Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **520,** 3529–3548 (2023) https://doi.org/10.1093/mnras/stac3810
Advance Access publication 2022 January 9

Euclid preparation – XXIII. Derivation of galaxy physical properties with deep machine learning using mock fluxes and *H*-band images

Euclid Collaboration: L. Bisigello [1,2,3★] C. J. Conselice,[4] M. Baes [5] M. Bolzonella [2] M. Brescia [6] S. Cavuoti [6,7,8] O. Cucciati [2] A. Humphrey,[9] L. K. Hunt [10] C. Maraston [11] L. Pozzetti [12] C. Tortora [7] S. E. van Mierlo [13] N. Aghanim,[14] N. Auricchio [2] M. Baldi [2,15,16] R. Bender [17,18] C. Bodendorf,[18] D. Bonino,[19] E. Branchini [20,21] J. Brinchmann [9] S. Camera [19,22,23] V. Capobianco [19] C. Carbone,[24] J. Carretero [25,26] F. J. Castander [27,28] M. Castellano [29] A. Cimatti,[10,30] G. Congedo [31] L. Conversi [32,33] Y. Copin [34] L. Corcione [19] F. Courbin [35] M. Cropper [36] A. Da Silva [37,38] H. Degaudenzi [39] M. Douspis,[14] F. Dubath,[39] C. A. J. Duncan,[4,40] X. Dupac,[32] S. Dusini [41] S. Farrens [42] S. Ferriol,[34] M. Frailis [43] E. Franceschi [2] P. Franzetti,[24] M. Fumana [24] B. Garilli [24] W. Gillard [44] B. Gillis [31] C. Giocoli [12,45] A. Grazian [46] F. Grupp,[17,18] L. Guzzo,[47,48,49] S. V. H. Haugan [50] W. Holmes,[51] F. Hormuth,[52] A. Hornstrup [53] K. Jahnke [54] M. Kümmel,[17] S. Kermiche [44] A. Kiessling [51] M. Kilbinger [42] R. Kohley,[32] M. Kunz [55] H. Kurki-Suonio [56] S. Ligori [19] P. B. Lilje [50] I. Lloro,[57] E. Maiorano [2] O. Mansutti [43] O. Marggraf [58] K. Markovic [51] F. Marulli [2,16,59] R. Massey [60] S. Maurogordato,[61] E. Medinaceli [2] M. Meneghetti,[2,16] E. Merlin [29] G. Meylan,[62] M. Moresco [2,59] L. Moscardini [2,16,59] E. Munari [43] S. M. Niemi,[63] C. Padilla [25] S. Paltani,[39] F. Pasian,[43] K. Pedersen,[64] V. Pettorino,[42] G. Polenta [65] M. Poncet,[66] L. Popa,[67] F. Raison,[18] A. Renzi [1,41] J. Rhodes,[51] G. Riccio,[7] H.-W. Rix [54] E. Romelli [43] M. Roncarelli,[2,59] C. Rosset,[68] E. Rossetti,[59] R. Saglia [17,18] D. Sapone,[69] B. Sartoris,[17,43] P. Schneider,[58] M. Scodeggio,[24] A. Secroun [44] G. Seidel [54] C. Sirignano [1,41] G. Sirri [16] L. Stanco,[41] P. Tallada-Crespí,[26,70] D. Tavagnacco [43] A. N. Taylor,[31] I. Tereno,[37,71] R. Toledo-Moreo [72] F. Torradeflot [26,70] I. Tutusaus [55] E. A. Valentijn,[13] L. Valenziano [2,16] T. Vassallo [43] Y. Wang [73] A. Zacchei [43] G. Zamorani [2] J. Zoubian,[44] S. Andreon [48] S. Bardelli [2] A. Boucaud,[68] C. Colodro-Conde,[74] D. Di Ferdinando,[16] J. Graciá-Carpio,[18] V. Lindholm [56] D. Maino,[24,47,49] S. Mei [68] V. Scottez,[75] F. Sureau,[76] M. Tenti,[16] E. Zucca [2] A. S. Borlaff [77] M. Ballardini [2,59,78] A. Biviano [43,79] E. Bozzo [39] C. Burigana [78,80,81] R. Cabanac [82] A. Cappi,[2,61] C. S. Carvalho,[71] S. Casas [83] G. Castignani,[2,59] A. Cooray,[84] J. Coupon,[39] H. M. Courtois [85] J. Cuby,[86] S. Davini [87] G. De Lucia [43] G. Desprez,[39] H. Dole,[14] J. A. Escartin,[18] S. Escoffier [44] M. Farina,[88] S. Fotopoulou,[89] K. Ganga [68] J. Garcia-Bellido [90] K. George [17] F. Giacomini [16] G. Gozaliasl [91] H. Hildebrandt [92] I. Hook [93] M. Huertas-Company [94,95] V. Kansal,[76] E. Keihanen,[91] C. C. Kirkpatrick,[56] A. Loureiro [31,96,97] J. F. Macías-Pérez [98] M. Maglioccheti [88] G. Mainetti,[99] S. Marcin,[100] M. Martinelli [29] N. Martinet [86] R. B. Metcalf,[2,59] P. Monaco [43,79,101,102] G. Morgante,[2] S. Nadathur [11] A. A. Nucita,[103,104] L. Patrizii,[16] A. Peel [62] D. Potter [105] A. Pourtsidou [31,106] M. Pöntinen [91] P. Reimberg,[107] A. G. Sánchez [18] Z. Sakr [82,108,109] M. Schirmer [54] E. Sefusatti [43,79,102] M. Sereno [2,45] J. Stadel [105] R. Teyssier,[110] C. Valieri,[16] J. Valiviita [111] and M. Viel [43,79,102,112]

Affiliations are listed at the end of the paper

Accepted 2022 December 22. Received 2022 December 22; in original form 2022 June 28

★ E-mail: laura.bisigello@inaf.it

ABSTRACT

Next-generation telescopes, like *Euclid*, *Rubin*/LSST, and *Roman*, will open new windows on the Universe, allowing us to infer physical properties for tens of millions of galaxies. Machine-learning methods are increasingly becoming the most efficient tools to handle this enormous amount of data, because they are often faster and more accurate than traditional methods. We investigate how well redshifts, stellar masses, and star-formation rates (SFRs) can be measured with deep-learning algorithms for observed galaxies within data mimicking the *Euclid* and *Rubin*/LSST surveys. We find that deep-learning neural networks and convolutional neural networks (CNNs), which are dependent on the parameter space of the training sample, perform well in measuring the properties of these galaxies and have a better accuracy than methods based on spectral energy distribution fitting. CNNs allow the processing of multiband magnitudes together with H_E-band images. We find that the estimates of stellar masses improve with the use of an image, but those of redshift and SFR do not. Our best results are deriving (i) the redshift within a normalized error of <0.15 for 99.9 per cent of the galaxies with signal-to-noise ratio >3 in the H_E band; (ii) the stellar mass within a factor of two ($\sim$0.3 dex) for 99.5 per cent of the considered galaxies; and (iii) the SFR within a factor of two ($\sim$0.3 dex) for $\sim$70 per cent of the sample. We discuss the implications of our work for application to surveys as well as how measurements of these galaxy parameters can be improved with deep learning.

Key words: galaxies: evolution – galaxies: general – galaxies: photometry – galaxies: star formation.

1 INTRODUCTION

Understanding the physical processes driving galaxy evolution is one of the most outstanding issues in astronomy today. It is clear that galaxies are assembling through star formation and mergers through cosmic time (e.g. Conselice et al. 2014; Madau & Dickinson 2014), and morphologically evolve from irregular/peculiar galaxies at $z > 1$ to more normal regular systems at lower redshifts (e.g. Mortlock et al. 2013). Whilst these basic features are now well understood in a generalized way within an evolving galaxy population, the exact details of this process are still unknown but there will be significant improvements in their understanding in the next decade with large telescope projects such as the *Euclid Space Telescope* (Laureijs et al. 2011), the *Vera C. Rubin Observatory* (*Rubin*/LSST; Ivezic et al. 2008), and the *Nancy Roman Space Telescope* (Akeson et al. 2019).

One of the most important ways for carrying out the analysis of galaxy evolution is through measuring properties of galaxies, like stellar mass and star-formation rates (SFRs), at different distances (or redshifts). These properties of galaxies can be difficult to measure accurately even through standard methodologies (e.g. Bisigello et al. 2016, 2017; Ciesla, Elbaz & Fensch 2017), which are generally based on fitting the galaxy spectral energy distribution (SED) with theoretical or empirical models. In the epoch of large data projects such as *Euclid* and LSST the use of these standard techniques will require huge computing power in order to measure these properties for the hundreds of millions of galaxies that will be observed within these data. Therefore, these large surveys will require the use of methods that go beyond the traditional ones for an efficient and accurate data analysis (i.e. statistical methods, including those based on machine learning). This shows the necessity of improving machine-learning algorithms in the near future.

However, except for some pioneering works (e.g. Tagliaferri et al. 2003; Hoyle 2016; Stensbo-Smidt et al. 2016; D'Isanto & Polsterer 2018; Delli Veneri et al. 2019; Surana et al. 2020; Mucesh et al. 2021; Razim et al. 2021), the measurements of redshift, stellar mass, and SFR in an automatic way with machine learning is still largely under development. Other measurements, such as the galaxy morphology and structure, e.g. CAS parameters (i.e. concentration, asymmetry, and clumpiness; Conselice 2003), have been more extensively tested and can indeed be retrieved through deep-learning methods (e.g. Cheng et al. 2020; Tohill et al. 2021).

In a recent work by the Euclid Collaboration: Desprez et al. (2020), a careful comparison was performed between the photometric redshift obtained with different standard and machine-learning techniques, showing the strengths and weaknesses of both methods. In particular, the photometric redshift measurement obtained with machine learning is challenging where the colour space regions of the training sample are not well covered; while, traditional methods have issues at very low z (i.e. $z < 0.5$), at least when considering optical and near-infrared (IR) filters, perhaps because of a lack of a valid set of templates or priors. Thus, various methods should be investigated to determine the optimal ways to measure these properties.

In general, galaxy images can contain more information than integrated magnitudes, as the morphology, size, and the presence of companions hold information about their nature. SED fitting methods and machine-learning networks used to derive physical properties are mainly based on integrated quantities. However, recent works (Hoyle 2016; Pasquet et al. 2019) have shown the power of using images to derive photometric redshifts and the improvement caused by adding morphological information when estimating stellar masses (Dobbels et al. 2019). A similar analysis on the direct use of images to estimate stellar masses and SFR is however still missing.

In this paper, we discuss if we can retrieve the most basic galaxy properties from deep-learning neural networks (DLNNs)[1] and from convolutional neural networks (CNNs), which indeed can make use of galaxy images. We make use of the Cosmos Evolution Survey (COSMOS; Scoville et al. 2007) field and the catalogue from Laigle et al. (2016) as well as imaging from the COSMOS-Drift And SHift (COSMOS-DASH; Mowla et al. 2019) survey with the *Hubble Space Telescope* (*HST*) Wide Field Camera 3 (WFC3). Thanks to these data, we verify if we are able to retrieve with *Euclid* (I_E previously called VIS, Y_E, J_E, H_E; Euclid Collaboration: Schirmer et al. 2022c) and *Rubin*/LSST filters (u, g, r, i, and z) the same SFRs and stellar masses derived through SED fitting, but based on a larger number of filters [i.e. 30 from ultraviolet (UV) to near-IR]. This would mean that the machine-learning networks are able to correctly interpolate between filters to retrieve the same output quantities using less input information. In this work, we perform a first step by deriving the point estimates of redshift, stellar mass, and SFR; while, we leave to a future work the complex analysis of the uncertainties and the probability distribution functions associate to each of them.

[1] A DLNN is similar to a multilayer perceptron; however, it has more than one fully connected layer, as it happens in our case.

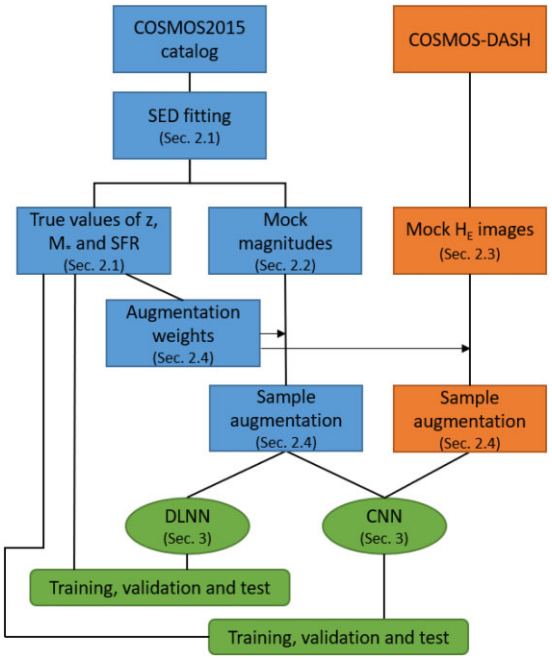

Figure 1. Workflow illustrating the different steps considered in this work. We highlight in blue the parts derived from the COSMOS2015 catalogue and in orange the part derived from the COSMOS-DASH survey. The green part indicates the two machine-learning networks considered (see Section 3).

This paper is organized as follows. In Section 2, we introduce the mock Euclid catalogue and the simulated H_E-band images. In Section 3, we describe the machine-learning algorithms considered; while, Section 4 contains the redshift, stellar mass, and SFR estimates. The main findings are summarized in Section 5. In this paper, we consider a Lambda cold dark matter cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_m = 0.27$, and $\Omega_\Lambda = 0.73$, a Chabrier (2003) initial mass function (IMF), and all magnitudes are in the AB system (Oke & Gunn 1983).

2 MOCK OBSERVATIONS

In the next sections, we report the procedure considered to derive mock *Euclid* magnitudes and Euclidized H_E-band images starting from observed galaxies. These are the inputs required by the two neural networks analysed in this work, which are described in details in Section 3. In Fig. 1, we report the full workflow to guide the reader.

