Universidad de La Laguna

Facultad de Ciencias

Searching for compaction in the TNG50 cosmological simulation using deep learning

Bachelor's Degree Final Project

Patricia Iglesias Navarro

Supervisor:
Marc Huertas-Company

Julio 2021

I am born into an environment – I know not whence I came nor whither I go nor who I am. This is my situation as yours, every single one of you. The fact that everyone always was in this same situation, and always will be, tells me nothing. Our burning question as to the whence and whither – all we can ourselves observe about it is the present environment. That is why we are eager to find out about it as much as we can. That is science, learning, knowledge; it is the true source of every spiritual endeavor of man. We try to find out as much as we can about the spatial and temporal surroundings of the place in which we find ourselves put by birth. And as we try, we delight in it, we find it extremely interesting (may that not be the end for which we are there?)

Ewin Scrhödinger, 'Science and Humanism'

Me gustaría agradecer a mi tutor por su dedicación durante todo el proyecto, así como a mi familia, por su apoyo en todas las etapas de este viaje.

We optimize a convolutional neural network, intending to study an astrophysical process known as 'blue nuggets' (BN), which consists of a compaction followed by a central quenching that occurs in young galaxies at high redshifts. This network is evaluated with mock 'observed' images of galaxies at three phases of evolution (Pre-BN, BN and Post-BN), generated by the zoom-in hydro-cosmological simulation VELA. We then use this to classify galaxies from the TNG50 simulation in these three phases, and finally, we study their physical properties such as the redshift, the effective radius and the star formation rate (SFR), as well as the masses of gas, of stars, and of the central supermassive black holes. The network successfully detects this compaction phase in the new simulation, consistent with the features observed in VELA galaxies. We highlight the existence of a temporal sequence, together with the fact that the BN phase forms stars while the Post-BN does not. Furthermore, the BN phase is associated with a gas mass peak at $z \sim 2$ and with a smaller radius.

El objetivo del trabajo es optimizar una red neuronal convolucional (CNN), para extraer información relativa a una posible fase de la evolución de las galaxias en el Universo temprano, denominada compactación o 'blue nuggets' (BN), a la que sigue una extinción (se frena la formación de estrellas). Esta técnica de procesamiento y clasificación de imágenes ha sido particularmente exitosa en la astrofísica, una disciplina en la que no son viables los experimentos controlados y que por ende se basa casi exclusivamente en la observación. Las nuevas generaciones de telescopios (Euclid, LSST, JWST), unidas a todos los proyectos pasados y actuales, contribuyen a una cantidad de datos sin precedentes, imposibles de procesar con métodos tradicionales. Por esta razón, el empleo de nuevos algoritmos es vital para el progreso, en particular en la rama de la evolución de las galaxias, donde este está condicionado por las escalas temporales involucradas, y por la degeneración entre los observables y los procesos físicos.

El proyecto consiste en optimizar el modelo de CNN presentado en [Huertas-Company et al., 2018], evaluándolo con galaxias jóvenes a altos redshifts simuladas por la tercera generación de la simulación hidrodinámica VELA, que clasificamos en tres fases: la Pre-BN (caracterizada por una apariencia grumosa), la BN (la compactación) y la Post-BN (la extinción central). Estas galaxias reproducen las características de CANDELS al ser procesadas con la respuesta al impulso (PSF por sus siglas en inglés), y el error del HST. El entrenamiento y evaluación de la CNN se realiza con 10.310 imágenes de 28 galaxias en el rango $1.0 \lesssim z \lesssim 2.2$ (generadas con diferentes orientaciones de la cámara y en los tres filtros F160W, F125W, y F105W). Una vez se obtiene una mejora significativa al modificar los hiperparámetros de la red (como el número de neuronas de las capas o el 'dropout'), se procede a clasificar imágenes de TNG50. Esta simulación, aunque presenta menor resolución (70-140 pc frente a los 17-35 pc de VELA), recoge un conjunto mucho más extenso de galaxias, incluyendo un mayor rango de masas, así como AGN, por lo que se emplea para estudiar las limitaciones de VELA, ya que además cuenta con modelos de sub-resolución distintos. Trabajamos con 11.047 imágenes de 10.803 galaxias en el rango $0.5 \lesssim z \lesssim 3.0$ con $\log (M_*/M_\odot) > 9.0$ en los mismos tres filtros.

Así, el trabajo comienza con una introducción, en la que se recoge la motivación del proyecto y su descripción, y con una sección teórica, en la que se explican los conceptos fundamentales relativos a las simulaciones, a las propiedades físicas que definen la compactación y al funcionamiento de una CNN. A continuación, encontramos una sección de objetivos, en la que brevemente se indica qué esperamos conseguir, en definitiva buscamos comprobar que la simulación TNG50 presenta una fase BN con unas propiedades determinadas. También se incluyen las especificaciones más técnicas de los datos con los que trabajamos, y en otra sección la metodología que se lleva a cabo en las dos partes bien diferenciadas del proyecto, la optimización de la red y el estudio de las fases en la simulación TNG50. Posteriormente se incluyen los resultados, que se exponen de forma concisa en los siguientes párrafos.

Pese a obtener una exactitud superior al 80% para las galaxias de prueba de VELA, el

conjunto de imágenes de esta simulación está desbalanceado entre las tres categorías, pues la clase Pre-BN es minoritaria. Esto repercute negativamente en la clasificación de TNG50. No obstante, trabajamos con diferentes umbrales de probabilidad para las tres clases (la red nos devuelve la probabilidad de pertenecer a cada una de ellas para cada galaxia), de forma que podemos continuar con el estudio de este fenómeno en TNG50. Analizamos cómo se distribuye el redshift entre estas tres fases, así como el radio efectivo, la densidad estelar proyectada (en 2D) en el kpc central (Σ_1) y la formación estelar (SFR por sus siglas en inglés). Concretamente, buscamos reproducir una distribución con forma de L en el plano SFR-Σ_1, introducida en trabajos previos. También estudiamos la evolución de las masas del gas, del agujero negro supermasivo y de las estrellas con el redshift, junto con la de la SFR.

De acuerdo con lo esperado, al analizar la distribución de las fases con el redshift en ambas simulaciones, observamos una secuencia temporal, de modo que las galaxias Pre-BN se asocian con mayores redshifts, es decir, con tiempos más lejanos, mientras que las Post-BN tienen los redshifts más bajos, correspondientes a épocas más recientes. La fase BN se localiza entre ambas. Además de verificar que las galaxias en fase BN son las más compactas, estudiando el radio, obtenemos que las galaxias clasificadas como BN presentan altas tasas de formación estelar, por lo que se encuentran formando estrellas, a diferencia de las Post-BN, que tienen SFR próximas a cero y por tanto no forman estrellas. También comprobamos que el pico de gas molecular, cuando se estudia la evolución de su masa con el redshift, se asocia exitosamente con la fase BN, y que la evolución de la masa del resto de componentes de las galaxias presenta el comportamiento observado en trabajos anteriores. Sin embargo, no se reproducen las distribuciones esperadas para Σ_1, pues la fase BN no se asocia con sus valores máximos. Este parámetro es difícil de medir, especialmente si se tiene en cuenta el ruido del HST con el que han sido procesadas las imágenes, por lo que es difícil concluir algo claro al respecto.

Tras el apartado de resultados incluimos las conclusiones de forma esquematizada. Luego discutimos los problemas planteados: las pocas galaxias clasificadas en la fase Pre-BN y la distribución obtenida para Σ_1. Como continuación del proyecto se propone emplear nuevos métodos para trabajar con conjuntos de imágenes desbalanceados (penalizando a las clases mayoritarias, generando muestras sintéticas, etc.), así como analizar estas fases con otras simulaciones, con modelos sub-resolutivos diferentes pero con las mismas condiciones iniciales, por ejemplo con VELA Generation 6, para evaluar el impacto de las limitaciones de la simulación en el estudio del fenómeno.

Contents

1   Introduction

En esta sección se introducen los obstáculos principales del estudio de la evolución de las galaxias, presentando el Deep Learning como una herramienta capaz de identificar los fenómenos físicos que tienen lugar. Se exponen brevemente los resultados presentados en [Huertas-Company et al., 2018], a la hora de clasificar la fase de compactación 'blue nuggets' en galaxias de la simulación hidrodinámica de alta resolución VELA. Por último se incluye la motivación del trabajo: evaluar el impacto de las limitaciones de dicha simulación en el análisis del fenómeno, mediante su estudio en TNG50.

In the ~ 14 billion years since the Big Bang, large molecular clouds in structures known as galaxies have collapsed to form stars, creating a heterogeneous universe from quasi-homogeneous initial conditions. The evolution of galaxies and their consequent morphological transformations remain actively debated, even though larger and better quality imaging and spectroscopic datasets have become available in the past years (e.g CANDELS, SDSS, etc.) The near future is even more promising, with big data sets becoming available such as Euclid (2022), Roman (2025) and the LSST (2023). The volume and complexity of these future datasets challenge traditional analysis techniques.

A major issue is the time scales involved, which do not allow to follow the evolution of each galaxy individually, but also the degeneracy between the observable features and the physical processes, since many processes may cause the same observable effects. These circumstances are further aggravated if we keep in mind that many of the images we have will never be seen by a human eye.

In this context, Deep Learning (DL) is presented as a useful tool that can help to constrain some of these physical processes. Its algorithms have been rapidly implemented in various scientific disciplines [Wei et al., 2017, Wei and Roberts, 2018, Heinzinger et al., 2019], but are especially useful in astrophysics, where the experimentation margin is very limited and the observation is so absolutely crucial. Machine Learning techniques have proven to be successful to extract physical relevant features from images in several situations [Shallue and Vanderburg, 2018, Armstrong and Fletcher, 2019, Sarmiento et al., 2021].

In particular, [Huertas-Company et al., 2018] applied supervised DL to detect and describe a process known as 'blue nuggets' (BN). It consists of a compact star-forming phase characteristic of the central regions of young galaxies at high redshifts. As the cosmological simulations show, it is preceded by an earlier phase characterized by a clumpy gas appearance, that we may call Pre-BN phase, and it is followed by a central quenching, a Post-BN phase.

Specifically, they used a sequential convolutional neural network (CNN) with a set of high-resolution hydro-cosmological simulations (VELA), which included 35 intermediate-mass galax-

ies. Once the mock images were generated, they got transformed under a process we may call CANDELization, whose purpose was to ensure that the galaxies were simulated with the same depth as the real CANDELS data [Koekemoer et al., 2011], and correctly reproduced the characteristic noise of HST (we will focus later on it). Thus, they used 3 infrared filters and 19 different camera orientations at steps of $\Delta a = 0.01$ of the expansion factor in order to generate a dataset of approximately 10.000 images in each filter. Finally, 70% of global accuracy in the classification was achieved and logical connections were found between these gaseous compactions and physical parameters such as stellar mass distributions, central mass densities, redshifts and gas fractions. Furthermore, a temporal sequence between the three phases was also distinguished.

Despite its success, the main limitation of their work is the size of the dataset they trained with, according to the volume of VELA simulation, too small to fully characterize all CANDELS data. For instance, AGNs are not included, and the range of masses is limited. To further investigate the impact of these limitations, we will use a new available simulated dataset, TNG50, with a volume of $51.7^3\ Mpc^3$. Even when VELA's maximum resolution is $\sim$17 pc while for TGN50 it is $\sim$70 pc, a detailed look at the structural properties of galaxies and the gas around them will be also possible. Obviously, we must keep in mind that simulations are still unable to capture the complexity of galaxy formation and then, the project is based on a working hypothesis. In the present work, we will try to identify the BN phase in this new dataset and to check if it is possible to establish the same connection with physical quantities.

2 Theoretical Background

En esta sección se incluye una descripción teórica de los conceptos con los que trabajamos. En primer lugar, se explica brevemente en qué consisten las simulaciones cosmológicas, los distintos enfoques posibles y las limitaciones que presentan. También se añaden las características principales de VELA y TNG50. A continuación, se introducen las magnitudes físicas que se emplean (R_e, SFR, Σ_1, etc.) Para terminar, se realiza un resumen con las nociones y la terminología básicas de las CNN, necesarias para una comprensión completa del proyecto.

2.1 Simulations

Most of the natural sciences are based on experimentation for the development of theories, using the process known as the scientific method. In summary, observation leads to a hypothesis, which must be tested several times with different conditions before being validated. But what happens when the closest laboratory is millions of light-years away? This is the day-to-day in galaxy studies, and in astrophysics in general. There are no controlled experiments and of course we can not go back in time to study how a single variable influences the evolution of the universe.

In this context, the first simulations emerge. The first recorded N-body simulation was done in 1941 by Erik Holmberg [Holmberg, 1941]. It consisted of two elliptical galaxies interacting in a plane that were simulated using an optical analog computer. Since then, there has been significant growth, both in size and resolution, of cosmological simulations, as well as in the implemented algorithms.

Modern cosmological simulations reproduce galaxy formation over the last 14 billion years from the first initial conditions after the Big Bang until today. The current cosmological model is called Lambda-CDM (Lambda Cold Dark Matter). The model establishes that the energy budget of the universe is dominated by dark energy (DE) and dark matter (DM). Baryons (all ordinary matter observed as atoms, chemical elements, gas and plasma), account only for $\sim 5\%$ of the total energy budget. In this model, structures originate from primordial fluctuations under the effect of gravity (see figure [1]).

