Galaxy stellar and total mass estimation using machine learning

Jiani Chu [1]★ Hongming Tang [1]★ Dandan Xu,[1] Shengdong Lu [2] and Richard Long [1,3]

[1]*Department of Astronomy, Tsinghua University, Beijing 100084, China*
[2]*Department of Physics, Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[3]*Department of Physics and Astronomy, Jodrell Bank Centre for Astrophysics, The University of Manchester, Oxford Road, Manchester M13 9PL, UK*

ABSTRACT

Conventional galaxy mass estimation methods suffer from model assumptions and degeneracies. Machine learning (ML), which reduces the reliance on such assumptions, can be used to determine how well present-day observations can yield predictions for the distributions of stellar and dark matter. In this work, we use a general sample of galaxies from the TNG100 simulation to investigate the ability of multibranch convolutional neural network (CNN) based ML methods to predict the central (i.e. within $1-2$ effective radii) stellar and total masses, and the stellar mass-to-light ratio (M_*/L). These models take galaxy images and spatially resolved mean velocity and velocity dispersion maps as inputs. Such CNN-based models can, in general, break the degeneracy between baryonic and dark matter in the sense that the model can make reliable predictions on the individual contributions of each component. For example, with r-band images and two galaxy kinematic maps as inputs, our model predicting M_*/L has a prediction uncertainty of 0.04 dex. Moreover, to investigate which (global) features significantly contribute to the correct predictions of the properties above, we utilize a gradient-boosting machine. We find that galaxy luminosity dominates the prediction of all masses in the central regions, with stellar velocity dispersion coming next. We also investigate the main contributing features when predicting stellar and dark matter mass fractions ($f_*, f_{\rm DM}$) and the dark matter mass $M_{\rm DM}$, and discuss the underlying astrophysics.

Key words: methods: data analysis – galaxies: kinematics and dynamics.

1 INTRODUCTION

Achieving a full understanding of galaxy evolution requires accurate measurements of the 'unseen' matter. This is why, among the many areas in astrophysical measurements and modelling, galaxy dynamics and gravitational lensing play unique roles, as they provide meaningful constraints on dark matter. On the other side of these measurements lies the distribution of the stellar component. Properties, such as stellar mass-to-light ratio (M_*/L) and initial mass function (IMF), are essential properties to solving the galaxy evolution puzzle. Therefore, a fundamental task in this regard comes down to accurate determinations of the different contributions of dark matter and baryons – a central goal of galaxy dynamics and gravitational lensing studies.

In this regard, recent integral field spectroscopic observations have provided good data sets to study the dynamical properties of galaxies for a large sample of galaxies across a wide range of Hubble types, both in the nearby Universe, for example, those from the ATLAS[3D] (Cappellari et al. 2011), MaNGA (Bundy et al. 2015), and SAMI (Fogarty et al. 2014) surveys, and at high redshifts, e.g. the KMOS Galaxy Evolution Survey (KGES, Turner et al. 2017). Historically, people have developed various dynamical modelling methods (e.g. Jeans 1922; Schwarzschild 1979; Syer & Tremaine

1996), which typically combine a single-band image of a galaxy with stellar Integral Field Unit (IFU) kinematic maps to constrain matter distributions across the galaxy. Specifically, these methods routinely split the total matter distribution into dark matter and baryonic matter components with each described by a pre-specified density profile. Modelling is then concerned with disentangling the individual contributions from both components, with specific objectives often set on estimating the central dark matter fraction or the IMF of the galaxy. The latter objective is often accomplished in combination with stellar population synthesis (SPS) utilizing multiband photometry or spectroscopic data. Such models have been widely applied to existing IFU galaxy surveys, for example, Zhu et al. (2018a) adopted the orbit-based Schwarzschild modelling techniques (Schwarzschild 1979) to the stellar kinematic data of 300 galaxies in the CALIFA survey (Sánchez et al. 2012). Long & Mao (2012) applied the made-to-measure (M2M, Long & Mao 2010) dynamical modelling technique to the Spectrographic Areal Unit for Research on Optical Nebulae (Cappellari et al. 2006) galaxies. Recently, Zhu et al. (2023) applied the Jeans anisotropic modelling (JAM, Cappellari 2008, 2020) to the complete MaNGA sample (over 10K nearby galaxies with different morphologies) and obtained their 'quality-assessed' dynamical properties.

One must be aware that even for the most sophisticated modelling techniques, there exist assumptions and approximations, which may lead to biased estimates on different levels. In terms of galaxy dynamical modelling, as full six-dimensional phase-space data can

★ E-mail: zjn20@mails.tsinghua.edu.cn (JC); hongmingt@tsinghua.edu.cn (HT)

not be obtained with current observational capabilities, dynamical properties are only inferred from projected light and kinematics. As a result, faithful estimation of dynamical, spatial, and orbital properties is not achievable over a wide range of cases. For example, Schwarzschild-based dynamical studies show that edge-on projections are preferred as more kinematic information is available (e.g. Zhu et al. 2018b, 2020), and inclined galaxies are more difficult to model accurately. In all cases, one must first make assumptions about a galaxy's matter geometry (e.g. spherical, elliptical, axisymmetric, or triaxial, etc.). In the case of Jeans-based methods, one must also make assumptions on the shape of velocity dispersion ellipsoid. The inferred results may then differ under different model assumptions. For example, as shown in fig. 10 of Xu et al. (2017), radial isotropy tends to underestimate the logarithmic slope of a galaxy's total density distribution in the inner region, while tangential isotropy tends to give overestimated results, when testing with simulated galaxies. Note that there are limitations introduced by Jeans modelling and by the weighting schemes used in the Schwarzschild and M2M methods. Jeans modelling may produce non-physical models (Cappellari 2008, section 3.1.1) while the weighting schemes determine weights that are numerically satisfactory in an optimization context but may not be so astrophysically.

In addition, one must also take into account that information about a galaxy's light distribution does NOT directly equate to that of the baryonic mass distribution, unless it is known exactly how one is related to the other. An average converting factor, the baryonic M_*/L, could be obtained from the stellar M_*/L with a correction for gas, for example, via an assumed stellar mass-to-gas mass scaling relation. The stellar M_*/L, therefore, is a fundamental attribute, upon which estimates of many other properties may depend, for example, the dark matter fraction. As we know, M_*/L is neither a fixed value across a single galaxy nor some universal distribution across the galaxy population (e.g. van Dokkum & Conroy 2010; Li et al. 2017; van Dokkum et al. 2017; Oldham & Auger 2018a; Zhou et al. 2019; Lu et al. 2023). It depends on many galaxy properties such as IMF, star formation history, and so on, which may vary spatially, evolve with time, and depend on galaxy types. From observations, estimating the 3D stellar mass distribution from the light distribution may be attempted using SPS with 2D spatially resolved, spectroscopic data such that IMF-sensitive absorption-line features can help indicate the underlying stellar population and help constrain the stellar M_*/L profile and thus approximate the stellar mass distribution (e.g. van Dokkum & Conroy 2010; Spiniello et al. 2014; Parikh et al. 2018; Bernardi et al. 2023). However, such analyses are not straightforward to achieve at low cost for the majority of galaxies at all redshifts and have their own degeneracies and shortcomings, for example, arising from lack of 3D observational data.

In conventional dynamical modelling, people commonly approximate M_*/L with a constant value across a galaxy (e.g. Cappellari et al. 2011; Zhu et al. 2023), though not always (e.g. Oldham & Auger 2018b). Unsurprisingly, observations indicate that constant M_*/L may not be a universally good assumption across the galaxy population (e.g. Tortora et al. 2011; García-Benito et al. 2019; Ge et al. 2021; Lu et al. 2023). Some studies, in particular those which combine stellar kinematics with gravitational lensing measurements for galaxies at higher redshifts, have adopted a power-law model to describe an imposed radial dependence of M_*/L, and attempted to estimate the power-law index from the observed data (e.g. Sonnenfeld et al. 2018; Oldham & Auger 2018b; Shajib et al. 2021). The situation becomes even more complicated when the choice of different light and dark matter density models is taken into account. As neither of these M_*/L assumptions (neither constant nor power law) represents

the true distribution, model fitting under such assumptions lacks the power to select the right density model that is truly responsible for generating the observational data. This leads to model degeneracies being artificially broken, and causes the results to sensitively depend on the specific choice of light and/or dark matter density model, reaching biased estimates on either the stellar mass and thus the derived IMF or the central dark matter fraction. What is more, the absence of 6D data means we know nothing about the precise 3D spatial distribution of matter in a galaxy. This bias has been indeed manifested when tested against simulation galaxies for which ground truth values are known (e.g. Li et al. 2016), or when tested against different model implementations on the same observational data (e.g. see fig. 12 of Zhu et al. 2023), and in some cases even by contradicting results obtained for similar galaxy populations (under the same M_*/L assumption), but through different choices of the light model adopted (see Sonnenfeld et al. 2018; Shajib et al. 2021).

Machine-learning provides an alternative way to start tackling the galaxy evolution problem, and has the advantage of making estimates of galaxy properties while eliminating many of the previous modelling assumptions. In addition, it has also been widely used as a powerful tool to understand the significant physical properties that link to cosmic structure formation and galaxy evolution. More and more studies have taken such approaches, from simply making predictions to certain properties (e.g. for galaxies, by Bonjean et al. 2019; for galaxy clusters, by e.g. Armitage, Kay & Barnes 2019; Ho et al. 2019), to inferring cosmological models and parameters (e.g. Arjona & Nesseris 2020), from emulating cosmic structure growth (e.g. He et al. 2019; Man et al. 2019; Chen et al. 2021; Tabor & Loeb 2021) to searching for physical connections between the predicted properties and input observational features (e.g. Dobbels et al. 2019; Lucie-Smith, Adhikari & Wechsler 2022). In a recent study by Angeloudi et al. (2023), galaxy populations from the TNG100[1] (Genel et al. 2018; Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018) and EAGLE simulations (Schaye et al. 2015) were used to calibrate an ML approach, which successfully predicted the fraction of accreted stars in galaxies from IFU-like observations. Gomer et al. (2023) trained a neural network as an emulator, massively speeding up likelihood evaluation for sophisticated and expensive dynamical modelling (around 200 times faster than similar emulations using JAM). In Herná ndez, González & Padilla (2023), the stellar mass and star formation rate (SFR) of galaxies from the TNG300 simulation (Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018) were predicted using a neural network, which took as input 12 properties of galactic haloes and their nearby environments. It was found that certain merger tree properties contribute significantly to the results from their ML model. Wu et al. (2023) used a random forest (RF) based ML algorithm on TNG100 to predict the total and dark matter masses of galaxies with several simple observables as input, and then tested their approach on real galaxies. The results of their RF-based algorithm are consistent with the dynamical masses of real samples, and show the great potential of ML to make realistic estimates of galaxy masses. The Euclid Collaboration et al. (2023) explored the potential of ML to estimate galaxy properties such as redshift, stellar mass, and SFR with data from the Euclid (Laureijs et al. 2011) and Rubin/LSST (Ivezic et al. 2008) surveys. They found that their models performed better in accuracy than spectral energy distribution modelling when predicting these properties.

---

[1] www.tng-project.org

Our goal in this study is not to develop any specific models to be applied to observational data, but to address the particular question as to whether existing or future observations might provide us with sufficient information for us to correctly disentangle individual mass contributions from baryons and dark matter. If yes, what are the reasons for such achievement; if no, again, what are the reasons? To do so, we take galaxies from the state-of-the-art cosmological hydrodynamical simulation – the TNG100 simulation, and make mock observations of photometric images and IFU-like velocity maps for these galaxies. We utilize a convolutional neural network (CNN; Fukushima & Miyake 1982; Lecun et al. 1998; Krizhevsky, Sutskever & Hinton 2012; He et al. 2016) model to predict the stellar mass M_* and total mass M_{tot} of galaxies that are enclosed within one half-stellar-mass spherical radius R_{hsm}, as well as the stellar M_*/L. We note that the detailed stellar and total mass density distributions of TNG100 galaxies are not precisely consistent with those observed (e.g. Lu et al. 2020; Romeo, Agertz & Renaud 2020). Therefore, we only employ the ML methods investigated in this study purely on the simulation data set as a proof-of-concept study. Implementation to real observational data will also require further investigations examining observational effects, selection rules and so on.

