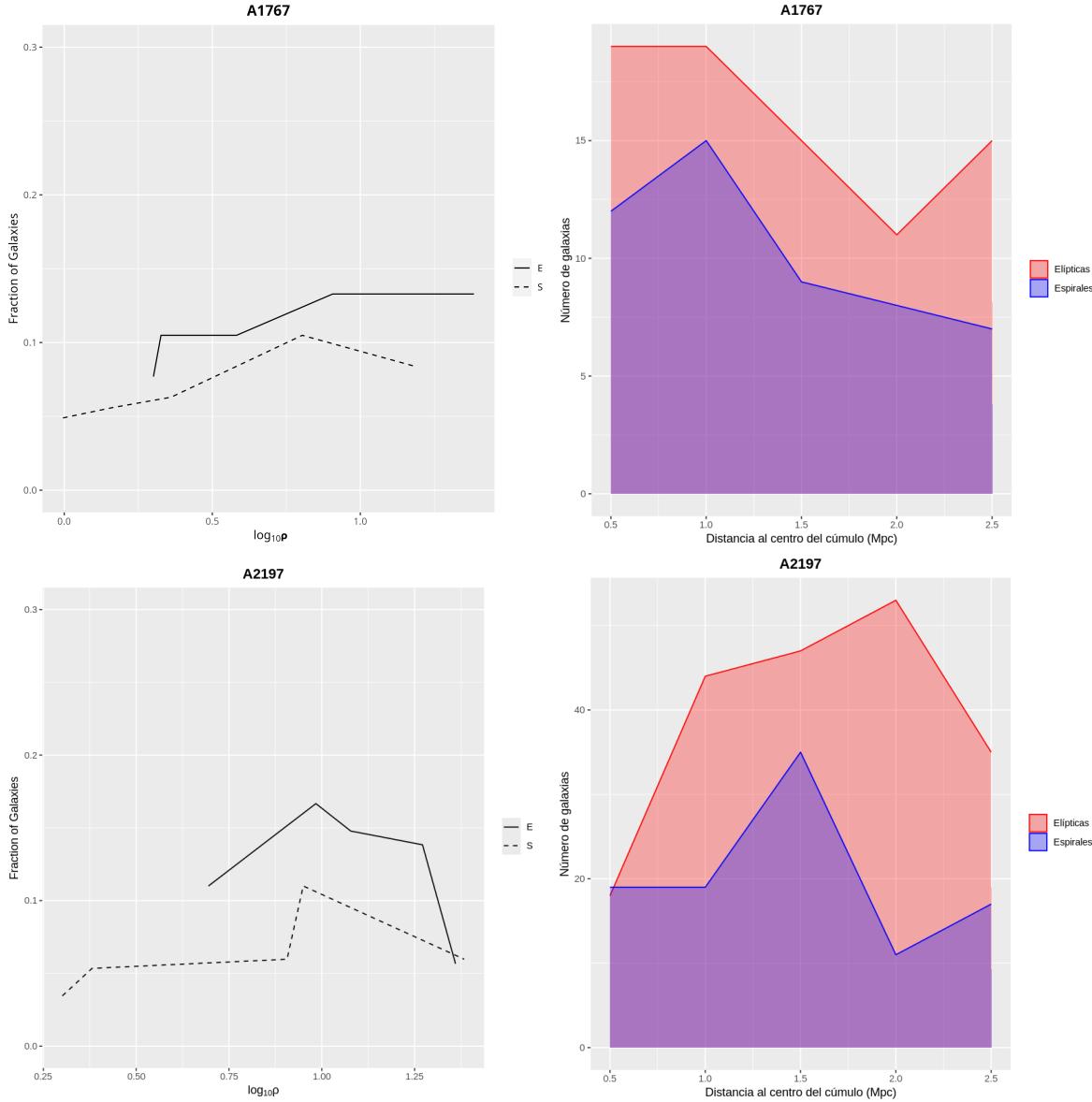
A pesar de seguir esta tendencia, es importante notar que en todos los cúmulos de este grupo persiste una presencia notable de galaxias elípticas incluso en densidades bajas, y ambas clases disminuyen sus fracciones hacia las regiones menos densas, como se observa en los diagramas de distribución radial (panel derecho) de las Figuras mencionadas. Esto podría estar asociado con la evolución dinámica de las distintas regiones intracumulares, que cambian con el tiempo (Postman et al., 2005; Fasano et al., 2015). Dado que los cúmulos analizados siguen de cierta medida la relación morfología-densidad, se consideran relativamente regulares, lo que justificaría la abundancia de galaxias elípticas como sugieren Dressler et al. (1997) y Cava et al. (2017).