2.1 Mock catalogue and SED fitting procedure

In this work, we made use of an updated version of the mock catalogue of the Euclid Wide Survey presented by Bisigello et al. (2020). This mock catalogue was created starting from multiwavelength observations of real galaxies in the 2 deg^2 survey of the COSMOS field. These multiwavelength observations consist of 30 filters, ranging from the UV to near-IR wavelengths, and they are part of the public COSMOS2015 catalogue (Laigle et al. 2016). We removed from the original COSMOS2015 catalogue stars and X-ray sources, the latter corresponding to <1 per cent of the galaxy sample.

Each observed galaxy was fitted comparing theoretical templates with the fluxes available in the 30 COSMOS15 filters, performing a χ^2 fitting procedure using the *LePhare* code (Arnouts et al. 1999; Ilbert et al. 2006). For these fits we considered a broad set of SED templates from Bruzual & Charlot (2003) with exponentially declining star-formation histories with e-folding time-scale τ between 0.1 and 10 Gyr, Solar and subsolar metallicity ($Z_\odot$, 0.04 $Z_\odot$), ages from 0.1 to 12 Gyr, 12 values of colour excess from $E(B-V) = 0$ to 1, and the Calzetti et al. (2000) reddening law. The redshift was fixed to the value reported in the COSMOS2015 catalogue.

This fitting procedure allowed us to derive, for each observed galaxy, the best theoretical template and the galaxy physical properties associated to it. In particular, for this work we are interested in the redshift, the stellar mass, and the SFR, while the best-fitting templates were used to derive mock *Euclid* magnitudes (see next section). We considered the previously mentioned associated physical properties to be the ground truth.

We highlight that the considered physical properties have an associate uncertainty derived from the SED fitting procedure. In particular, as mentioned by Laigle et al. (2016), the normalized median absolute deviation (NMAD) of the redshift, derived comparing the photometric redshift with available spectroscopic ones, varies from 0.007 to 0.057 moving from bright galaxies ($16 < i < 21$) to faint galaxies ($25 < i < 26$). At the same time, the median error on the stellar mass and SFR, which are derived from the output probability distribution of each object, is 0.07 and 0.16 dex, respectively. These errors are stated here, but are not considered when showing, in the next sections, the results for the different networks, as they depend on the number of filters and the method used to derived the physical properties used as the ground truth.

2.2 Mock magnitudes

The original catalogue from Bisigello et al. (2020) was derived to mimic the Euclid Wide Survey (Euclid Collaboration: Scaramella et al. 2022b) and consists of five filters, i.e. I_E, Y_E, J_E, H_E, and the Canada–France Imaging Survey (CFSI) u band. To these filters we add complementary magnitudes in the Sloan Digital Sky Survey (SDSS; Gunn et al. 1998), the g, r, i, and z filters. Observations in similar filters, such as the ones that will be used by *Rubin*/LSST, will be available to complement *Euclid* observations (Euclid Collaboration: Scaramella et al. 2022b).

The inclusion of filters bluewards of the *Euclid* bands, such as the previously mentioned u, g, r, i, and z filters, may improve the reconstruction of the overall SED template by broadening the wavelength coverage. In particular, such filters are expected to improve the derivation of the SFR, by tracing UV wavelengths (Pforr, Maraston & Tonini 2012, 2013), as well as the Lyman break (i.e. 912 Å), which is one of the most prominent feature in a galaxy spectra.

We included photometric errors by scattering each magnitude around its true value, derived from its best-fitting SED, and considering the respective survey noise. We did not include any other source of error. In Table 1 we report the observational depths considered, as expected for the Euclid Wide Survey, to perturb the original COSMOS2015 photometry and to perform the SED fitting procedure using *Euclid* and *Rubin*/LSST filters. These depths in the *Euclid* bands correspond to the values presented in the *Euclid* definition study report (Laureijs et al. 2011) and are different from the more recent values presented in Euclid Collaboration: Scaramella et al. (2022b).[2] However, as photometric errors are not included in the machine-learning networks, we do not expect these differences to

[2]Their median 10σ values are 25.45, 23.55, 23.74, and 23.65 for I_E, Y_E, J_E, and H_E, respectively.

Table 1. Observational depth (point source, 10σ) in AB magnitude, central wavelength, and full width at half-maximum (FWHM) of the filters considered in our Euclid Wide mock catalogue.

Band	10σ Depth	λ_{cen} (Å)	FWHM (Å)
I_E	24.50	7150	3640
NISP/Y_E	23.24	10 850[a]	2660[a]
NISP/J_E	23.24	13 750[a]	4040[a]
NISP/H_E	23.24	17 725[a]	5020[a]
CFSI/u	24.20	3715	510
SDSS/g	24.50	4700	1263
SDSS/r	23.90	6174	1149
SDSS/i	23.60	7534	1239
SDSS/z	23.40	8782	994

Note. [a]The central wavelengths and FWHMs of the NISP filters are slightly (<0.6 per cent and <1.3 per cent, respectively) different from the more recent values reported in Euclid Collaboration: Scaramella et al. (2022c).

impact the results presented in this paper. Photometric errors are not included as inputs because the feature analysis on a similar machine-learning algorithm have shown they provide little or no information compared to just using magnitudes (Euclid Collaboration: Humphrey et al. 2022a) . At this stage, we did not apply any magnitude cut, as this will be performed later on the images; however, we assigned a value of -1 to all magnitudes below $S/N < 3$.

2.3 H_E-band Euclidized images

In this work, we consider machine-learning networks that have as inputs not only magnitudes, but also images. In particular, we make use of simulated images in the H_E filter instead of images in the I_E filter. On one hand, the I_E filter is more sensitive, as it covers a wide 530-920 nm wavelength range (equivalent to r, i, and z together), and it has three times higher angular resolution, together implying more complex modelling. On the other hand, observations in filters similar to the *Euclid* H_E band are already available from *HST*.

We derived our simulated H_E-band images from the *HST*-WFC3 Imaging Survey in the COSMOS field (COSMOS-DASH; Mowla et al. 2019), which covers a large fraction of the COSMOS field. We consider the *HST*/*F*160*W* images because they correspond to a filter close in wavelength to the H_E filter. We did not apply any k-correction to convert from the *HST*/*F*160*W* filter to the H_E band. This would imply the use of a SED model and would change the flux values of each pixel but not the morphological features, which are the ones relevant for the CNN analysis.

We created H_E-band images starting from *HST*/*F*160*W* thumbnails of 51 × 51 pixels centred around each galaxy. We derive the signal-to-noise ratio (S/N) of each image by retrieving the background flux and the noise from the median and standard deviation of the fluxes on an area of 51 × 51 pixels, with the central 18 × 18 pixels masked to remove the source. For the noise, we derive the value present in the image by calculating the standard deviation after applying a 3σ-clipping procedure. We then add an additional noise, applying a scatter from a Gaussian distribution, in order to reproduce the expected *Euclid* noise. The signal on the source was then roughly derived from the central 6 × 6 pixels, which correspond to a square of ∼0.7 arcsec × 0.7 arcsec.[3] The S/N was then derived by subtracting the background from the signal and dividing the result by the retrieved noise. We restricted our analysis to galaxies with an S/N > 3, to avoid training the networks with images dominated by noise.

As a second step, we apply the *Euclid* NISP H_E-band point spread function (PSF),[4] photometric noise, and spatial resolution. Both the *Euclid* (∼0.7 arcsec) and *HST* PSFs were approximated by two-dimensional (2D) Gaussian functions. In reality, both PSFs are highly non-Gaussian in the wings, we do not expect this to impact extremely our results, as the central small spatial resolution should dominate the training; however, it is recommended to train the network on real images in the future. In this way, we start from an *HST*/*F*160*W* image of 51 × 51 pixels centred on each observed galaxy and we obtain an H_E-band-simulated image of 25 × 25 pixels. These sizes generally include the entire galaxy in each image and allow for a sample augmentation through rotation without image loss. In Fig. 2, we give an example of the transformation from an observed *HST*/*F*160*W* image to a H_E-band-simulated image.

In the last step, we matched the COSMOS-DASH catalogue with the COSMOS2015 catalogue, considering a matching radius of 1 arcsec. In this way, we linked each H_E-band simulated image to the set of mock magnitudes described in the previous section. We then separate the catalogue into two subsamples with different S/N cuts, i.e. S/N > 3 and S/N > 10, with the S/N derived, as mentioned before, on the H_E-band-simulated images. These subsamples correspond to 27 340 and 9799 COSMOS-DASH galaxies with an S/N > 3 and 10, respectively. The redshift, stellar mass, and SFR distributions of both samples are reported in Fig. 3.

2.4 Sample augmentation

Sample augmentation is necessary to increase the number of objects and improve the training of the different algorithms. This was done for the magnitudes in the training and validation samples (see Section 3) randomly extracting their values from a Gaussian distribution centred on the true flux values and with a dispersion equal to the photometric noise, as expected for the Euclid Wide Survey. We then convert fluxes to magnitudes. On the other hand, to increase the number of images available, we rotated each of them by 10° for a maximum of 35 times, this change is sufficient to make the network recognize each image as a new one. This method has been often applied in the literature (e.g. Dieleman, Willett & Dambre 2015; Huertas-Company et al. 2015) and it has been demonstrated to improve machine-learning classifications (Cheng et al. 2020).

For the redshift derivation, we increased the number of sources in the catalogue by a factor of 10; while, for the stellar mass and SFR, whose measurements are more challenging, we increased the number of objects up to a factor of 35 in order to obtain a flat distribution (Fig. 3). This is performed to avoid biases on the training (i.e. most present galaxy having the best estimation) and to obtain an estimation that is similar over a range in mass or SFR as large as possible. However, the number of galaxies with S/N > 3 (10) and $\log_{10}[SFR/(M_\odot\,yr^{-1})] < -2.5$ (−3) or $\log_{10}(M_*/M_\odot) < 8$ (8.5) are very low and the resulting SFR and stellar mass distributions are flat only above these values. In addition, we applied a rough conversion of the 3σ I_E photometric depth (i.e. $I_E = 25.81$, see Table 1) to SFR by using the relation by Kennicutt (1998), which links the UV luminosity to the unobscured component of the SFR, and converting it to Chabrier (2003) IMF. This conversion is not possible at $z < 1$, as the I_E does not trace the UV light; however, it is useful

[3]This procedure, applied to the original *HST* images, is expected to underestimate the flux of local extended galaxies, for which a specific analysis on the estimation of physical properties will be considered in a future work.

[4]In order to have a final image consistent with the *Euclid* PSF, we apply a PSF with a standard deviation $\sigma^2 = \sigma^2_{Euclid} - \sigma^2_{HST}$.

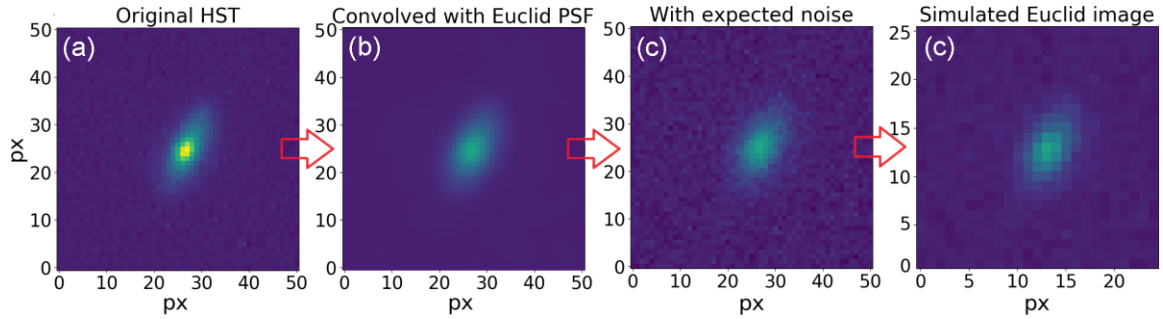

Figure 2. Example of the transformation from an *HST*/*F*160*W* image (a) to a *Euclid* H_E-band-simulated image (d) of a galaxy at $z = 0.2$. In the first step, we include the *Euclid* PSF (b), we then include the expected photometric noise (c), and finally we apply the *Euclid* angular resolution (d). The colour scale is linear and it is the same for all panels.

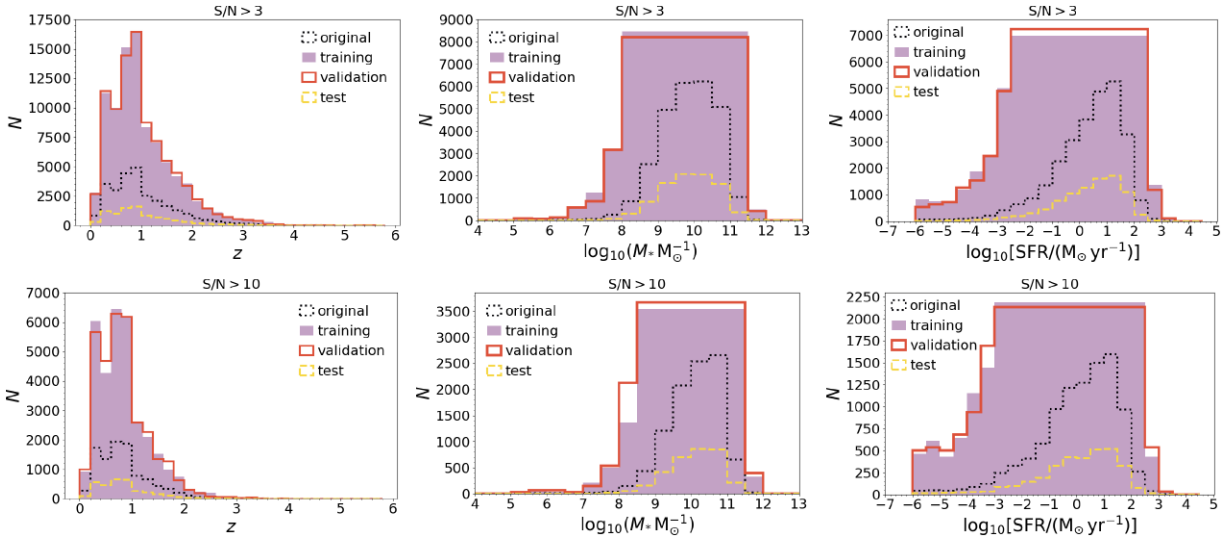

Figure 3. Distribution of the galaxies in the COSMOS field with imaging S/N > 3 (top panel) and S/N > 10 (bottom panel) in redshift (left-hand panel), stellar mass (centre panel), and SFR (right-hand panel). The distributions are for the training sample (filled purple) and the validation sample (solid red) after applying sample augmentation, and for the test sample (dashed yellow) and the original sample (dotted black) without augmentation. Sample augmentation for SFR and stellar mass is performed to obtain a distribution which is as flat as possible, to avoid biases when training.

to give a rough estimate of the limits imposed by the photometric noise to the SFR estimation. Indeed, we expect the photometric noise to limit the SFR estimates to $\log_{10}[\mathrm{SFR}/(\mathrm{M}_\odot\,\mathrm{yr}^{-1})] > -0.9$ (0.3) at $z = 1$ (3). In the following analysis we take into account these effects, showing the results for the full sample and for the subsample with a completely flat distribution in stellar mass. For the SFR, we instead show the results for the full sample and for the sample with $\log_{10}[\mathrm{SFR}/(\mathrm{M}_\odot\,\mathrm{yr}^{-1})] > 0$.