Typically, DM evolution can be followed into the non-linear regime using collision-less N-body simulations. However, these simulations cannot predict the distribution of galaxies made up of baryonic matter. There are two approaches to achieve it: a post-processing procedure, known as semi-analytic modeling, in which the output of N-body simulations is complemented with simple physical prescriptions to estimate the distribution of galaxies; or directly accounting for the baryonic component (gas, stars, supermassive black holes, etc.), in cosmological simulations that include both hydrodynamics and gravity. This second approach is self-consistent and

fully predictive, but the computational challenges involved have hampered its development. So, large-scale predictions of the galaxy population have been typically obtained with semi-analytic methods. However, hydrodynamical simulations of galaxy formation have become more prominent in the last years with successful results [Springel et al., 2005, Li et al., 2021].
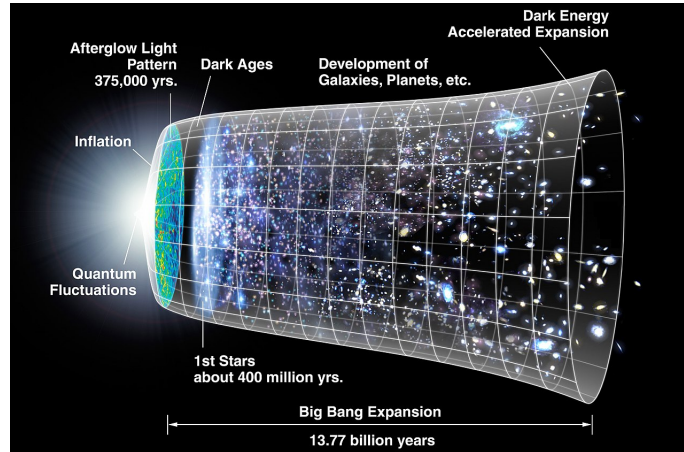

Figure 1: Timeline of the universe. A representation of the evolution of the universe over 13.77 billion years. The far left depicts the earliest moment we can now probe, when a period of "inflation" produced a burst of exponential growth in the universe (size is depicted by the vertical extent of the grid in this graphic.) For the next several billion years, the expansion of the universe gradually slowed down as the matter in the universe pulled on itself via gravity. More recently, the expansion has begun to speed up again as the repulsive effects of dark energy have come to dominate the expansion of the universe. The afterglow light seen by WMAP was emitted about 375,000 years after inflation and has traversed the universe largely unimpeded since then. The conditions of earlier times are imprinted on this light; it also forms a backlight for later developments of the universe. [NASA / WMAP Science Team, 2006].

Starting from initial conditions deduced from the cosmic microwave background radiation and its power spectrum, they are performed with powerful algorithms, and refinements (P^3M, ART, TreePM) are used to increase the resolution. The main difficulty these simulations face is solving at the same time the formation of large-scale structures (clusters, filaments) and the smaller relevant scales (interstellar medium, molecular clouds, star formation, feedback, galactic structure). For this reason, they are divided into two approaches: low resolution - big volume simulations (OWLs, EAGLE, Horizon-AGN, Illustris, TNG) and high-resolution zoom-in simulations (ART / VELA, FIRE, AGORA). While the first ones attempt to simulate a certain volume (box) with multiple galaxy populations to study their interactions, the second ones focus on a specific set of galaxies and trace their evolution from the big bang. Until the computing capacity increases sufficiently, we must work with the two in equilibrium to learn more about the complex problem of galaxy formation. It should be noted that even zoom-in
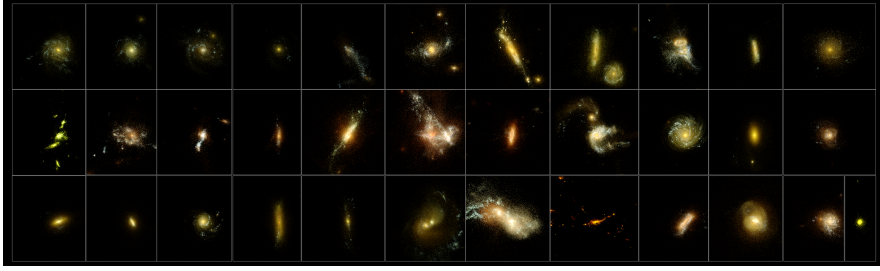
Figure 2: 35 galaxies of VELA simulation. [Moody, 2013].

simulations require a subgrid of many physical processes that occur in the evolution of the universe, based on sub-resolution models for star formation, radiative cooling processes, stellar evolution and chemical enrichment procedures, stellar feedback, and feedback from active galactic nuclei (AGN).

In the present work, we attempt to transfer some results obtained for a zoom-in simulation (VELA) to a much more extensive one (TNG50). In the following, we will briefly expose the characteristics that each one presents:

- The VELA simulation [Moody, 2013] spans the cosmic time evolution of 35 galaxies (with a typical mass at z∼1 of $10^{11} - 10^{12} M_\odot$) over 10-50 timesteps ($1 \lesssim z \lesssim 4$). Each simulation represents the time evolution of an individual galaxy forming, which are labeled from VELA01 through VELA35. They were performed in a WMAP5 cosmology with the gas dynamics and N-body ART code. In addition to gas dynamics and gravity, the simulations implement sub-grid physical models such as cooling, background photoionization, star formation, metal enrichment, thermal feedback from supernovae, and feedback from radiation pressure and heating from young stars. In particular, its high resolution (between 17-35 pc at all times), allows to study the streams of high redshift galaxies that determine the disk and bulge evolution, and therefore crucial to identify the BN phase. Among its main limitations, we find not only its small size, which entails a limited range of masses, but also the assumption of an SFR without taking into account the effects of the formation of molecules and metallicity. AGNs are not included either. For a more detailed description, the following works are proposed [Ceverino et al., 2014, Tomassetti et al., 2016].

- The IllustrisTNG project [Vogelsberger et al., 2014] is a suite of state-of-the-art cosmological galaxy formation simulations. It consists of three box simulations with cubic volumes of roughly 50, 100, and 300 Mpc side length, which we refer to as TNG50, TNG100, and TNG300, respectively. Even when TNG50 simulation [Pillepich et al., 2019] has a comparatively limited sampling of rare objects, we will work with it for a detailed look at the structural properties of galaxies and the gas around them. It contains AGNs and reaches
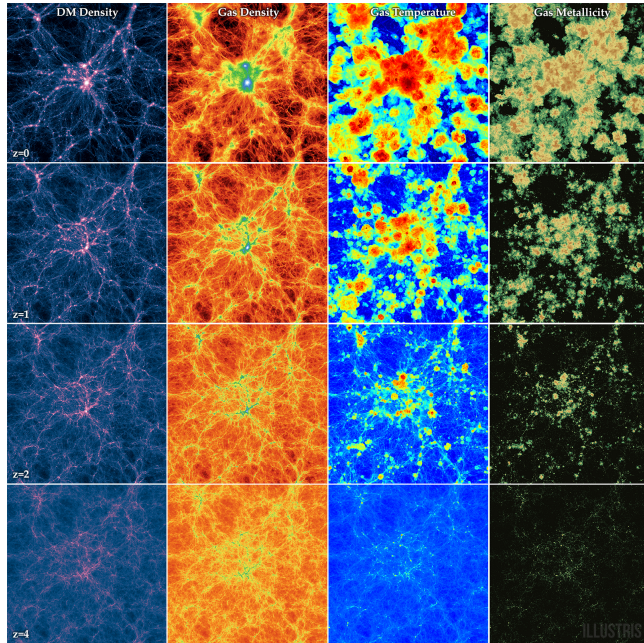
Figure 3: Redshift evolution of a whole box slice from z=4 to z=0, showing four projections: dark matter density, gas density, gas temperature, and gas metallicity. [Vogelsberger et al., 2014].

(almost) a numerical resolution typical of zoom-in simulations, with an average resolution of $70 - 140$ pc in the star-forming regions of galaxies. TNG50 samples galaxies with stellar masses above $10^8 M_\odot$ at z ~ 1 across cosmic time $0 \lesssim z \lesssim 6$. TNG50, as with the other IllustrisTNG simulations, has been evolved with AREPO, a massively-parallel simulation code used to solve the coupled equations of ideal magnetohydrodynamics (MHD) and self-gravity. A TreePM algorithm was also needed. As we can see, although it has several limitations, it seems that TNG50 has an adequate tradeoff between volume and resolution, so it is convenient to choose it as the next step to study the compaction of the BN phase.

2.2 The Compaction Phase

2.2.1 Definition of Compaction

Several previous works [Barro et al., 2017, Tacchella et al., 2016, Zolotov et al., 2015] have found that stream-fed, highly perturbed, gas-rich discs undergo phases of dissipative contraction into compact, star-forming systems ('blue nuggets') at $z \sim 4 - 2$. Such galaxies have a high density in their cores, both in stellar and gas mass. The compaction is triggered by an intense inflow episode, involving mergers and counter-rotating streams [Dekel and Burkert, 2013], and it can be associated with violent disc instability (VDI) [Gammie, 2001, Cacciato et al., 2012]. The

BN phase is also responsible for transitions in the galaxy structural, kinematic and compositional properties, for example from central dark-matter dominance to baryon dominance, which induces a transition of global shape from a prolate to an oblate stellar system [Dekel et al., 2019].

So, these numerical simulations predict a generic pattern of evolution for high-redshift galaxies through several characteristic phases, which can be observed in figure [4]. First, the early formation of a gas-rich, star-forming, highly perturbed and prolate disc, then, the dissipative, quick compaction of the gas disc into a compact, star-forming 'blue nugget', and immediately following, the rather fast quenching of star formation into a compact 'red nugget'. Thus, the peak of gas compaction marks the onset of central gas depletion and inside-out quenching. As the labels we work within VELA simulation have been obtained by detecting these cold gas peaks, it is interesting to analyze how gas mass, dark matter mass and stellar mass evolve with time in TNG50. As shown on the left in figure [5], we expect a drastic peak of gas mass, that defines the BN phase. The compaction should correspond to the moment when the central parts of the galaxy become dark matter dominated, and it should also lead to a strong increment of stellar mass.

2.2.2 Observational Properties of Compaction

- R_e: The effective radius or half-light radius of a galaxy is the radius in which half of the total light of the system is emitted assuming some spherical symmetry.

- Σ_1: The central stellar mass density is the projected mass per unit of area at the central kpc. For this reason, it can be calculated as the projected stellar mass divided by π, in the adequate units.

- SFR: The star formation rate is the total mass of stars formed per unit of time. It is usually studied in a certain radius such as the central kpc.

- sSFR: The specific star formation rate is calculated by dividing the SFR between the total mass of stars. This quantity allows exploring the relation between stellar mass and SFR directly.

According to the compaction features studied in VELA, we expect the BN phase to present a smaller effective radius and a larger density Σ_1 than the Pre-BN and Post-BN phase. Furthermore, the SFR of the BN galaxies must be high, since these are forming stars, in contrast to the Post-BNs, which we expect to have SFRs close to zero.

This produces a characteristic L-shaped track in the $sSFR - \Sigma_1$ plane when plotting the compaction and quenching sequence. This is an evolutionary path from star-forming galaxies (SFG) to quiescent. On it, Pre-BN galaxies tend to be in the main sequence (MS), then the compaction occurs at a roughly constant sSFR and at higher Σ_1 (BN phase), and the quenching
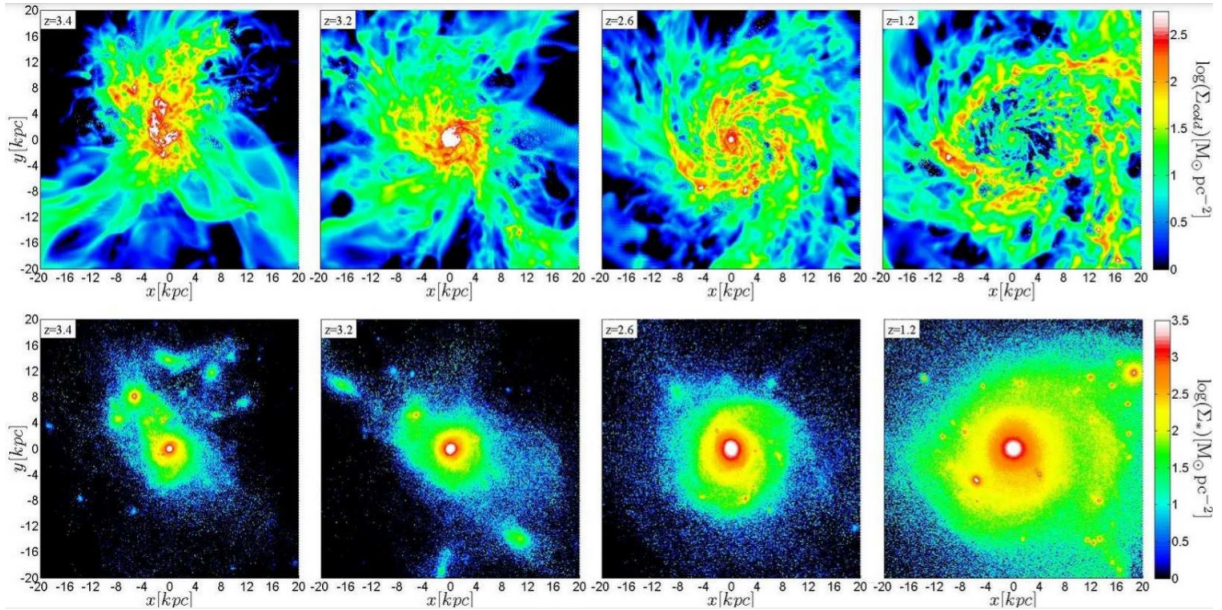
Figure 4: Compaction and quenching in VELA07. Shown are images of face-on projected density of the cold component made of gas and stars younger than 100 Myr (top) and of the stellar component (bottom), in a cubic box of side 40 kpc. The snapshots from left to right correspond to (a) prior to or during the compaction phase, (b) the 'blue nugget' (BN) phase near maximum gas compaction, (c) the 'green nugget' (GN) phase during the quenching process, and (d) the 'red nugget' (RN) phase after quenching. A BN with a dense core of gas and stars develops via dissipative compaction. It leads to gas depletion in the core while an extended ring develops. The dense stellar core remains intact from the BN to the RN phase, while in this case an extended stellar envelope develops around the RN core. [Zolotov et al., 2015].

at a constant Σ_1 but lower sSFR (Post-BN phase). This distribution is shown on the right in figure [5].