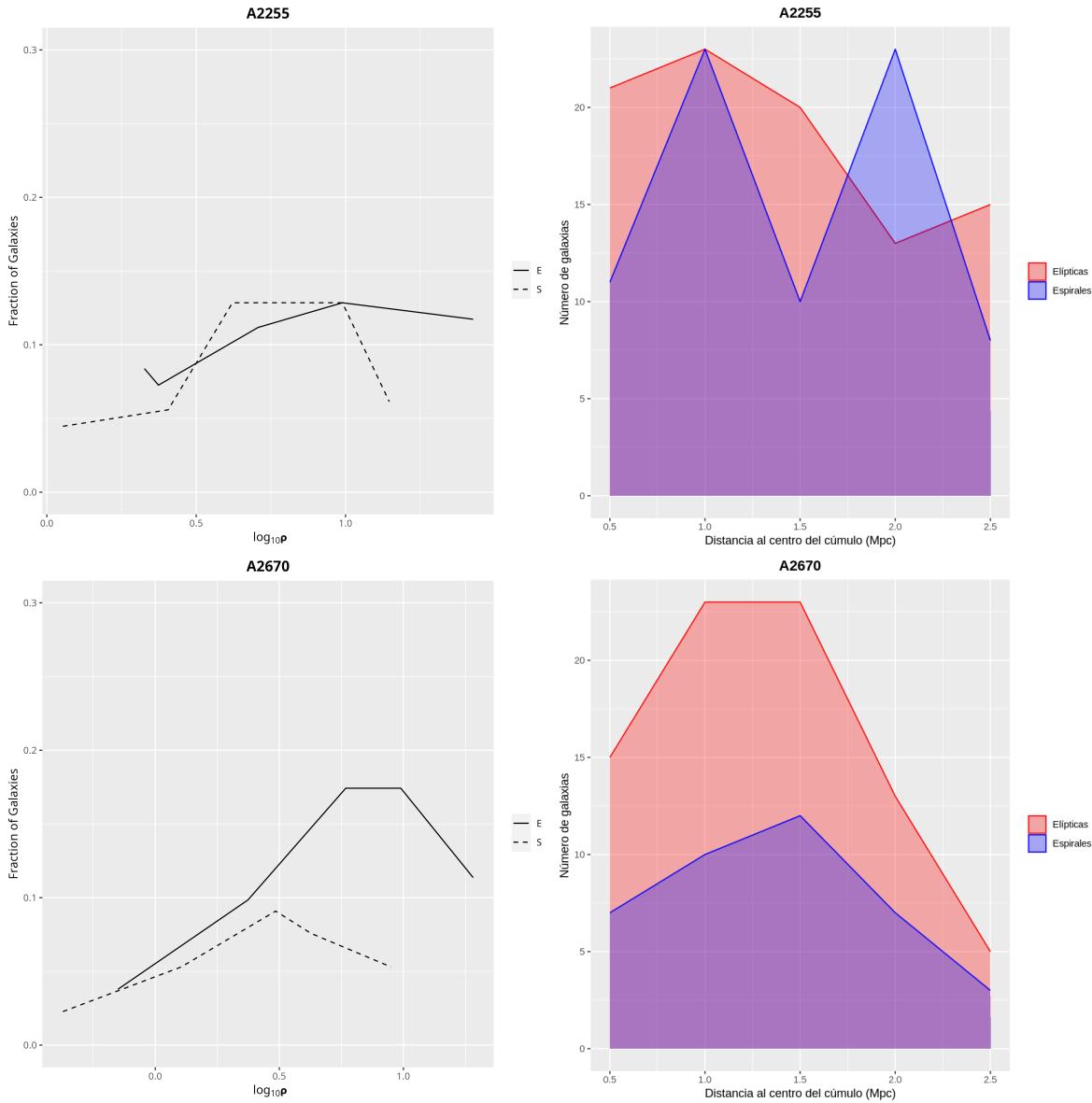
**Figura 4.11:** Diagrama morfología-densidad del cúmulo A0085. En el gráfico de la izquierda muestra la fracción de galaxias  $f_g$  como función del logaritmo de la densidad proyectada  $\log_{10}\rho$ . Se observa un claro predominio de galaxias elípticas (E) en las regiones de alta densidad, mientras que las espirales (S) se vuelven más frecuentes en densidades bajas, aunque sin un incremento marcado. El panel derecho presenta la distribución radial de galaxias (radio en Mpc frente al número de galaxias), donde se evidencia nuevamente la concentración dominante de elípticas a lo largo del cúmulo.



**Figura 4.12:** Diagramas para los cúmulos A0168 y A1139. Apreciamos la representación gráfica de la relación entre la densidad proyectada  $\rho$  y la fracción de galaxias según tipo morfológico  $f_g$  respecto al total de galaxias de cada cúmulo (izquierda). En los entornos de densidad alta, es decir,  $\log_{10}\rho \approx 1$  existen más galaxias tipo E en comparación con las galaxias tipo S que prevalecen en  $\log_{10}\rho \approx 0$  y que a su vez, existen a lo largo de las densidades intermedias y altas, como se muestra en los gráficos de distribución radial de la derecha.



**Figura 4.13:** Diagramas morfología-densidad de los cúmulos alineados con la tendencia esperada. En cada caso, se grafica la fracción de galaxias E y S respecto a la densidad proyectada  $\log_{10}\rho$  (izquierda). También se incluye la distribución radial de galaxias (derecha) del radio vs el número de galaxias, lo cual destaca la diferencia que existe entre la población de las galaxias elípticas con relación a las espirales. Cabe mencionar el caso de A2197, donde se nota una población de espirales parecida a la población de elípticas en densidades altas aunque estas últimas siguen dominando todo el rango.