Results from our GPU-based CNN models suggest that the ML approach, for our simulated galaxies, is able to untwine the individual contributions from both dark and baryonic matter, from input maps of a galaxy's surface brightness distribution and its first and second line-of-sight velocity moments. To reveal any key factors which lead the CNN model to be able to make successful predictions, we use summary statistics from the images and maps as input to a gradient-boosting decision tree model (GBDT; Friedman 2001; Ke et al. 2017) to predict the values of the same galaxy properties.

The structure of the rest of this paper is as follows. In Section 2, we introduce the IllustrisTNG galaxy sample that we use for this study, and how we build the data sets suitable for ML model training, testing, and interpretation. In Section 3, we give detailed descriptions of the two ML methods (CNN and GBDT) that we use. We show our results in Sections 4 (for CNN) and 5 (for GBDT). Finally, we present discussions of our results and our overall conclusions in Section 6.

2 SAMPLE SELECTION AND DATA SET FOUNDATION

As mentioned in Section 1, this work aims to test the feasibility and fidelity of two supervised ML methods (CNN and GBDT) in predicting the stellar, dark-matter, and total masses of galaxies, and the galaxies' M_*/Ls. In order to do this, we require a galaxy sample where these properties are known. Given that observations of real galaxies may suffer from various systematic biases, we choose to extract observationally equivalent data values from realistic galaxy simulations. In Section 2.1, we introduce the simulation-based galaxy sample used in this study. In Sections 2.2 and 2.3, we describe how we organize the necessary input data and targets suitable for CNN and GBDT modelling.

2.1 Sample selection

Our galaxy sample comes from the TNG100 simulation (Genel et al. 2018; Marinacci et al. 2018; Naiman et al. 2018; Nelson et al. 2018; Pillepich et al. 2018; Springel et al. 2018), which is a set of magnetohydrodynamical (MHD) cosmological simulations of galaxy formation and evolution, using the AREPO software (Springel 2010). The simulation has been shown to broadly agree with many observed galaxy properties and general scaling relations, including

the bimodal colour distribution (Nelson et al. 2018), the mass–size relation (Genel et al. 2018), the galaxy mass density profiles (Wang et al. 2019, 2020), the fundamental plane relation (Lu et al. 2020), the dark matter fractions (Lovell et al. 2018), as well as the stellar orbit compositions (Xu et al. 2019). Specifically, the simulation has a box volume of $(110.7\,\mathrm{Mpc})^3$, a mass resolution of 1.4×10^6 and $7.5 \times 10^6\,\mathrm{M}_\odot$ for baryons and dark matter, respectively, and a force softening length of $0.5\,h^{-1}\,\mathrm{kpc}$. The SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009) is used to identify galaxies and their dark matter haloes. General galaxy properties are available from Nelson et al. (2019).

We take all galaxies at redshift $z = 0$ which have stellar mass within 30 kpc greater than $5 \times 10^9\,\mathrm{M}_\odot$, and total subhalo mass (as calculated by SUBFIND) less than $10^{14}\,\mathrm{M}_\odot$ The lower limit is to guarantee sufficient resolution, and the upper limit is to exclude systems in galaxy cluster environments, which are beyond the galaxy–mass range investigated in this study. To mimic the random orientation effect of observed galaxies, we just use the orientation of galaxies in the simulation. We project individual galaxies along the three principal axes of the simulation box, and take each projection as an independent galaxy in our sample. This projection operation enlarges our sample size, balancing data set complexity (Barella et al. 2021) and model complexity (Hu et al. 2021). The final $z = 0$ data set contains a total of 28 110 galaxies (i.e. 9370 unique galaxies with three different projections of each).

2.2 Data input, target generation, and pre-processing for CNN

For CNN modelling, our input data for a galaxy comprises its r-band image, its $g - r$ colour map, the spatial distributions of stellar line of sight (along the direction of projection, i.e. the simulation axes) mean velocity and velocity dispersion for a given projection as they all contribute to the estimation of galaxy masses and stellar M_*/L (Binney & Tremaine 2008; Dobbels et al. 2019). Note that the first and second moments of line-of-sight velocities are directly calculated from stellar particles in the simulated galaxies. In this sense, these quantities do not have the same kinds of measurement errors as those derived from spectral line fittings. For simplicity, we do not consider the third and fourth velocity moments (h_3 and h_4). We note that all the data used in this work were extracted using pipelines developed for various previous studies by the authors (Xu et al. 2017, 2019; Lu et al. 2020, 2021, 2022). In particular, the spatial range and resolution of the kinematic maps of the simulated galaxies generally resemble typical SDSS and MaNGA–IFU observations. For MaNGA galaxies, the IFU observations for the stellar kinematic maps typically have a spatial coverage within $1.5-2.5$ effective radii from the galaxy centre (Bundy et al. 2015). For all images and maps of the simulated galaxies, the spatial range was set to be within $\pm 3 R_{hsm}$ from the galaxy centre, where R_{hsm} is the 3D half-stellar-mass radius of the galaxy, roughly equivalent to the effective radius for an observed galaxy. Here below, we give a brief recapitulation of the techniques used to extract the data for the simulated galaxies.

For the g- and r-band images of the simulated galaxies, the luminosities of the stellar particles were processed for dust attenuation effects. This was carried out through a simple semi-analytical approach (see Xu et al. 2017 for details). Specifically, the r-band images and colour maps were produced in a mesh of 300×300, corresponding to $0.02 R_{hsm}$ per pixel. This high spatial resolution allows the adoption of a cubic spline kernel as used in the smoothed particle hydrodynamics (SPH; Monaghan 1992; Hultman & Pharasyn 1999). The dust-attenuated luminosities are then assigned and smoothed via the SPH scheme into mesh pixels,
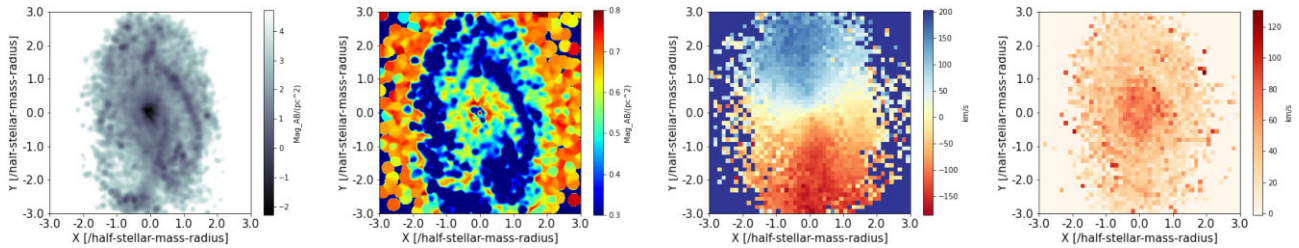
Figure 1. From left to right panels, the four panels present the *r*-band image, the $g − r$ colour map, the line-of-sight mean velocity map, and velocity dispersion map of an example galaxy (subhalo-ID: 501761). All images and velocity maps are produced in the range of $\pm 3R_{\text{hsm}}$ from the galaxy centre. They are the basic input data set fed to our CNN-based model as Fig. 3 shows.

with a smoothing length that encloses the nearest 32 neighbouring stellar particles. For our CNN resolution tests, the original SPH-smoothed images are re-binned to resolutions of 150×150 and 60×60. The last setting matches the median resolution for the SDSS galaxy survey.

For the kinematic maps, instead of adopting a Voronoi binning scheme, for simplicity, we directly projected stellar particles on to a mesh using the near-grid-point (NGP) scheme. These velocity maps have dimensions of 48×48 pixels, corresponding to a spatial resolution of 8 pixel per R_{hsm}, which corresponds to resolving a typical MaNGA galaxy at redshift $z = 0.05$, and is slightly below the median value of 14 pixel per R_{hsm} for the entire MaNGA galaxy sample (derived from Bundy et al. 2015). Fig. 1 shows the above-mentioned images and maps for one example galaxy in our data set. Note that we do not add any observational errors to the generated images and maps.

For the training process on our CNN network, we must provide target galaxy data values for the properties we wish our network model to predict. In our case, we provide the central stellar mass (M_*), total mass (M_{tot}), and M_*/L ($M_*/L \equiv M_*^{2D}/L$, where L is the *r*-band dust-attenuated luminosity). For modelling real galaxies, estimating these values has always been the goal for conventional studies in SPS, stellar kinematics, and gravitational lensing. In particular, the stellar mass (or the stellar M_*/L) and total mass are the two most commonly derived basic quantities. Once they are obtained, the dark matter fraction f_{dm} can, in principle, be further derived. Here, the mass values are determined using particles of the corresponding type, located within a 3D sphere of radius R_{hsm} from the galaxy centre. The M_*/L value is a projected quantity and is calculated using the *r*-band luminosity and stellar mass of stellar particles projected within a radius of R_{hsm} for a given line of sight.

Prior to using the data in CNN modelling, the data are cleaned (for example, to ensure it does not contain any spurious values) and formatted according to the requirements of the modelling software being used (PyTorch in our case). To ensure auto-diff (automatic differentiation) functions properly, we perform a normalization operation, which is to linearly scale all pixels of images and maps to ensure the numerical value of the pixels is between 0 and 1. In our models, we split our galaxy samples into three parts: a training set (16000 samples), a validation set (4000 samples), and a test set (8110 samples). Operationally, it is convenient, for comparison purposes, to ensure that the sets always contain the same galaxies. This is achieved by setting a random seed to the same value in all modelling runs.

2.3 Data input and target generation for GBDT

For GBDT modelling (Friedman 2001), the model inputs are a number of summary statistics extracted from the particles of the

Table 1. Feature inputs of our GBDT model.

input	description
M_r	SDSS *r*-band absolute AB magnitude
$g − r$	SDSS $g − r$ colour
SFR	Star-forming rate over the past 1 Gyr within a projected radius of $2R_{\text{hsm}}$ from the galaxy centre
n_{Ser}	Sérsic index from a Sérsic profile (Sérsic 1963) fitting to the light distribution within a projected radius of $5R_{\text{hsm}}$
c/a	Shortest-to-longest axis ratio *c/a* of the stellar mass distribution, calculated using the inertial tensor method (Allgood et al. 2006), through an iterative approach started within a 3d radius of $3R_{\text{hsm}}$ (see Emsellem et al. 2007 for detailed definition)
σ_{v}	The the *r*-band luminosity weighted line-of-sight velocity dispersion within a projected radius of $2R_{\text{hsm}}$
λ	dimensionless spin calculated within projected $2R_{\text{hsm}}$ (see Emsellem et al. 2007 for detailed definition)
B/T	A kinetic bulge-to-total ratio within $2R_{\text{hsm}}$, defined as two times the mass fraction of stellar particles with negative circularity, i.e. $\epsilon < 0$ (see Xu et al. 2019 for definition)
f_{cold}	Mass fraction of stellar particles with cold orbits ($\epsilon > 0.8$) within a 3d radius of $2R_{\text{hsm}}$ (see Xu et al. 2019 for definition)
f_{hot}	Mass fraction of stellar particles with hot orbits ($|\epsilon| < 0.25$) within a 3d radius of $2R_{\text{hsm}}$ (see Xu et al. 2019 for definition)

simulated galaxies. We use the following quantities: the *r*-band absolute magnitude M_r and $g − r$ colour of a galaxy, the star-formation rate, the Sérsic index n_{Ser}, the stellar axis ratio *c/a*, the velocity dispersion σ_{v}, the dimensionless spin attribute λ (quantifying the degree of stellar rotation, see Emsellem et al. 2007 for a detailed definition), the kinetic bulge-to-total ratio (B/T), and the cold and hot orbital fractions f_{cold} and f_{hot}. Notice that orbit fractions can not be directly obtained without dynamical modelling. and are only used in GBDT tasks. Detailed descriptions of the quantities can be found in Table 1.

Using the above-mentioned quantities, we make predictions on the stellar mass M_*, dark matter mass M_{DM}, and total mass M_{tot} of our galaxies, as well as on the M_*/L and dark matter fraction f_{dm}. For training purposes, all these quantities are determined within one R_{hsm}.

We exclude galaxies whose Sérsic index n_{Ser} is larger than 100 and whose B/T is larger than 1. Such galaxies only make ~ 3 per cent of the total sample size.