On the other hand, the redshift also provides information about the compaction phase. We expect to find the time sequence analyzed in [Huertas-Company et al., 2018], so the Pre-BN phase should correspond to higher redshifts (younger galaxies), while the Post-BN to lower redshifts (older galaxies). The BN phase should show intermediate redshifts.

2.3   Convolutional Neural Networks for Supervised Machine Learning

A convolutional neural network (CNN) is a type of Artificial Neural Network (ANN). It processes its layers by mimicking the visual cortex of the human eye to identify different characteristics in the inputs, that ultimately make possible to identify objects. They play a fundamental
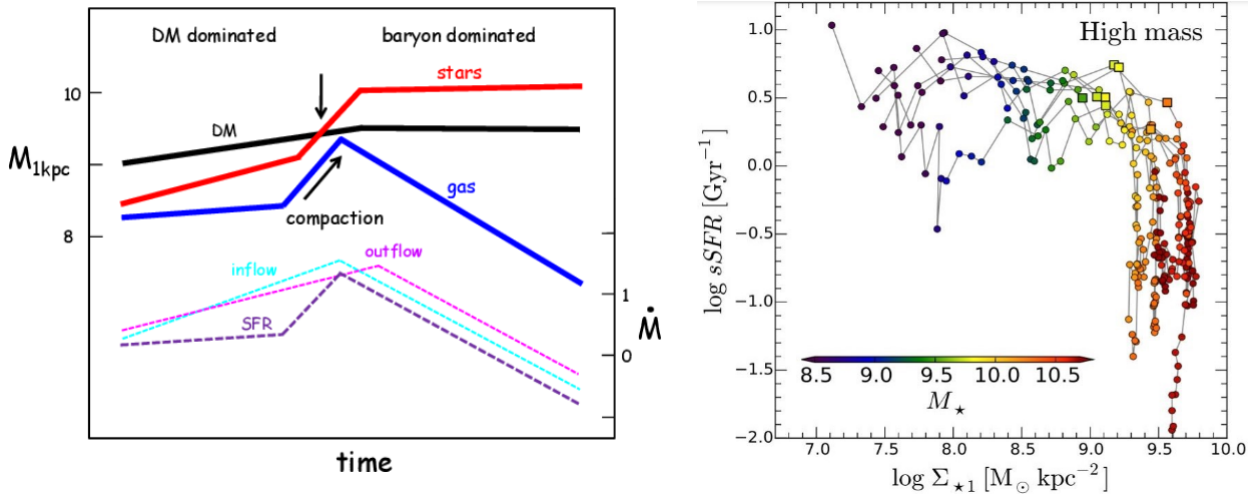
Figure 5: Compaction and quenching in cosmological simulations. **Left:** A cartoon describing the typical wet-compaction event, showing the evolution of masses (left axis in $log(M/M_\odot)$) within the inner 1 kpc. The compaction is the steep rise of gas mass (in blue), reaching a peak as a BN, and soon after declining as the central gas is depleted by star formation and outflows with no replenishment. The SFR (in purple, right axis in $log(M_\odot yr^{-1})$) follows closely, demonstrating a compaction-triggered central quenching. The central stellar mass (in red) is rising accordingly during the compaction, and it flattens off later. The inner 1 kpc is dominated by dark matter (in black) during pre-compaction and by baryons (stars, in red) during post-compaction. *Dekel private communication, based on* [Zolotov et al., 2015]. **Right:** The universal L-shape evolution track of eight galaxies from the VELA simulation in the plane of sSFR and stellar surface density within 1 kpc, Σ_1, which serves as a measure of compactness. The compactness is growing at a roughly constant sSFR (horizontally) before and during the compaction event, turning over at the BN phase (the 'knee', marked by a square symbol) to quenching at a constant Σ_1 (vertically). [Dekel et al., 2019].

role in DL algorithms, the bleeding edge of artificial intelligence. One of its most important applications is image classification, as we do in the present work. We will roughly introduce its operation and structure to ensure a better understanding of the project.

CNN are an extension of ANN by adding two new types of layers at the beginning of the network, the convolution layers and the pooling layers (see figure [6]). These are used to manipulate the pixels of the input images as we expose below:

- Convolution layer: extracts the characteristic features of an image by applying filters (kernels). These are generated by initializers which seek not to have elements (weights) that are too small or large (worse for learning). The set of matrices resulting from multiplying the images by the kernels are the feature maps. Once a feature map is created, we can pass each of its values through a nonlinearity, such as ReLU. This is a very simple activation function: $f(x) = \max(0, x)$, commonly used in DL. It helps to

prevent vanishing/exploding gradients (training unstable, very slow, or no convergent).

- Pooling Layer: extracts the most representative pixels from a delimited strip of an image. It is an operation of downsampling used to reduce the overall size of tensors.
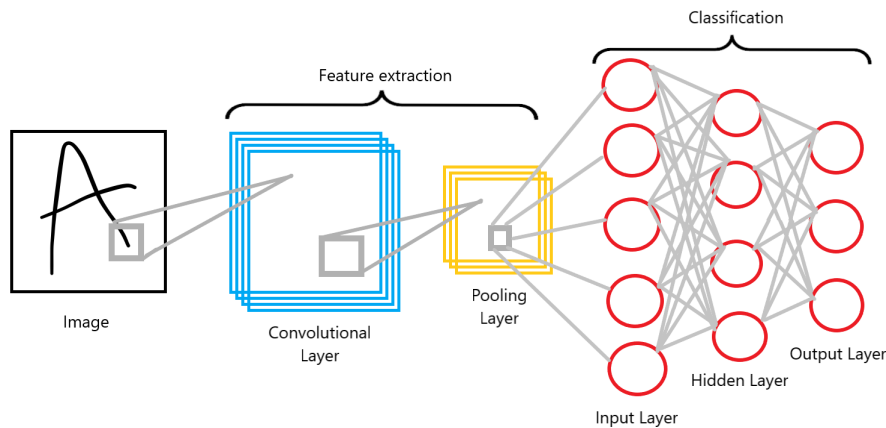

Figure 6: Structure of a CNN. The convolution and pooling layers extract the most important features of the images, while the FCC (in this diagram we have three of them), carry out the classification.

A subsequent dropout is also common once the layers of convolution and pooling have been applied. As relatively small datasets can overfit the training data, which results in poor performance when the model is evaluated on new data, some layer outputs are randomly ignored or 'dropped out.'

To continue, this last layer is flattened (the feature maps are combined) to obtain a layer of 'traditional' neurons called fully connected layer (FCC). The neurons on it have connections to all activations in the previous layer. In these connections, the output value of the previous neuron is multiplied by a weight value. These bond weights can increase or inhibit the activation state of adjacent neurons. To obtain the ideal ones, an attempt is made to minimize a loss function that evaluates the network as a whole. The values of the weights of the neurons are updated seeking to reduce the value of the loss function using an optimizer with a particular learning rate. This process is called 'backpropagation'.

At this point it is important to distinguish between three different datasets: the training dataset, the sample of data used to fit the model, the validation dataset, used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters (helps during the development stage of the model), and the test dataset, used to provide an unbiased evaluation of a final model. So, once we have a model, we can evaluate its accuracy with test data, not used in training, and confusion matrices.

3   Aim

En esta sección se señala de forma concisa el objetivo del proyecto: estudiar si tiene lugar el fenómeno de compactación en la simulación TNG50 con las mismas propiedades físicas. Para ello el trabajo se divide en dos partes, la optimización de la CNN y el estudio de la distribución de las magnitudes físicas en TNG50 entre las tres fases (Pre-BN, BN, Post-BN).

As previously introduced, the aim of this project is to study if the compaction phase (BN) takes also place in other simulations, since it has only been observed in VELA before. We will try to find logical connections between BN phases and physical observables in the galaxies of the TNG50 simulation. In particular, we will focus on the effective radius (R_e), the stellar-mass surface density (Σ_1), the star formation rate (SFR), the redshift and the gas profiles. The latter turned out to be very relevant in VELA, so we will pay special attention to them. Finally, we will compare the conclusions obtained for both simulations.

In order to achieve this goal, we will start optimizing the model of CNN proposed in [Huertas-Company et al., 2018], using the same dataset (28 galaxies of VELA simulation). To get started, we will modify manually different hyperparameters that determine the model such as the initializer, the dropout rate or the learning rate of the optimizer algorithm. In all this process, confusion matrices will be studied to evaluate its performance. This not only allows us to gain greater sensitivity and understanding of how CNN works in relation to these parameters, but it will also guide us when using Keras Tuner [O'Malley et al., 2019], our next step. This library is used to search the optimal set of hyperparameters (hypertuning). Specifically, we will seek to verify the conclusions we reach manually to generate the definitive model we are looking for.

Later, we will use the model weights to classify the ~ 11.000 galaxies from TNG50 simulation according to the phase they are in: Pre-BN, BN, Post-BN. Once we have it, we will study how these phases are distributed around the previously mentioned features for both VELA and TNG50 simulations. Basically, we will reproduce the distributions obtained in [Huertas-Company et al., 2018] for the first one with an improved model, so we hope that the conclusions will be confirmed; while we will study these distributions for the first time for TNG50, to check if the BN compaction phase exists with the assumed properties.

4 Data

*Se incluyen las especificaciones técnicas de las imágenes de VELA Y TNG50 con las que tra-
bajamos, en especial se describe el proceso de CANDELización (convolución con la PSF del
HST, y adición del ruido característico). También se señala el número total de imágenes y de
galaxias de cada simulación.*

4.1 VELA Simulation

We will work with the same dataset used in [Huertas-Company et al., 2018] from VELA Gener-
ation 3. The full output of the simulation was saved at many time steps and analyzed at steps
of $\Delta a = 0.01$ of the expansion factor, which roughly correspond to ~ 100 Myr at $z \sim 2$. For
every snapshot, mock 2D images were generated as they would be observed by the HST. It was
possible by using the radiative transfer code SUNRISE [Jonsson, 2006, Jonsson and Primack,
2010] which propagates the light of stars through the dust. We refer to [Snyder et al., 2015] for
details on the procedure followed.

Very briefly, a spectral energy distribution (SED) is assigned to every star particle in the
simulation based on its mass, age, and metallicity. The dust density is assumed to be directly
proportional to the metal density predicted by the simulations. It was fixed a dust-to-metals
mass ratio of 0.4 [Dwek and Arendt, 1998, James et al., 2002] and the dust grain size distri-
bution that was updated by [Draine and Li, 2007]. SUNRISE then performed dust radiative
transfer using a Monte Carlo ray-tracing technique. As each multiwavelength ray emitted by
every star particle and H II region (according to its SED) propagates through the ISM and
encounters dust mass, its energy is probabilistically absorbed or scattered until it exits the grid
or enters one of the viewing apertures (cameras).

The output of this process was then the SED at each pixel in all cameras. For this run,
19 cameras were set, of which five were fixed with respect to the angular momentum vector of
each galaxy, seven were fixed in the simulation coordinates, and the remaining seven were fully
random at each time step.

Eventually, from these data cubes, raw mock images were created by integrating the SED in
each pixel over the spectral response functions of the CANDELS WFC3 filters (F160W, F125W,
and F105W) in the observer frame. Images were then convolved with the corresponding HST
point spread function (PSF) at a given wavelength, and it was added a random real-noise stamp
taken from the CANDELS data. This ensures that the galaxies were simulated at the same
depth as the real CANDELS data and the correlated noise from the HST pipeline was well
reproduced. We refer to this process as CANDELization.

Finally, we work with 10.310 images of 28 galaxies in the range $1 \lesssim z \lesssim 2.2$ with $\log\left(M_*/M_\odot\right) > 9.0$ at $z \sim 2$. The VELA simulations 16, 18, 23, 24, 31 and 35 have been discarded because some of them are low-mass galaxies that have a quenching attempt that brings them considerably below the MS in the plane of SFR vs stellar mass, and the others have not been simulated down to $z = 2.2$.