After augmentation, the training samples used for the redshift measurements have 91 130 galaxies with S/N > 3 and 32 660 with S/N > 10. The augmented samples for the stellar mass (SFR) have instead 63 295 (82 899) galaxies with S/N > 3 and 24 296 (29 065) objects with S/N > 10. Each S/N cut refers to the H_E-band images, as mentioned in the previous section. The sizes of the samples correspond to the maximum size possible given the input data set and the augmentation procedure explained.

3 MACHINE-LEARNING ALGORITHMS

In this work, we considered two different machine-learning algorithms, a DLNN, which has as input exclusively tabular data, and a CNN, which includes also images. We investigate these two different machine-learning algorithms to determine the performance of both for measuring different physical parameters of galaxies and how they compare with each other. In the future, this work may be extended considering a 1D CNN instead of a DLNN, which may better capture features in the SED, and a more complex CNN with multiple input images.

3.1 Deep-Learning Neural Network

In the DLNN, we used two different sets of inputs, namely (i) the four *Euclid* magnitudes and (ii) the four *Euclid* magnitudes complemented with the magnitudes in the *u* CHFT and *g*, *r*, *i*, and *z* SDSS ground-based filters. These ground-based filters will be available through ancillary surveys with different facilities (Euclid Collaboration: Scaramella et al. 2022b). In both cases, we considered the sample with H_E-band images with S/N > 3 and the subsample with S/N > 10, training the network separately for each S/N cut.

The architecture of the DLNN is summarized in Table 2 and consists of four linear layers with the number of neurons ranging from 500 to 2000. Each neuron of each layer is fully connected with the neurons of the previous and the following layer and these connections are updated during the training in order to derive the

Table 2. The DLNN architecture used in this paper. Linear is a fully connected layer that applies a linear transformation. We also include a ReLU function between each linear layer.

Layer	N_{input}	N_{output}
Linear	N^a	2000
Linear	2000	1000
Linear	1000	500
Linear	500	1

Note. [a]N is equal to the number of input filters (i.e. four or nine) for the single runs, and it is equal to the number of runs, i.e. 10, for the Meta-learner.

Table 3. The CNN architecture used in this paper. Conv2d indicates a 2D convolutional layer, Max pool corresponds to a pooling layer using the maximum value to downsample each image, Linear is a fully connected layer that applies a linear transformation. We included a ReLU function between each convolutional or linear layer. Galaxy images are introduced in the first Conv2d layer, while fluxes are included in the first Linear layer.

Layer	Kernel size	N_{input}	N_{output}
Conv2d	3×3	$1 \times 18 \times 18$	$64 \times 16 \times 16$
Conv2d	3×3	$64 \times 16 \times 16$	$64 \times 14 \times 14$
Max pool	2×2	$64 \times 14 \times 14$	$64 \times 13 \times 13$
Conv2d	3×3	$64 \times 13 \times 13$	$128 \times 11 \times 11$
Conv2d	3×3	$128 \times 11 \times 11$	$128 \times 9 \times 9$
Max pool	2×2	$128 \times 9 \times 9$	$128 \times 8 \times 8$
Conv2d	3×3	$128 \times 8 \times 8$	$256 \times 6 \times 6$
Conv2d	3×3	$256 \times 6 \times 6$	$256 \times 4 \times 4$
Conv2d	3×3	$256 \times 4 \times 4$	$256 \times 2 \times 2$
Max pool	2×2	$256 \times 2 \times 2$	$256 \times 1 \times 1$
Linear	–	$256 + N^a$	2000
Linear	–	2000	1000
Linear	–	1000	500
Linear	–	500	1

Note. [a]N is equal to the number of input filters (i.e. four or nine).

optimal way to map the input integrated magnitudes into the desired physical properties. Among each linear layer there is a Rectified Linear Unit (ReLU; Nair & Hinton 2010), such that $f(x) = 0$ if $x < 0$ and $f(x) = x$ if $x \geq 0$. A similar architecture has been previously used, even if with different inputs, number of neurons and hidden layers, to derive photometric redshift (e.g. Firth, Lahav & Somerville 2003; Collister & Lahav 2004).

As a simple and direct test, in this work we kept the same architecture when changing the set of inputs (e.g. *Euclid* only versus *Euclid* + LSST filters). However, we have not fully explored all the possible combinations of number of neurons and hidden layers, so there may be some architectures that optimize the use of the different set of inputs separately. This would have an impact on the absolute precision of the networks, but this is expected to leave the qualitative statements of the paper untouched. A full optimization will be performed in the future, once real data become available.

3.2 Convolutional Neural Network

The CNN has as input either of the sets of magnitudes of the DLNN (i.e. only *Euclid* filters or *Euclid* and ancillary filters), but it also includes the simulated H_E-band images. The architecture of this second network is summarized in Table 3. It consists of a series of convolutional layers applied to the H_E-band images, whose outputs are then combined with the magnitudes and processed through a set of linear layers. The convolutional layers are key for identifying features and shapes inside each H_E-band image. We chose a deep network with 3×3 kernels, instead of a network with less layers but larger kernels, as this architecture makes the decision function more discriminative and reduces the number of free parameters (Simonyan & Zisserman 2015).

After every two convolutional layers we applied a max-pooling layer, which is used for down-sizing the images which reduces the number of parameters of the network. As for the DLNN, the linear layers are interspersed with ReLU functions. The CNN is trained separately with the two sub-samples with different S/N cuts in the H_E-band images.

As a preliminary approach, we decided to derive, in both the CNN and DLNN runs, the redshift, stellar mass, and SFR independently. In this way we can investigate the challenges of the derivation of the three properties separately. We leave the combined analysis to a future work, but as reported in Appendix A (available online), we do not find evidence of galaxies with unrealistic combinations of physical properties, e.g. very-low-mass galaxies at very high z.

3.3 Re-scaling

We re-scaled all input parameters, i.e. stellar mass, SFR, redshift, and magnitudes, in order to have values between 0 and 1. This is performed by subtracting from each parameter its minimum value and dividing it by the difference between its maximum and minimum values. This is performed in logarithmic scale for the stellar mass and SFR and in linear scale for the redshift. We do not consider magnitudes below S/N < 3 in the re-scaling, as we assigned a value of -1 to all of them (Section 2.2). A similar re-scaling is also applied to each simulated H_E-band image in order to have pixel values between 0 and 1. The same re-scaling is applied to the entire sample; however, it is calculated using only the galaxies considered for the training (see later). This re-scaling is an important step in machine learning as the inputs and outputs may differ over orders of magnitude and, therefore, the largest one may dominate the training process. It is necessary to keep in mind, when comparing the CNN to the DLNN in the next sections, that because of the re-scaling performed separately for each galaxy, each H_E-band image has lost information about the overall galaxy flux and mainly contains the information on features and shapes, which is what we aim to train on.

3.4 Hyperparameters

We also divide the full sample into batches of 200 objects, which are used serially to update the training process, to increase stochasticity, and reduce the problem of local minima. In both the CNN and DLNN runs we implement an Adam algorithm (Kingma & Ba 2014), which is an optimization function based on stochastic gradient descent, to optimize the hyperparameters of the networks. To evaluate the difference between the data and the predictions, for each update of the network we derived a loss function based on the mean squared error (MSE), i.e. $l(x, y) = \sum_{n=0}^{N}(x_n - y_n)^2/N$, where x are the predicted values, y the target ones, and N is the total number of galaxy in input.

3.5 The training, the validation, and the test samples

We randomly split all the samples in three subsamples, of equal numbers, and we then apply augmentation only to the training and validation samples (see Fig. 3 and Section 2.4). Each network is trained with the first subsample (training sample), while the second subsample (validation sample) is used to estimate in an independent way the loss function and stop the training when it converges to avoid overfitting (i.e. over learning features specific of the training

sample), which may happen if we only analyse the loss function derived with the subsample used for training. The third subsample (test sample) is never seen by the networks, and is only used to evaluate the results. This subsample is not augmented so that the final statistics corresponds to a realistic galaxy sample. The split is performed once for all networks and is done before augmentation to avoid having the same object present in the training and in the test sample.

3.6 Networks combination methods

All networks are run 10 times using different random seeds. This number was chosen as a compromise between computational time and stochasticity. We then combine these runs with three different approaches. We report the results for the following:

(i) The best network, defined as the network with the smallest outlier fraction considering the full sample for the redshift estimation, then the subsample with a flat stellar mass distribution for the stellar mass derivation, and then the subsample above SFR $> 1\ M_{\odot}\ yr^{-1}$ for the SFR measurements.

(ii) The median of the outputs of the 10 networks.

(iii) A Meta-learner (Wolpert 1992; Euclid Collaboration: Humphrey et al. 2022a) that is an additional machine-learning network used as a linear discriminant among the different runs. This allows us to take into account the fact that some runs may have identified features peculiar to a subset of data. This Meta-learner consists of a DLNN with the architecture shown in Table 2; however, it uses the results of the 10 runs, instead of the magnitudes, as inputs.

4 GALAXY PROPERTIES DERIVED WITH MACHINE LEARNING

In this section, we report and discuss the results for the redshift, stellar mass, and SFR estimates based on machine-learning methods. We highlight that in order to train the machine-learning algorithms we need to have a sample with known output values. In this work, we rely on simulated data and with real data we could rely on spectroscopic redshifts and SFRs derived from a combination of different tracers (e.g. UV and IR stellar continuum). However, there is not an equivalent method to derive the true stellar mass of galaxies and we need to rely on the SED fitting applied to a subsample of galaxies with plenty of ancillary data. The power of machine learning is the capability of deriving the properties with a better accuracy.

For comparison, we also report the results derived with the same SED fitting procedure used to retrieve mock magnitudes, but using both the four *Euclid* filters and the nine *Euclid* and ancillary filters, as inputs. This of course corresponds to an ideal situation, as the same code and the same set of templates are used to retrieve the mock magnitudes and to estimate physical properties. It is, however, necessary to take into account that other SED fitting codes may perform differently, not only because of different SED libraries, but also because of the use of priors, which are instead not used here. This test is anyway useful for a direct comparison with the machine-learning algorithms considered.

4.1 Computational performance

One of the main advantages of machine-learning algorithms is the time necessary to derive the desired results. In general, the time necessary to apply a SED fitting procedure, regardless of the considered code, depends on the number of templates considered. For example, with the set-up considered in this work, i.e. 14 SED templates (see Section 2.2), 12 dust extinction values, 23 age values, and with redshift steps of 0.05 up to $z = 6$, *LePhare* takes 0.23 s per object,[5] requiring more than 4 h for $\sim 63 \times 10^3$ objects. Conversely, the DLNN training requires around 20–40 min, depending on the sample size (e.g. ~ 24 or 63×10^3 objects); while, the evaluation requires less than a minute in total for the same 63×10^3 objects and using the same machine. The CNN, using a graphics processing unit, requires a longer time for training, up to 13 h, and for evaluating ($\sim$15 min), given the larger complexity of the set of inputs. There is in any case a huge improvement on time cost moving from SED fitting codes to DLNN or CNN, as machine-learning networks, once the training is performed, require only to apply a set of linear transformations, or convolution for the CNN, to calculate the output values; while, a SED fitting procedure requires more complex steps, e.g. chi-square derivation for each combination of SED template and object.

4.2 Redshift derivation

In Table 4, we report the fraction of outliers ($f_{\rm out}$), defined as objects with $|z_{\rm out} - z_{\rm in}| > 0.15\,(1 + z_{\rm in})$ as commonly defined in the literature (e.g. Ilbert et al. 2010; Laigle et al. 2016), the bias

$$\langle \Delta z \rangle = \mathrm{median}[(z_{\rm out} - z_{\rm in})/(1 + z_{\rm in})], \quad (1)$$

and the NMAD[6]

$$\mathrm{NMAD} = 1.48\,[|z_{\rm out} - z_{\rm in}|/(1 + z_{\rm in})], \quad (2)$$

of the recovered redshifts for all networks. In the same table, we list results for the best and the median of the 10 runs of each network, as well as the results derived considering the Meta-learner. The latter are also shown in Figs 4 and 5, for the four (i.e. two sets of inputs and two S/N cuts) DLNN and the four CNN runs, respectively.