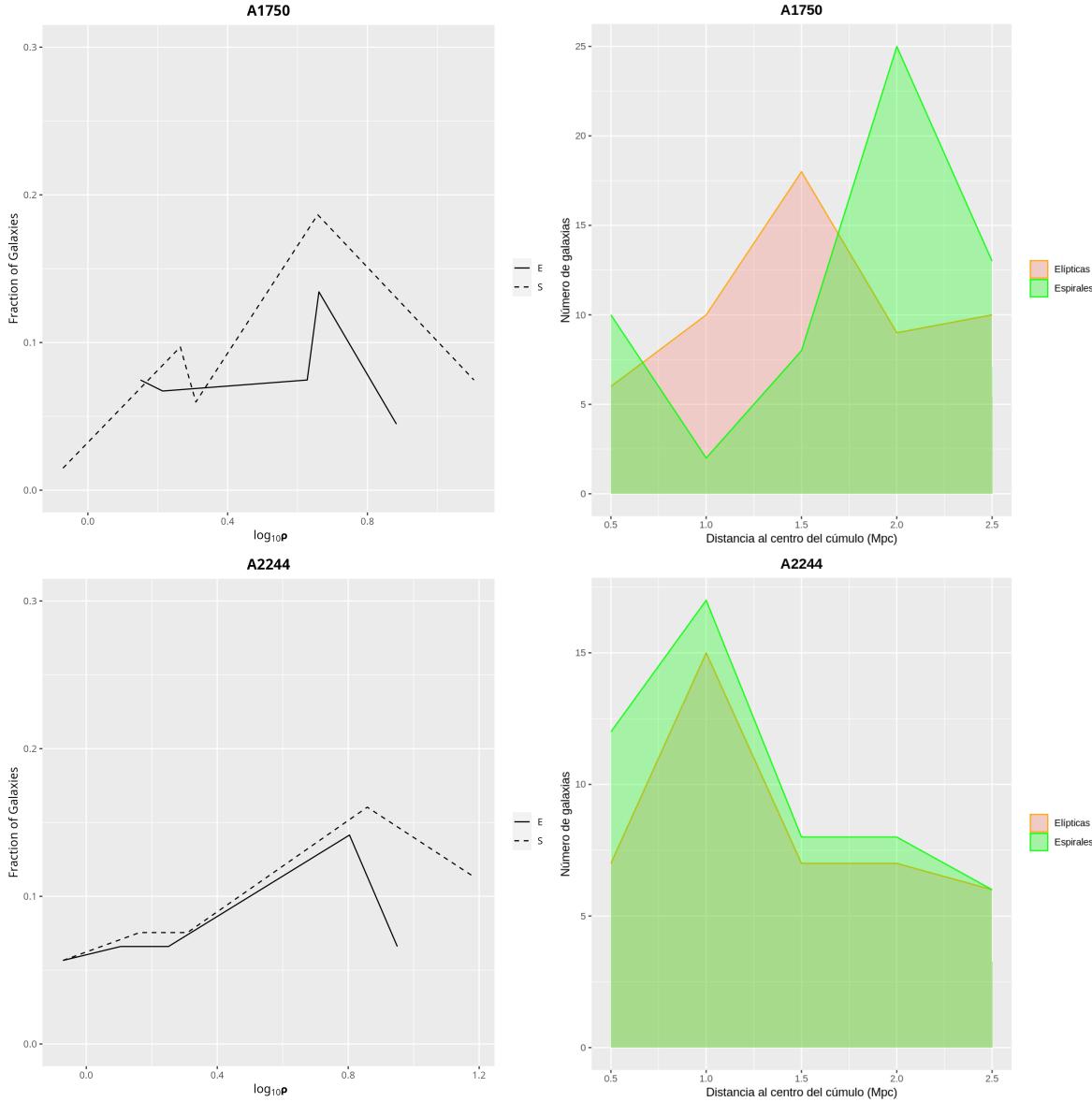


**Figura 4.14:** Diagramas de cúmulos que siguen la tendencia esperada pues en densidades altas la población de galaxias tipo E es mayor que la población de galaxias tipo S y en densidades bajas ocurre lo contrario. Destaca A2255, donde se manifiesta un aumento en la fracción de galaxias espirales en la densidad intermedia. A pesar de esta variación, en las regiones de alta y baja densidad, se mantiene el patrón observado en los demás cúmulos, conservando la característica de dominio elíptico en todo el cúmulo pero que disminuye a medida que se aleja del centro.

## 2. Dominio de espirales

Los cúmulos **A1750** y **A2244** exhiben un comportamiento morfológico distinto al del resto de la muestra. En ambos casos, las galaxias espirales dominan la población galáctica tanto en regiones de alta como de baja densidad, sin que se observe una tendencia clara de segregación morfológica (Figura 4.15). Al analizar las distintas regiones de densidad, se identifica un predominio de espirales incluso en los entornos más densos, aunque tanto su fracción como la de las elípticas tienden a decrecer simultáneamente.

En regiones menos densas, ambas poblaciones coexisten, pero las espirales siguen siendo la clase dominante a lo largo de todo el cúmulo. Este hallazgo sugiere que, en estos casos, la evolución del cúmulo podría estar en una etapa temprana, donde la relación morfología-densidad aún no se ha establecido plenamente, ya que no son dinámicamente estables (Fasano et al., 2015; Cava et al., 2017), evidenciando de alguna manera un efecto Butcher-Oemler a *redshifts* cercanos como menciona Vollmer (2013).

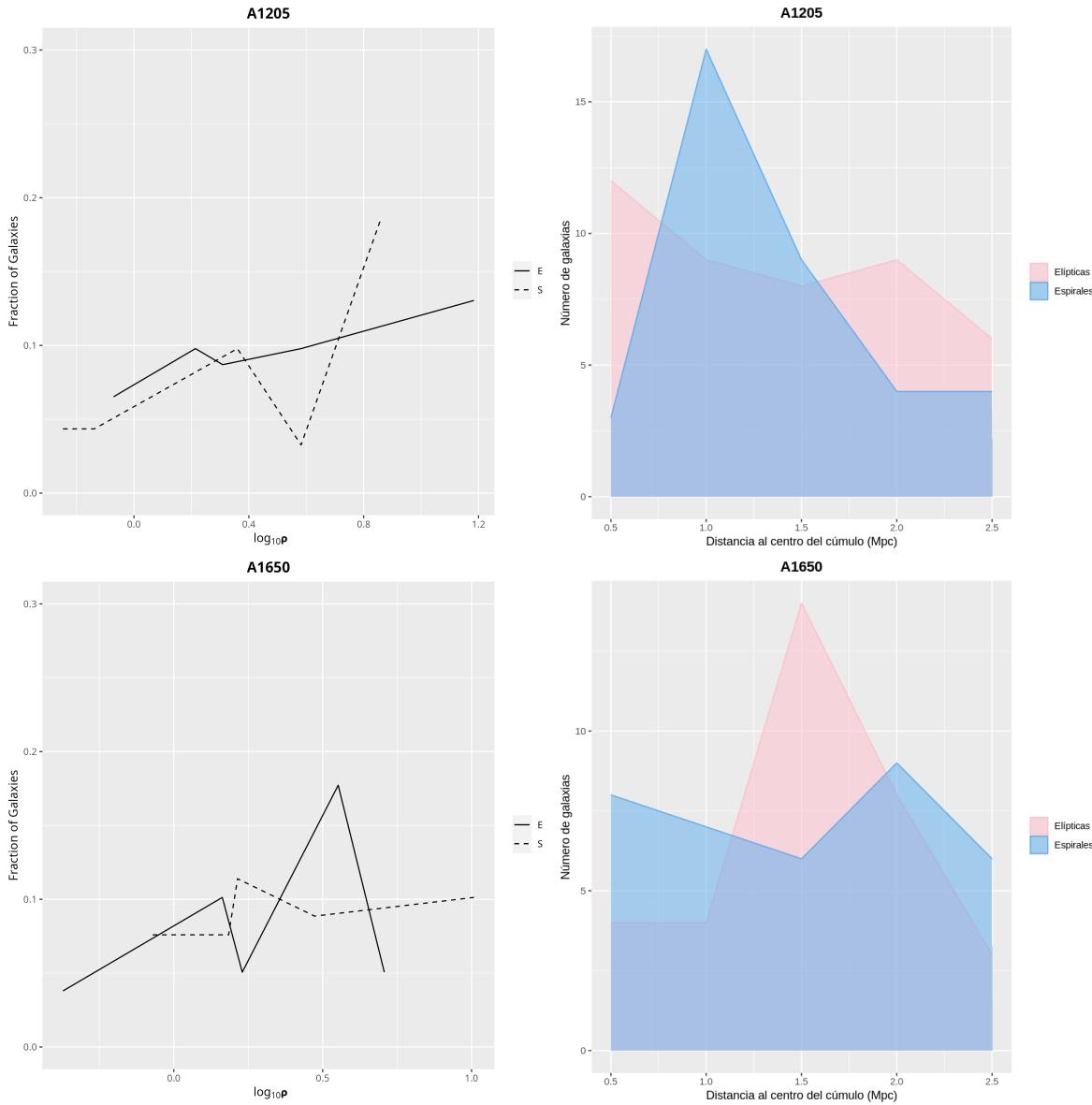


**Figura 4.15:** Diagramas de dos cúmulos en los que las galaxias espirales son mayoritarias a lo largo de todo el rango de densidad. El gráfico de la izquierda describe cómo se relaciona la fracción de cada tipo morfológico con la densidad proyectada, lo que evidencia que las curvas espirales predominan en comparación con las elípticas, tal y como se reafirma en los gráficos de distribución radial de la derecha, donde se observa el número de galaxias vs el radio en Mpc. Además, en A1750, se detecta una variación en las regiones de densidad intermedia, donde las elípticas son más abundantes, aunque en el resto del cúmulo se conserva el patrón visto en A2244.