3 METHODOLOGY

The workflow for our modelling is displayed in Fig. 2. Section 2 covers the galaxy sample selection and data-processing aspects of the workflow. In this section, we introduce the model architecture and
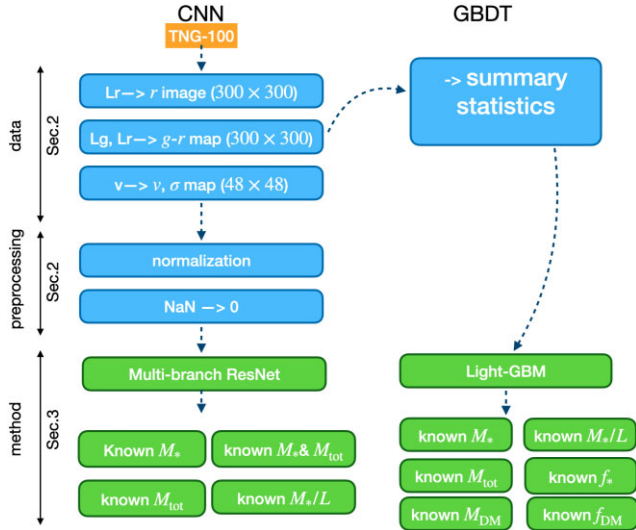
Figure 2. Workflow of this study: data, pre-processing, and model tasks from top to bottom (left panel: CNN; right panel: GBDT).

Figure 3. Structure of multibranch ResNet in this work: $g - r$ map, r-band image, and velocity maps are processed by normal ResNet-18 backbone independently until the last fully connected layer. At last, all intermediate outputs are combined to give predictions via a fully connected layer.

training setup of the two ML algorithms we use: CNN is described in Section 3.1 and GBDT in Section 3.2.

Given that we make use of both algorithms to predict numerical values, we utilize the same mean square error (MSE) loss function (equation 1) when training our models.

$$MSE = \frac{1}{N} \sum_i^N (y_i^{\text{pred}} - y_i^{\text{true}})^2, \qquad (1)$$

where y_i^{pred} and y_i^{true} are the predicted and the true values of the galaxy attributes, and N is the sample size.

3.1 Convolutional neural network: multibranch ResNet

CNNs are a type of artificial neural network (ANN) making use of convolution filters (kernels) that enable them to capture features directly from input images, and are widely used in image processing (Arena et al. 2003; Bialopetravičius & Narbutis 2020; Han et al. 2020; Nishimoto et al. 2022; Shi et al. 2022). In practice, a typical CNN would follow a top-down structure: some convolutional layers in linear sequence first extract features from model inputs, squeeze these features into a linear format, and forward them to sequenced fully connected layers for further feature abstraction and then make a model prediction.

A common technique to improve a CNN model's performance is to increase the number of layers in the model. However, it was found that a model cannot always improve its performance by simply adding more network layers. Model classification accuracy may saturate and eventually suffer from rapid degradation (He et al. 2016). To resolve this so-called 'degradation problem', He et al. (2016) proposed deep residual networks. This architecture introduced a 'residual block' (see schematic in fig. 2 of He et al. 2016) – instead of optimizing the output of a stacked two-layer block $F(x)$, a 'residual block' asked the network to optimize the combination of block output $F(x)$ and block input x, which gives $H(x) = F(x) + x$. This optimization was believed to be easier to achieve (He et al. 2016). Such an innovation helped ResNet win the 2015 ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al. 2015). ResNet has been used in earlier astronomical studies such as finding galaxy–galaxy strong lenses
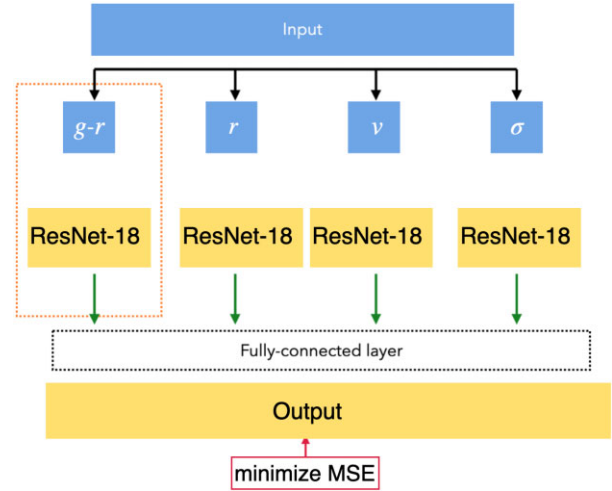
(Lanusse et al. 2018), and classifying galaxy clusters (Su et al. 2020). With both model performance and computation power limitations in mind, we chose a modified version of ResNet-18 (ResNet-18 hereafter for simplicity; He et al. 2016; Su et al. 2020) as our CNN backbone: The backbone contains one convolutional layer and 8 'residual blocks'.

While a classic top-down CNN structure can extract features from a single image or map and make predictions, it cannot solve our requirement to use multiple images and maps with different mesh sizes. An approach to address such a requirement is to use a multibranch neural network (Al Rahhal et al. 2018). Multibranched network architectures allow one to simultaneously utilize multiple input data sets in one model for feature extraction and model prediction. Multibranched networks have previously been used to identify lensed supernovae (Morgan et al. 2022) and giant radio galaxies (Tang et al. 2022). Tang et al. (2022) suggested implementing such architectures could boost model performance. In this work, we adopt a multibranched network architecture for our CNN-based models. It can be seen from Fig. 3 that our input images and maps are individually fed into ResNet-18 backbones. These backbones extract features from images/maps, and forward their outputs to a fully connected layer to produce a model prediction.

When training models, we determine model hyperparameters through a manual selection process (Bergstra & Bengio 2012). This is achieved by training our model using the training data set, and manually selecting model hyperparameters by looking at the model's behaviours when making predictions using samples in the validation set. The model's ability to generalize is evaluated using the test set of samples, as may be seen in Section 4.

After experimentation, we choose the Adam optimizer, which performs better than the stochastic gradient descent optimizer. To avoid overfitting, we utilize a cosine learning rate scheduler (Schaul, Zhang & LeCun 2013) ($lr = 0.0001 \cos(n/200)$, where n is the training epoch number) instead of a constant learning rate. Table 2 lists all the hyperparameters of our multibranch ResNet. A training batch size of 40 was chosen after consideration of the available GPU memory. As is common practice, we shuffle our sample input sequences before each model training epoch.

Table 2. Hyperparameters of our multibranch ResNet model.

Hyperparameter	Setup
Batch size	40
Initial learning rate	0.0001
Learning rate step	200
Training epoch	200

The CNN-based model implementation in this work uses the PYTORCH python library,[2] where ResNet backbones (i.e. ResNet-18 and ResNet-50) are inbuilt and ready to use.

3.2 Gradient-boosting machine: Light–GBM

A GBDT is an ML method with an efficient catalogue data-processing capability and high model interpretability. It builds a series of decision trees to address either classification or regression problems, and has been used in recent astronomical studies (e.g. Coronado-Blázquez 2022; Sahakyan, Vardanyan & Khachatryan 2023). GBDT attempts to build a 'strong' model using multiple 'weak' models (i.e. decision trees). The general procedure for GBDT training is as follows:

(i) The first tree is formed to fit the given training data and make predictions.

(ii) A second tree is then formed to fit the residuals between the first tree's predicted values and the truth values.

(iii) The next step is iterative where successive trees are trained to fit the residuals of the previous one.

(iv) The model training process stops when some customized stopping criteria have been met.

Although GBDT has been widely implemented, training a GBDT model can be time-consuming when the training data set has a large sample size or a considerable number of features. This is because the classic GBDT needs to scan all data points and estimate the information gain for every possible split point (Ke et al. 2017). A representative approach that tackled this time issue is LightGBM (Ke et al. 2017). This is an improved version of GBDT with two novel techniques introduced, gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB) (see Ke et al. (2017) for full technical details). In short, GOSS allows LightGBM to estimate the information gain at its split points using those data samples with larger gradients, and EFB bundles mutually exclusive features to decrease the number of features. Compared with the original GBDT, LightGBM speeds up the model training process by up to 20 times and achieves similar accuracy Ke et al. (2017). Here, we implement LightGBM to predict the attributes of our galaxies (M_*, M_{tot}, M_{DM}, M_*/L, f_* and f_{DM}). We also investigate which summarized image or map features have contributed to the prediction of which attributes by performing feature importance evaluations of these LightGBM models (see Section 5).

To train and test LightGBM-based models, we split our samples into two parts: the training set (18833 samples) and the test set (9277 samples). We do not consider validation sets or model cross-validation as these models are developed for proof-of-concept purposes. Rather than emphasizing model accuracy and stability, we would like to investigate its ability to (a) break the degeneracy between baryonic and dark matter in galaxies, and (b) explore why different galaxies have different properties.

Table 3. Hyper parameters of our GBDT model.

Hyperparameter	Setup
num_leaves	5
num_boost_round	500
objective	regression
min_data_in_leaf	20 (default)
max_depth	No limit
Evaluation metrics	MAE and MSE

Instead of tuning hyperparameters, we conduct experiments with different input and output combinations, while maintaining the same set of hyperparameter values, to help us evaluate the behaviour of our GBDT models and to understand what summary statistics of images and maps have specific physical meanings. By comparing model behaviours trained with inputs of different ranges with fixed model training hyperparameters, we find that when using input features at $2R_{hsm}$ apertures instead of $1R_{hsm}$, the resulting mass prediction uncertainty decreased by 9 per cent. This is why our inputs are mainly galaxy properties from $2R_{hsm}$ apertures, while our outputs are all at $1R_{hsm}$ apertures.

Table 3 lists the hyperparameters we use (after experimentation) in our LightGBM models. We use 500 trees for training, with the maximum number of leaves in any node (the *num_leaves* parameter) set to 5. We do not include further constraints on the minimum data sample number in any one leaf, the maximum depth of a tree (the typical depth in this work is 4), or apply other regularization methods as we do not observe any sign of model overfitting. Notably, we have evaluated MAE and MSE loss functions to understand whether model robustness is affected by possible outliers in our data samples. We find that training algorithms in this work behave well when looking at both loss curves (see Fig. 4 for an example).

4 CNN-BASED MODEL RESULTS

In this section, we present the results from our CNN-based modelling. In all our models, we train the network using the multibranch ResNet taking some combination of r-band images (300×300 SPH-smoothed images within $\pm 3R_{hsm}$) and the two velocity maps (48×48 NGP-smoothed maps within $\pm 3R_{hsm}$) for mean velocity and dispersion simultaneously as input, and a variety of galaxy mass and light quantities as targets.

As an example, Fig. 5 shows the model training and validation loss curves using galaxy stellar mass M_* as the target. As can be seen, while the model validation loss is high and oscillates considerably initially, it gradually decreases and becomes stable after ~ 140 training epochs. We train our model for a total of 200 epochs and select the epoch where the validation loss minimizes our model. Doing so prevents a model from overfitting the data. We save this model and evaluate its performance using the test data set. Instead of using MSE loss to evaluate our model performance, we use $1-\sigma$ of $\log y^{pred}/y^{true}$ to evaluate the uncertainty of the prediction, which can be understood as the scatter of the prediction. Here, the uncertainty is intended to describe the performance of the model prediction, and is not related to observational errors.