4.2 TNG50 Simulation

We work with 6 different snapshots at $z = 0.5, 1.0, 1.5, 2.0, 2.4$ and 3.0, and galaxies with $\log\left(M_*/M_\odot\right) > 9.0$. This results in a sample of 11.047 images, that were generated in three different HST filters ($F160W, F125W, F105W$), the same ones used for VELA.

For that purpose, the radiative transfer code SKIRT [Baes et al., 2011] was implemented. Briefly, each simulated galaxy was viewed in a unique XY - plane, with the size of the images being equivalent to 15 times the 3D stellar half-mass radius. The stellar populations were modeled with the [Bruzual and Charlot, 2003] stellar population models for stars older than 10 Myrs. Younger stars are considered starbursting regions and were modeled with the MAPPINGS-III photoionization code [Allen et al., 2008]. All details of how parameters are set can be found in [Rodriguez-Gomez et al., 2018]. For computational reasons, dust was considered only if the fraction of star-forming gas was above 1% of the total baryonic mass. It was assumed for these objects that the dust content is traced by the star-forming gas. A constant dust-to-metal ratio of 0.3 was also assumed.

The final output was a 3D data cube for each galaxy, consisting of a full rest-frame SED for each pixel. In order to make the mock images a more accurate representation of the observed CANDELS ones, as it was done with VELA dataset, each mock image was convolved by the PSF, and real CANDELS background noise was added.

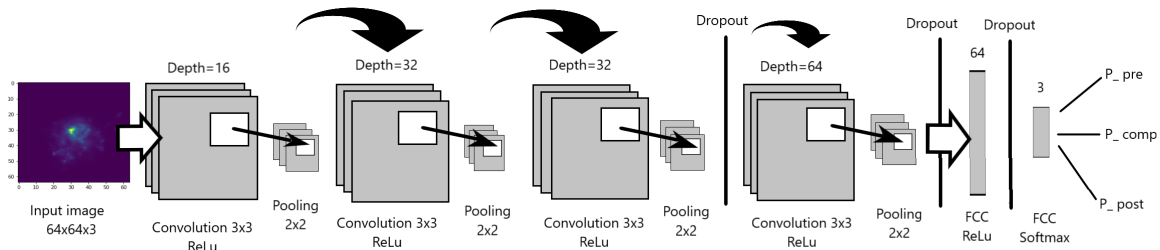In the end, we work with 11.047 images of 10.803 different galaxies.

Figure 7: Architecture of the CNN.

5   Methodology

Se explica la metodología a seguir en las dos partes bien diferenciadas del trabajo. Primero optimizamos la CNN mediante la modificación de los hiperparámetros del modelo y del algo-ritmo, tanto manualmente como empleando la librería Keras Tuner. También se señala cómo se realiza la evaluación de los modelos, a partir de las matrices de confusión. A continuación, se introducen las distribuciones a reproducir una vez se clasifican las galaxias de TNG50 con el modelo optimizado.

In this section we will specify the methodology followed throughout the work, dividing the process into the optimization of CNN and the prediction of BN phases for TNG50, in order to study their characteristic physical properties.

5.1   Optimization of the Network

As we have seen, a CNN depends on many hyperparameters. These can be divided into two types: model hyperparameters, which influence the model selection, such as the number and width of hidden layers; and algorithm hyperparameters, which influence the speed and quality of the learning algorithm, such as the learning rate.

Variation of both can significantly alter the results of the training, so it is important to choose the most appropriate ones based on the dataset that we have to classify.

5.1.1   Architecture

We are going to use the network structure used in [Huertas-Company et al., 2018]. It is a very simple sequential CNN architecture with four convolutional layers and two fully connected layers implemented in Keras [Chollet et al., 2015] (see figure [7]). The architecture was inspired by previous configurations that were successful in detecting strong lenses in space-based images [Metcalf et al., 2019], and in classical morphological classification [Domínguez Sánchez et al.,

2018]. For the first fully connected layer ReLu activation function is used, and for the last layer Softmax activation function (make the outputs behave like probabilities). On the other hand, Categorical Crossentropy and Adam are respectively the loss function and the optimizer used.

To prevent overfitting, a dropout is applied before and after the last convolutional layer, as well as in between the fully connected layers. Moreover, a Gaussian noise layer at the entrance of the network is included to avoid the model learning from the noise pattern. Finally, we use data augmentation: images are randomly rotated (within 45°), flipped, and slightly off-centered by 5 pixels maximum at every iteration. That ensures the network never sees exactly the same image.

This is a crucial point since the main limitation of the training is that we only work with 28 galaxies, even though the different orientations of the cameras generate a considerably longer dataset of images.

Of the 28 galaxies, 24 are used for training, 2 for validation, and 2 for testing (~ 400 images per galaxy). We train the model for 250 epochs, and every 50 epochs we change the validation galaxies (exchanging them for two from training). We repeat the process 5 times changing the test galaxies at each time. The final test dataset thus contains 10 galaxies, classified with 5 slightly different models.

5.1.2   Modifying Model Hyperparameters

The first parameter that we modify is the batchsize, that is, the number of samples that will be propagated through the network at each time. In general terms, small batches require less computational memory and speed up training. However, the smaller the batch the less accurate the estimate of the gradient will be. We will train in addition to the original batchsize, 32, with 64, 128, and 512. Another option would be not to divide the sample, but the computational time would be significantly longer.

Next, we try to vary the number of neurons (units) of the first fully connected layer. Initially, they were 64 and we will try with the values 32, 128 and 512. Modifying the number of neurons can have a considerable effect on the model, since too many connections can also contribute to overfitting. A model with more parameters has more "memorization capacity" and therefore it can easily learn a perfect mapping of the training samples, but without any capacity of generalization. It should be noted that the number of units of the last layer cannot be changed, since it must coincide with the number of categories in the classification, in our case 3.

On the other hand, we will work with different dropouts: 0.2, 0.5, and 0.8. At every training step, each neuron will have these probabilities of being dropped out of the collated contribution from connected neurons. We will use the same value for the three dropouts that are performed

on the network.

To evaluate the performance of the model when varying these hyperparameters, we will resort to confusion matrices, which compute the number of images classified in a certain category (the prediction of the model), and the number of images that actually belong to that category. The performances are calculated from the sums of the traces of these matrices (which collect the images classified in the correct category). These traces are averaged over 10-15 trainings with the same model and then compared with the means of the other models.

5.1.3 Modifying Algorithm Hyperparameters

The learning rate is a measure of how often the model weights are updated during training. This controls how quickly the model is adapted to the problem. Smaller learning rates require more training epochs (more computational time), whereas larger learning rates result in rapid changes and require fewer training epochs. However, learning rates too large may cause a quick convergence to a suboptimal solution. While [Huertas-Company et al., 2018] worked with the default learning rate for the Adam optimizer, 0.001, we will also try with 0.01 and 0.0001.

Moreover, we will train with different initializers for the kernels of the convolutional layers. These define the way to set the initial random weights. It is initially set Orthogonal, this initializer is very convenient since orthogonal matrices eigenvalues have absolute value 1 and no matter how many times we perform matrix multiplication, the resulting matrix doesn't explode or vanish. However, we will also try Kaiming initialization, which has shown very good results together with ReLu activation function as detailed in [He et al., 2015].

5.1.4 Using Keras Tuner Library

As we have introduced, a CNN depends on many parameters, and its behavior is not linear. Among other consequences, this implies that by varying the hyperparameters separately, we may not achieve the maximum performance of the model. In order to overcome this limitation, we will work with Keras Tuner [O'Malley et al., 2019], which improves the hypertuning.

First, we select the hyperparameters we want to tune and the range in which they can vary. We will tune the units in the first densely connected layer (between 32 and 512 units with steps of 32), the dropout (between 0 and 0.8 with steps of 0.05), and the learning rate (between the three values 0.01, 0.001, and 0.0001).

Then, we specify the objectives to be optimized, in our case, we will seek to maximize the accuracy when classifying the validation set. We also create a callback to stop training early after reaching the same value for the validation loss after 10 epochs.

Finally, we repeat the search for the optimal hyperparameters 10 times and then we select the most repeated set of values, that will make up the final model.

5.2   TNG50 Galaxy Classification and Study of Physical Properties

Once we obtain a better precision for the model, we proceed to predict a classification for the ~ 11.000 images from the TNG50 dataset.

To continue, we load physical information about the images from the catalog: the galaxy ID, the redshift, the effective radius, the SFR, and Σ_1. We also require information about the stellar, gas, and supermassive black hole masses.

We load both TNG50 and VELA data. We seek to check whether with the new model for VELA the distributions of the phases as a function of the different physical quantities remain the same, including the L-shaped track (see section 2.2). On the other hand, when studying these relations in the TNG50 phases we try to generalize the conclusions obtained for VELA to more extensive simulations.

To refine the results, we restrict the study to images that were classified with a high probability (greater than 0.4-0.9) in any of the three categories. We also establish mass and SFR ranges, among others.

Later, we plot the evolution of the mass for stars, gas, and the black hole in the galaxies, as a function of redshift. We look for a drastic peak of gas mass corresponding to the BN phase, a sudden increase of Σ_1, the measure of stellar mass in the central kpc, and a mild increment for the black hole mass.

6 Results

Se exponen los resultados de las dos partes del proyecto. Mediante la modificación del 'initializer' del modelo se obtiene una exactitud superior al 80%. Además, se incluyen las distribuciones del redshift, del radio y de Σ_1 para VELA y TNG50. También se representan las galaxias en el plano sSFR - Σ_1 y la evolución de las masas del agujero negro supermasivo, del gas, de las estrellas, y de la SFR, con el redshift.

6.1 Optimization of the Network

In this section, we analyze the classification accuracy with the test dataset (10 galaxies that were not used in the training), using the different hyperparameters mentioned in 5.1.2 and 5.1.3.

We evaluated the mean traces (computed with 10-15 trainings with each particular model) of the confusion matrices of the classification done with different hyperparameters. First, we obtained that increasing batchsize did not have a significant effect on accuracy (mean traces did not improve), but it did have higher computational costs. Likewise, when evaluating the performance for different numbers of neurons in the first fully connected layer, it was obtained that while using 32 neurons did not significantly modify the accuracy, increasing to 128 and 512 had a negative impact on it (lower traces). On the other hand, both increasing the dropout to 0.8 and reducing it to 0.2 impaired the performance of the model.

Modifying the optimizer learning rate had a very significant effect on the traces, it was found that 0.001 was preferable to 0.01 and 0.0001. It is interesting to note that for 0.01 the model tended to classify everything in the intermediate category (BN). As we expected, it causes a quick convergence to this suboptimal classification. Alternatively, when training when 0.0001 the computational time was considerably higher, and even the traces of the matrices were lower.

Eventually, we try to change the initializers of the kernels of the convolutional layers. It was fixed as Orthogonal but by using Kaiming initialization (normal-distributed) we achieved a significant improvement in the performance of the model (a mean trace 15% higher).

The hyperparameters obtained with Keras Tuner are consistent with these results. In the 10 searches, highly variable values were obtained for the number of neurons in the first fully connected layer. The absence of a clear tendency confirms that the accuracy in the test set is practically independent of this parameter (in the range in which we worked, between 32 and 512 units with steps of 32.) In addition, the selected dropouts were mostly in the 0.4-0.6 interval, which corresponds to the value 0.5 that we obtained manually. Finally, Keras Tuner selected the learning rates 0.001 (6 times) and 0.0001 (4 times), of the 10 trainings that were carried out. Therefore, the choice of a learning rate of 0.001 seems consistent, which also implies a

computational saving.

So, we have finally set a model with batchsize 32, 64 units in the first fully connected layer, three dropouts of 0.5, a learning rate of 0.001 for Adam optimizer, and Kaiming normal-distributed initialization for the kernels of the convolutional layers. We compare the final accuracy for the test data in figure [8]. For a probability threshold equal to 0.5 we achieve an 81% of accuracy [1] , versus the 63% of the original model. We also get a higher accuracy when the threshold is 0.8, an 83% versus an 80%. It is important to note that the accuracy for the Pre-BN category does not improve with the threshold. In fact, it is lower than expected. It could be explained if the fraction of the total training galaxies is taken into account. We have a smaller number of Pre-BNs in all the simulations, so it is not surprising that the model tends to classify some of them in the other two categories. In total we worked with 10.310 images (distributed between the training set, the validation set, and the test set) of which 13% were Pre-BNs (1.349 images), 34% were BNs (3.458 images), and 53% were Post-BNs (5.503 images). It is understandable if we take into account that the Pre-BN galaxies correspond theoretically to higher redshifts, that is, to more distant times.

6.2   TNG50 Galaxy Classification and Study of Physical Properties

When classifying the 11.047 images of TNG50 simulation using the model trained with VELA, we get 66 galaxies classified as Pre-BN, 4.875 as BN, and 6.106 as Post-BN (some examples are shown in figure [9]). Clearly, the imbalance of VELA images has had a magnified effect in the TNG50 classification, even if it should be noted that our sample is limited to galaxies more massive than $10^9 M_\odot$, and most of the Pre-BN galaxies are expected to be less massive.