### 3. Tendencia inusual

En los cúmulos **A1205** y **A1650**, el comportamiento no se ajusta a la tendencia morfología-densidad clásica. En A1205, ambas poblaciones morfológicas coexisten en regiones de alta densidad ( $\log_{10} \rho \approx 1.0$ ), y se identifican algunos picos en la fracción de espirales en torno a densidades intermedias. En las zonas de baja densidad ( $\log_{10} \rho \approx 0$ ), las fracciones de espirales y elípticas son similares en número, aunque las elípticas presentan una ligera dominancia y una disminución progresiva del centro hacia el exterior del cúmulo. Por su parte, A1650 presenta un patrón inusual: las espirales son más frecuentes en densidades altas, mientras que las elípticas predominan en densidades bajas, con picos significativos en valores intermedios de densidad, como se expone en la Figura 4.16.

Estos resultados podrían deberse a interacciones tanto entre galaxias como con el cúmulo en sí. Además, la presencia de subestructuras podría alterar el equilibrio dinámico y, en consecuencia, el grado de regularidad del cúmulo, influyendo en los procesos de transformación morfológica de las galaxias, como mencionan Park y Hwang (2009), Fasano et al. (2015) y Sazonova et al. (2020).



**Figura 4.16:** Estos diagramas evidencian una tendencia inusual en la relación morfología-densidad conforme a lo observado en los gráficos de densidad proyectada respecto a la fracción de galaxias por tipo morfológico (panel izquierdo). A1205 muestra una distribución de galaxias elípticas y espirales en todas las densidades, mientras que A1650 contradice la tendencia esperada, con una mayor proporción de espirales en regiones de alta densidad y elípticas en bajas densidades con picos significativos en densidades intermedias, como evidencian los gráficos de distribución radial contra el número de galaxias de cada tipo en el radio (panel derecho).

## Comparación con clasificaciones visuales

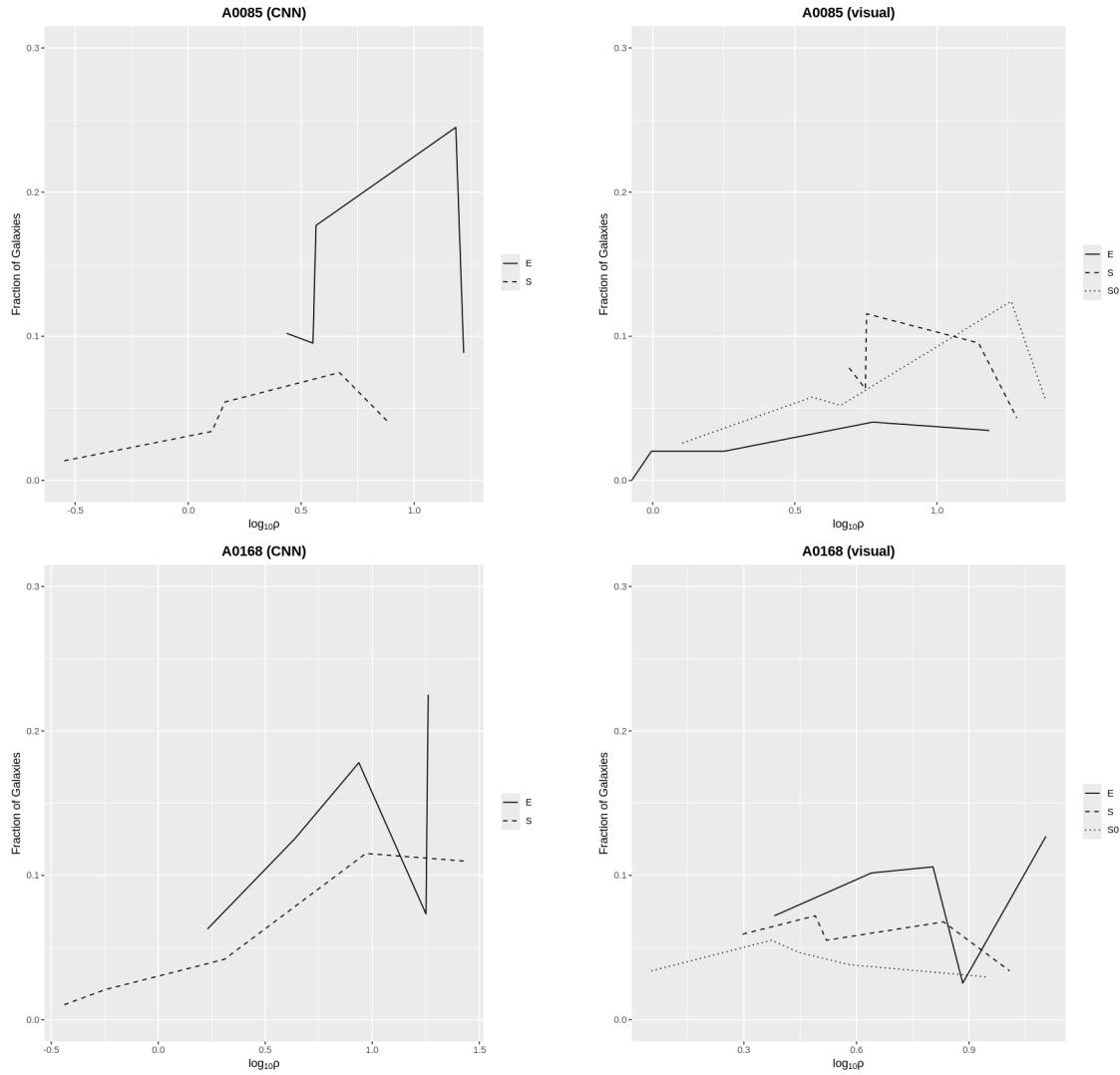
Como parte de un análisis más completo, se decidió examinar de forma independiente los diagramas de los cúmulos A0085 y A0168. Esto se hizo con el propósito de comparar las clasificaciones obtenidas a través de la CNN desarrollada en este trabajo con las obtenidas a partir de inspección visual por expertos en la materia<sup>8</sup>. Esto nos permite evaluar si hay variaciones importantes entre las técnicas de clasificación automatizadas y las humanas, así como de discutir la influencia de dichas diferencias en la comprensión de la relación morfología-densidad.