First, when focusing on each network to compare the 10 different runs, it is evident that the Meta-learner (see Section 3) gives in general better results than both the best of the 10 runs and the median of them. The fraction of outliers of the Meta-learner is always the smallest, even if for some networks with nine input filters the other two approaches give comparable results. The improvement in the fraction of outliers goes up to $\Delta\, f_{\rm out} = 0.016$ (0.027), when comparing with the best (median) of the 10 runs with four input filters. When we considered the networks with nine input filters, both for the DLNN and the CNN, the fraction of outliers are very small in all cases and the difference is, at maximum, $\Delta f_{\rm out} = 0.003$. In addition, the NMAD of the Meta-learner is always the smallest, showing that this approach not only generally decreases the fraction of outliers, but also improves the overall redshift accuracy. As an additional test, we analyse the impact of including at the end of the network a dropout layer, which randomly set to zero some of the elements of the inputs during training with probability 0.5 using samples from a Bernoulli distribution. This method, whose results are not shown here, even if it can identify different trends present in the data, as the Meta learner, does not have the advantage of using the results from multiple runs. Indeed, it performs worse than the Meta learner (e.g. $f_{\rm out} = 0.129$, $\langle \Delta z \rangle = 0.005$, and NMAD $= 0.066$ for the DLNN with four input filters and images with S/N > 3) even when doubling the nodes of the last hidden layer (i.e. $f_{\rm out} = 0.111$, $\langle \Delta z \rangle = -0.005$, and NMAD $= 0.058$).

[5]This is using a machine with 12 central processing units of 3.20 GHz and a 16 GB random access memory.

[6]This is equivalent to the standard deviation for a normal distribution.

Table 4. Statistics of the redshift derivation. Columns are (1) considered algorithms, (2) numbers of input filters N_{in}, (3) S/N cuts, (4) method used to combine the 10 runs of each network, (5) fraction of outliers, (6) bias, (7) NMAD, and (8) MSE. For the combination method, we include the results of the best run, the median among the 10 runs, and the results for the Meta-learner applied to the 10 runs. The first four lines correspond to results derived with SED fitting.

Algorithm (1)	N_{in} (2)	S/N (3)	Combination (4)	f_{out} (5)	$\langle\Delta z\rangle$ (6)	NMAD (7)	MSE (8)
SED	4	3	–	0.604	0.090	0.327	0.442
SED	4	10	–	0.596	0.105	0.315	0.280
SED	9	3	–	0.127	−0.002	0.045	0.081
SED	9	10	–	0.040	−0.003	0.029	0.028
DLNN	4	3	Best	0.099	0.011	0.052	0.014
			Median	0.103	0.011	0.050	0.014
			Meta-learner	0.088	0.005	0.050	0.014
DLNN	4	10	Best	0.076	0.005	0.050	0.010
			Median	0.081	0.003	0.050	0.010
			Meta-learner	0.068	0.004	0.048	0.010
CNN	4	3	Best	0.133	0.015	0.073	0.017
			Median	0.138	0.009	0.071	0.017
			Meta-learner	0.119	0.008	0.064	0.015
CNN	4	10	Best	0.133	−0.012	0.077	0.014
			Median	0.144	−0.001	0.081	0.015
			Meta-learner	0.117	0.013	0.071	0.013
DLNN	9	3	Best	0.001	−0.002	0.008	0.000
			Median	0.002	−0.001	0.010	0.001
			Meta-learner	0.001	0.001	0.006	0.000
DLNN	9	10	Best	0.002	0.001	0.013	0.001
			Median	0.002	0.001	0.014	0.001
			Meta-learner	0.002	0.000	0.010	0.000
CNN	9	3	Best	0.002	0.005	0.028	0.001
			Median	0.003	−0.001	0.022	0.001
			Meta-learner	0.002	−0.003	0.017	0.001
CNN	9	10	Best	0.003	−0.009	0.030	0.001
			Median	0.005	0.000	0.027	0.001
			Meta-learner	0.002	−0.003	0.023	0.001

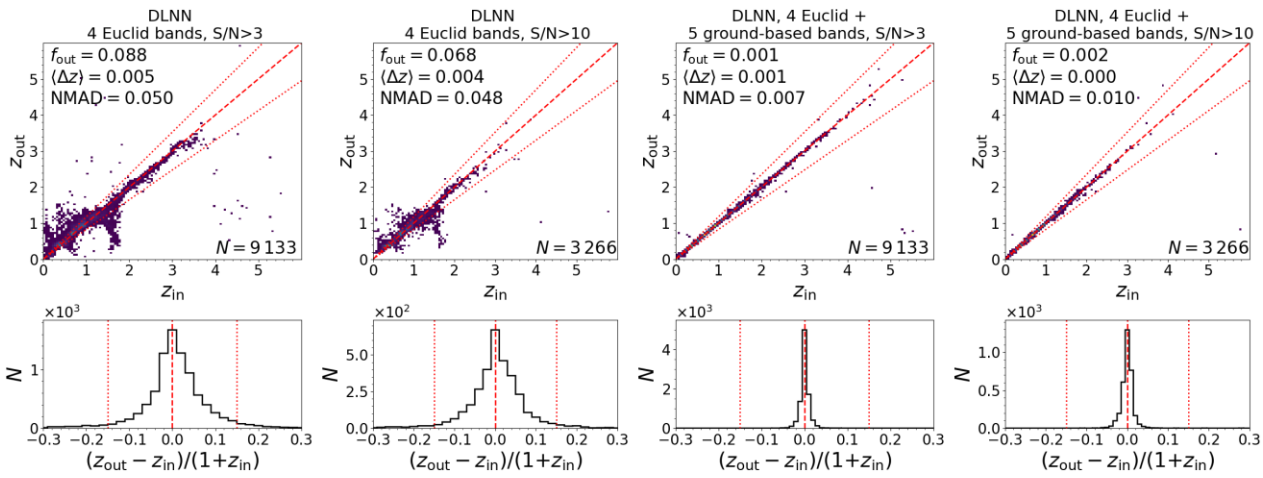

Figure 4. Top panel: Comparison between the recovered redshift and the input one for the DLNN methods. Points are coloured depending on the number of galaxies with the same combination of input and output redshift, following a linear scale from blue to yellow corresponding to 1 and 450 (200) galaxies with S/N > 3 (S/N > 10). The red dashed line is the identity and the red dotted lines indicate the outlier limits, i.e. $|\Delta z| = 0.15\,(1 + z_{in})$. On the top left of each panel, we report the fraction of outliers, the bias, and the NMAD. On the bottom right, we report the number of objects in the test sample. Bottom panel: Distribution of the absolute normalized redshift difference. The red vertical dashed line shows a null difference and the red dotted lines correspond to values of 0.15 and −0.15. From left-hand to right-hand panel: Redshift recovered using DLNN with four *Euclid* filters considering objects with S/N > 3 and with S/N > 10, and redshift recovered using DLNN with nine input filters considering objects with S/N > 3 and with S/N > 10. The 10 runs of each network are combined using a Meta-learner.

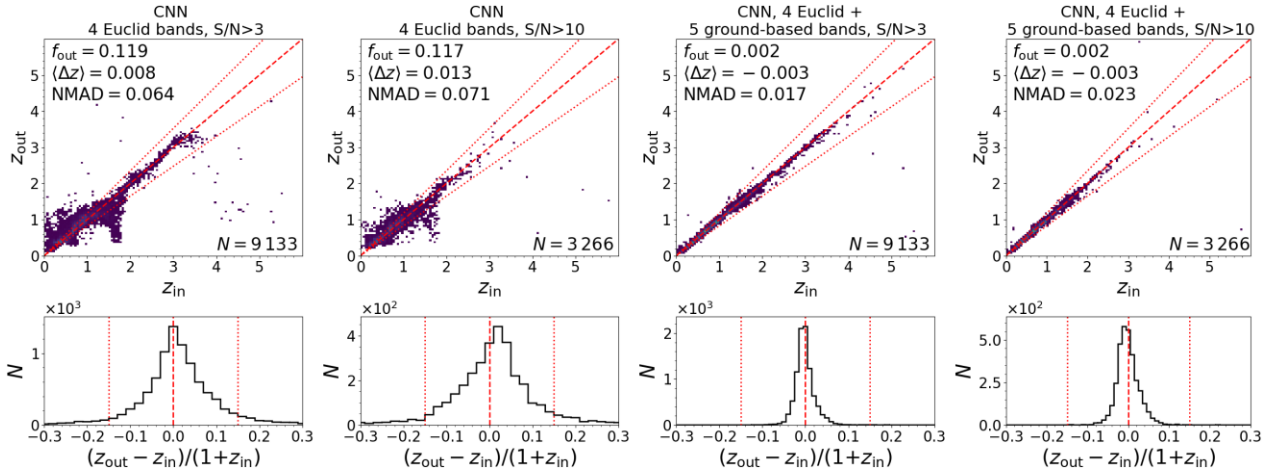

Figure 5. Same as Fig. 4, but for the runs using the CNN.

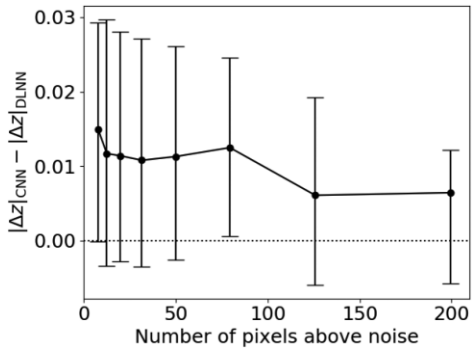

Figure 6. Difference between the recovered redshift in the CNN and DLNN with respect to the number of pixels that are three times above the noise level. Solid circles show the median difference, while the error bars show the central 25 per cent of the distribution.

We now compare the results of the DLNN and the CNN methods. First, if we run the CNN without including any additional flux, but only the H_E-band images, the fraction of outliers, averaging the 10 runs, is quite large: i.e. $f_{out} = 0.601$ when limiting the sample to images with S/N $>$ 3. Second, the CNN does not show an improvement with respect to the DLNN with any combination of S/N cuts or the number of input filters. Even in the cases where the fraction of outliers remains similar, which happens in networks with nine input filters, the NMAD increases. The inclusion of the H_E-band image adds information about the size of the objects, which could in principle improve the redshift estimation; however, this is probably limited by the H_E-band spatial resolution (0.3 arcsec, $\sim$2.5 kpc at $z = 1.5$). Indeed, as can be seen in Fig. 6, as the number of pixels above the noise level increases there is an increase of objects for which the CNN gives better results than the DLNN. In the future, the inclusion of images in multiple filters could be tested to allow the CNN to identify features in the SED, like the 4000 Å break, and to take advantage of the higher angular resolution of the I_E filter (0.1 arcsec, $\sim$0.8 kpc at $z = 1.5$).

We now focus on the results of the DLNN runs combined using the Meta-learner, which gives the best redshift estimation. By comparing the two samples of galaxies with S/N $>$ 10 and S/N $>$ 3, the redshift estimation is improved only when four filters are considered as input. This shows that, in the case with nine input filters, the improvement in data quality given by selecting only S/N $>$ 10 is shadowed by a decrease of the number of objects in the training sample (see Section 4.3 for further discussion). The inclusion of the additional five ground-based filters, i.e. *u*, *g*, *r*, *i*, and *z*, decreases the fraction of outliers from 0.066–0.088 to 0.001–0.002, depending on the S/N limit. The bias is always very small, below 0.001, while the NMAD decreases from $\sim$0.05 to $<$0.01, when changing from four to nine input magnitudes.

When four input filters are considered as input, there are galaxies at $z \sim 1.7$ for which the redshift is underestimated. In particular, in this redshift range the 4000 Å break is inside the I_E filter, so galaxies generally have a red I_E-I_Y colour. These outliers are intermediate-mass ($\langle M_* \rangle = 10^{9.8}\,M_\odot$) star-forming galaxies that, given they have high relative sSFRs, have I_E-I_Y colours similar to galaxies at lower redshifts. For these galaxies, LSST filters are probably necessary to add information bluewards the 4000 Å break. The importance of the optical filters is highlighted also by the sensitivity analysis reported in Appendix B (available online).

4.2.1 *Comparison with the* Euclid*photometric-redshift challenge*

The *Euclid* photometric-redshift challenge presented in Euclid Collaboration: Desprez et al. (2020) compared the photometric redshift estimation derived using 13 different methods, nine of which are based on machine-learning techniques. The considered machine-learning networks are based on the nearest neighbour (i.e. Directional Neighborhood Fitting by De Vicente, Sánchez & Sevilla-Noarbe 2016; FRANKENZ and the Nearest-Neighbour Photometric Redshift by Tanaka et al. 2018), boosted decision trees, random forest (Pedregosa et al. 2011), Gaussian processes, and neural networks (ANNz by Collister & Lahav 2004; Machine-learning Estimation Tool for Accurate PHOtometric Redshifts by Cavuoti et al. 2017; Amaro et al. 2019). We refer to each specific papers and the work by Euclid Collaboration: Desprez et al. (2020) for all the details about these methods.

A precise comparison between our work and their results needs to be considered with caution, given the differences in the considered input samples and filters; however, it can still be used to put our work into contest. In particular, their work uses-in input magnitudes in eight optical-to-near-IR filters (no *u* band) derived from observations available in the COSMOS field. Their analysis is restricted to galaxies with available spectroscopic redshifts and the derived one-point statistics, such as outlier fraction and NMAD, are calculated

weighting the spectroscopic sample in order to match the colour-space of the parent photometric catalogue. Therefore, on one hand, their sample may be prone to biases due to the spectroscopic selection; however, on the other hand, our input sample may be missing some galaxy populations not included in the considered SED templates.

Taking these differences in mind, results obtained with their machine-learning algorithms correspond to a fraction of outliers and NMAD varying from 0.031 to 0.326 and from 0.053 to 0.114, respectively. Both quantities are smaller than the ones derived in this work considering nine input filters; however, at least some of them are better than our results derived with only *Euclid* filters.