There are numerous techniques to avoid these biases related to unbalanced training sets, for example by applying weights to the loss function when training with VELA [Buda et al., 2018]. However, seeking to simplify the process, we will adapt the probabilities threshold of the prediction of the TNG50 classification. Note that the Softmax function returns the probability for each image to be in each category, for example, 40% of being Pre-BN, 50% of being BN, and 10% of being Post-BN. With these results, technically we should say that it is a galaxy in a BN phase, but according to what has been previously commented, it is probably a galaxy in a Pre-BN phase. So, when getting the distributions of the phases for different physical quantities, we will require at least a probability greater than 90 % to classify a galaxy as Post-BN and 80 % as BN. However, 40 % will be enough for a Pre-BN classification. A more detailed explanation is included in the appendix 8.

---

[1]Accuracy is computed by dividing the trace of the confusion matrix between the number of classes (3), equivalent to divide the sum of the galaxies correctly classified between the number of total galaxies.
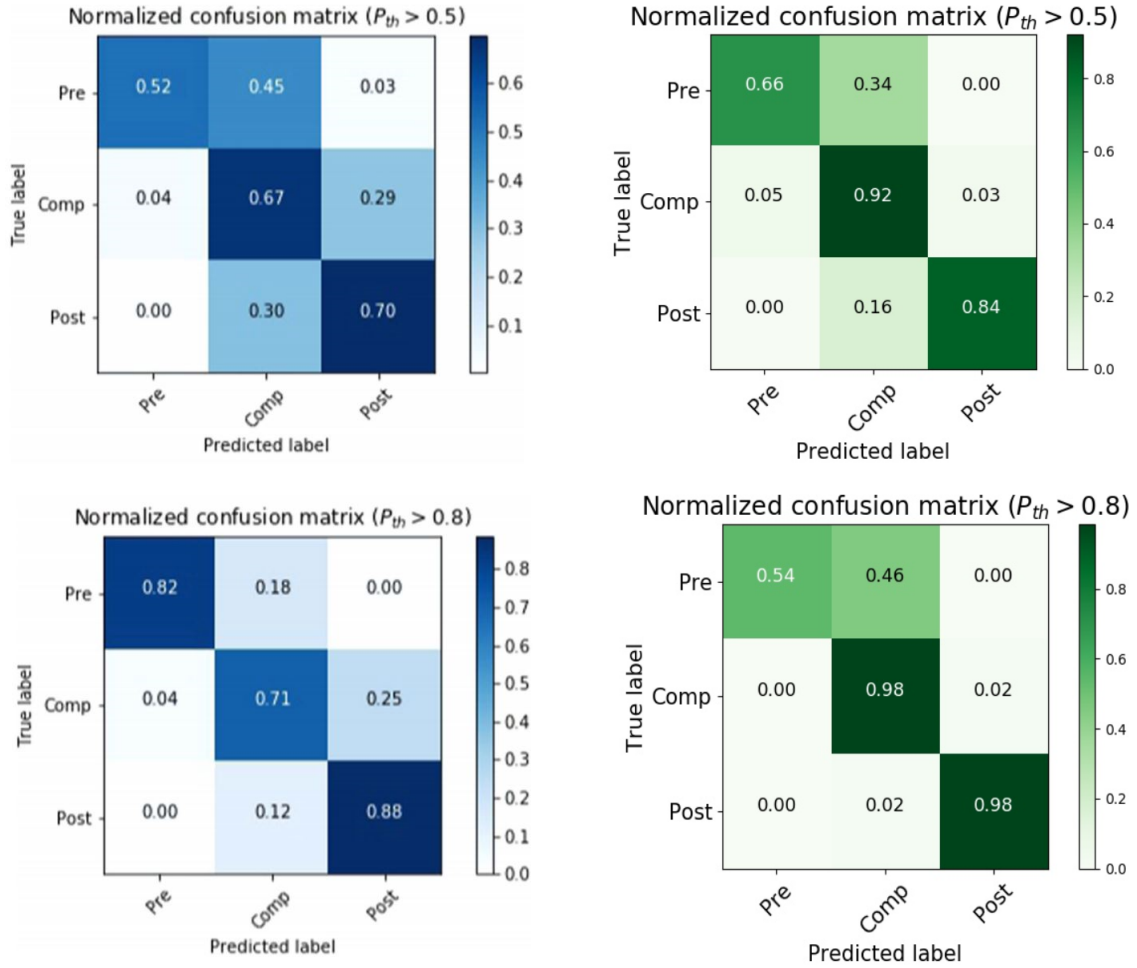
Figure 8: Normalized confusion matrices of the three-label classification of test data (not used for training or validation). The y-axis shows the true class from the simulation metadata, and the x-axis shows the predicted class. The two matrices on blue correspond to the classification predicted used the original model, while the two on green with the classification predicted with the new model. With the two rows, we show the effect of increasing the probability threshold (from 0.5 to 0.8) to select the galaxies belonging to a given class.
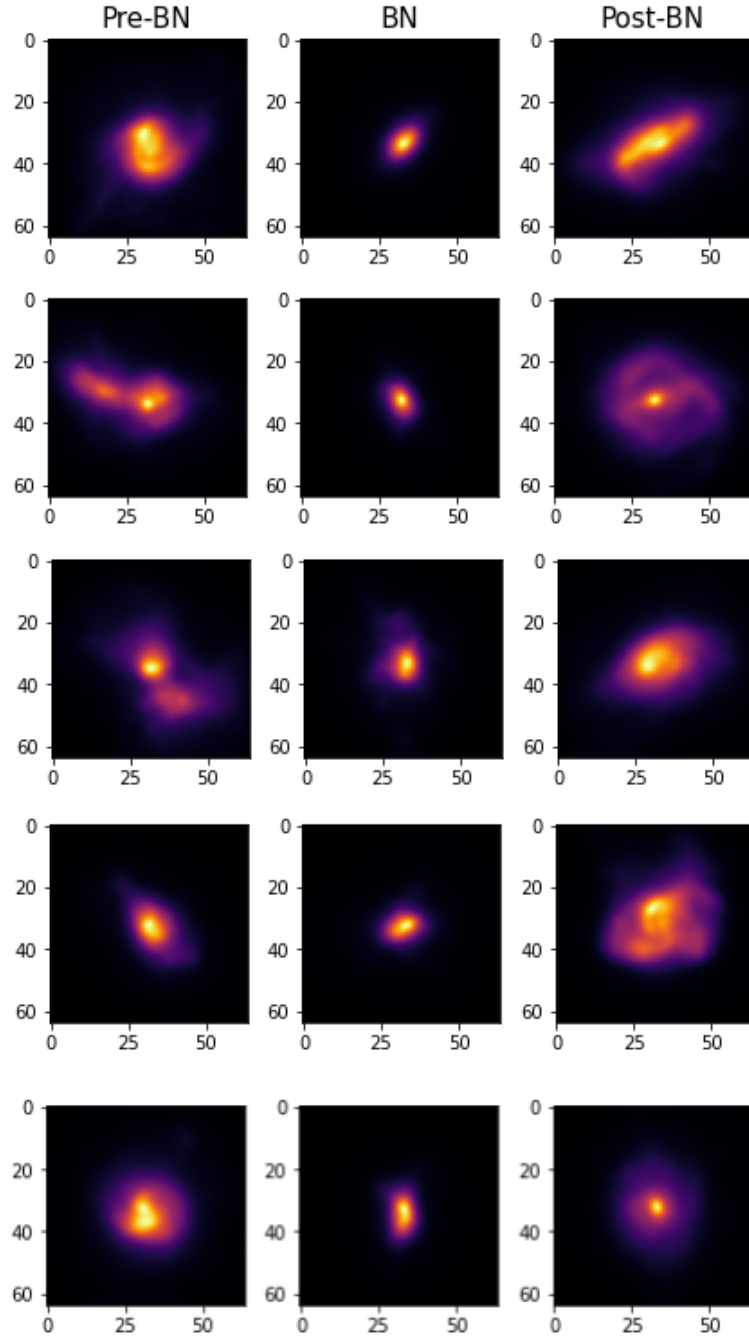
Figure 9: The three columns include, from left to right, examples of TNG50 galaxies classified as Pre-BN (with a $P_{th} = 0.4$), as BN (with a $P_{th} = 0.8$), and as Post-BN (with a $P_{th} = 0.9$). The features associated with each of the phases are visually verified: Pre-BN galaxies usually have a clumpy gas appearance, BN galaxies are the most compact, and Post-BN galaxies are extensive and diffuse.
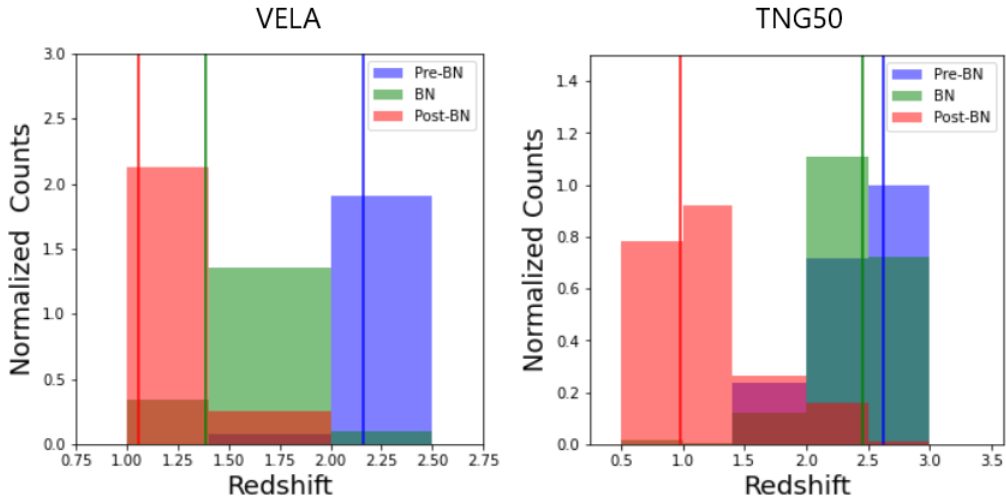
Figure 10: The histograms show the redshift distribution of the three different classes for both VELA and TNG50 simulations. The red, green, and blue histograms show the Post-BN, BN, and Pre-BN phases respectively. The vertical lines indicate the median for each class with the same color code. Despite some overlap, especially in the TNG50 simulation, the classifier is able to establish a temporal sequence on the three phases. Note that the range of redshift is different for each simulation, working from 1.0 to 2.5 in VELA and from 0.5 to 3.0 in TNG50.

As we introduced in section 2.2, we expect the Pre-BN phase to correspond to an earlier step in the evolutionary sequence, and therefore to be assigned high redshifts. On the contrary, the Post-BN phase must correspond to a later step and have low redshifts. We can see in figure [10] how there is indeed a time sequence involved in the three phases. It is remarkable that the BN phase is situated at higher redshifts for TNG50, overlapped with the Pre-BN phase. In figure [11], the mean probability of a galaxy to be classified as BN or as Post-BN for the TNG50 dataset is plotted as a function of redshift.

On the other hand, in figure [12] we can see the distribution of the radius for the three phases. It is important to notice that even when we work with the disc radius for VELA and with the effective radius for TNG50, some conclusions can be drawn anyway. The BN phase has the smallest radii, closely followed by the Pre-BN phase, while the Post-BN phase has the largest radii, as expected from theory.