En la Figura 4.17 se comprueba que, en general, ambas metodologías muestran una tendencia en la que las galaxias elípticas se encuentran en regiones de mayor densidad, mientras que las galaxias espirales son más abundantes en regiones de menor densidad. Sin embargo, sí hay diferencias entre ambas clasificaciones: la CNN categoriza una mayor fracción de galaxia a la clase elíptica en todo el rango de densidades en comparación con la clasificación visual, lo que indica que podría tener un sesgo hacia la detección de estructuras más uniformes. Además, en la clasificación visual, la fracción de galaxias lenticulares (S0), ausente en la CNN, muestra una contribución significativa en densidades intermedias.

A pesar de tener una clasificación visual, A0085 no muestra una tendencia tan marcada como la que se menciona en la investigación de Dressler (consultar Figura 2.4). Esto podría ser resultado de la complicada dinámica presente en cúmulos masivos como A0085 y A0168, la cual se ve afectada por la existencia de subestructuras e interacciones gravitacionales entre galaxias y cúmulos que no hacen claramente notoria la tendencia. Además, los mecanismos de formación y evolución que aún están sucediendo tienen un impacto notable en la transformación morfológica (Boschin et al., 2006; Yu et al., 2016). A esto se suman factores observacionales que pueden afectar la interpretación, como la proyección en el cielo, la selección de miembros y la definición de densidad proyectada.

En resumen, la relación morfología-densidad se confirma en términos generales. No obstante, en cúmulos como A0085 y A0168, esta tendencia puede mostrar variaciones, comportamientos particulares o segregaciones menos evidentes, principalmente por factores como la evolución dinámica, variaciones entre ambos métodos de clasificación y limitaciones propias de las observaciones.

<sup>8</sup>Se agradece al Dr. Juan Manuel Islas Islas por proporcionar los datos de clasificación de las galaxias de ambos cúmulos de forma privada.



**Figura 4.17:** En los diagramas se ilustra la relación morfología-densidad para diferenciar lo obtenido mediante la clasificación de la CNN (izquierda) y la clasificación visual (derecha). El eje x representa el logaritmo de la densidad proyectada ( $\rho$ ) a 2.5 Mpc del centro. El eje y refleja la fracción de galaxias de cada tipo morfológico respecto al total del cúmulo, respectivamente. En ambos casos, se sigue la tendencia general o esperada. No obstante, en **A0085**, la fracción de galaxias elípticas estimada por la CNN es mayor que la determinada visualmente en densidades altas (hacia la derecha del gráfico) mientras que en densidades intermedias y bajas (hacia la izquierda del gráfico) las galaxias espirales dominan pero no crecen significativamente, lo que nos sugiere una importante diferencia. A su vez, en **A0168**, la CNN sugiere una mayor proporción de elípticas en densidades en comparación con la clasificación visual, al mismo tiempo que las galaxias espirales parecen seguir un patrón similar pero sin aumentar en densidades bajas.

# CAPÍTULO 5

---

## CONCLUSIONES

---

Los resultados de este estudio muestran tres patrones principales en los diagramas morfología-densidad: (1) **Esperado**, donde concuerda con la relación prevista aunque las elípticas predominan en todo el rango de densidades (probablemente debido a la sobreestimación de esta clase por la CNN); (2) **Dominio de espirales**, donde estas galaxias superan en número a las elípticas tanto en regiones de baja como de alta densidad, por un posible efecto Butcher-Oemler a *redshifts* cercanos; (3) **Inusual**, caracterizado por la falta de una segregación clara y la presencia inesperada de espirales en entornos de alta densidad que puede asociarse a la irregularidad del cúmulo, la presencia de subestructuras o a la influencia de interacciones gravitacionales e hidrodinámicas.

Estas variaciones respecto a la tendencia clásica de Dressler reflejan la complejidad de los procesos dinámicos dentro de los cúmulos; destaca el caso de A0085 porque, aun cuando ha sido estudiado mediante clasificación visual, la relación morfología-densidad no es tan clara. En este contexto, el uso de una CNN entrenada con datos de GZ y SDSS ha permitido una clasificación estructurada y replicable. Sin embargo, el sesgo inherente a la red introduce un error estimado del  $\sim 22.5\%$ , lo que destaca la importancia de seguir refinando estos modelos.

Futuras investigaciones, podrían profundizar en los procesos físicos que afectan la distribución morfológica y continuar desarrollando métodos automáticos más precisos para analizar amplias muestras. Mejorar la arquitectura de la CNN, ampliar el conjunto de entrenamiento con imágenes más nítidas y etiquetas que incluyan la clase S0, así como integrar información en distintas longitudes de onda (radio, infrarrojo y rayos X), brindaría nuevas perspectivas sobre la evolución de las galaxias en los cúmulos.

---

## BIBLIOGRAFÍA

---

- Abell, G. (1958). "The distribution of rich clusters of galaxies". En: *Astrophysical Journal Supplement* 3, págs. 211-288. DOI: [10.1086/190036](https://doi.org/10.1086/190036).
- Alzubaidi, L. et al. (2021). "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions". En: *Journal of Big Data* 8.1, pág. 53. DOI: [10.1186/s40537-021-00444-8](https://doi.org/10.1186/s40537-021-00444-8).
- Annunziatella, M. et al. (2023). "Lack of influence of the environment in the earliest stages of massive galaxy formation". En: *Monthly Notices of the Royal Astronomical Society* 519.1, págs. 1476-1493. DOI: [10.1093/mnras/stac2731](https://doi.org/10.1093/mnras/stac2731).
- Bahcall, N. (1999). "Clusters and superclusters of galaxies". En: *Formation of Structure in the Universe*, pág. 135.
- Barchi, P. et al. (2020). "Machine and Deep Learning applied to galaxy morphology - A comparative study". En: *Astronomy and Computing* 30, 100334, pág. 100334. DOI: [10.1016/j.ascom.2019.100334](https://doi.org/10.1016/j.ascom.2019.100334).
- Böhringer, H. y N. Werner (2010). "X-ray spectroscopy of galaxy clusters: studying astrophysical processes in the largest celestial laboratories". En: *Astronomy and Astrophysics Reviews* 18.1-2, págs. 127-196. DOI: [10.1007/s00159-009-0023-3](https://doi.org/10.1007/s00159-009-0023-3).
- Bond, J., L. Kofman y D. Pogosyan (1996). "How filaments of galaxies are woven into the cosmic web". En: *Nature* 380.6575, págs. 603-606. DOI: [10.1038/380603a0](https://doi.org/10.1038/380603a0).
- Boschin, W. et al. (2006). "Internal dynamics of the radio halo cluster Abell 2744". En: *Astronomy & Astrophysics* 449.2, págs. 461-474. DOI: [10.1051/0004-6361:20054408](https://doi.org/10.1051/0004-6361:20054408).
- Boselli, A., M. Fossati y M. Sun (2022). "Ram pressure stripping in high-density environments". En: *Astronomy and Astrophysics Reviews* 30.1, pág. 3. DOI: [10.1007/s00159-022-00140-3](https://doi.org/10.1007/s00159-022-00140-3).
- Bower, R. (2013). "Clusters of Galaxies". En: *Planets, Stars and Stellar Systems. Volume 6: Extragalactic Astronomy and Cosmology*. Ed. por T. Oswalt y W. Keel. Springer, pág. 265. DOI: [10.1007/978-94-007-5609-0\\_6](https://doi.org/10.1007/978-94-007-5609-0_6).
- Buta, R. (2013). "Galaxy Morphology". En: *Secular Evolution of Galaxies*. Ed. por J. Falcón-Barroso y J. Knapen. Cambridge University Press & Assessment, pág. 155. DOI: [10.48550/arXiv.1304.3529](https://doi.org/10.48550/arXiv.1304.3529).