Table 4 summarizes all the CNN models we run, and indicates the subsections in which the results are described. We wish to point out that some of the tests are concerned with predicting attribute values within $2R_{hsm}$ rather than just the $1R_{hsm}$ we usually employ.
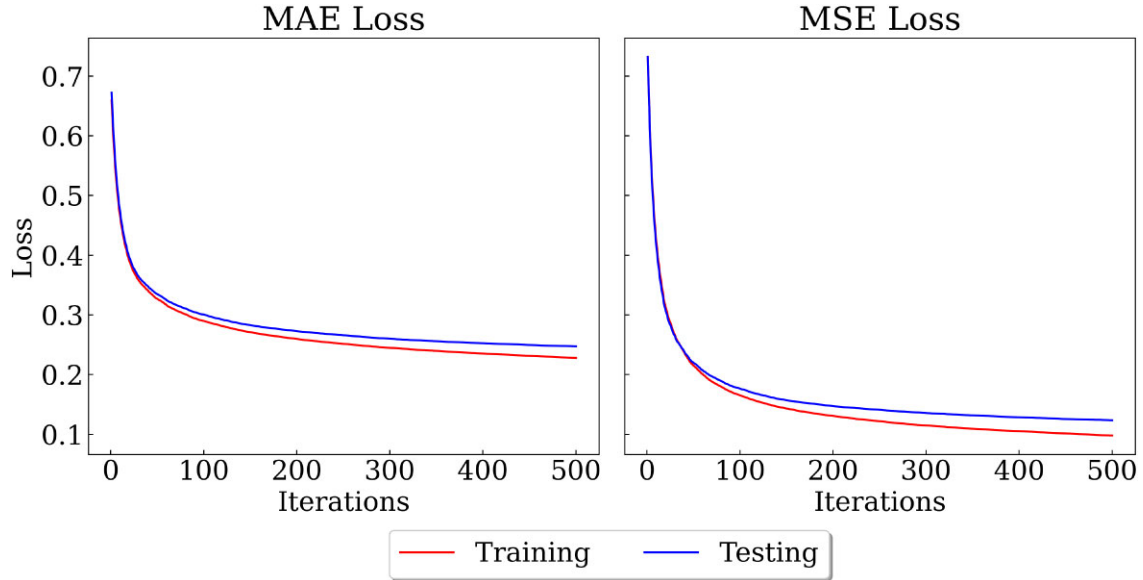
Figure 4. Loss curves of LightGBM-based model as a function of training iterations with known M_*/L [left panel: mean absolute error (MAE) loss; right panel: mean square error (MSE) loss]. In each panel, the red and blue curves indicate the training and testing losses, respectively.
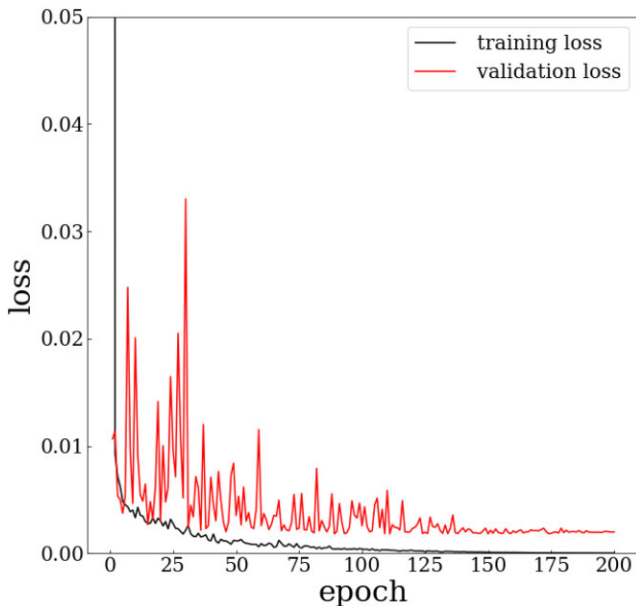

Figure 5. MSE loss (see Section 4.1) curve of CNN-based model with known M_* as function of training epoch. The black and red curves indicate the training and validation loss, respectively.

4.1 Training CNN on galaxies with known stellar mass as target

Using observations, the stellar mass can be estimated through many methods, including SPS analysis through spectrum or SED fitting, multicomponent modelling with stellar kinematics, and gravitational lensing measurements. In our CNN model, we train a multibranch ResNet using r-band images and two velocity maps as input, and using known stellar mass M_* ($\leqslant R_{hsm}$) (i.e. the stellar mass enclosed within a 3D sphere with radius of R_{hsm}) as target. This specific setup is to imitate the situation where we only have good knowledge about stellar mass M_* for the training sample. For instance, galaxies in the southern sky may lack high-quality spectroscopic and

multiwaveband photometric data, lowering SED estimated galaxy stellar mass reliability. The top panel of Fig. 6 shows the CNN-predicted stellar mass M_* ($\leqslant R_{hsm}$) versus the true values for the test set. As shown in Table 4, the $1-\sigma$ scatters in the predicted $\log M_*^{pred}/M_*^{true}$ values, either within R_{hsm} or within $2R_{hsm}$, are 0.04 dex (i.e. ~ 10 per cent). Under the default image and map conditions, if only r-band images, or g- and r-images combined, or solely two velocity maps are used as input, the uncertainties become slightly larger, of 0.06 (i.e. $13-15$ per cent), 0.05, and 0.05 dex, respectively. We note that changing the image resolution also affects the $1-\sigma$ scatters. Uncertainties increase to 0.06 and 0.07 dex (i.e. $15-17$ per cent) when only using r-band images with resolutions of 150×150 and 60×60 (MaNGA-like), respectively.

It is interesting to ask, when making a prediction, whether the CNN model simply picks up some general scaling relation between a galaxy's stellar mass and some summary statistics, such as total luminosity, or has it actually used higher order information encoded in the light and kinematic maps? As a simple test, we fit a power-law relation to the stellar mass and luminosity of galaxies in the training sample and use the best-fitting $\log M_* - \log L$ relation to predict stellar masses for the test set luminosities. The relationship is shown at the bottom panel of Fig. 6. The scatter in $\log M_*^{pred}/M_*^{true}$ is 0.16 dex in this case, significantly larger than the scatter of our CNN-based results (see the top panel of Fig. 6). In addition, a decision-tree based regression method, which takes the total luminosity L and velocity dispersion σ_v as input, also results in a scatter of 0.05 dex (details are presented in Section 5). The much smaller uncertainty from our multibranch ResNet CNN model indicates that the network has actually made better predictions using spatial distributions of light and kinematics. This is similar to the findings of the Euclid Collaboration et al. (2023), where they also found that their model predictions of stellar masses improved with the inclusion of image data.

Since 3D stellar mass can not be obtained in observation without dynamical modelling, we also use 2D cylindrical/projected masses as targets. As shown in Table 4, the $1-\sigma$ scatters in the predicted $\log M_*^{2D, pred}/M_*^{2D, true}$ values are similar to cases where targets are 3D spherical stellar mass.

Table 4. Results of CNN-based models, where multibranch ResNet takes r-band images and two velocity maps simultaneously as input (see Section 4 for details).

Section	Input	Training on/output	Prediction	Mean $\pm 1 - \sigma$ (dex)
4.1	r image	M_*	M_*	0.00 ± 0.06
4.1	r image	M_*^{2D}	M_*^{2D}	-0.01 ± 0.07
4.1	g, r image	M_*	M_*	0.00 ± 0.05
4.1	g, r image	M_*^{2D}	M_*^{2D}	0.00 ± 0.06
4.1	2 V maps	M_*	M_*	0.00 ± 0.05
4.1	2 V maps	M_*^{2D}	M_*^{2D}	0.00 ± 0.05
4.1	r image + 2 V maps	M_*	M_*	0.00 ± 0.04
4.1	r image + 2 V maps	M_*^{2D}	M_*^{2D}	0.00 ± 0.05
4.1	r image + 2 V maps	$M_*(\leq 2R_{hsm})$	$M_*(\leq 2R_{hsm})$	0.00 ± 0.04
4.1	r image + 2 V maps	$M_*^{2D}(\leqslant 2R_{hsm})$	$M_*^{2D}(\leqslant 2R_{hsm})$	0.00 ± 0.04
4.2	r image	M_{tot}	M_{tot}	0.00 ± 0.07
4.2	g, r image	M_{tot}	M_{tot}	0.00 ± 0.07
4.2	2 V maps	M_{tot}	M_{tot}	0.00 ± 0.09
4.2	r image + 2 V maps	M_{tot}	M_{tot}	0.00 ± 0.06
4.2	r image + 2 V maps	$M_{tot}(\leq 2R_{hsm})$	$M_{tot}(\leq 2R_{hsm})$	0.00 ± 0.06
4.3	r image + 2 V maps	M_*, M_{tot}	M_*	0.01 ± 0.04
4.3	r image + 2 V maps	M_*, M_{tot}	M_{tot}	-0.01 ± 0.06
4.3	r image + 2 V maps	M_*, M_{tot}	f_{DM}	0.02 ± 0.05
4.3	r image + 2 V maps	M_*, M_{tot}	f_{DM} of ETGs	-0.02 ± 0.04
4.4	r image	M_*/L	M_*/L	0.01 ± 0.10
4.4	2 V maps	M_*/L	M_*/L	0.01 ± 0.11
4.4	r image + 2 V maps	M_*/L	M_*/L	0.01 ± 0.07
4.4	g, r image	M_*/L	M_*/L	0.01 ± 0.05
4.4	g, r image + 2 V maps	M_*/L	M_*/L	0.00 ± 0.04

Notes. The fifth column indicates the mean and standard deviation of the logarithmic ratio between the predicted and the true values for quantities given in the fourth column. All properties are evaluated within a radius of R_{hsm} from the galaxy centre, except for those explicitly specified with parentheses.

Having obtained stellar masses M_* from CNN-based models, we can make further predictions on stellar M_*/Ls by utilizing r-band luminosities L^{true} directly calculated from images (within the same aperture radius) assuming that the luminosity can be well measured. In this case, the uncertainties in $\log[(M_*/L)^{pred}/(M_*/L)^{true}]$ are then dominated by the uncertainties in CNN-based stellar mass predictions. Note that in Section 4.4, we compare between the M_*/L predictions made by CNN models which take M_* as the target (as presented here) and the predictions made by CNN models that directly take M_*/L as the target. We find that the latter has larger uncertainties than the former.

4.2 Training CNN on galaxies with known total mass as target

Galaxies live in dark matter haloes. The total mass M_{tot} is a fundamental property of a galaxy. From a dynamical modelling perspective, unlike stellar mass, where the results are degenerate with that of dark matter, the total mass can be more reliably determined through dynamical or lensing modelling approaches (e.g. Treu 2010; Li et al. 2016; Zhu et al. 2020). This specific CNN model is to mimic the situation where total masses M_{tot} (i.e. the total mass enclosed within a 3D sphere with a radius of R_{hsm}) are known and available in the training sample, together with their r-band images and IFU-like kinematic maps. The predictions on $M_{tot}(\leqslant R_{hsm})$ for the test set are given in Fig. 7. As shown in Table 4, the $1 - \sigma$ uncertainties in $\log M_{tot}^{pred}/M_{tot}^{true}$, as predicted within R_{hsm} and $2R_{hsm}$, are both 0.06 dex (~ 15 per cent). It is important to realize that taking images and velocity maps together works better than if individual input maps are used alone. Specifically, if only 2D photometric information is used, either taking r-band images or taking both r- and g-band images together, the scatter is 0.07 dex (~ 17 per cent). If only

stellar kinematic maps are used, the scatter is larger at 0.09 dex (~ 23 per cent).

We note that singleband or multiband images and velocity maps, taking each kind on their own, contain information about the stellar mass and the total mass. However, it is hard for us to answer which kind of map has actually provided more information. This is because the input image and velocity maps have different spatial resolutions. As recorded in Li et al. (2016), higher resolution maps result in smaller uncertainties in the estimated dynamical masses of galaxies. Here, without carrying out a further resolution test, we cannot make a concrete assessment on this point. However, as we will see in Section 5, a decision tree-based method helps us to address this question to the first order, revealing that, by comparison with other galaxy properties, luminosity plays a dominant role in predicting the stellar and total masses.

It is also interesting that, given the same input, the uncertainty in predicted stellar masses is always smaller than in the predicted total masses. This essentially reflects a tighter correlation between a galaxy's stellar mass and its morphology and kinematics, by comparison with the total mass.