Additionally, we can plot the distribution of Σ_1 as shown in figure [13]. As explained in [Zolotov et al., 2015], we expected that in early times the galaxies presented a more diffuse stellar light distribution (Pre-BN), where it would form stars at a low surface density. Then, a rather quick compaction event that would increase the surface density Σ_1 (BN). Finally, a quenching would start, but Σ_1 would remain high (Post-BN). However, we see how in both
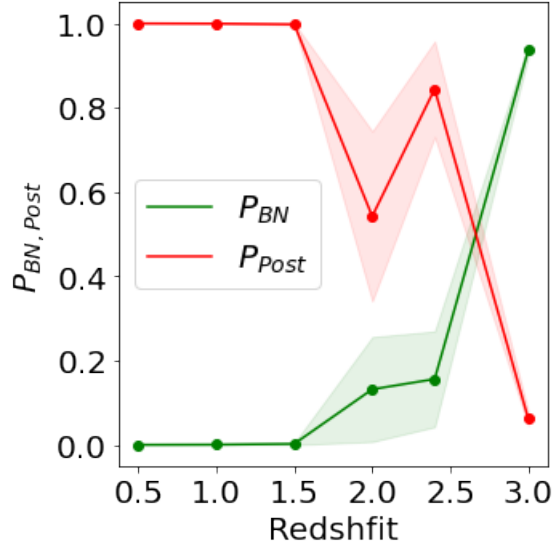
Figure 11: The plot represents the evolution of the mean probabilities (Softmax output) of the BN phase in green and of the Post-BN phase in red, as a function of the redshift for TNG50 simulation. It is observed how after a compaction phase for redshifts greater than 3.0 the probability of the Post-BN phase increases drastically, while the one of the BN phase decreases.
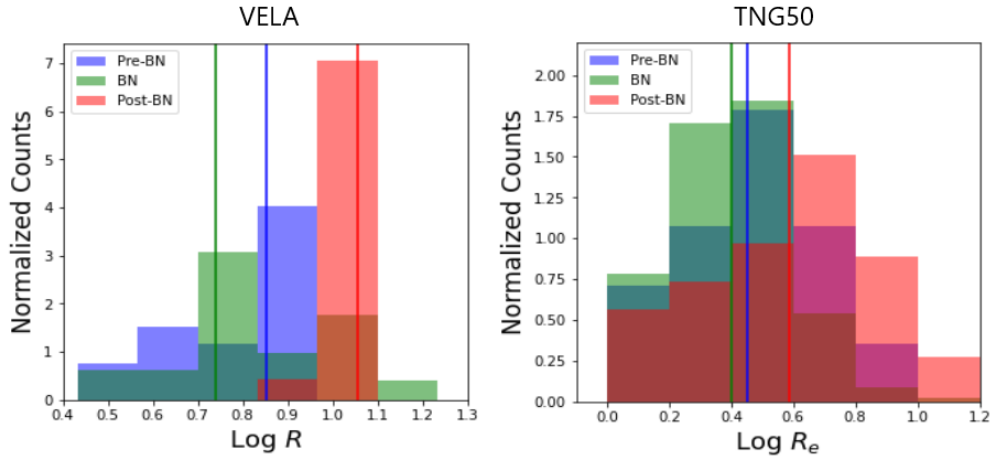

Figure 12: The histograms show the distribution of the decimal logarithm of the disc radius [kpc] for VELA and the effective radius for TNG50 [kpc]. The blue, green, and red histograms correspond to the Pre-BN, BN, and Post-BN phases respectively. The vertical lines indicate the median for each class with the same color code. There is a higher level of overlap for the three phases in the TNG50 simulation, but even so, it is observed how the BN phase is associated with a compaction, presenting the smallest radii.
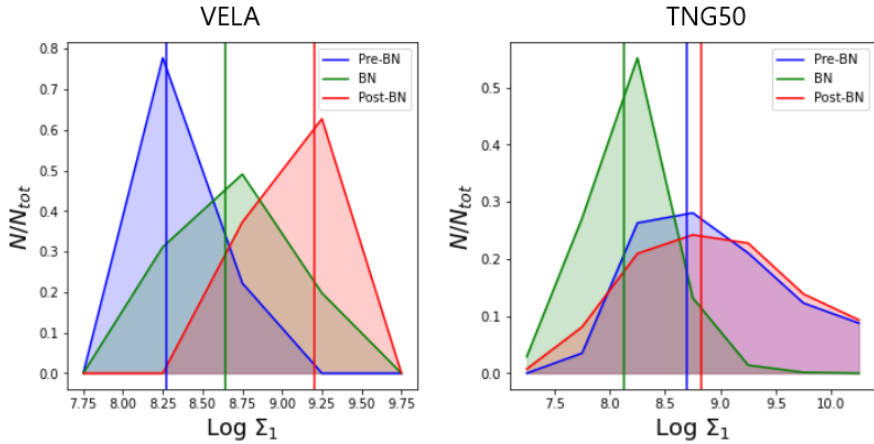
Figure 13: The plot shows the distribution of the decimal logarithm of Σ_1 (in $[M_\odot/kpc^2]$) for VELA and TNG50 simulations. The blue, green, and red lines correspond to the Pre-BN, BN, and Post-BN phase respectively. The vertical lines indicate the median for each class with the same color code. There is a high level of overlap for the three phases in both simulations.

simulations Σ_1 for the BN phase is not maximum. While in VELA a more typical behavior is observed, with a Pre-BN phase corresponding to a minimum of Σ_1, in TNG50 we observe how the Pre-BN phase and the Post-BN phase present maximum central surface densities, while the BN phase presents minimum ones. In general terms, this quantity is difficult to measure, especially if we take into account that the noise from the HST has been added, so we will continue studying these phases in TNG50 to better understand this behavior.

We also try to reproduce the L-shaped track in the sSFR - Σ_1 plane in figures [14] and [15]. We achieve the expected behavior for VELA, but in TNG50 simulation, the Pre-BN and BN phases are interchanged compared with previous results [Zolotov et al., 2015, Tacchella et al., 2016]. It should be noted that galaxies in the Pre-BN phase are very scattered, and since they are also few, the position of the median in the plane is non-representative. In this plane for the TNG50 simulation we see how the BN phase is located in a very specific region ($8.0 < \log \Sigma_1 < 8.5$ and $-0.25 < \log sSFR < 0.25$). This can be related to the plot [13], overall, the galaxies of TNG50 in the BN phase have a lower central surface density than expected. For this reason, we also study the stellar mass distribution for the BN phase in TNG50 in a 30 kpc radius: these galaxies have a characteristic stellar mass of $10^{9.5}M_\odot$.

Finally, we represent in figure [16] the evolution of the masses of gas, of the supermassive black hole, and of those of stars (from Σ_1) with the redshift for the TNG50 simulation. We also trace the evolution of the SFR and the sSFR in the central kpc. This information is crucial because the classifications that were used to train the CNN model were made for VELA from

this analysis (specifically of the gas peaks). Thus, the essential condition for the phase to be well defined in TNG50 is that there is an associated peak of gas. As we can see, the mass of the black hole increases smoothly as the redshift decreases. In turn, there is a gas peak for $z = 2.0$, which is effectively associated with the BN phase. Stellar-mass increases, while SFR peaks at $z = 2.0$ too, until it goes down to zero (i.e. quenching). On the other hand, the sSFR progressively decreases. These last two quantities indicate that galaxies in the BN phase are forming stars, while those in the Post-BN phase are not. It is worth stressing that the BN phase is longer than expected, so probably the predominant phase at $z = 3.0$ should be the Pre-BN, and this result is due to the limitations of the classification.
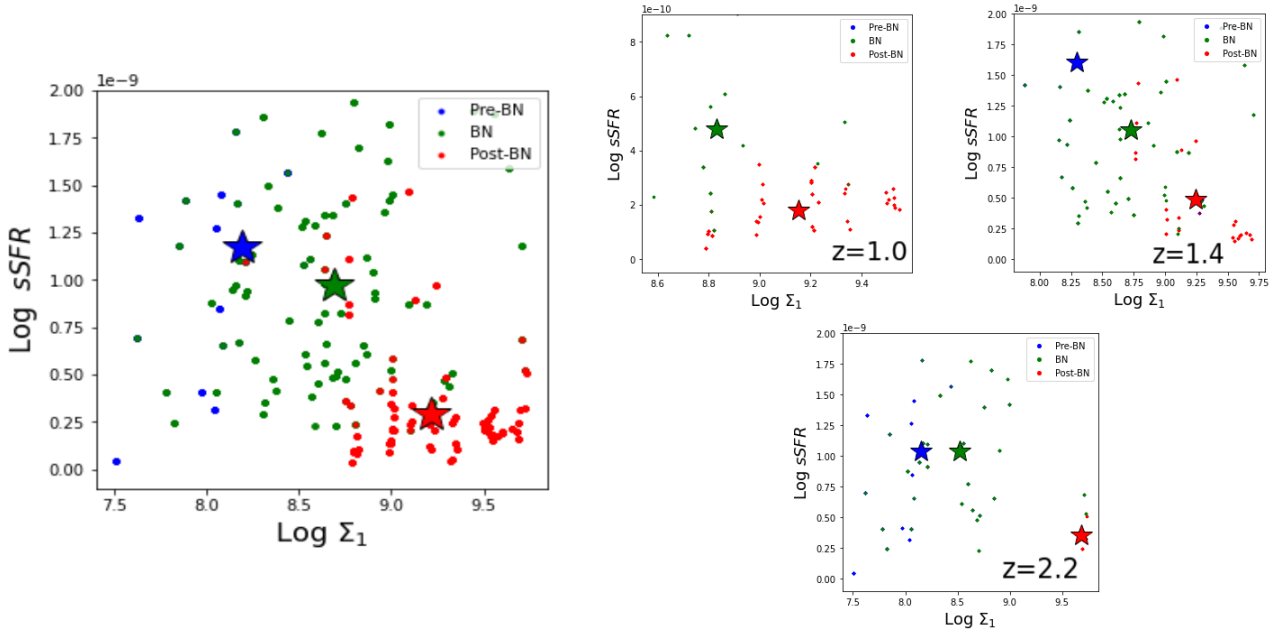

Figure 14: Track of evolution in the sSFR-Σ_1 plane for VELA simulation. Units are $[Gyr^{-1}]$ and $[M_\odot/kpc^2]$ respectively. Blue, green, and red dots are galaxies in phase Pre-BN, BN, and Post-BN, and the stars with the same color code indicate the median for each class. The plot on the left contains all the classifications regardless of redshift, while the three on the right include only the galaxies with a given redshift. The median for the Pre-BN phase in $z = 1.0$ is not included because there are very few galaxies classified in this category.
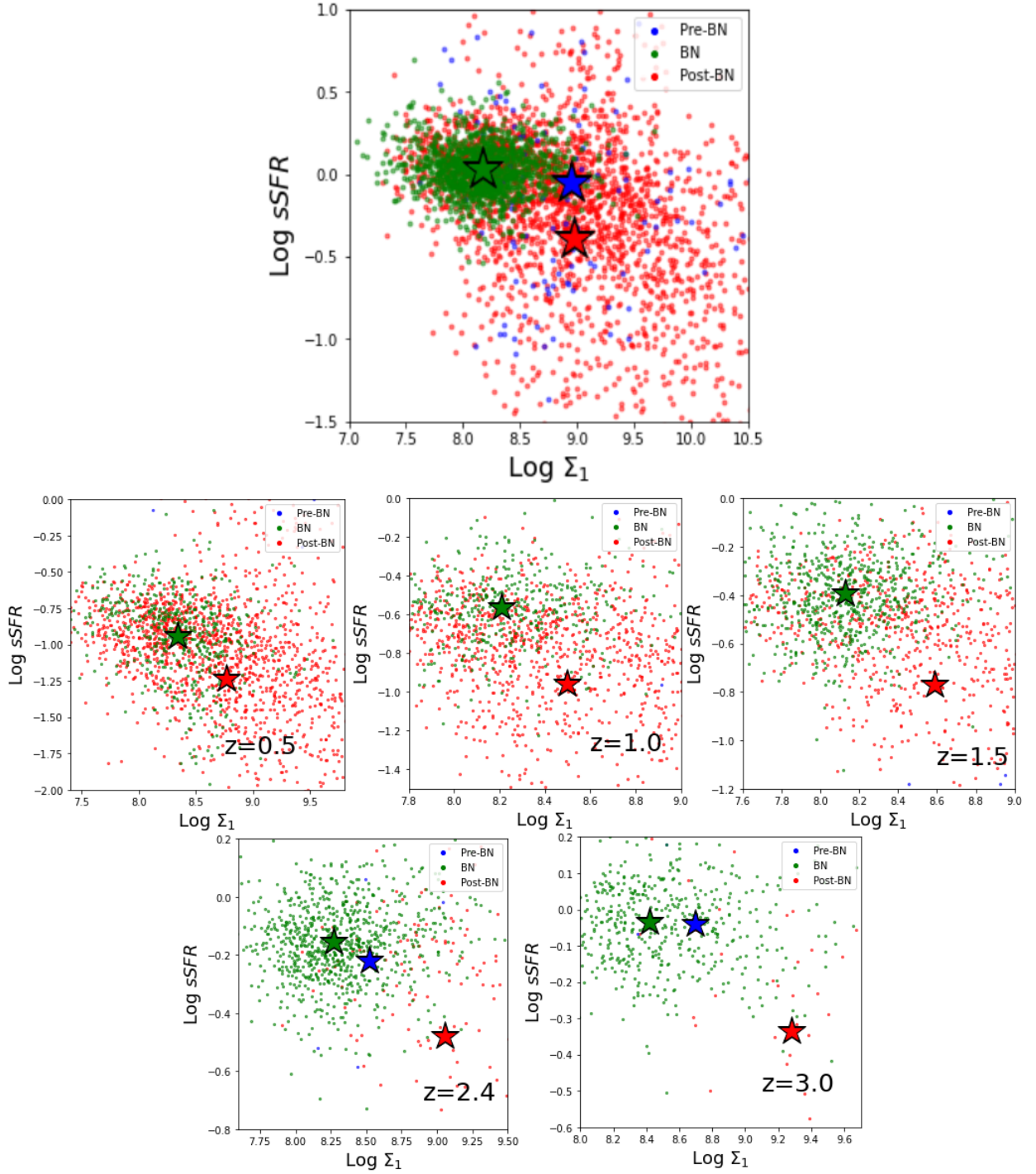
Figure 15: Track of evolution in the sSFR-Σ_1 plane for TNG50 simulation. Units are $[Gyr^{-1}]$ and $[M_\odot/kpc^2]$ respectively. Blue, green, and red dots are galaxies in phase Pre-BN, BN, and Post-BN, and the stars with the same color code indicate the median for each class. The first plot contains all the classifications regardless of redshift, while the next five include only the galaxies with a given redshift. In the range $0.5 \leq z \leq 1.5$ the median for the Pre-BN phase is not included because there are very few galaxies classified in this category.
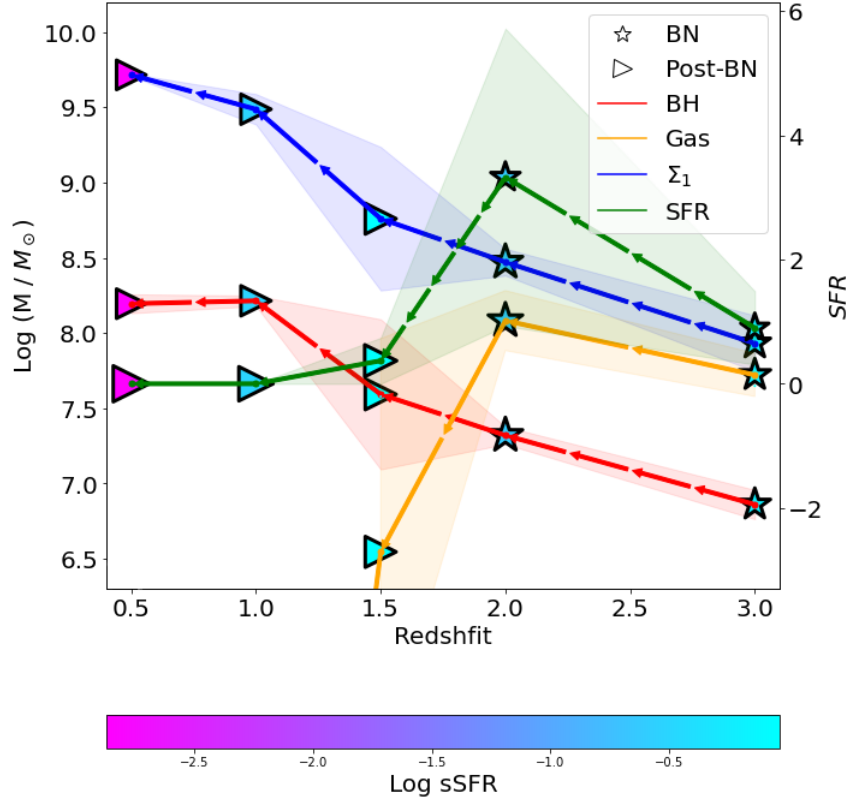
Figure 16: TNG50 redshift evolution of the average (between all the galaxies) of the supermassive black hole mass in red, of the gas mass in the central kpc in yellow, of Σ_1 in blue (as a measure of the stellar mass in the central kpc), and of the SFR in the central kpc in green. All the masses are in $[M_\odot]$, while the SFR units are $[M_\odot Gyr^{-1}]$. Arrows indicate the direction of evolution, and the shaded regions the deviation between galaxies. It is also plotted the averaged sSFR for the five redshifts analyzed, as shown in the colorbar, in $[Gyr^{-1}]$. Finally, the averaged class associated with each redshift is plotted, with a star for the BN phase and a triangle for the Post-BN phase. The Pre-BN class is not predominant for any redshift in the range $0.5 \leq z \leq 3.0$.