4.2.2 *Comparison with the* LePhare *SED fitting*

Results obtained with the CNN or DLNN all outperformed results from the SED fitting, using their same set of input magnitudes, even in the ideal case where both the code and the SED templates are the same, and are thus used to create the mock magnitudes. The fraction of outliers with the SED fitting corresponds to 0.604 and 0.127, when considering the sample with S/N > 3 and four and nine filters as input, respectively. For comparison, Euclid Collaboration: Desprez et al. (2020) used eight input filters (no *u* filter), the same SED fitting code, but a different input sample not corresponding to the templates used to generate photometric magnitudes (see previous section), finding an outlier fraction of 0.134 and a NMAD of 0.056. The outlier fractions we find range between four and 100 times more than the fraction of outliers derived with any CNN or DLNN runs. The improvement of the machine-learning algorithms over SED fitting has also been shown by Euclid Collaboration: Humphrey et al. (2022a) when selecting passive galaxies. These authors argued that this is due to the machine-learning networks' capability to optimally weight the different input data points; while, the SED fitting methods generally use a more direct weighting method, based on the S/N. More details on the SED fitting results are reported in Appendix C (available online).

Overall, among the different cases tested here, the best network for redshift estimation consists of the DLNN with nine input filters and S/N > 3, combined using a Meta-learner.

4.3 Stellar masses

The results for the stellar mass retrieval with machine learning are summarized in Table 5 for both the CNN and the DLNN methods, considering all the different inputs, both S/N cuts and the different methods to combine the 10 runs of each network. We remind the reader that the redshift is not among the inputs when deriving the stellar mass, as the two quantities are derived with separate networks. We estimate for the entire sample the fraction of outliers, arbitrary defined as galaxies for which the stellar mass is overestimated or underestimated by a factor of two ($\sim$0.3 dex). In addition, we estimate for each method the bias

$$\langle \Delta M_* \rangle = \mathrm{median}[\log_{10}(M_{*,\mathrm{out}}/M_{*,\mathrm{in}})] \quad (3)$$

and the NMAD of the recovered stellar mass

$$\mathrm{NMAD} = 1.48\,\mathrm{median}[|\log_{10}(M_{*,\mathrm{out}}/M_{*,\mathrm{in}})|]. \quad (4)$$

Figs 7 and 8 show the results for the DLNN and CNN runs, after combining the results using a Meta-learner.

As for the redshift, we first focus on the three methods to combine the 10 runs of each network. In general, the differences among the methods are less evident than for the redshift, with the three methods alternating on what gives the best results. However, the fraction of outliers derived with the Meta-learner is the smallest, except for the DLNN with nine filters as input and S/N > 3, with a difference in the fraction of outliers $\Delta f_{\mathrm{out}} \leq 0.042$ with respect to the best and the median of the 10 runs. Given the improvement, even if small, offered by the Meta-learner, we will focus on the results obtained with this method in the rest of this section.

We now compare the results of the DLNN, which includes only integrated magnitudes, and the CNN, which contains both integrated magnitudes and H_{E}-band images. We remind the reader that the H_{E}-band images include information about the features and shapes, but not the overall H_{E}-band magnitude. The use of only H_{E}-band images, without any integrated flux, is not sufficient to estimate the stellar mass, as it results, e.g. in a large outlier fraction $f_{\mathrm{out}} = 0.668$ when averaging the results of the 10 runs of the sample limited to images with S/N > 3 and $\log_{10}(M_*/\mathrm{M}_\odot) > 8$. Using the H_{E}-band images together with the integrated fluxes reduces instead the outlier fraction in the stellar mass with respect to the results obtained with the DLNN (see Table 5). An exception is the case with only four input filters and the sample limited to images with S/N > 10. The improvement in the fraction of outliers using the CNN is generally $\Delta f_{\mathrm{out}} < 0.014$; however, it is present even when the fraction is already very small. This happens, e.g. in the networks with nine input filters, for which the fractions of outliers in the DLNN are below 0.007; however, the inclusion of the H_{E}-band images produces an improvement by $\Delta f_{\mathrm{out}} = 0.001$–$0.002$.

The H_{E}-band filter traces light from a relatively old stellar population, at least at low redshift, so we expect it to be a good tracer of the stellar mass and drive the improvement when adding the H_{E}-band images. To verify this point, we analyse the fraction of outliers which are evolved galaxies, also called quiescent, (i.e. number of galaxies that are outlier and quiescent divided by the total number of quiescent) and the fraction of outliers which are galaxies currently forming stars and, therefore, including a younger stellar population (i.e. number of galaxies that are outlier and star-forming divided by the total number of star-forming galaxies). The first population is defined as galaxies with input specific star-formation rates (sSFRs) $\log10[\mathrm{sSFR}/(\mathrm{yr}^{-1})] < -10.5$, while the second has $\log10[\mathrm{sSFR}/(\mathrm{yr}^{-1})] \geq -10.5$.

The comparison between star-forming and quiescent outlier galaxies is shown for the stellar mass and for the redshift (Fig. 9). The inclusion of H_{E}-band images results on an improvement on the measurements of the stellar mass, but not of the redshift (see Section 4.2). While for redshift the outlier fraction of evolved galaxies generally increases for CNN with respect to DLNN, the opposite happens for the stellar mass. Moreover, there is no improvement in the stellar mass measurement of star-forming galaxies between DLNN and CNN when there are four input filters and S/N > 10. There is instead an improvement, even if small, in the mass outlier fraction for evolved galaxies, even if they are less than 30 per cent of star-forming galaxies in the training sample. This explains the improvement on the stellar mass measures introduced by the CNN.

In addition, we investigate the impact of galaxy size in the H_{E}-band images on the stellar mass derivation, by examining the number of pixels that are above three times each image's noise level, or S/N > 3. It is necessary to consider that a compact and unresolved structure is information that the network is using, therefore the introduction of the H_{E}-band images may also improve the stellar mass derivation of unresolved galaxies. Indeed, comparing DLNN and CNN with S/N > 3, galaxies for which the stellar mass measurement improved adding the H_{E}-band images have on average 35 pixels that are three

Table 5. Same as Table 4, but for the stellar mass. The results correspond to the mass range where the sample distribution is flat, i.e. $\log_{10}(M_*/\mathrm{M}_\odot) > 8$ for S/N > 3 and 8.5 for S/N > 10; while, values in parentheses correspond to the full sample. The first five lines correspond to results derived with a constant M/L_H ratio and with a SED fitting.

Algorithm (1)	N_in (2)	S/N (3)	Combination (4)	f_out (5)	$\langle\Delta M_*\rangle$ (6)	NMAD (7)	MSE (8)
$M/L_\mathrm{H} = 0.62$	1	3	–	0.298(0.300)	0.000[a](0.003)	0.300(0.302)	0.084(0.085)
SED	4	3	–	0.403(0.412)	0.134(0.140)	0.341(0.348)	0.215(0.268)
SED	4	10	–	0.432(0.436)	0.193(0.196)	0.375(0.378)	0.181(0.217)
SED	9	3	–	0.128(0.135)	0.001(0.002)	0.120(0.121)	0.112(0.130)
SED	9	10	–	0.048(0.051)	0.012(0.012)	0.094(0.095)	0.040(0.051)
DLNN	4	3	Best	0.132(0.139)	−0.037(−0.036)	0.146(0.148)	0.073(0.089)
	–	–	Median	0.123(0.129)	−0.025(−0.024)	0.129(0.131)	0.067(0.082)
	–	–	Meta-learner	0.121(0.128)	−0.033(−0.031)	0.133(0.135)	0.068(0.085)
DLNN	4	10	Best	0.217(0.228)	−0.034(−0.031)	0.208(0.212)	0.079(0.098)
	–	–	Median	0.223(0.231)	−0.057(−0.052)	0.221(0.223)	0.082(0.099)
	–	–	Meta-learner	0.217(0.228)	−0.076(−0.071)	0.209(0.213)	0.093(0.103)
CNN	4	3	Best	0.128(0.136)	−0.034(−0.032)	0.141(0.144)	0.079(0.097)
	–	–	Median	0.131(0.139)	−0.024(−0.022)	0.134(0.136)	0.068(0.083)
	–	–	Meta-learner	0.111(0.119)	−0.022(−0.020)	0.127(0.129)	0.062(0.079)
CNN	4	10	Best	0.252(0.262)	−0.086(−0.082)	0.235(0.240)	0.093(0.112)
	–	–	Median	0.263(0.273)	−0.087(−0.082)	0.239(0.242)	0.099(0.117)
	–	–	Meta-learner	0.221(0.230)	−0.020(−0.015)	0.220(0.226)	0.088(0.103)
DLNN	9	3	Best	0.005(0.008)	−0.017(−0.017)	0.054(0.054)	0.008(0.013)
	–	–	Median	0.005(0.008)	0.003(0.003)	0.041(0.041)	0.008(0.014)
	–	–	Meta-learner	0.006(0.010)	0.011(0.011)	0.042(0.042)	0.009(0.014)
DLNN	9	10	Best	0.007(0.012)	−0.001(0.000)	0.066(0.066)	0.013(0.022)
	–	–	Median	0.009(0.012)	−0.019(−0.019)	0.068(0.070)	0.009(0.016)
	–	–	Meta-learner	0.007(0.010)	0.004(0.004)	0.054(0.054)	0.007(0.015)
CNN	9	3	Best	0.006(0.011)	0.006(0.006)	0.050(0.050)	0.008(0.015)
	–	–	Median	0.006(0.010)	−0.001(−0.001)	0.045(0.045)	0.007(0.012)
	–	–	Meta-learner	0.005(0.009)	−0.024(−0.023)	0.051(0.051)	0.009(0.015)
CNN	9	10	Best	0.006(0.010)	−0.013(−0.013)	0.057(0.058)	0.008(0.015)
	–	–	Median	0.023(0.025)	−0.030(−0.030)	0.081(0.082)	0.019(0.026)
	–	–	Meta-learner	0.005(0.009)	−0.022(−0.022)	0.056(0.057)	0.008(0.015)

Note. [a]The bias is null by construction, as the used M/L_H is equal to the median value of the sample.

times above the noise level. On the other hand, galaxies for which the H_E-images worsen the stellar mass derivation have on average 36 pixels above the noise. Such a small difference indicates that the stellar mass estimation is not affected by the galaxy size. This is further visible in Fig. 10, where we analyse the improvement in the stellar mass measurement as a function of the number of pixels that are three times above the noise level. The median difference is always quite small, and is almost constant with the number of pixels above the noise level, except for the largest galaxies, which are probably only partially included inside the cut-out image (i.e. 18 × 18 pixel). We will analyse these extended galaxies in a future work focused on local galaxies. In the same figure it is also shown that when four filters are used as input the stellar mass improves when adding the H_E-band images for more than 50 per cent of galaxies at each size bin, except for the largest galaxies, justifying the additional computational effort of including images. The improvement is below $\Delta \log_{10}(M_*/\mathrm{M}_\odot) = 0.08$ for most (68 per cent) of the galaxies.

Limiting the sample to those objects with a H_E-band image at S/N > 10 produces different results, depending on the number of input filters. Indeed, when only the four *Euclid* filters are considered as inputs, the fraction of outliers increases from $f_\mathrm{out} \sim 0.107$–$0.128$ to $f_\mathrm{out} \sim 0.217$–$0.230$ for samples with S/N > 10. On the contrary, the fraction of outliers remains stable when nine filters are used as input; while, the bias improves by 0.002–0.007, depending on the considered network architecture.

To investigate the cause of the different impacts of limiting the sample to S/N > 10, we explore how the size of the training sample impacts the resulting fraction of outliers (Fig. 11), using the DLNN as an example and varying the size of the training sample down to 1 per cent of the complete one. This test is performed by randomly removing galaxies from the training sample after augmentation, therefore the stellar mass distribution, with some limitation once the sample size is very small, should be similar to the one of the training sample (see in Fig. 3, top central panel).

First, the fractions of outliers in both the S/N > 10 and the S/N > 3 samples with nine filters as inputs are consistent within the errors with the fractions of samples of similar size, but with S/N > 3. Second, the outlier fraction increases with decreasing sample size. However, this decrease, when nine magnitudes are used as input, becomes relevant (e.g. $\Delta f_\mathrm{out} > 0.02$) at smaller sample sizes (i.e. $<3 \times 10^3$ objects) than in the case of four input magnitudes (i.e. $<3 \times 10^4$ objects). Therefore, when only four filters are used as input, there is not enough information available, and it is preferable to have a larger, even if more noisy, sample. On the contrary, when more information is available, i.e. nine input filters, it is possible to focus on quality over quantity.

We now focus on the improvement of the stellar mass retrieval given by the inclusion of the *u*, *g*, *r*, *i*, and *z* ground-based filters, focusing again on the results obtained with the Meta-learner. The improvement is evident by looking at the fraction of outliers that

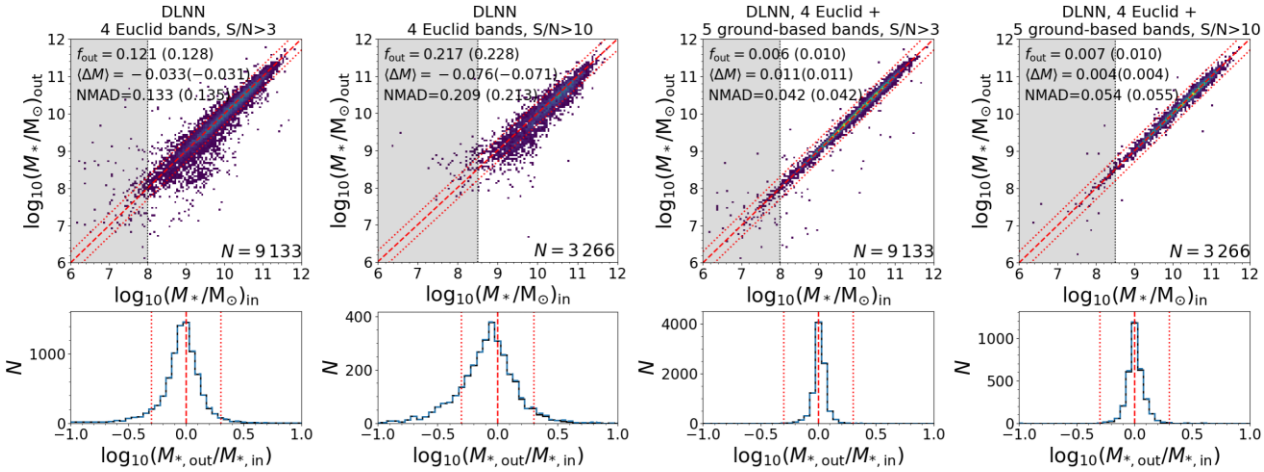

Figure 7. Top panel: Comparison between the recovered stellar mass and the input one for the DLNN methods. Points are coloured depending on the number of galaxies with the same combination of input and output stellar mass, following a linear scale from blue to yellow corresponding to 1 and 100 (40) galaxies with S/N > 3 (S/N > 10). The grey shaded area indicate the stellar mass range in which the input stellar mass distribution is not flat but underrepresented in the training sample, i.e. $M_* < 10^{8.5}$ $M_\odot$ for S/N > 3 and $M_* < 10^8$ $M_\odot$ for S/N > 10. The red dashed line is the identity and the red dotted lines indicate output stellar mass equal to twice or half the input one, which corresponds to the definition of an outlier. On the top-left side of each panel we report the fraction of outliers, bias, and NMAD of the sample with $M_* > 10^{8.5}$ $M_\odot$ or $M_* > 10^8$ $M_\odot$, depending on the S/N cut. In parentheses, we reported the same values for the full sample. On the bottom-right side, we report the number of objects in the test sample. Bottom panel: Distribution of the difference between the output and input stellar mass, for the full sample (blue dashed line) and for galaxies with $M_* > 10^{8.5}$ $M_\odot$ or $M_* > 10^8$ $M_\odot$ (black solid line), depending on the S/N cut. The red vertical dashed line shows a null difference and the red dotted lines indicate output stellar mass equal to twice or half the input one. From left-hand to right-hand panel: Stellar mass recovered using DLNN with four *Euclid* filters considering objects with S/N > 3 and with S/N > 10, and stellar mass recovered using DLNN with four *Euclid* filters and five ancillary bands considering objects with S/N > 3 and with S/N > 10. The 10 runs of each network are combined using a Meta-learner.