- Butcher, H. y A. Oemler Jr. (1978). “The evolution of galaxies in clusters. I. ISIT photometry of Cl 0024+1654 and 3C 295.” En: *The Astrophysical Journal* 219, págs. 18-30. DOI: [10.1086/155751](https://doi.org/10.1086/155751).
- Carroll, B. (2014). *An introduction to Modern Astrophysics Second Edition*. Pearson Education Limited.
- Cava, A. et al. (2017). “Structural and dynamical modeling of WINGS clusters. I. The distribution of cluster galaxies of different morphological classes within regular and irregular clusters”. En: *Astronomy & Astrophysics* 606, A108. DOI: [10.1051/0004-6361/201730785](https://doi.org/10.1051/0004-6361/201730785).
- Chandra X-ray Observatory (2008). *Dark Energy and the Universe*. URL: <https://chandra.harvard.edu/photo/2008/darkenergy/>.
- Cheng, T. et al. (2021). “Galaxy morphological classification catalogue of the Dark Energy Survey Year 3 data with convolutional neural networks”. En: *Monthly Notices of the Royal Astronomical Society* 507.3, págs. 4425-4444. DOI: [10.1093/mnras/stab2142](https://doi.org/10.1093/mnras/stab2142).
- Desai, V. et al. (2007). “The Morphological Content of 10 EDisCS Clusters at  $0.5 < z < 0.8$ ”. En: *The Astrophysical Journal* 660.2, págs. 1151-1164. DOI: [10.1086/513310](https://doi.org/10.1086/513310).
- Diaferio, A. (1999). “Mass estimation in the outer regions of galaxy clusters”. En: *Monthly Notices of the Royal Astronomical Society* 309.3, págs. 610-622. DOI: [10.1046/j.1365-8711.1999.02864.x](https://doi.org/10.1046/j.1365-8711.1999.02864.x).
- Domínguez Sánchez, H. et al. (2018). “Improving galaxy morphologies for SDSS with Deep Learning”. En: *Monthly Notices of the Royal Astronomical Society* 476.3, págs. 3661-3676. DOI: [10.1093/mnras/sty338](https://doi.org/10.1093/mnras/sty338).
- Dressler, A. (1980). “Galaxy morphology in rich clusters-Implications for the formation and evolution of galaxies”. En: *Astrophysical Journal* 236, págs. 351-365. DOI: [10.1086/157753](https://doi.org/10.1086/157753).
- Dressler, A. et al. (1997). “Evolution since  $z = 0.5$  of the Morphology-Density Relation for Clusters of Galaxies”. En: *The Astrophysical Journal* 490.2, págs. 577-591. DOI: [10.1086/304890](https://doi.org/10.1086/304890).
- Fasano, G. et al. (2015). “Morphological fractions of galaxies in WINGS clusters: revisiting the morphology-density paradigm”. En: *Monthly Notices of the Royal Astronomical Society* 449.4, págs. 3927-3944. DOI: [10.1093/mnras/stv500](https://doi.org/10.1093/mnras/stv500).
- Gifford, D. et al. (2013). “A Systematic Analysis of Caustic Methods for Galaxy Cluster Masses”. En: *The Astrophysical Journal* 773.2, pág. 116. DOI: [10.1088/0004-637X/773/2/116](https://doi.org/10.1088/0004-637X/773/2/116).
- Goodfellow, I., Y. Bengio y A. Courville (2016). *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press.
- Goto, T. et al. (2003). “The morphology-density relation in the Sloan Digital Sky Survey”. En: *Monthly Notices of the Royal Astronomical Society* 346.2, págs. 601-614. DOI: [10.1046/j.1365-2966.2003.07114.x](https://doi.org/10.1046/j.1365-2966.2003.07114.x).
- Houghton, R. (2015). “Revisiting the original morphology-density relation”. En: *Monthly Notices of the Royal Astronomical Society* 451.4, págs. 3427-3436. DOI: [10.1093/mnras/stv1113](https://doi.org/10.1093/mnras/stv1113).