7 Conclusions and Discussion

Se incluyen de forma esquematizada las principales conclusiones del trabajo. Más adelante se discuten las limitaciones y problemas: el desequilibrio entre las tres categorías en las que clasificamos las imágenes y la distribución obtenida para Σ_1. A continuación, se proponen alternativas para continuar con el proyecto, como técnicas para balancear las categorías u otras simulaciones en las que estudiar el fenómeno. Finalmente, se insiste en la necesidad de una simbiosis entre la astrofísica, la estadística, la informática y el DL, aludiendo a proyectos inminentes como Euclid y el JWST.

We have optimized a CNN trained with CANDELized galaxy images from VELA simulation to classify TNG50 galaxies according to evolution patterns. Specifically, we seeked to detect a BN phase characteristic of high-redshift galaxies: a compaction preceded by a diffuse phase and proceeded by a quenching.

DL Conclusions

- Model performance doesn't improve by modifying the batchsize, the number of neurons in the first fully connected layer, or the dropout in the ranges described in section 5.1.2. Neither by changing the learning rate of the optimizer. We verify it both manually and by using the Keras Tuner library.

- Kaiming initialization (normal-distributed) is more efficient than Orthogonal initializer when using ReLu activation function. Initial random weights for the kernels of the convolutional layers provide better accuracy when the Kaiming initializer is fixed, getting mean traces for the confusion matrices a 15% higher.

- The optimized model gets high accuracy when it is evaluated in VELA test data. For a probability threshold of 0.5, it achieves an 81% of accuracy (versus the 63% of the original model), while it achieves an 83% of accuracy (versus the 81 % of the original model), for a probability threshold of 0.8.

- Accuracy for the Pre-BN class doesn't improve with the probability threshold, getting a relatively low accuracy as a consequence of the imbalance of the dataset among the three classes.

- This imbalance is magnified when using the model on the galaxies of the TNG50 simulation, since only 0.6 % of the images are classified as Pre-BN. However, by working with different probability thresholds for the classes we can study their physical properties anyway.

Physical Conclusions

- The three phases predicted for TNG50 galaxies are in time sequence, as it is the case for VELA galaxies. While the galaxies in the Pre-BN phase correspond to larger redshifts, that is, to more distant times, the Post-BN have smaller redshifts, so they are from closer times. BNs are in the middle in both simulations, although they are associated with larger redshifts in TNG50.

- When evaluating the radii (of the disc for VELA and effective for TNG50), we observe that effectively the BN phase is the most compact of the three, followed by the Pre-BN and finally by the Post-BN.

- For none of the simulations, the BN phase corresponds to a maximum of Σ_1. We get a particularly strange distribution for TNG50 galaxies. On it, the BN phase corresponds to a minimum of Σ_1, while the other two phases have similar and higher densities. This issue will be investigated in future work.

- This problem is also observed when studying the sSFR - Σ_1 plane. While for VELA the L-shaped track is reproduced for all redshifts, in the TNG50 simulation it is not found. Although it is verified that the sSFR is approximately constant for the Pre-BN and BN phases, and decreases for the Post-BN phase, the BN phase has a lower central density than expected, and as consequence, the Pre-BN and BN phases are swapped in the plane. Even so, it is verified that Pre-BN and BN galaxies are forming stars, while Post-BN ones are not.

- The BN phase in TNG50 has a characteristic stellar mass of $10^{9.5} M_\odot$ in the inner 30 kpc, and corresponds to both a gas peak and a SFR peak. After this phase, the mass of gas in the galaxy decreases considerably, as does the sSFR, while the SFR also decreases, stabilizing around zero for the Post-BN phase (which does not form stars). On the other hand, the stellar mass in the central kpc increases sharply, and the mass of the supermassive black hole increases subtly until it stabilizes around $10^8 M_\odot$.

The fact of having an unbalanced training dataset, together with the mass threshold ($10^9 M_\odot$), has caused that very few galaxies were classified with certainty as Pre-BN in TNG50 simulation. In the appendix 8, we focus on this fact by studying other metrics from the confusion matrix. We had two majority classes in the training, BN and Post-BN, and a minority, Pre-BN. We used a simple technique to be able to extract physical information from the galaxies classified in these classes, working with different threshold probabilities. However, there are other more complex but more efficient techniques to balance the dataset before training [Buda et al., 2018, Ali et al., 2015].

For example, it would be better to evaluate the performance with the validation dataset during training by analyzing other different metrics (the accuracy does not veraciously reflect

the quality of the model if the data is unbalanced). The most suitable metric would be the F1-score, whose definition is provided in the appendix 8. Another frequently used option is to penalize the majority classes. Although the process itself is laborious, it can be applied in Keras by adding the argument *class_weight = 'balanced'* to the model fitting. It is also possible to reshape the dataset by undersampling, which consists of sampling from the majority class in order to keep only a part of these points, by oversampling, which consists of replicating some points from the minority class in order to increase its cardinality (using data augmentation), and by generating synthetic data, creating new synthetic points from the minority class to increase its cardinality (for example with SMOTE method [Chawla et al., 2002]). These alternatives are proposed as a continuation of the project. It is important to emphasize that each set of images behaves differently, and just as the network architecture and hyperparameters must be adjusted, the method that works best for each specific dataset must also be searched.

When plotting the Σ_1 distribution for the TNG50 simulation, we obtained that the BN phase was associated with a minimum, while the Pre-BN and Post-BN phases had higher densities and were similar to each other. However, we expected the opposite: that the BN phase would correspond to a maximum. In general, it is a difficult parameter to measure, since 1 kpc is a relatively small measure compared to the simulated dimensions. The difference between the resolution of TNG50 (70 - 140 pc) and that of VELA (17-35 pc) may explain this error. On the other hand, we must take into account that the noise of the HST, with which the images have been processed, has a negative effect on the accuracy of this quantity.

Despite the issue with Σ_1, we have studied the impact of the main limitations of the VELA simulation when studying compaction and quenching in galaxies at high redshift. Neither the absence of AGN nor the small number of galaxies have been a real problem, because in TNG50, where these two are solved, the same distributions are reproduced for the redshift, the radius, the SFR, the sSFR, and for the masses of gas and the supermassive black hole. The assumptions of the subgrid astrophysics have also been subtly assessed since TNG50 employs other physical prescriptions such as dust density or ISM metal enrichment. However, these subresolution models remain the weak point of the study of physical phenomena from cosmological simulations. There is still considerable room for improvement, but while advancing with computational capacity, it is proposed to study the BN phase with other simulations, preferably with the same initial conditions, but with different subgrid astrophysics, such as VELA Generation 6.

Finally, it should be noted that the success of this project is not only based on the physical results but also on the methodology followed. DL can be used to study other complex processes, in astrophysics and science in general. These algorithms allow the user to classify images without any dimension reduction or extra preprocessing, but above all, they do not require a priory assumption of the optimal observables for a given physical process. They are particularly promising for comparing future imaging surveys such as Euclid or the JWST with forthcoming

simulations. This last survey [Gardner et al., 2006] will be launched on October 2021, and it will allow resolving less massive galaxies (up to $10^8 M_\odot$), thus sorting out the mass limitations of our project, and opening a future line for research. In accordance with this idea, further steps should be taken to apply new data management methods to astronomy, since significant progress in the face of these new challenges can be achieved only via a true symbiosis of astrophysics, statistics, computing, and DL, among other disciplines.

8   Appendix

En este apéndice se realiza un análisis de las matrices de confusión con distintas P_{umbral} para el modelo optimizado. Para ello se recurre a otras métricas como la precisión o la exhaustividad, y finalmente se concluye que el desequilibrio de las imágenes con las que entrenamos entre las fases Pre-BN y BN provoca que parte de las imágenes de la primera categoría se clasifiquen en la segunda.

Analyzing the Phases in VELA Test Data from Confusion Matrices

We analyze other classification parameters such as the precision and the recall for the classes from the confusion matrices ($P_{th} = 0.4, 0.8, 0.9$) included in figure [17]. This can help to better understand the issues when classifying the Pre-BN phase, and in general the limitations of the model. First of all, we give a brief definition of the parameters we work with:

- Precision: for a given class, the precision can be determined as the ratio between the number of images well classified in this class and the number of all the images classified in it.

- Recall: for a given class, it is calculated by dividing the number of images well classified in this class by the number of images that actually belong to it.

- F1-score: the F1 score of a class is given by the harmonic mean of precision and recall, combining the two of them in one single metric.

According to this, we can conclude that if a class has high recall and high precision, then it is perfectly handled by the model, while if both are low, then it is poorly handled. However, the most interesting cases occur when one of these has a high value and the other a low one. In particular, when we get low precision but high recall the class is well detected but the model also includes features from other classes, while, when there is high precision but low recall, the model cannot detect the class well but is highly trustable when it does. These last two situations usually take place when the data we train with is unbalanced (the class is a majority in the first case and a minority in the second one).