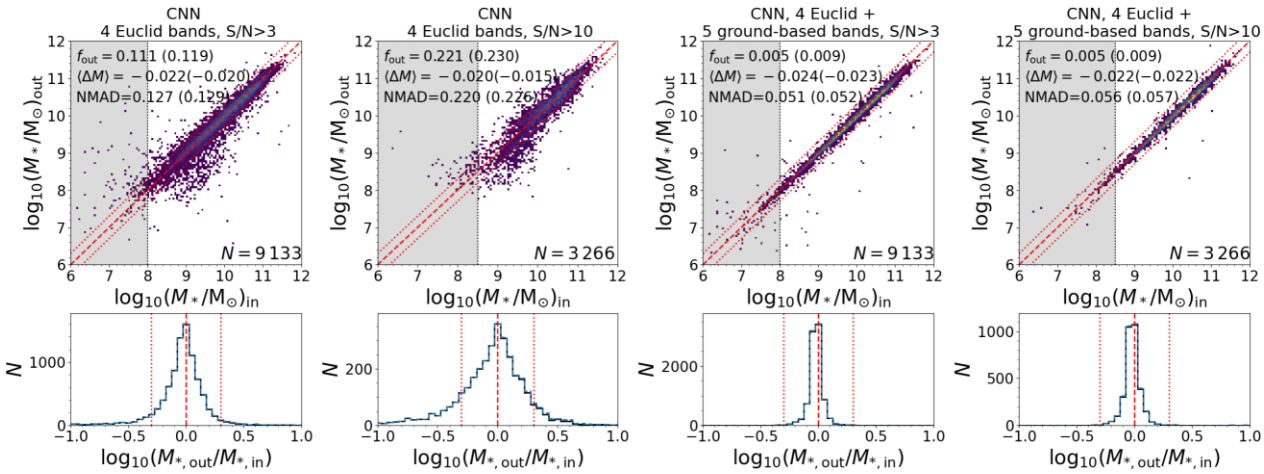

Figure 8. Same as Fig. 7, but for the runs using the CNN.

varies between $f_{out} = 0.107$–0.230 when only *Euclid* filters are used as input; while, it never exceeds $f_{out} = 0.010$ when all nine filters are used as input. The presence of long-wavelength filters, such as the H_E band at least up to $z \sim 1.5$, is indeed fundamental for obtaining a reliable stellar mass, as evident by the relative good stellar mass estimation when only four *Euclid* filters are used as input. However, the inclusion of shorter wavelength filters probably helps anchor the overall SED template to estimate very accurately the stellar mass. As for the redshift, the importance of the optical filters is highlighted also by the sensitivity analysis reported in Appendix B (available online).

4.3.1 *Variation with redshift and I_E magnitude*

In Figs 12 and 13, we show the variation of the stellar mass measures with redshift and I_E magnitude, respectively. There is a clear trend with redshift, as the fraction of outliers is below 0.28 and the NMAD is below 0.06 at $z < 1.5$. This trend with redshift is at least partially driven by the small number of galaxies available at $z > 3$ for the training, i.e. galaxies at lower redshift are more numerous and, therefore, they dominate the training process of the network. A similar effect is seen when looking at the variation with the I_E magnitude, as the fraction of outliers and the NMAD value increases at the brightest magnitudes, i.e. $I_E < 18$, that are

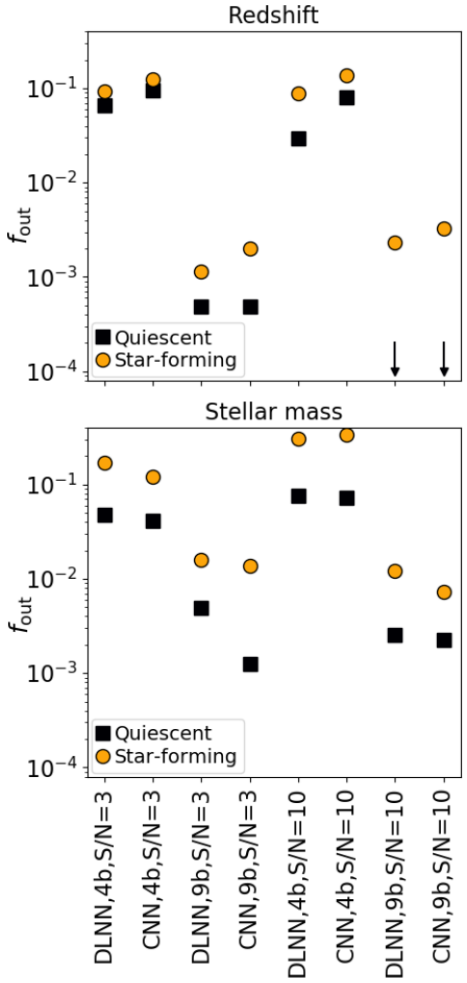

Figure 9. Fraction of outliers which are quiescent (i.e. $\log_{10}[\mathrm{SFR}/(\mathrm{M}_\odot\,\mathrm{yr}^{-1})] < -10.5$, black squares) and star-forming galaxies (i.e. $\log_{10}[\mathrm{SFR}/(\mathrm{M}_\odot\,\mathrm{yr}^{-1})] > -10.5$, yellow circles) for different networks, as derived for the redshift (top panel) and the stellar mass (bottom panel). Arrows in the top panel correspond to $f_{\rm out} = 0$.

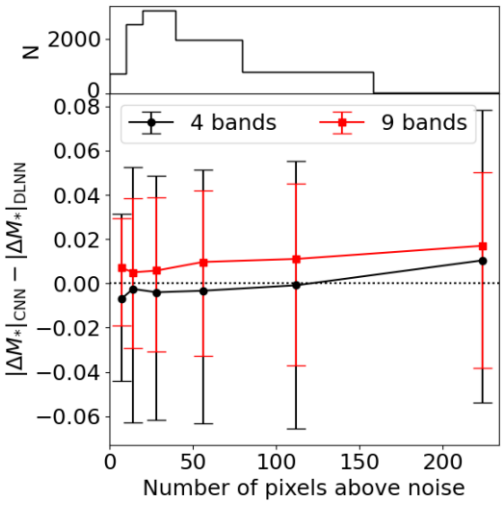

Figure 10. In the main panel, we have shown the difference between the absolute stellar mass errors in the CNN and DLNN with respect to the number of pixels that are three times above the noise level. Solid symbols show the median errors, while the error bars show the central 68 per cent ($\pm 1\sigma$) of the distribution. The difference is shown for the networks when using four (black circles) and nine input filters (red squares) and for the sample with images having S/N > 3. In the top panel, we report the distribution of the objects as a function of the number of pixels that are three times above the noise level.

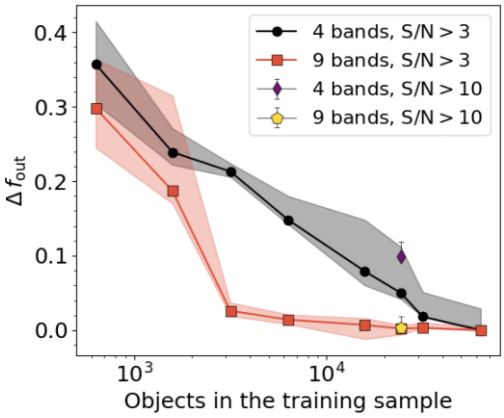

Figure 11. Difference in the outlier fraction of the stellar mass with decreasing size of the training sample. The difference is derived for the DLNN with four input magnitudes and S/N > 3 (black circles) and for the DLNN with nine input magnitudes and S/N > 3 (red squares), considering as zero point the fraction of outlier of the complete sample. For comparison, we report also the difference in the outlier fraction of the sample with S/N > 10 with respect to the sample with S/N > 3, considering the DLNN with four (purple diamond) and nine input magnitudes (yellow pentagon). Fractions are derived averaging the results of the 10 runs and the shaded areas show the standard variation within the 10 runs.

poorly represented in the sample. In the future, a larger sample with a flat multidimensional (i.e. z, stellar mass, SFR, and magnitude) distribution may help improve the measurement of these objects.

The deterioration of the stellar mass measurement with redshift may also be explained with the $H_{\rm E}$ band, which is the filter at the longest wavelength among the analysed ones, tracing shorter wavelengths at larger redshift and, therefore, are less sensitive to the light emitted by the old stellar populations which make up most of the stellar mass. The stellar mass measurement at magnitude $I_{\rm E} > 18$ is instead generally constant, except for the networks with four input filters and at S/N > 10, for which it becomes worse between $I_{\rm E} = 22$ and 24. The decrease visible at fainter magnitudes is probably spurious and driven by the limited number of galaxy in the sample (<100 at $I_{\rm E} > 25$) caused by the S/N cut.

4.3.2 *Comparison with the* LePhare *SED fitting*

Finally, to put these results into context, we derived the stellar mass directly from $H_{\rm E}$-band magnitudes assuming a single mass-to-light ratio, which is a simplistic but direct method, and using a SED fitting procedure. For the first case, we considered an ideal situation where we calculate the $H_{\rm E}$-band luminosity from the true redshift and we assumed a mass-to-light ratio equal to the median value (i.e. $\mathrm{M}/L_{\rm H} \sim 0.6$), obtained by comparing the $H_{\rm E}$-band luminosity directly with the true stellar mass. For the full sample with S/N > 3 and $\log_{10}(M_*/\mathrm{M}_\odot) > 8$, we obtained a fraction of outliers of 0.298, $\langle \Delta \log_{10}(M_*\,\mathrm{M}_\odot^{-1})\rangle = 0$, by construction, and NMAD = 0.3 (Table 5), which is overall a worse result than that obtained with both the DLNN and CNN methods for the same sample.

With the SED fitting (more details in Appendix C, available online), when considering only the four *Euclid* filters as input, results

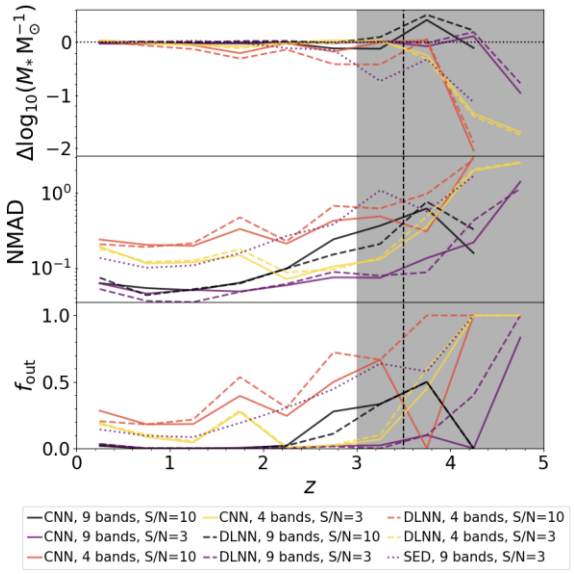

Figure 12. Redshift variation of the bias (top panel), NMAD (centre), and outlier fraction (bottom panel) of the recovered stellar mass. Different symbols indicate different algorithm architectures and different inputs (see legend). The grey area (black vertical dashed line) shows the redshift bins with <100 objects for the training samples which have S/N > 10 (S/N > 3).

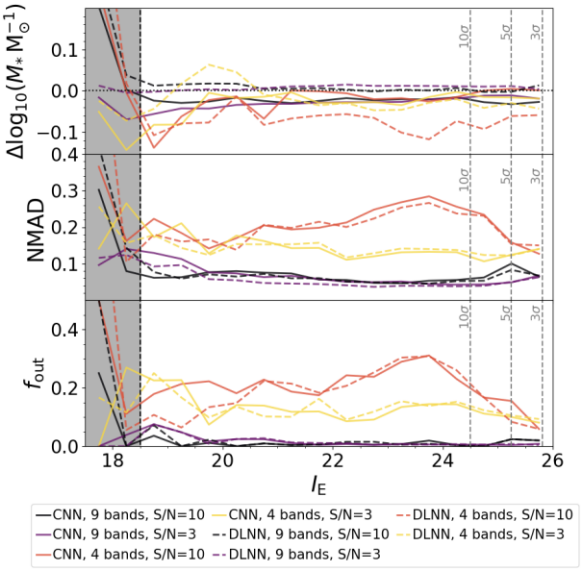

Figure 13. Same as Fig. 12, but focusing on the variation with I_E magnitude.

are even worse than with a constant M/L_H ratio, which is, however, derived considering the true median M/L_H. These results, however, improve when nine filters are used as input, but these still perform less well than the best CNN or DLNN results. Indeed, with the SED fitting we obtain $f_{out} = 0.128$–0.048 compared with $f_{out} \leq 0.02$, considering all combinations methods, or $f_{out} \leq 0.007$ focusing on results derived with the Meta-learner. In the SED fitting, redshift is kept free and the improvement when adding the *u*, *g*, *r*, *i*, and *z* filters is also driven by the improvement in the redshift estimation (see Section 4.2).