- Khan, A. et al. (2020). “A survey of the recent architectures of deep convolutional neural networks”. En: *Artificial Intelligence Review* 53.8, págs. 5455-5516. DOI: [10.1007/s10462-020-09825-6](https://doi.org/10.1007/s10462-020-09825-6).
- Kingma, D. y J. Ba (2014). “Adam: A Method for Stochastic Optimization”. En: *arXiv e-prints*, arXiv:1412.6980. DOI: [10.48550/arXiv.1412.6980](https://doi.org/10.48550/arXiv.1412.6980).
- Kodama, T. y R. Bower (2001). “Reconstructing the history of star formation in rich cluster cores”. En: *Monthly Notices of the Royal Astronomical Society* 321.1, págs. 18-36. DOI: [10.1046/j.1365-8711.2001.03981.x](https://doi.org/10.1046/j.1365-8711.2001.03981.x).
- Lourenço, A. et al. (2023). “The effect of cluster dynamical state on ram-pressure stripping”. En: *Monthly Notices of the Royal Astronomical Society* 526.4, págs. 4831-4847. DOI: [10.1093/mnras/stad2972](https://doi.org/10.1093/mnras/stad2972).
- Max Planck Society (2019). *Heavyweight black hole in Abell 85*. URL: <https://www.mpg.de/14210061/heavyweight-black-hole-abell-85>.
- Mei, S. et al. (2023). “Morphology-density relation, quenching, and mergers in CARLA clusters and protoclusters at  $1.4 < z < 2.8$ ”. En: *Astronomy & Astrophysics* 670, A58. DOI: [10.1051/0004-6361/202243551](https://doi.org/10.1051/0004-6361/202243551).
- Mo, H., F. Van den Bosch y S. White (2010). *Galaxy formation and evolution*. Cambridge University Press.
- NASA Science (2024). *Types of Galaxies*. URL: <https://science.nasa.gov/universe/galaxies/types/>.
- Nuijten, M. et al. (2005). “The Canada-France-Hawaii Telescope Legacy Survey: The Morphology-Density Relation of Galaxies out to  $z \sim 1$ ”. En: *The Astrophysical Journal* 626.2, págs. L77-L80. DOI: [10.1086/431755](https://doi.org/10.1086/431755).
- O’Shea, K. y R. Nash (2015). “An Introduction to Convolutional Neural Networks”. En: *arXiv e-prints*, arXiv:1511.08458. DOI: [10.48550/arXiv.1511.08458](https://doi.org/10.48550/arXiv.1511.08458).
- Park, C. y H. Hwang (2009). “Interactions of Galaxies in the Galaxy Cluster Environment”. En: *The Astrophysical Journal* 699.2, págs. 1595-1609. DOI: [10.1088/0004-637X/699/2/1595](https://doi.org/10.1088/0004-637X/699/2/1595).
- Pérez-Millán, D. et al. (2023). “The relation between morphology, star formation history, and environment in local Universe galaxies”. En: *Monthly Notices of the Royal Astronomical Society* 521.1, págs. 1292-1315. DOI: [10.1093/mnras/stad542](https://doi.org/10.1093/mnras/stad542).
- Pfeffer, J. et al. (2023). “The galaxy morphology–density relation in the EAGLE simulation”. En: *Monthly Notices of the Royal Astronomical Society* 518.4, págs. 5260-5278. DOI: [10.1093/mnras/stac3466](https://doi.org/10.1093/mnras/stac3466).
- Popesso, P. et al. (2004). “RASS-SDSS Galaxy clusters survey-I. The catalog and the correlation of X-ray and optical properties”. En: *Astronomy & Astrophysics* 423.2, págs. 449-467. DOI: [10.1051/0004-6361:20035818](https://doi.org/10.1051/0004-6361:20035818).
- Postman, M. et al. (2005). “The Morphology-Density Relation in  $z \sim 1$  Clusters”. En: *The Astrophysical Journal* 623.2, págs. 721-741. DOI: [10.1086/428881](https://doi.org/10.1086/428881).
- Reza, M. (2021). “Galaxy morphology classification using automated machine learning”. En: *Astronomy and Computing* 37, págs. 100492. DOI: [10.1016/j.ascom.2021.100492](https://doi.org/10.1016/j.ascom.2021.100492).
- Sarazin, C. (1988). *X-ray emission from clusters of galaxies*. Cambridge University Press.

- Sazonova, E. et al. (2020). "The Morphology-Density Relationship in  $1 < z < 2$  Clusters". En: *The Astrophysical Journal* 899.1, pág. 85. DOI: [10.3847/1538-4357/aba42f](https://doi.org/10.3847/1538-4357/aba42f).
- Serra, A. et al. (2011). "Measuring the escape velocity and mass profiles of galaxy clusters beyond their virial radius". En: *Monthly Notices of the Royal Astronomical Society* 412.2, págs. 800-816. DOI: [10.1111/j.1365-2966.2010.17946.x](https://doi.org/10.1111/j.1365-2966.2010.17946.x).
- Spinrad, H. (2005). *Galaxy formation and evolution*. Springer Science & Business Media.
- Thomas, D. (2011). "And the Winner Is: Galaxy Mass". En: *Environment and the Formation of Galaxies: 30 years later*. Ed. por I. Ferreras y A. Pasquali. Vol. 27. Astrophysics and Space Science Proceedings, pág. 89. DOI: [10.1007/978-3-642-20285-8\\_17](https://doi.org/10.1007/978-3-642-20285-8_17).
- Thomas, D. et al. (2003). "Stellar population models of Lick indices with variable element abundance ratios". En: *Monthly Notices of the Royal Astronomical Society* 339.3, págs. 897-911. DOI: [10.1046/j.1365-8711.2003.06248.x](https://doi.org/10.1046/j.1365-8711.2003.06248.x).
- Torchvision Contributors (2024). *Torchvision Models Documentation*. <https://pytorch.org/vision/0.18/models/generated/torchvision.models.resnet50.html>.
- Urechiatu, R. y M. Frincu (2024). "Improved Galaxy Morphology Classification with Convolutional Neural Networks". En: *Social Science Research Network*. DOI: [10.2139/ssrn.4699260](https://doi.org/10.2139/ssrn.4699260).
- van der Wel, A. (2008). "The morphology-density relation: a constant of nature". En: *Formation and Evolution of Galaxy Bulges*. Vol. 245. IAU Symposium, págs. 59-62. DOI: [10.1017/S1743921308017286](https://doi.org/10.1017/S1743921308017286).
- Voges, W. et al. (1999). "The ROSAT all-sky survey bright source catalogue". En: *Astronomy & Astrophysics* 349, págs. 389-405. DOI: [10.48550/arXiv.astro-ph/9909315](https://doi.org/10.48550/arXiv.astro-ph/9909315).
- Voit, G. (2005). "Tracing cosmic evolution with clusters of galaxies". En: *Reviews of Modern Physics* 77.1, págs. 207-258. DOI: [10.1103/RevModPhys.77.207](https://doi.org/10.1103/RevModPhys.77.207).
- Vollmer, Bernd (2013). "The Influence of Environment on Galaxy Evolution". En: *Planets, Stars and Stellar Systems. Volume 6: Extragalactic Astronomy and Cosmology*. Ed. por T. Oswalt y W. Keel. Vol. 6. Springer, pág. 207. DOI: [10.1007/978-94-007-5609-0\\_5](https://doi.org/10.1007/978-94-007-5609-0_5).
- Whitmore, B. et al. (1993). "What Determines the Morphological Fractions in Clusters of Galaxies?" En: *The Astrophysical Journal* 407, pág. 489. DOI: [10.1086/172531](https://doi.org/10.1086/172531).
- Willett, K. et al. (2013). "Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey". En: *Monthly Notices of the Royal Astronomical Society* 435.4, págs. 2835-2860. DOI: [10.1093/mnras/stt1458](https://doi.org/10.1093/mnras/stt1458).
- Wright, E. (2006). "A Cosmology Calculator for the World Wide Web". En: *The Astronomical Society of the Pacific* 118.850, págs. 1711-1715. DOI: [10.1086/510102](https://doi.org/10.1086/510102).
- York, D. et al. (2000). "The Sloan Digital Sky Survey: Technical Summary". En: *The Astronomical Journal* 120.3, págs. 1579-1587. DOI: [10.1086/301513](https://doi.org/10.1086/301513).
- Yu, H. et al. (2016). "The Unrelaxed Dynamical Structure of the Galaxy Cluster Abell 85". En: *The Astrophysical Journal* 831.2, 156, pág. 156. DOI: [10.3847/0004-637X/831/2/156](https://doi.org/10.3847/0004-637X/831/2/156).
- Zhang, A. et al. (2023). *Dive into Deep Learning*. <https://D2L.ai>. Cambridge University Press.