4.3 Training CNN on galaxies with known stellar and total masses as targets

From observations, we can obtain both the stellar mass M_* and the total mass M_{tot} for a galaxy, either through individual estimates as mentioned above, or jointly through multiple dynamical tracers. Alternatively, such information can be acquired by using galaxy formation and evolution models. This specific CNN model is to imitate this situation, where both quantities are available in the training sample. The top panels in Fig. 8 show the performance
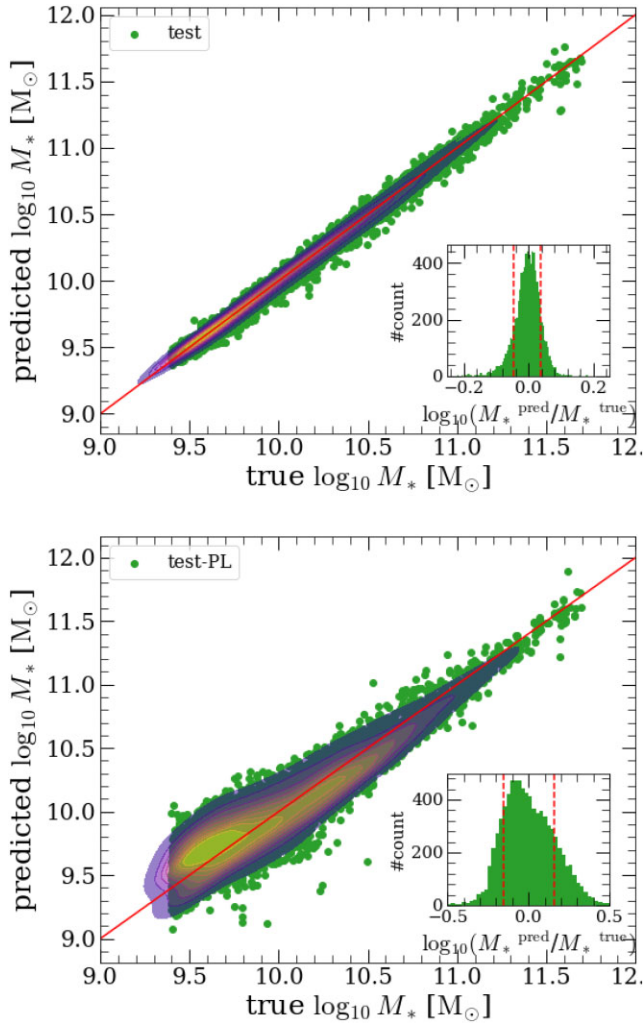
Figure 7. Central M_{tot} prediction of our CNN-based model (r-band image and two velocity maps as input, trained by known central M_{tot}), as a function of their true value. The symbols are the same as Fig. 6.

Figure 6. Top panel: central M_* prediction of our CNN-based model (r-band image and two velocity maps as input, trained by known central M_*), as a function of their true value. Bottom panel: central M_* prediction through power-law fitting, as a function of their true value. In both panels, the red line indicates the prediction equals ground truth, and the green dots are the test set of our samples. The contours indicate the density distribution of the green dots. The histogram in the lower right of each panel shows the distribution of M_* prediction over ground truth ratio, respectively, with the red-dashed lines indicating the $1-\sigma$ range.

of the trained multibranch ResNet in simultaneously predicting 3D spherical $M_*(\leq R_{hsm})$ and $M_{tot}(\leq R_{hsm})$ for the test set. The $1-\sigma$ uncertainties in $\log M_*^{pred}/M_*^{true}$ and $\log M_{tot}^{pred}/M_{tot}^{true}$ are 0.04 and 0.06 dex, respectively. However, some biases of ± 0.01 dex are noticed in this model. It is interesting to note that employing both quantities at the same time as targets in the model does NOT increase prediction accuracy significantly, by comparison with only utilizing one type of target at a time (see Sections 4.1 and 4.2).

When reliably quantified observational systems are used as training data, the overall model performance cannot exceed the accuracy on the training data. However, it is interesting to make comparisons between model predictions and conventional estimates over a generalized statistical population. As the latter suffer from various systematics that vary differently from galaxy to galaxy (as already discussed in Section 1), here, we compare the mass estimation
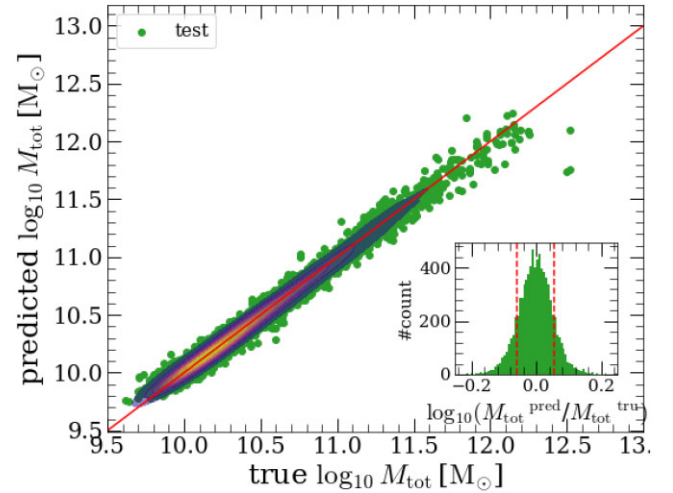
accuracy between CNN-based models and JAM-based models, given that the same kinds of input information are used, i.e. single-band images and IFU-like kinematic maps. Li et al. 2016 evaluated the performance of JAM using galaxies from the Illustris cosmological simulations (Nelson et al. 2015), assuming MaNGA-like image and IFU observational conditions. The typical scatter of JAM-based total masses is about 11–16 per cent, i.e. the scatter on total mass estimates from the two approaches is comparable. While CNN models in general predict stellar masses with higher accuracies ($\sim$10 per cent) than total masses, JAM-based predictions are the opposite. The stellar masses predicted by JAM modelling often suffer from much larger uncertainties with $\sim$30 per cent scatter due to model degeneracies between the stellar component and dark matter. Zhu et al. (2023) adopted six different composite models describing dark matter and baryon distributions and fitted the models to the MaNGA galaxies for which reliable measurements for IFU kinematics are available. The mean standard deviation in predicted stellar masses across the six models over the full galaxy sample is 0.19 dex (~ 50 per cent). We treat this scatter between different models as possible model uncertainties due to unknown degeneracies and hidden systematics. By comparison, our CNN models predict stellar masses with an uncertainty of 0.07 dex ($15 \sim 17$ per cent) for the entire population. Our superiority on CNN stellar mass accuracy might be as a result of the complexity of the neural network which encodes knowledge of the stellar masses for the training sample. We note that, however, a fair comparison between conventional methods and our CNN methods should be made with the same data sample in order to draw more concrete conclusions.

We calculate dark matter fractions $f_{DM}(\leq R_{hsm})$ based on our CNN model predictions for M_* and M_{tot}, and compare the fractions with their true values. Here, we assume that the dark matter fraction is simply given by $f_{DM} = 1 - f_*$, where $f_* \equiv M_*/M_{tot}$. As expected, such an approximation on the dark matter faction would be an overestimate for galaxies with a significant amount of central gas. This is manifested by $\log f_{dm}^{pred}/f_{dm}^{true} \sim 0.02 \pm 0.05$ dex for the overall sample. When we select only early-type galaxies (as defined in Wang et al. 2020) and carry out the same estimate (but without re-training the CNN model), as can be seen in the bottom panels of Fig. 8, the scatter becomes markedly narrower by comparison with the overall sample. In this latter case, however, the dark matter fraction
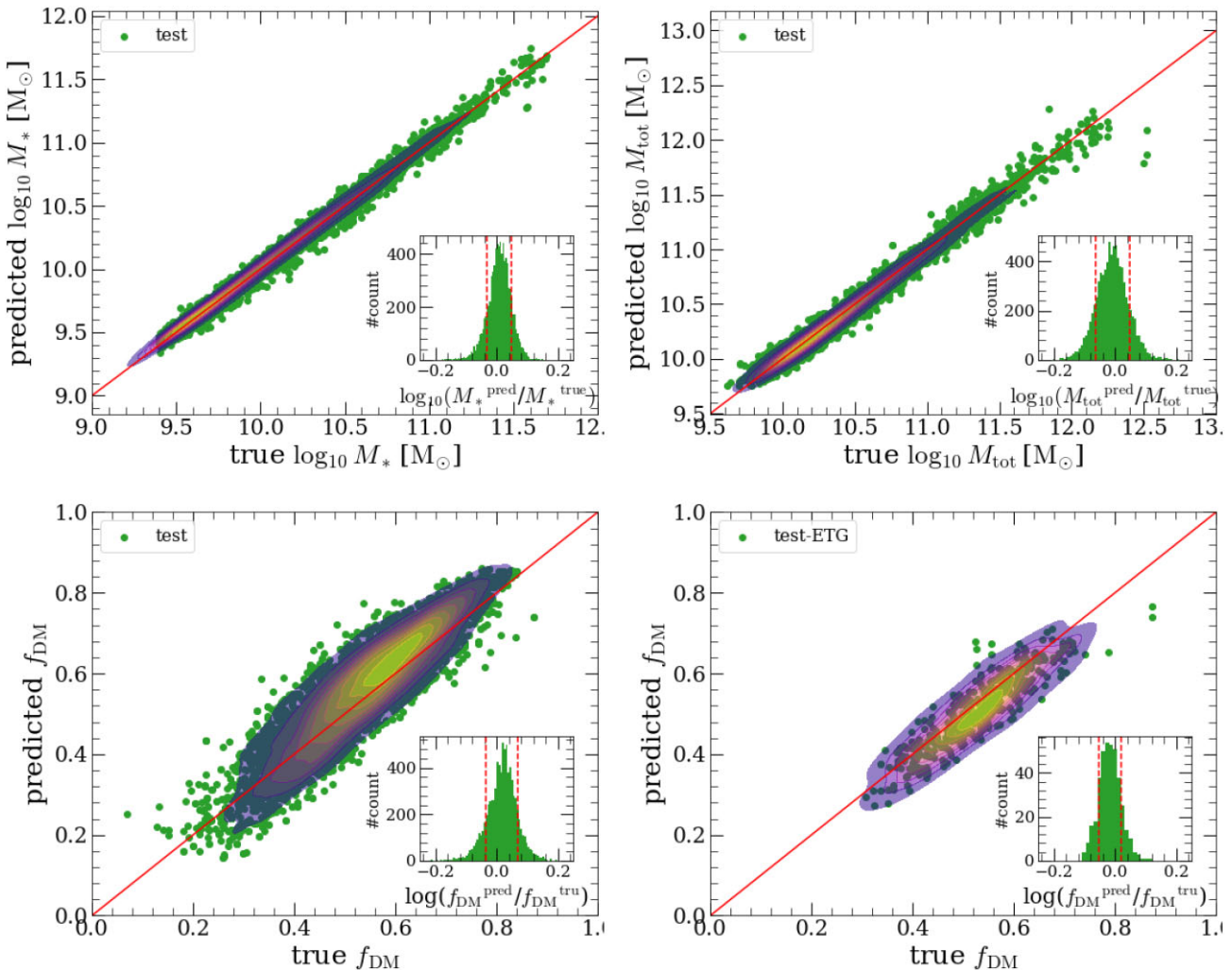
Figure 8. Central M_* (top left panel), M_{tot} (top right), f_{DM} (bottom left panel), and f_{DM} of selected early-type galaxies (bottom right panel) predictions of our CNN-based model (*r*-band image and two velocity maps as input, trained by known central M_* and M_{tot}), as a function of their true values. Here, f_{DM} is the dark matter fraction, calculated by $f_{DM} = 1 - M_*/M_{tot}$. The symbols are the same as Fig. 6. We note that the bottom panels present f_{DM} in linear scales and therefore the distributions appear wider than those for logarithmic masses in the upper panels.

for this sub-sample is underestimated by 0.02 dex. This is because the model-predicted M_{tot} was underestimated by 0.02 dex for these galaxies.

4.4 Training CNN on galaxies with known stellar mass to light ratio as target

A recent study by Dobbels et al. (2019) using an ML approach showed that morphological information from galaxy *g*-band images can noticeably improve the determination of galaxies' M_*/Ls, by comparison with those obtained from only one or two colours. Specifically, they used CNNs to learn key morphological features in the *g*-band images, which were then fed into a gradient-boosting model to predict the stellar M_*/Ls. This two-step algorithm was trained on a sample of more than 80 000 galaxies from the GALEX-SDSS-WISE Legacy Catalogue version 2 (GSWLC; Salim et al. 2016; Salim, Boquien & Lee 2018). The ground-truth M_*/Ls were determined by global spectral energy distribution fitting. The uncertainty in M_*/Ls for the observed galaxy sample was ~ 0.15 dex. Their investigation has already shed light on a feasible way to use ML to find connections between M_*/L and galaxy 2D mass and light distributions.