Finally, in Fig. 12, we also report the variation of bias, NMAD, and outlier fraction with redshift for one of the best SED estimates, i.e. nine filters used as input and S/N > 3. Surprisingly, machine learning results are more precise and accurate than SED fitting ones at all redshifts, even in the situation when the training sample is limited in number. This may also be linked to a difficult redshift estimation (see Appendix C, available online).

4.4 SFR derivation

In this section, we report the results for the SFR retrieval with machine learning, which are summarized in Table 6 for all networks, and are plotted in Figs 14 and 15, for the DLNN and CNN runs combined with the Meta-learner. Neither the redshift nor the stellar mass is among the inputs when deriving the SFR. As for the stellar mass, outliers are arbitrary defined as galaxies with SFR which are incorrect by, at least, a factor of two (∼0.3 dex). The bias is defined as

$$\langle \Delta \mathrm{SFR} \rangle = \mathrm{median}[\log_{10}(\mathrm{SFR_{out}}/\mathrm{SFR_{in}})] \quad (5)$$

and the NMAD for the SFR corresponds to

$$\mathrm{NMAD} = 1.48\,\mathrm{median}[|\log_{10}(\mathrm{SFR_{out}}/\mathrm{SFR_{in}})|]. \quad (6)$$

We find that the SFR is much more challenging to estimate than the stellar mass, as is evident by looking at the fraction of outliers, which ranges from 0.310 to 0.715, and the NMAD, which is always above 0.28, i.e. ∼32 per cent of the sample have a SFR wrong by at least 0.28 dex. The results obtained with the three methods to combine the 10 runs of each network are generally similar, with the Meta-learner and the best of the 10 runs giving slightly better results than the median of 10 runs. This is probably due to the large variation between the different runs, whose output SFRs have a mean standard deviation between 0.16 and 0.43. In the rest of this section, we focus on results obtained with the Meta-learner, for consistency with the redshift and stellar mass measures.

The DLNN gives in general a more precise value of the SFR than the CNN. This improvement is mainly driven by a reduction in the outlier fraction, down to $\Delta f_{out} = 0.11$, but also by a small decrease of the NMAD (i.e. Δ NMAD ≤ 0.15). One exception is the case with S/N > 3 and using nine input filters, for which the CNN results slightly improves (i.e. $\Delta f_{out} = 0.015$) over the DLNN ones. Not surprisingly, the use of only H_E-band images alone is not sufficient to estimate the SFR, as it results, e.g. in an outlier fraction of $f_{out} = 0.887$ when averaging the output of the 10 runs of the sample limited to images with S/N > 3 and SFR > $1\,M_\odot\,yr^{-1}$. In the future, the inclusion of images at wavelength shorter than the H_E band, which are more sensitive to the SFR and which will likely improve the predictions of this physical property, can be tested.

On one hand, for the SFR estimation it is not useful to limit the sample to galaxies with a high S/N, as the fraction of outliers increases by 0.07–0.16 when comparing the results of the samples with S/N > 3 and S/N > 10, similarly to what has been seen for the stellar mass (Section 4.3). On the other hand, it is evident that the inclusion of filters at short wavelengths, like the *u*, *g*, *r*, *i*, and *z* ground-based filters, improves the SFR estimates, lowering the outlier fraction by $\Delta f_{out} = 0.15$–0.25. The importance of the optical filters is also evident by performing a sensitivity analysis of the input features (Appendix B, available online). This is not surprising considering that UV wavelengths are better tracers of the SFR than near-IR ones (Pforr et al. 2012, 2013). However, even with nine input filters, the SFR measures remain more challenging than measuring the stellar mass or the redshift with the set-up we use.

Table 6. Same as Table 4, but for the SFR. The results correspond to all galaxies with SFR $>$ 1 $M_\odot$ yr^{-1}, while values in parentheses correspond to the full sample. The first four lines correspond to results derived with SED fitting.

Algorithm (1)	N_{in} (2)	S/N (3)	Combination (4)	f_{out} (5)	$\langle\Delta SFR\rangle$ (6)	NMAD (7)	MSE (8)
SED	4	3	–	0.752(0.814)	0.151(0.500)	0.997(1.412)	6.331(7.758)
SED	4	10	–	0.784(0.856)	0.286(0.807)	1.142(1.890)	5.036(10.887)
SED	9	3	–	0.560(0.622)	−0.115(−0.065)	0.521(0.637)	1.577(5.680)
SED	9	10	–	0.511(0.629)	−0.053(−0.028)	0.459(0.669)	1.313(9.075)
DLNN	4	3	Best	0.515(0.587)	−0.155(−0.136)	0.467(0.592)	0.810(0.956)
	–	–	Median	0.530(0.598)	−0.157(−0.125)	0.484(0.605)	0.791(0.920)
	–	–	Meta-learner	0.512(0.581)	−0.079(−0.056)	0.456(0.563)	0.697(0.904)
DLNN	4	10	Best	0.677(0.737)	−0.183(−0.143)	0.715(0.931)	1.207(1.463)
	–	–	Median	0.708(0.756)	−0.300(−0.233)	0.840(0.989)	1.462(1.559)
	–	–	Meta-learner	0.607(0.699)	−0.250(−0.248)	0.626(0.859)	1.263(1.479)
CNN	4	3	Best	0.553(0.619)	−0.162(−0.157)	0.516(0.634)	0.956(1.067)
	–	–	Median	0.559(0.625)	−0.185(−0.141)	0.528(0.656)	0.841(0.986)
	–	–	Meta-learner	0.582(0.628)	−0.333(−0.278)	0.562(0.657)	0.803(0.976)
CNN	4	10	Best	0.682(0.749)	−0.355(0.255)	0.775(0.997)	1.137(1.503)
	–	–	Median	0.715(0.761)	−0.459(−0.334)	0.891(1.065)	1.273(1.473)
	–	–	Meta-learner	0.692(0.746)	−0.381(−0.301)	0.777(0.972)	1.180(1.533)
DLNN	9	3	Best	0.310(0.411)	−0.023(−0.021)	0.280(0.350)	0.235(0.453)
	–	–	Median	0.319(0.419)	−0.033(−0.0046)	0.292(0.363)	0.246(0.461)
	–	–	Meta-learner	0.349(0.432)	−0.145(−0.144)	0.314(0.375)	0.249(0.478)
DLNN	9	10	Best	0.419(0.545)	0.028(0.016)	0.374(0.495)	0.369(0.730)
	–	–	Median	0.453(0.563)	−0.039(−0.072)	0.393(0.529)	0.367(0.752)
	–	–	Meta-learner	0.415(0.526)	−0.073(0.087)	0.365(0.475)	0.369(0.780)
CNN	9	3	Best	0.325(0.440)	−0.046(0.016)	0.293(0.383)	0.265(0.546)
	–	–	Median	0.342(0.446)	−0.056(−0.065)	0.304(0.390)	0.249(0.493)
	–	–	Meta-learner	0.334(0.434)	0.040(0.028)	0.300(0.380)	0.245(0.524)
CNN	9	10	Best	0.443(0.547)	−0.111(−0.112)	0.386(0.505)	0.395(0.822)
	–	–	Median	0.472(0.588)	−0.047(−0.067)	0.417(0.562)	0.378(0.795)
	–	–	Meta-learner	0.523(0.593)	−0.262(−0.259)	0.471(0.572)	0.412(0.888)

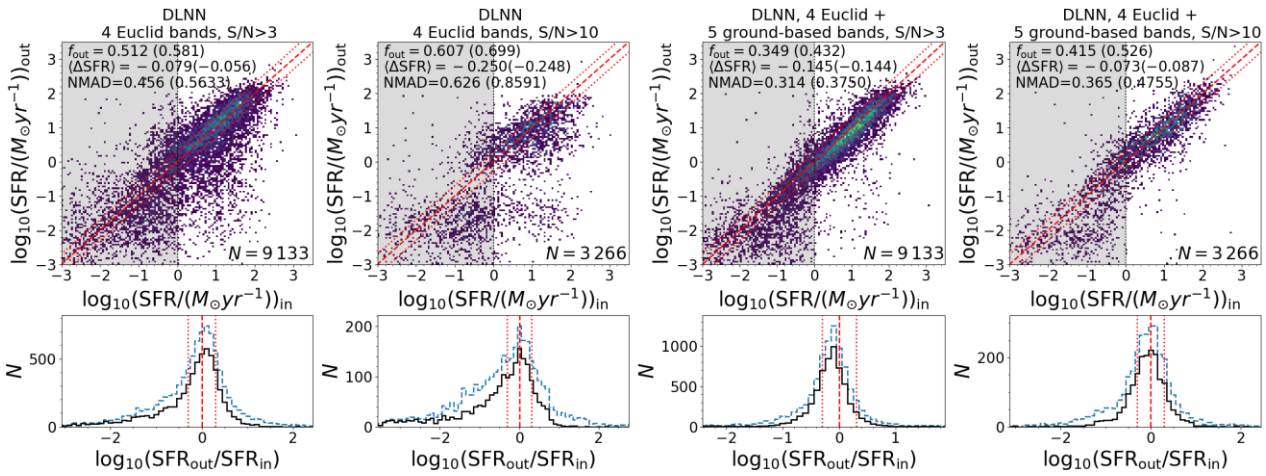

Figure 14. Top panel: Comparison between the recovered SFR and the input one for the DLNN methods. Points are coloured depending on the number of galaxies with the same combination of input and output SFR, following a linear scale from blue to yellow corresponding to 1 and 25 (10) galaxies with S/N $>$ 3 (S/N $>$ 10). The grey shaded area indicates SFR $<$ 1 $M_\odot$ yr^{-1}. The red dashed line is the identity and the red dotted lines indicate output SFR equal to twice or half the input SFR. On the top-left side of each panel, we report the fraction of outliers, bias, and NMAD of the sample with SFR $>$ 1 $M_\odot$ yr^{-1} and, in parentheses, the values for the full sample. On the bottom-right side, we report the number of objects in the test sample. Bottom panel: Distribution of the difference between the output and input SFR, for the full sample (blue dashed line) and for galaxies with SFR $>$ 1 $M_\odot$ yr^{-1} (black solid line). The red vertical dashed line shows a null difference and the red dotted lines indicate output SFR equal to twice or half the input SFR. From left-hand to right-hand panel: SFR recovered using DLNN with four *Euclid* filters considering objects with S/N $>$ 3, SFR recovered using DLNN with four *Euclid* filters and five ancillary bands considering objects with S/N $>$ 3, SFR recovered using DLNN with four *Euclid* filters considering objects with S/N $>$ 10, and SFR recovered using DLNN with four *Euclid* filters and five ancillary bands considering objects with S/N $>$ 10. The 10 runs of each network are combined using a Meta-learner.

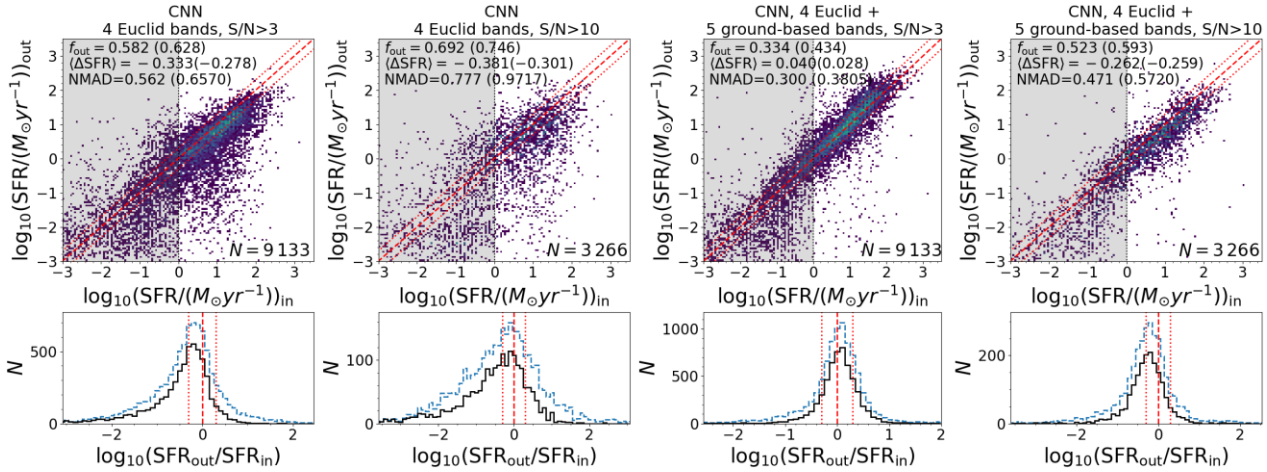

Figure 15. Same as Fig. 14, but for the runs using the CNN.