---

## APÉNDICES

---

### Apéndice 1: Código en Python de la CNN

```
1 # Cargar librerías y módulos
2 import torch
3 import torch.nn as nn
4 import torch.optim as optim
5 import torch.nn.functional as F
6 import matplotlib.pyplot as plt
7 import pandas as pd
8 import numpy as np
9 import csv
10 import time
11 import os
12 from sklearn.model_selection import train_test_split
13 from torch.utils.data import DataLoader, Dataset
14 from torchvision import datasets, transforms
15 from torchvision.datasets import ImageFolder
16 from PIL import Image
17
18 # Definir modelo
19 class GalaxyClassifier(nn.Module):
20     def __init__(self):
21         super(GalaxyClassifier, self).__init__()
22         self.conv1 = nn.Conv2d(in_channels=3, out_channels=64,
23                             kernel_size=3, stride=1, padding=1)
24         self.conv2 = nn.Conv2d(in_channels=64, out_channels=32,
25                             kernel_size=3, stride=1, padding=1)
26         self.relu = nn.ReLU(inplace=True)
27         self.pool = nn.MaxPool2d(kernel_size=3, stride=2, padding=0,
28                                ceil_mode=False)
29         self.flatten = nn.Flatten()
30         self.fc1 = nn.Linear(32 * 36 * 36, 96)
31         self.fc2 = nn.Linear(96, 2) # 2 clases: E y S
```

```

29
30     def forward(self, x):
31         # Extracción de características
32         x = F.relu(self.conv1(x))
33         x = self.pool(x)
34         x = F.relu(self.conv2(x))
35         x = self.pool(x)
36         # Aplanar la salida
37         x = self.flatten(x)
38         # Clasificación
39         x = F.relu(self.fc1(x))
40         x = torch.sigmoid(self.fc2(x))
41         return x
42
43 # Definir las clases a partir del dataset
44 class GalaxyDataset(Dataset):
45     def __init__(self, data_folder, transform=None):
46         self.dataset = ImageFolder(data_folder, transform=transform)
47     def __len__(self):
48         return len(self.dataset)
49     def __getitem__(self, idx):
50         return self.dataset[idx]
51
52 # Definir transformaciones para el dataset
53 trans = transforms.Compose([
54     transforms.Resize((150, 150)),
55     transforms.ToTensor(),
56     transforms.Normalize([0.485, 0.456, 0.406], [0.229, 0.224, 0.225])
57 ])
58
59 # Cargar el dataset desde Google Drive
60 drive.mount('/content/drive')
61 data_folder = '/content/drive/MyDrive/dataset'
62 !ls $data_folder
63
64 dataset = GalaxyDataset(data_folder, transform=trans)
65 dataloader = torch.utils.data.DataLoader(dataset, batch_size=32,
66                                         shuffle=True)
67
68 # Separar en 'entrenamiento' y 'validación'
69 train_size = int(0.8 * len(dataset))
70 val_size = len(dataset) - train_size
71 train_dataset, val_dataset = torch.utils.data.random_split(dataset, [
72     train_size, val_size])
73
74 # Crear los data loaders
75 batch=64
76 train_loader = DataLoader(train_dataset, batch_size=batch, shuffle=
77                           True)
78 val_loader = DataLoader(val_dataset, batch_size=batch, shuffle=False)
79 print("The number of images in training set is: ", len(train_loader)*
80       batch)
81 print("The number of images in validation set is: ", len(val_loader)*
82       batch)

```

```

78 # Inicializar el modelo, la función de pérdida, y el optimizador
79 model = GalaxyClassifier()
80 learning_rate = 0.0001
81 criterion = nn.CrossEntropyLoss()
82 optimizer = optim.Adam(model.parameters(), lr=0.0001)
83
84 # Loop de entrenamiento
85 train_losses = []
86 val_losses = []
87 val_accuracies = []
88 training_time = 0
89 num_epochs = 30
90
91 for epoch in range(num_epochs):
92     start_time = time.time()
93     model.train()
94     epoch_train_losses = []
95     for images, labels in train_loader:
96         optimizer.zero_grad()
97         outputs = model(images)
98         loss = criterion(outputs, labels)
99         loss.backward()
100        optimizer.step()
101        epoch_train_losses.append(loss.item())
102    # Calcular las pérdidas promedio
103    avg_train_loss = sum(epoch_train_losses) / len(epoch_train_losses)
104    train_losses.append(avg_train_loss)
105
106    # Loop de validación
107    model.eval()
108    with torch.no_grad():
109        epoch_val_losses = []
110        correct = 0
111        total = 0
112        for images, labels in val_loader:
113            outputs = model(images)
114            loss = criterion(outputs, labels)
115            epoch_val_losses.append(loss.item())
116
117            _, predicted = torch.max(outputs.data, 1)
118            total += labels.size(0)
119            correct += (predicted == labels).sum().item()
120
121    # Determinar la precisión y la pérdida
122    avg_val_loss = sum(epoch_val_losses) / len(epoch_val_losses)
123    val_losses.append(avg_val_loss)
124    accuracy = correct / total
125    val_accuracies.append(accuracy)
126
127    end_time = time.time()
128    elapsed_time = end_time - start_time
129    training_time += elapsed_time
130    print(f'Epoch {epoch+1}/{num_epochs}, Validation Accuracy: {accuracy * 100:.2f}%')

```

```

131 # Imprimir el tiempo transcurrido
132 print(f'Total Training Time: {training_time:.2f} seconds')
133
134 # Guardar el modelo entrenado
135 torch.save(model.state_dict(), 'galaxy_classifier.pth')
136
137 # Graficar las curvas de aprendizaje
138 def learning_curve(train_losses, val_losses, val_accuracies):
139     epochs = range(1, len(train_losses) + 1)
140     # Pérdidas
141     plt.figure(figsize=(12, 5))
142     plt.subplot(1, 2, 1)
143     plt.plot(epochs, train_losses, 'r', label='Training loss')
144     plt.plot(epochs, val_losses, 'b', label='Validation loss')
145     plt.title('Training and Validation Loss')
146     plt.xlabel('Epochs')
147     plt.ylabel('Loss')
148     plt.legend()
149
150     # Precisiones
151     plt.subplot(1, 2, 2)
152     plt.plot(epochs, val_accuracies, 'g', label='Validation accuracy')
153     plt.title('Validation Accuracy')
154     plt.xlabel('Epochs')
155     plt.ylabel('Accuracy')
156     plt.legend()
157     plt.tight_layout()
158     plt.show()
159
160 learning_curve(train_losses, val_losses, val_accuracies)

```