In this work, we train a CNN model to directly predict M_*/Ls in the case where both *r*-band images and IFU-like kinematic maps are available. In this case, M_*/L is defined as the ratio between the projected 2D stellar mass and *r*-band luminosity within a given radius. Our results are given in Fig. 9. As can be seen, the scatter in $\log[(M_*/L)^{pred}/(M_*/L)^{true}]$ is about 0.07 dex (~ 15 per cent, for within both R_{hsm} and $2R_{hsm}$). By comparison, the scatter is about 0.1 dex if we only take *r*-band images, or only take two velocity maps, as input.

The scatters in $\log[(M_*/L)^{pred}/(M_*/L)^{true}]$ we obtain are generally much larger than those estimated via M_* in the previous sections (see Sections 4.1 and 4.3). This indicates that, under the input conditions used and with the same network complexity, using M_* as the CNN model target yields better predictions on M_*/L than directly using M_*/L as the target.

An additional investigation is to add images in another band such that colour information is also available to the model network. As Bell & de Jong (2001) reveal, galaxy M_*/L strongly correlates with the galaxy's colour. Indeed, as is shown in Section 5.3, our GBDT results also reveal that a galaxy's colour is a key contributing factor
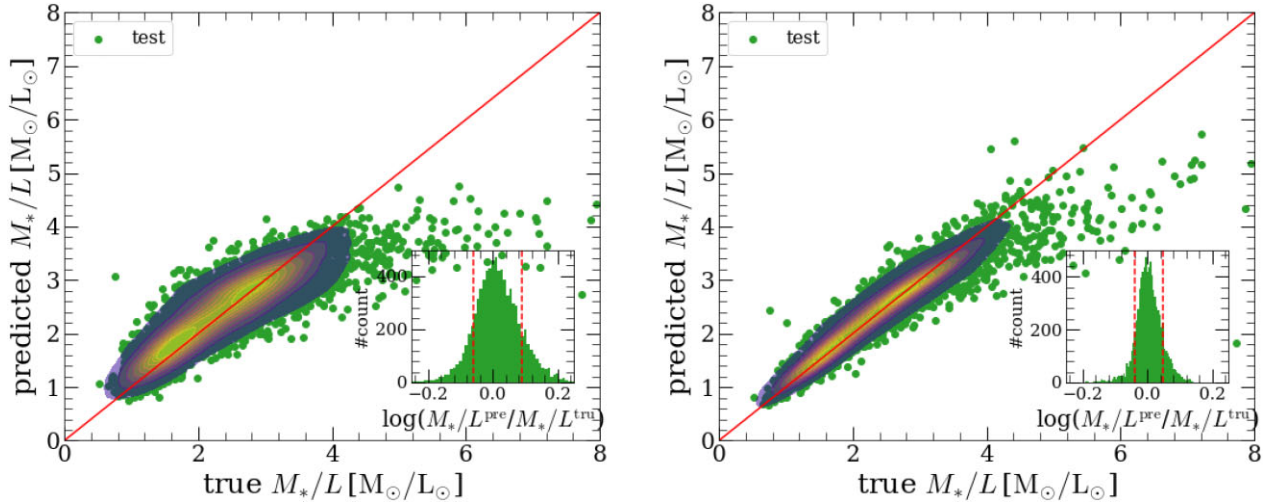
Figure 9. Central M_*/L prediction (r-band image and two velocity maps as input, trained by known central M_*/L; left panel) and central M_*/L prediction (g- and r-band images and two velocity maps as input, trained by known central M_*/L; right panel) of our CNN-based model, as a function their true value, respectively. The symbols are the same as Fig. 6.

in making a correct prediction for M_*/L. We took a galaxy's g-band image as an additional input to our multibranch ResNet. When both g- and r-band images are used, the scatter in $\log[(M_*/L)^{\text{pred}}/(M_*/L)^{\text{true}}]$ is reduced to 0.05 dex. If the colour information is further combined with two velocity maps, the scatter then reduces to 0.04 dex. In both cases, the scatters are significantly smaller than 0.1 dex when only r-band images, or only kinematic maps, were used, or smaller than 0.07 dex when r-band images and kinematic maps combined were used.

5 LIGHT–GBM RESULTS

Having demonstrated in the previous section the abilities of our CNN models to predict galaxy masses, we apply a GBDT method to investigate the driving factors in making successful predictions based on spatially resolved light and kinematic distributions. To do so, we take a gradient boosting machine (Light–GBM), and train a model to compute feature importance. We use 'gain' importance – the total gains of conditions in the model which use a feature.[3] In case any pair of linearly correlated features/targets might bias the feature importance evaluation, we also compute linear correlation coefficients between features and targets (Section 5.1). Fig. 2 summarizes our GBDT training workflow, starting from calculating input summary statistics (listed in Table 1) from images and maps, to model training and final property predictions. Table 5 shows the mean and standard deviation (uncertainty) of the predicted properties from different GBDT models. Detailed results are presented below.

5.1 Linear correlation between features and targets

Before assessing non-linear feature importance using the Light–GBM algorithms when predicting masses (Section 5.2) and ratios/fractions (Section 5.3), we first evaluate possible linear correlations between pairs of features or targets. Similarly to von Marttens et al. (2022), we compute the Pearson correlation coefficient (R). In statistics, computing R (ranging from –1 to 1) between two variables

Table 5. Results of GBDT mode.

Section	Input	Predictions	Mean $\pm 1 - \sigma$ (dex)
5.2	Default	M_*	0.00 ± 0.05
5.2	Default w/o mag	M_*	0.00 ± 0.08
5.2	Default	M_{DM}	0.00 ± 0.13
5.2	Default w/o mag	M_{DM}	0.00 ± 0.15
5.2	Default	M_{tot}	0.00 ± 0.08
5.2	Default w/o mag	M_{tot}	0.00 ± 0.11
5.3	Default	f_*	0.01 ± 0.08
5.3	Default	f_{DM}	0.00 ± 0.06
5.3	Default	M_*/L	0.00 ± 0.06

Notes. Performance of GBDT methods. The fourth column indicates the mean and standard deviation of the logarithmic ratio between the predicted and the true values for quantities given in the third column. All properties are evaluated within a radius of R_{hsm} from the galaxy centre.

in a sample can help assess if there exists a linear correlation between them. The two variables are more likely to be positively linearly correlated if the R is close to 1, or negatively correlated when close to –1. The results of such computations can be seen as a correlation matrix in Fig. 10.

Looking at the input features, we find that pairs of features share reasonably strong linear (anti) correlations ($|R| > 0.75$) including σ_v – luminosity, λ – B/T, λ – f_{cold}, λ – f_{hot}, B/T – f_{hot} and f_{cold} – f_{hot}. The R value between B/T and f_{cold} is also close to 0.75 ($R = 0.74$). In other words, B/T, λ, f_{cold} and f_{hot} are generally intercorrelated with each other. Physically, a fast-rotating galaxy is expected to have a higher stellar spin λ as well as a higher cold-orbit fraction f_{cold}. The galaxy would also have a smaller kinematic B/T and a smaller hot-orbit fraction f_{hot}. The strong correlation between σ_v and luminosity essentially reflects the Tully–Fisher relationship and the Faber–Jackson relationship obeyed by the simulated galaxies (see Lu et al. 2020 for a detailed discussion on the fundamental plane properties of TNG100 galaxies). As can be seen in Fig. 10, other features do not show strong evidence of linear correlations.

When we look at targets and features together, we find luminosity and σ_v are generally strongly correlated (mostly with their $|R| >$ 0.75, though R for σ_v – M_{DM} equals 0.74) with the masses (M_*,

[3] https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.Booster.html#lightgbm.Booster.feature_importance
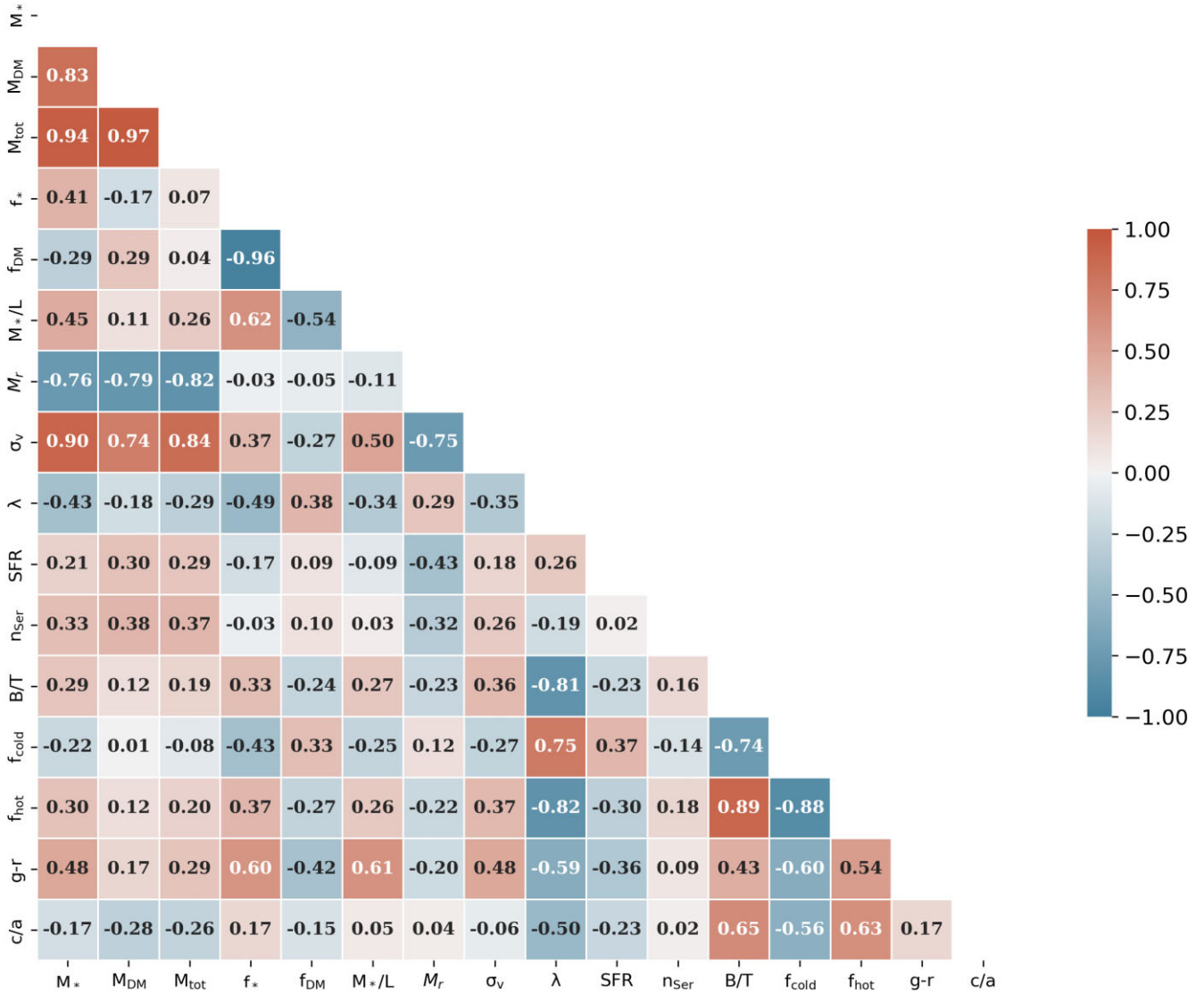
Figure 10. The correlation matrix of features and targets we used for LightGBM-based model training (Section 3.2) and evaluation (Section 5). A variable-pair on the diagram would have a higher possibility of being linearly correlated if the absolute value of its correlation coefficient is closer to 1.

$M_{\rm DM}$, and $M_{\rm tot}$), indicating they play a dominant role in predicting these mass values. However, as stated in von Marttens et al. (2022), R can only indicate a linear relationships between two variables. The evaluation of the non-linear relationships requires other methods, which are discussed in the following section.

5.2 Dependencies of the total, stellar, and dark matter masses

We train our GBDT model to predict separately the 3D spherical M_*, $M_{\rm DM}$, and $M_{\rm tot}$ values within a radius of $R_{\rm hsm}$ from the galaxy centre. The outcome is shown in the left column of Fig. 11. The scatter for the stellar and total masses are 0.05 and 0.08 dex, respectively. For dark matter mass, the uncertainty is about 0.13 dex. It is interesting to note that, among all the properties investigated, galaxy luminosity (magnitude) contributes the most to all three mass predictions. Having trained the GBDT model without using magnitude as input, the results are presented in the right column of Fig. 11. As can be seen, the second most significant feature that contributes to mass predictions is velocity dispersion.