As it is collected in table [1], the Post-BN class typically has both high precision and high recall. Nevertheless, for the Pre-BN class the precision is high and the recall low, which evidences the fact that there are few galaxies classified as Pre-BN for training. In contrast, the BN class has low precision and high recall, effectively it is a majority, and for this reason, the network tends to classify galaxies in the Pre-BN phase in this class. We can see how from these metrics we are able to extract information not contained in the accuracy, a particularly misleading parameter when working with unbalanced classes. As explained in section 6.2, we selected different probability thresholds for the three categories: $P_{th} = 0.4$ for the Pre-BN

class, $P_{th} = 0.8$ for the BN class, and $P_{th} = 0.9$ for the Post-BN class. This is justified because we are looking for a high F1-score for each class, but also seeking to have enough galaxies classified into them (the number of galaxies is included in figure [17]), so the analysis of the physical properties is representative. It is important to highlight how increasing the probability considerably reduces the number of classified galaxies, especially for the Pre-BN and BN classes.
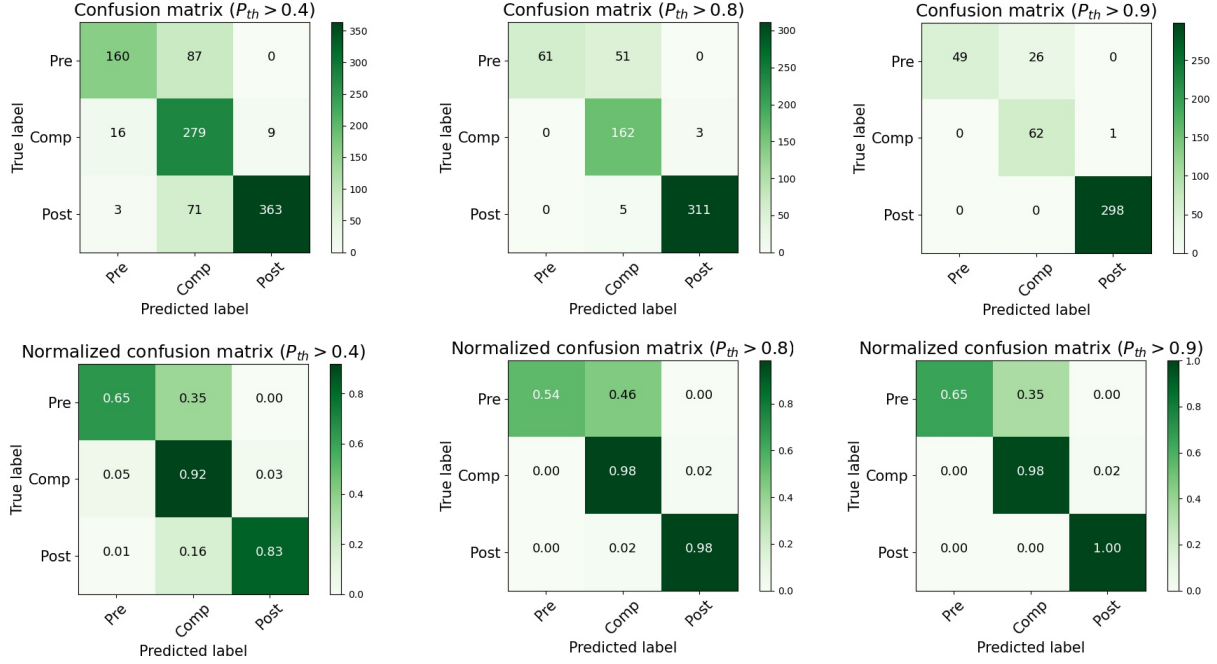

Figure 17: Confusion matrices for VELA test data, with a probability threshold equal to 0.4, 0.8, and 0.9 from left to right. The first row uses the real number of galaxies classified in each class, while the second corresponds to the normalized matrices. The y-axis shows the true class from the simulation metadata, and the x-axis shows the predicted class.

	$\mathbf{P_{th} = 0.4}$			$\mathbf{P_{th} = 0.8}$			$\mathbf{P_{th} = 0.9}$		
	Pre-BN	**BN**	**Post-BN**	**Pre-BN**	**BN**	**Post-BN**	**Pre-BN**	**BN**	**Post-BN**
Precision	$\frac{0.65}{0.65+0.05+0.01} = 0.92$	0.64	0.97	1.00	0.67	0.98	1.00	0.74	0.98
Recall	$\frac{0.65}{0.65+0.35+0.00} = 0.65$	0.92	0.83	0.54	0.98	0.98	0.65	0.98	1.00
F1	$\frac{2 \cdot 0.92 \cdot 0.65}{0.92+0.65} = 0.76$	0.75	0.89	0.70	0.80	0.98	0.79	0.84	0.99

Table 1: Precision, recall and F1-score for the three classes in VELA test data with a fixed probability threshold of 0.4, 0.8, and 0.9 from left to right. Calculations for the first probability in the Pre-BN class are included.

References

[Ali et al., 2015] Ali, A., Shamsuddin, S. M., and Ralescu, A. (2015). Classification with class imbalance problem: A review.

[Allen et al., 2008] Allen, M. G., Groves, B. A., Dopita, M. A., Sutherland, R. S., and Kewley, L. J. (2008). The MAPPINGS III library of fast radiative shock models. *The Astrophysical Journal Supplement Series*.

[Armstrong and Fletcher, 2019] Armstrong, J. A. and Fletcher, L. (2019). Fast solar image classification using Deep Learning and its importance for automation in solar physics. *Solar Physics*.

[Baes et al., 2011] Baes, M., Verstappen, J., De Looze, I., Fritz, J., Saftly, W., Pérez, E., Stalevski, M., and Valcke (2011). Efficient three-dimensional NLTE dust radiative transfer with SKIRT. *The Astrophysical Journal Supplement Series*.

[Barro et al., 2017] Barro, G., Faber, S. M., Koo, D. C., Dekel, A., Fang, J. J., Trump, J. R., Pérez-González, P. G., Pacifici, C., Primack, J. R., Somerville, R. S., and et al. (2017). Structural and star-forming relations since z ∼ 3: connecting compact star-forming and quiescent galaxies. *The Astrophysical Journal*.

[Bruzual and Charlot, 2003] Bruzual, G. and Charlot, S. (2003). Stellar population synthesis at the resolution of 2003. *Monthly Notices of the Royal Astronomical Society*.

[Buda et al., 2018] Buda, M., Maki, A., and Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*.

[Cacciato et al., 2012] Cacciato, M., Dekel, A., and Genel, S. (2012). Evolution of violent gravitational disc instability in galaxies: late stabilization by transition from gas to stellar dominance. *Monthly Notices of the Royal Astronomical Society*.

[Ceverino et al., 2014] Ceverino, D., Klypin, A., Klimek, E. S., Trujillo-Gomez, S., Churchill, C. W., Primack, J., and Dekel, A. (2014). Radiative feedback and the low efficiency of galaxy formation in low-mass haloes at high redshift. *Monthly Notices of the Royal Astronomical Society*.

[Chawla et al., 2002] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*.

[Chollet et al., 2015] Chollet, F. et al. (2015). Keras.

[Dekel and Burkert, 2013] Dekel, A. and Burkert, A. (2013). Wet disc contraction to galactic blue nuggets and quenching to red nuggets. *Monthly Notices of the Royal Astronomical Society*.

[Dekel et al., 2019] Dekel, A., Lapiner, S., and Dubois, Y. (2019). Origin of the golden mass of galaxies and black holes.

[Domínguez Sánchez et al., 2018] Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., Tuccillo, D., and Fischer, J. L. (2018). Improving galaxy morphologies for SDSS with Deep Learning. *Monthly Notices of the Royal Astronomical Society*.

[Draine and Li, 2007] Draine, B. T. and Li, A. (2007). Infrared emission from interstellar dust. IV. The Silicate-Graphite-PAH model in the Post-Spitzer era. *The Astrophysical Journal*.

[Dwek and Arendt, 1998] Dwek, E. and Arendt, R. G. (1998). A tentative detection of the cosmic infrared background at 3.5 μm from COBE / DIRBE observations. *The Astrophysical Journal*.

[Gammie, 2001] Gammie, C. F. (2001). Nonlinear outcome of gravitational instability in cooling, gaseous disks. *The Astrophysical Journal*.

[Gardner et al., 2006] Gardner, J. P., Mather, J. C., Clampin, M., Doyon, R., Greenhouse, M. A., Hammel, H. B., Hutchings, J. B., Jakobsen, P., Lilly, S. J., Long, K. S., and et al. (2006). The James Webb Space Telescope. *Space Science Reviews.*

[He et al., 2015] He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: surpassing human-level performance on ImageNet classification.

[Heinzinger et al., 2019] Heinzinger, M., Elnaggar, A., Wang, Y., Dallago, C., Nechaev, D., Matthes, F., and Rost, B. (2019). Modeling aspects of the language of life through transfer-learning protein sequences. *BMC Bioinformatics.*

[Holmberg, 1941] Holmberg, E. (1941). On the clustering tendencies among the Nebulae. II. A study of encounters between laboratory models of stellar systems by a new integration procedure.

[Huertas-Company et al., 2018] Huertas-Company, M., Primack, J. R., Dekel, A., Koo, D. C., Lapiner, S., Ceverino, D., Simons, R. C., Snyder, G. F., Bernardi, M., Chen, Z., and et al. (2018). Deep Learning identifies high-z galaxies in a central blue nugget phase in a characteristic mass range. *The Astrophysical Journal.*

[James et al., 2002] James, A., Dunne, L., Eales, S., and Edmunds, M. G. (2002). SCUBA observations of galaxies with metallicity measurements: a new method for determining the relation between submillimetre luminosity and dust mass. *Monthly Notices of the Royal Astronomical Society.*

[Jonsson, 2006] Jonsson, P. (2006). SUNRISE: polychromatic dust radiative transfer in arbitrary geometries. *Monthly Notices of the Royal Astronomical Society.*

[Jonsson and Primack, 2010] Jonsson, P. and Primack, J. R. (2010). Accelerating dust temperature calculations with graphics-processing units. *New Astronomy.*

[Koekemoer et al., 2011] Koekemoer, A. M., Faber, S. M., Ferguson, H. C., Grogin, N. A., Kocevski, D. D., Koo, D. C., Lai, K., Lotz, J. M., Lucas, R. A., McGrath, E. J., and et al. (2011). CANDELS: the cosmic assembly near-infrared deep extragalactic legacy survey the Hubble Space Telescope observations, imaging data products, and mosaics. *The Astrophysical Journal Supplement Series.*

[Li et al., 2021] Li, Y., Ni, Y., Croft, R. A. C., Di Matteo, T., Bird, S., and Feng, Y. (2021). AI-assisted superresolution cosmological simulations. *Proceedings of the National Academy of Sciences.*

[Metcalf et al., 2019] Metcalf, R. B., Meneghetti, M., Avestruz, C., Bellagamba, F., Bom, C. R., Bertin, E., Cabanac, R., Courbin, F., Davies, A., Decencière, E., and et al. (2019). The strong gravitational lens finding challenge. *Astronomy Astrophysics.*

[Moody, 2013] Moody, C. E. (2013). Galaxy simulations: kinematics and mock observations.

[NASA / WMAP Science Team, 2006] NASA / WMAP Science Team (2006). Timeline of the Universe.

[O'Malley et al., 2019] O'Malley, T., Bursztein, E., Long, J., Chollet, F., Jin, H., Invernizzi, L., et al. (2019). Keras Tuner.

[Pillepich et al., 2019] Pillepich, A., Nelson, D., Springel, V., Pakmor, R., Torrey, P., Weinberger, R., Vogelsberger, M., Marinacci, F., Genel, S., van der Wel, A., and et al. (2019). First results from the TNG50 simulation: the evolution of stellar and gaseous discs across cosmic time. *Monthly Notices of the Royal Astronomical Society.*

[Rodriguez-Gomez et al., 2018] Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., Nelson, D., Pillepich, A., Springel, V., Genel, S., Weinberger, R., Tacchella, S., Pakmor, R., and et al. (2018). The optical morphologies of galaxies in the IllustrisTNG simulation: a comparison to Pan-STARRS observations. *Monthly Notices of the Royal Astronomical Society.*

[Sarmiento et al., 2021] Sarmiento, R., Huertas-Company, M., Knapen, J. H., Sánchez, S. F., Sánchez, H. D., Drory, N., and Falcón-Barroso, J. (2021). Capturing the physics of MaNGA galaxies with self-supervised Machine Learning.

[Shallue and Vanderburg, 2018] Shallue, C. J. and Vanderburg, A. (2018). Identifying exoplanets with Deep Learning: a five-planet resonant chain around Kepler-80 and an eighth planet around Kepler-90. *The Astronomical Journal*.

[Snyder et al., 2015] Snyder, G. F., Lotz, J., Moody, C., Peth, M., Freeman, P., Ceverino, D., Primack, J., and Dekel, A. (2015). Diverse structural evolution at z > 1 in cosmologically simulated galaxies. *Monthly Notices of the Royal Astronomical Society*.

[Springel et al., 2005] Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., and et al. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*.

[Tacchella et al., 2016] Tacchella, S., Dekel, A., Carollo, C. M., Ceverino, D., DeGraf, C., Lapiner, S., Mandelker, N., and Primack Joel, R. (2016). The confinement of star-forming galaxies into a main sequence through episodes of gas compaction, depletion and replenishment. *Monthly Notices of the Royal Astronomical Society*.

[Tomassetti et al., 2016] Tomassetti, M., Dekel, A., Mandelker, N., Ceverino, D., Lapiner, S., Faber, S., Kneller, O., Primack, J., and Sai, T. (2016). Evolution of galaxy shapes from prolate to oblate through compaction events. *Monthly Notices of the Royal Astronomical Society*.

[Vogelsberger et al., 2014] Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., and Hernquist, L. (2014). Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *Monthly Notices of the Royal Astronomical Society*.

[Wei and Roberts, 2018] Wei, L. and Roberts, E. (2018). Neural network control of focal position during time-lapse microscopy of cells. *Scientific Reports*.

[Wei et al., 2017] Wei, T., Wang, Y., and Zhu, Q. (2017). Deep Reinforcement Learning for building HVAC control.

[Zolotov et al., 2015] Zolotov, A., Dekel, A., Mandelker, N., Tweed, D., Inoue, S., DeGraf, C., Ceverino, D., Primack, J. R., Barro, G., and Faber, S. M. (2015). Compaction and quenching of high-z galaxies in cosmological simulations: blue and red nuggets. *Monthly Notices of the Royal Astronomical Society*.