5.3 Dependencies of ratios and fractions: M^*/L, f_*, and $f_{\rm DM}$

We train GBDT models to predict M_*/L, f_*, and $f_{\rm DM}$ within a radius of $R_{\rm hsm}$ from the galaxy centre. Fig. 12 shows the results from the best-trained models on the test set. For the stellar and dark matter mass fractions f_* and $f_{\rm DM}$, the uncertainties are 0.08 and 0.06 dex, and the uncertainty of the M_*/L is 0.06 dex. It is important to note that CNN model predictions are much more accurate than GBDT predictions as produced in this study. This is unsurprising because the former provides information on the spatial distribution of galaxy properties, while the latter only takes low-order summary statistics into account.

In our models, feature importance shows that the contributions are complicated. Unlike mass predictions, there is no dominant feature in the ratio predictions. Generally, velocity dispersion contributes the most, ~ 37 per cent in predicting M_*/L and ~ 28 per cent in $f_{\rm DM}$, with the other features having smaller contributions. For f_* and M_*/L, velocity dispersion and a galaxy's colour are the top two contributing features. The stellar spin parameter λ, which reflects
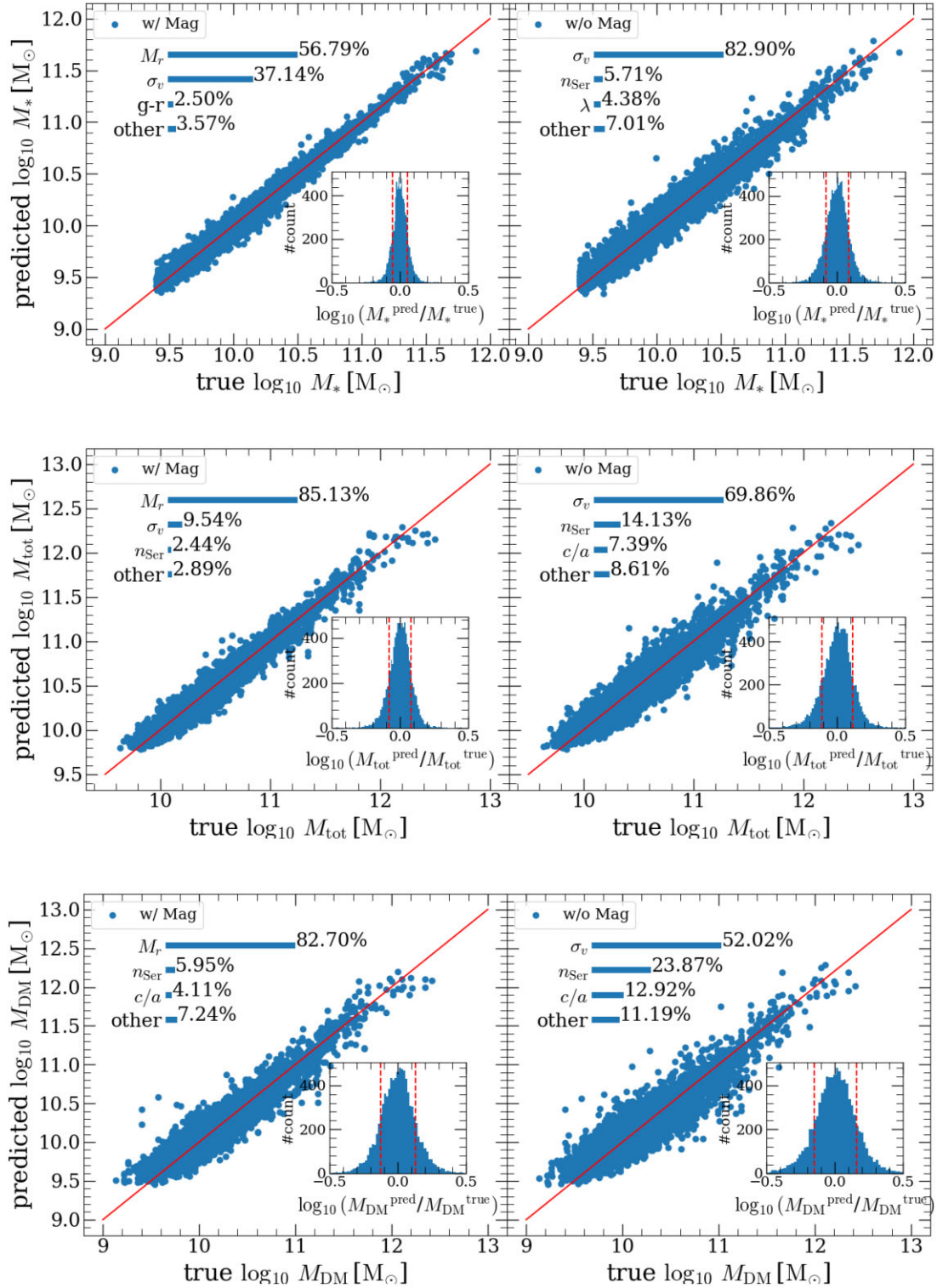
Figure 11. From top to bottom panels: central M_*, $M_{\rm tot}$, and $M_{\rm DM}$ predictions (trained by known M_*, $M_{\rm tot}$, and $M_{\rm DM}$, respectively), as function of their true value. In each row, the left panel has taken r-band magnitude as one of the input features, while the right panel has not. In all six panels, our GBDT models are trained independently. The red line indicates that the prediction equals ground truth, and the blue dots are the samples of our test set. The histogram at the lower right of each panel shows the distribution of mass prediction over the ground truth ratio, respectively, with the red-dashed line indicating 1σ range. The bar graph at the upper left of each panel shows the important features and their contributions to the GBDT predictions.
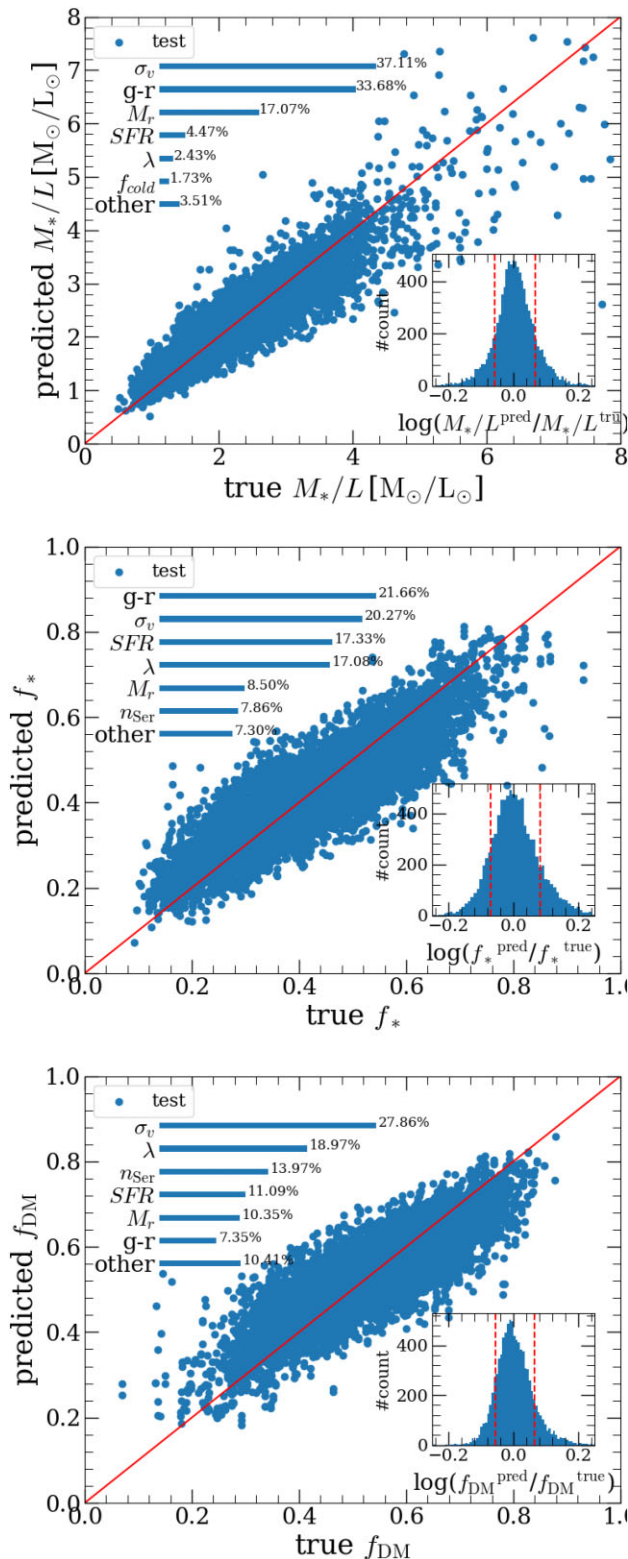
Figure 12. From top to bottom panels: central M_*/L, f_*, and f_{DM} predictions (with r-band magnitude and other summary statistics as input, trained by known M_*/L, f_*, and f_{DM}, respectively), as a function of their true value. The symbols are same as Fig. 11.

the dynamical status of a galaxy, is the second most important feature in predicting f_{DM}.

6 DISCUSSION AND CONCLUSIONS

In this study, we use a general sample of galaxies from the TNG100 simulation to investigate the ability of our CNN-based models to predict the central (i.e. within $1 - 2R_{hsm}$) stellar mass, total mass, stellar M_*/L, and to estimate the dark matter fraction. Specifically, we take galaxy images, spatially resolved mean velocity and velocity dispersion maps as input to our multibranch ResNet CNN models (see Section 2.2 for detailed data input and target generation; the detailed method is given in Section 3.1). In particular, the IFU-like kinematic maps have spatial resolution typical of the MaNGA galaxy sample, and cover a square region of $[-3R_{hsm}, 3R_{hsm}]^2$ around the galaxy centre. The CNN-based models, with the help of the training data set, can in general break the degeneracy between the baryon and dark matter distributions and make reliable mass predictions. In order to understand which (global) features contribute the most to our predictions, we utilize a gradient-boosting machine Light-GBM, which takes global galaxy properties as input, including luminosity, colour, SFR, Sersic index, axis ratio, stellar velocity dispersion, spin parameter, kinetic B/T, and orbital fractions (see Section 2.3 for detailed data input and target generation; the detailed method is given in Section 3.2).

Our main results are listed as follows:

(i) Our multibranch ResNet CNN models can predict (central) stellar and total masses of galaxies with $1 - \sigma$ uncertainties of 0.04 and 0.06 dex, respectively, when taking r-band images and two velocity maps as input. Under such conditions, the prediction for M_*/L has an uncertainty of 0.07 dex. However, when combined with galaxy colour information, for example, taking both g- and r-band images together with kinematic maps as input, the uncertainty decreases to 0.04 dex (for more details, see Table 4 in Section 4).

(ii) Given the default input to the GBDT models, the stellar and total masses of galaxies can be reproduced with uncertainties of 0.05 and 0.08 dex, respectively. The predicted dark matter mass uncertainty is somehow larger at 0.13 dex. The uncertainties on the central stellar (f_*) and dark matter (f_{DM}) fractions are 0.08 and 0.06 dex, respectively; while that for M_*/L is 0.06 dex (for more details see Table 5 in Section 5).

(iii) We find from our GBDT models that galaxy luminosity is the dominating feature (contributed > 50 per cent) in predicting all masses in the central $1 - 2R_{hsm}$ regions (see the left column of Fig. 11). When galaxy luminosity is not considered as an input of our GBDT models, the dominating feature is velocity dispersion. In the case of f_*, f_{DM}, and M_*/L predictions, we do not observe the existence of a dominating feature (see Fig. 12). Velocity dispersion and galaxy's colour are the top two contributing features when predicting f_* and M_*/L. Regarding f_{DM} prediction, we find velocity dispersion contributed the most. At the same time, stellar spin parameter λ should also be valued, given it ranked as the second most important feature on the diagram (the bottom panel of Fig. 12).

We note that a galaxy's luminosity is the dominant feature in predicting all masses. In particular, the correlation between luminosity and stellar mass is even tighter than that with the total mass. This can be seen from both the CNN and GBDT model results such that predictions on the stellar mass always have smaller uncertainties than those on the total mass. The tighter connection between the luminosity and the stellar mass can be understood as a consequence of a straightforward conversion through the stellar M_*/L

ratio, which is governed by stellar evolution physics and typically spans less than one order of magnitude in value. The connection to the total mass can be understood as a consequence of the fact that observed galaxies obey a certain fundamental plane relation, and, through successful simulation calibration, the galaxy sample has thus implicitly reinforced a correlation between the luminosity and the total mass, which additionally is further subtly influenced by the dynamical interplay between baryons and dark matter.

Predictions on fractional masses (f_* and f_{DM}) and on the M_*/Ls show a significant dependence on stellar velocity dispersion (as the leading feature), which reflects the fact that the detailed balance between baryons and dark matter and among different stellar populations, to the first order, have a mass dependence – a consequence of the hierarchical galaxy assembly history. We also found colour significantly contributed to the stellar predictions (f_* and M_*/L) and the stellar spin parameter (which reflects the dynamical nature of a galaxy) to the central f_{DM} prediction, essentially reflect the different physical mechanisms that shape the target properties of the baryonic and dark matter components.

The investigation in this study is in a way reassuring that galaxy images and stellar kinematic maps can provide sufficient information to disentangle the individual dynamical effects from baryons and dark matter. However, one must note that any training sample-based CNN, in principle, cannot reach an accuracy that exceeds that for the training set itself. It is not only hard to obtain an observational galaxy sample with unbiasedly estimated properties, but also impossible to reach accurate predictions for a given sample of observed galaxies by directly applying models that are trained using simulation data, and without taking observational effects and selection rules into account. In addition, data uncertainties and uncertainties in the IMF and galaxy formation physics may cause biases and systematics when making predictions. One potential way to help bridge the gap between simulations and observations may be to test models trained on one simulation with another simulation where different galaxy formation physics have been implemented, though this is yet to be considered in any detail. In this regard, great efforts are still required to find ML models that can unbiasedly estimate matter composition for observed galaxies, especially for those ML methods using image data as input.

ACKNOWLEDGEMENTS

DATA AVAILABILITY

The data underlying this paper will be shared after a reasonable request to the corresponding author.

REFERENCES

Al Rahhal M. M., Bazi Y., Abdullah T., Mekhalfi M. L., AlHichri H., Zuair M., 2018, Remote Sens., 10, 1890
Allgood B., Flores R. A., Primack J. R., Kravtsov A. V., Wechsler R. H., Faltenbacher A., Bullock J. S., 2006, MNRAS, 367, 1781
Angeloudi E., Falcón-Barroso J., Huertas-Company M., Sarmiento R., Pillepich A., Walo-Martín D., Eisert L., 2023, MNRAS, 523, 5408
Arena P., Basile A., Bucolo M., Fortuna L., 2003, Nucl. Instrum. Methods Phys. Res. A, 497, 174
Arjona R., Nesseris S., 2020, Phys. Rev. D, 101, 123525
Armitage T. J., Kay S. T., Barnes D. J., 2019, MNRAS, 484, 1526
Barella V. H., Garcia L. P., de Souto M. C., Lorena A. C., de Carvalho A. C., 2021, Inform. Sci., 553, 83
Bell E. F., de Jong R. S., 2001, ApJ, 550, 212
Bergstra J., Bengio Y., 2012, J. Mach. Learn. Res., 13, 281
Bernardi M., Domínguez Sánchez H., Sheth R. K., Brownstein J. R., Lane R. R., 2023, MNRAS, 518, 4713
Bialopetravičius J., Narbutis D., 2020, A&A, 633, A148
Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition, revised, 2 edn. Princeton Univ. Press, Princeton
Bonjean V., Aghanim N., Salomé P., Beelen A., Douspis M., Soubrié E., 2019, A&A, 622, A137
Bundy K. et al., 2015, ApJ, 798, 7
Cappellari M., 2008, MNRAS, 390, 71
Cappellari M., 2020, MNRAS, 494, 4819
Cappellari M. et al., 2006, MNRAS, 366, 1126
Cappellari M. et al., 2011, MNRAS, 413, 813
Chen Y., Mo H. J., Li C., Wang K., Wang H., Yang X., Zhang Y., Katz N., 2021, MNRAS, 507, 2510
Coronado-Blázquez J., 2022, MNRAS, 515, 1807
Dobbels W., Krier S., Pirson S., Viaene S., De Geyter G., Salim S., Baes M., 2019, A&A, 624, A102
Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497
Emsellem E. et al., 2007, MNRAS, 379, 401
Euclid Collaboration et al., 2023, MNRAS, 520, 3529
Fogarty L. M. R. et al., 2014, MNRAS, 443, 485
Friedman J. H., 2001, Ann. Stat., 29, 1189
Fukushima K., Miyake S., 1982, Pattern Recognit., 15, 455
García-Benito R., González Delgado R. M., Pérez E., Cid Fernandes R., Sánchez S. F., de Amorim A. L., 2019, A&A, 621, A120
Ge J., Mao S., Lu Y., Cappellari M., Long R. J., Yan R., 2021, MNRAS, 507, 2488
Genel S. et al., 2018, MNRAS, 474, 3976
Gomer M. R. et al., 2023, A&A, 679, 13
Han F., Yao J., Zhu H., Wang C., 2020, J. Sensors, 2020, 6707328
He K., Zhang X., Ren S., Sun J., 2016, in Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Las Vegas, NV, USA, p. 770
He S., Li Y., Feng Y., Ho S., Ravanbakhsh S., Chen W., Póczos B., 2019, Proc. Natl. Acad. Sci., 116, 13825
Hernández C. A., González R. E., Padilla N. D., 2023, MNRAS, 524, 4653
Ho M., Rau M. M., Ntampaka M., Farahi A., Trac H., Póczos B., 2019, ApJ, 887, 25
Hu X., Chu L., Pei J., Liu W., Bian J., 2021, Knowl. Inf. Syst., 63, 2585
Hultman J., Pharasyn A., 1999, A&A, 347, 769
Ivezic Z. et al., 2008, Serb. Astron. J., 176, 1
Jeans J. H., 1922, MNRAS, 82, 122
Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y., 2017, Adv. Neur. Inf. Proc. Syst., 30, 3146
Krizhevsky A., Sutskever I., Hinton G. E., 2012, in Proceedings of the 25th International Conference on Neural Information Processing Systems, Vol. 1, NIPS'12. Curran Associates Inc., Red Hook, NY, p. 1097
Lanusse F., Ma Q., Li N., Collett T. E., Li C.-L., Ravanbakhsh S., Mandelbaum R., Póczos B., 2018, MNRAS, 473, 3895
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lecun Y., Bottou L., Bengio Y., Haffner P., 1998, Proc. IEEE, 86, 2278

Li H., Li R., Mao S., Xu D., Long R. J., Emsellem E., 2016, MNRAS, 455, 3680

Li H. et al., 2017, ApJ, 838, 77

Long R. J., Mao S., 2010, MNRAS, 405, 301

Long R. J., Mao S., 2012, MNRAS, 421, 2580

Lovell M. R. et al., 2018, MNRAS, 481, 1950

Lu S. et al., 2020, MNRAS, 492, 5930

Lu S. et al., 2021, MNRAS, 503, 726

Lu S., Xu D., Wang S., Wang Y., Mao S., Xia X., Vogelsberger M., Hernquist L., 2022, MNRAS, 509, 5062

Lu S., Zhu K., Cappellari M., Li R., Mao S., Xu D., 2023, MNRAS, 526, 1022

Lucie-Smith L., Adhikari S., Wechsler R. H., 2022, MNRAS, 515, 2164

Man Z.-Y., Peng Y.-J., Shi J.-J., Kong X., Zhang C.-P., Dou J., Guo K.-X., 2019, ApJ, 881, 74

Marinacci F. et al., 2018, MNRAS, 480, 5113

Monaghan J. J., 1992, ARA&A, 30, 543

Morgan R. et al., 2022, ApJ, 927, 109

Naiman J. P. et al., 2018, MNRAS, 477, 1206

Nelson D. et al., 2015, Astron. Comput., 13, 12

Nelson D. et al., 2018, MNRAS, 475, 624

Nelson D. et al., 2019, Comput. Astrophys. Cosmol., 6, 2

Nishimoto S. et al., 2022, in Ibsen J., Chiozzi G., eds, Proc. SPIE, Vol. 12189, Proceedings of Software and Cyberinfrastructure for Astronomy VII. SPIE, Bellingham, p. 121891Q

Oldham L. J., Auger M. W., 2018a, MNRAS, 474, 4169

Oldham L. J., Auger M. W., 2018b, MNRAS, 476, 133

Parikh T. et al., 2018, MNRAS, 477, 3954

Pillepich A. et al., 2018, MNRAS, 475, 648

Romeo A. B., Agertz O., Renaud F., 2020, MNRAS, 499, 5656

Russakovsky O. et al., 2015, Int. J. Comp. Vis., 115, 211

Sahakyan N., Vardanyan V., Khachatryan M., 2023, MNRAS, 519, 3000

Salim S. et al., 2016, ApJS, 227, 2

Salim S., Boquien M., Lee J. C., 2018, ApJ, 859, 11

Sánchez S. F. et al., 2012, A&A, 538, A8

Schaul T., Zhang S., LeCun Y., 2013, in Dasgupta S., McAllester D., eds, Proceedings of Machine Learning Research Vol. 28, Proceedings of the 30th International Conference on Machine Learning. PMLR, Atlanta, GA, p. 343

Schaye J. et al., 2015, MNRAS, 446, 521

Schwarzschild M., 1979, ApJ, 232, 236

Sérsic J. L., 1963, Boletín de la Asociación Argentina de Astronomía, 6, 41

Shajib A. J., Treu T., Birrer S., Sonnenfeld A., 2021, MNRAS, 503, 2380

Shi J.-H., Qiu B., Luo A. L., He Z.-D., Kong X., Jiang X., 2022, MNRAS, 516, 264

Sonnenfeld A., Leauthaud A., Auger M. W., Gavazzi R., Treu T., More S., Komiyama Y., 2018, MNRAS, 481, 164

Spiniello C., Trager S., Koopmans L. V. E., Conroy C., 2014, MNRAS, 438, 1483

Springel V., 2010, MNRAS, 401, 791

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Springel V. et al., 2018, MNRAS, 475, 676

Su Y. et al., 2020, MNRAS, 498, 5620

Syer D., Tremaine S., 1996, MNRAS, 282, 223

Tabor E., Loeb A., 2021, preprint (arXiv:2105.08081)

Tang H., Scaife A. M. M., Wong O. I., Shabala S. S., 2022, MNRAS, 510, 4504

Tortora C., Napolitano N. R., Romanowsky A. J., Jetzer P., Cardone V. F., Capaccioli M., 2011, MNRAS, 418, 1557

Treu T., 2010, ARA&A, 48, 87

Turner O. J. et al., 2017, MNRAS, 471, 1280

van Dokkum P. G., Conroy C., 2010, Nature, 468, 940

van Dokkum P., Conroy C., Villaume A., Brodie J., Romanowsky A. J., 2017, ApJ, 841, 68

von Marttens R. et al., 2022, MNRAS, 516, 3924

Wang Y. et al., 2019, MNRAS, 490, 5722

Wang Y. et al., 2020, MNRAS, 491, 5188

Wu S., Napolitano N. R., Tortora C., von Marttens R., Casarini L., Li R., Lin W., 2023, preprint (arXiv:2310.02816)

Xu D., Springel V., Sluse D., Schneider P., Sonnenfeld A., Nelson D., Vogelsberger M., Hernquist L., 2017, MNRAS, 469, 1824

Xu D. et al., 2019, MNRAS, 489, 842

Zhou S. et al., 2019, MNRAS, 485, 5256

Zhu L. et al., 2018a, Nat. Astron., 2, 233

Zhu L. et al., 2018b, MNRAS, 473, 3000

Zhu L. et al., 2020, MNRAS, 496, 1579

Zhu K., Lu S., Cappellari M., Li R., Mao S., Gao L., 2023, MNRAS, 522, 6326

This paper has been typeset from a TEX/LATEX file prepared by the author.