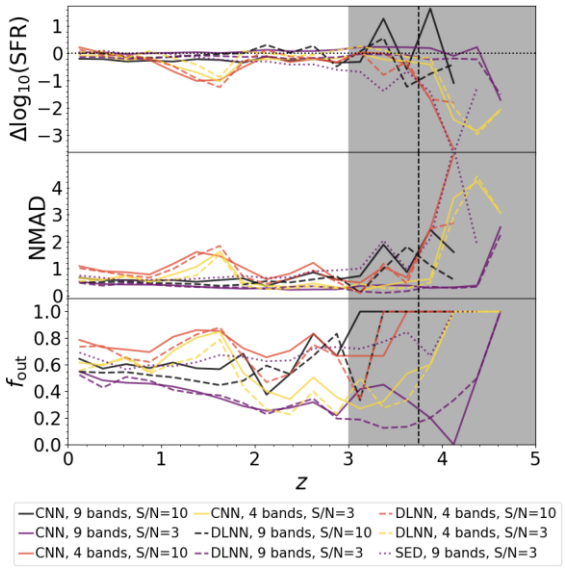

Figure 16. Redshift variation of the bias (top panel), NMAD (centre), and outlier fraction (bottom panel) of the recovered SFR. Solid and dashed lines indicate the statistics of CNN and DLNN, respectively, with different colours depending on the set of inputs (see legend). The grey area (black vertical dashed line) shows the redshift bins with <100 objects in the training samples which have S/N > 10 (S/N > 3).

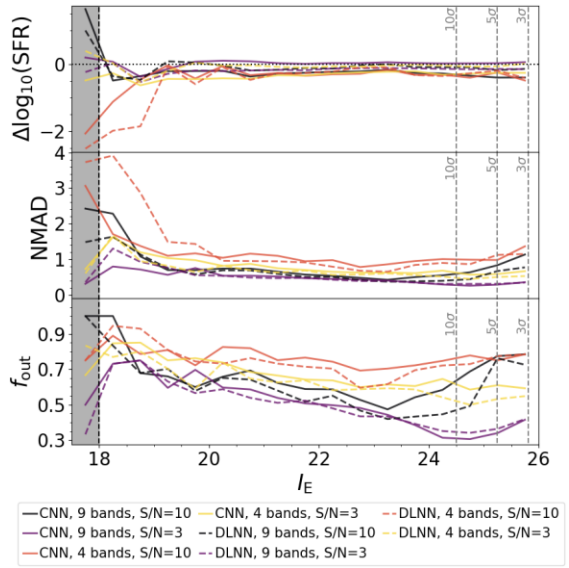

Figure 17. Same as Fig. 16, but focusing on the variation with the I_E magnitude.

4.4.1 *Variation with redshift and I_E magnitude*

In Figs 16 and 17, we show the variation with redshift and I_E magnitude of the fraction of outliers, the bias, and the NMAD of the recovered SFR. As for the stellar mass, there is a rapid deterioration of the SFR estimation as soon as the number of objects available for training is relative small, i.e. $z > 4$ and $I_E < 18$. In addition, for some of the most accurate networks, i.e. CNN and DLNN with nine input filters and S/N > 3 (purple solid and dashed lines line in Fig. 16), the fraction of outliers decreases at increasing redshift, ranging from $f_{out} = 0.52$–0.55 at $z = 0.125$ to $f_{out} \sim 0.13$–0.33 at $z = 3.6$. For these networks, the outlier fraction also decreases towards fainter I_E magnitudes. We can speculate on different effects driving these dependencies. First, at increasing redshifts our filters trace shorter rest-frame wavelengths, which are more sensitive to SFR. Second, the average SFR of star-forming galaxies increases with redshift (i.e. Brinchmann et al. 2004; Noeske et al. 2007; Bisigello et al. 2018), making the SFR easier to estimate for the networks. In the future the SFR estimation at low redshift can be further analysed with a sample more focused on low-z galaxies than the one analysed in this work. The dependence of the outlier fraction with I_E magnitude can be linked with the dependence with redshift, as high-redshift galaxies are expected to be fainter than lower redshift ones.

4.4.2 *Comparison with the* LePhare *SED fitting*

Finally, as done for redshift and stellar mass, we compare the results obtained with the CNN and DLNN methods with the results derived with a SED fitting procedure, for which we give more details in Appendix C (available online). Our machine-learning algorithms perform better than the SED fitting; however, the improvement is not as pronounced as it is for the redshift and the stellar mass, with a difference $\Delta f_{out} < 0.250$. It is, however, worth noticing that the fraction of outliers derived with the DLNN with S/N > 3 and four filters as input ($f_{out} = 0.512$) is lower than the fraction of outliers

derived with the SED fitting method for the same sample, but when using nine input filters ($f_{\rm out} = 0.560$).

Finally, in Fig. 16, we investigate the variation of bias, NMAD, and outlier fraction with redshift, in the case of SED fitting applied to galaxies with S/N > 3 and nine input filters. As for the stellar mass, the SFR estimation derived with SED fitting is always worse than the one derived with DLNN or CNN at any given redshift. This, however, may be due to a wrong redshift estimation (see Appendix C, available online).

5 SUMMARY

This paper is a general exploration of using machine learning to determine and measure the most basic properties of galaxies, particularly those at higher redshifts. This will be a critical process for the next generation of galaxy surveys as *Euclid*, *Rubin*/LSST, and the *Roman Space Telescope*. We investigate this problem in several ways, including different machine-learning methods and by using as input different forms of data. We use information from the *Euclid Space Telescope* as a baseline for understanding how these estimates can be done on other telescopes with similar data. We thus investigate how well machine learning does in retrieving three main features of galaxies – redshifts, stellar masses, and SFRs. This work presents only point estimates for all these quantities and the inclusion of probability distribution functions or statistical errors will be investigated on a future work.

Our main results are the following:

(i) Our machine-learning algorithms perform better than traditional methods. In particular, to estimate the stellar mass we consider a simple but direct method consisting on a constant $M/L_{\rm H}$, which is derived from the true median mass-to-light ratio of the sample. As a second method we test a SED fitting procedure, using the same code and templates considered to derive the mock magnitudes, for redshift, stellar mass, and SFR. The redshift and stellar mass machine-learning runs outperformed the other methods; while, the improvement in the SFR estimation is more limited. It is, however, necessary to keep in mind that machine-learning networks, on the contrary of SED fitting procedure, are limited to the parameter space of the sample used for training.

(ii) We verify that it is preferable to combine the results of different runs using a Meta-learner, i.e. an additional DLNN which uses the results of the other networks as inputs. The Meta-learner outperforms the median of the results and the best among the different runs for the redshift predictions, with an improvement in the outlier fraction even up to $\Delta\ f_{\rm out} = 0.029$.

(iii) The inclusion of $H_{\rm E}$-band images, in addition to the integrated magnitudes, is particularly useful for the stellar mass estimation, due to the fact that the $H_{\rm E}$-band filter traces the light from relatively old stellar populations, at least at low redshift, and it is therefore a good tracer of stellar mass through structure. The inclusion of images in this filter has a small impact on the redshift estimation; while, it mainly introduces noise in the SFR derivation. In the future, the impact of images on the SFR and redshift retrieval may be further tested by including images at shorter wavelengths, which are more sensitive to on-going star formation, and SED features useful for redshift estimation, such as the 4000 Å, or images with a smaller angular resolution than the ones tested.

(iv) Limiting the input sample only to galaxies with S/N > 10 in the $H_{\rm E}$ band improves the results only in a few cases. This selection improves the quality of the input data; however, at the same time, it reduces the number of galaxies in the training sample.

(v) We compare results obtained using only the four *Euclid* filters and complementing them with additional five LSST-like filters. The improvement is evident in all cases, with the fraction of outliers decreasing by $\Delta\ f_{\rm out} = 0.15$–$0.25$ for the SFR estimation; while, it decreases to below $f_{\rm out} \leq 0.020$ and 0.005 for the stellar mass and the redshift, respectively. These results indicate the necessity of an eventual coordinated effort from *Rubin*/LSST and *Euclid* to improve the measurement of physical properties such as redshift, stellar mass, and SFR.

ACKNOWLEDGEMENTS

LB and CC acknowledge the support of the STFC Cosmic Vision funding. LB acknowledges the financial support of Agenzia Spaziale Italiana (ASI) under the research contract 2018-31-HH.0. SvM acknowledges funding from the European Research Council through the award of the Consolidator Grant ID 681627-BUILDUP. HH is supported by a Heisenberg grant of the Deutsche Forschungsgemeinschaft (Hi 1495/5-1) as well as an ERC Consolidator Grant (No. 770935). MB acknowledges financial contributions from the agreement ASI/INAF 2018-23-HH.0, Euclid ESA mission - Phase D. The Euclid Consortium acknowledges the European Space Agency and a number of agencies and institutes that have supported the development of *Euclid*, in particular the Academy of Finland, the Agenzia Spaziale Italiana, the Belgian Science Policy, the Canadian Euclid Consortium, the French Centre National d'Etudes Spatiales, the Deutsches Zentrum für Luft- und Raumfahrt, the Danish Space Research Institute, the Fundação para a Ciência e a Tecnologia, the Ministerio de Ciencia e Innovación, the National Aeronautics and Space Administration, the National Astronomical Observatory of Japan, the Netherlandse Onderzoekschool Voor Astronomie, the Norwegian Space Agency, the Romanian Space Agency, the State Secretariat for Education, Research and Innovation (SERI) at the Swiss Space Office (SSO), and the United Kingdom Space Agency. A complete and detailed list is available on the Euclid web site (http://www.euclid-ec.org). In this work, we made use of the NUMPY (Harris et al. 2020) package for Python.

DATA AVAILABILITY

Data included in this paper will be available on request.

REFERENCES

Akeson R. et al., 2019, preprint (arXiv:1902.05569)
Amaro V. et al., 2019, MNRAS, 482, 3116
Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, MNRAS, 310, 540
Bisigello L. et al., 2016, ApJS, 227, 19
Bisigello L. et al., 2017, ApJS, 231, 3
Bisigello L. et al., 2020, MNRAS, 494, 2337
Bisigello L., Caputi K. I., Grogin N., Koekemoer A., 2018, A&A, 609, A82
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Bruzual G., Charlot S., 2003, MNRAS, 344, 1000
Calzetti D., Armus L., Bohlin R. C., Kinney A. L., Koornneef J., Storchi-Bergmann T., 2000, ApJ, 533, 682
Cavuoti S., Amaro V., Brescia M., Vellucci C., Tortora C., Longo G., 2017, MNRAS, 465, 1959
Chabrier G., 2003, PASP, 115, 763
Chen X.-w., Jeong J. C., 2007, Proc. Sixth Int. Conf. Mach. Learn. Appl. (ICMLA 2007), Enhanced Recursive Feature Elimination. IEEE, Cincinnati, OH, USA, p. 429

Cheng T.-Y. et al., 2020, MNRAS, 493, 4209
Ciesla L., Elbaz D., Fensch J., 2017, A&A, 608, A41
Collister A. A., Lahav O., 2004, PASP, 116, 345
Conselice C. J., 2003, ApJS, 147, 1
Conselice C. J., Bluck A. F. L., Mortlock A., Palamara D., Benson A. J., 2014, MNRAS, 444, 1125
D'Isanto A., Polsterer K. L., 2018, A&A, 609, A111
De Vicente J., Sánchez E., Sevilla-Noarbe I., 2016, MNRAS, 459, 3078
Delli Veneri M., Cavuoti S., Brescia M., Longo G., Riccio G., 2019, MNRAS, 486, 1377
Dieleman S., Willett K. W., Dambre J., 2015, MNRAS, 450, 1441
Dobbels W., Krier S., Pirson S., Viaene S., De Geyter G., Salim S., Baes M., 2019, A&A, 624, A102
Euclid Collaboration: Desprez G. et al., 2020, A&A, 644, A31
Euclid Collaboration: Humphrey A. et al., 2022a, A&A, in press
Euclid Collaboration: Scaramella R. et al., 2022b, A&A, 662, A112
Euclid Collaboration: Schirmer M. et al., 2022c, A&A, 662, A92
Firth A. E., Lahav O., Somerville R. S., 2003, MNRAS, 339, 1195
Gunn J. E. et al., 1998, AJ, 116, 3040
Guyon I., Elisseeff A., 2003, J. Mach. Learn. Res., 3, 1157
Harris C. R. et al., 2020, Nature, 585, 357
Hoyle B., 2016, Astron. Comput., 16, 34
Huertas-Company M. et al., 2015, ApJS, 221, 8
Ilbert O. et al., 2006, A&A, 457, 841
Ilbert O. et al., 2010, ApJ, 709, 644
Ivezic Z. et al., 2008, Serb. Astron. J., 176, 1
Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q., Liu T.-Y., 2017, Adv. Neural Inf. Process. Syst., 30, 3149
Kennicutt Robert C. J., 1998, ARA&A, 36, 189
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Kursa M. B., Jankowski A., Rudnicki W. R., 2010, Fundam. Inform., 101, 271
Laigle C. et al., 2016, ApJS, 224, 24
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Lipovetsky S., Conklin M., 2001, Appl. Stochastic Models Bus. Ind., 17, 319
Lundberg S. M., Lee S.-I., 2017, Adv. Neural Inf. Process. Syst., 30
Madau P., Dickinson M., 2014, ARA&A, 52, 415
Mortlock A. et al., 2013, MNRAS, 433, 1185
Mowla L. A. et al., 2019, ApJ, 880, 57
Mucesh S. et al., 2021, MNRAS, 502, 2770
Nair V., Hinton G. E., 2010, Proc. 27th Int. Conf. Mach. Learn. ICML'10, Rectified Linear Units Improve Restricted Boltzmann Machines. Omnipress, Madison, WI, USA, p. 807
Noeske K. G. et al., 2007, ApJ, 660, L43
Oke J. B., Gunn J. E., 1983, ApJ, 266, 713
Pasquet J., Bertin E., Treyer M., Arnouts S., Fouchez D., 2019, A&A, 621, A26
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Pforr J., Maraston C., Tonini C., 2012, MNRAS, 422, 3285
Pforr J., Maraston C., Tonini C., 2013, MNRAS, 435, 1389
Razim O., Cavuoti S., Brescia M., Riccio G., Salvato M., Longo G., 2021, MNRAS, 507, 5034
Sagi O., Rokach L., 2021, Inform. Sci., 572, 522
Schreiber C., Elbaz D., Pannella M., Ciesla L., Wang T., Koekemoer A., Rafelski M., Daddi E., 2016, A&A, 589, A35
Scoville N. et al., 2007, ApJS, 172, 1
Simonyan K., Zisserman A., 2015, Very Deep Convolutional Networks for Large-scale Image Recognition, 3rd International Conference on Learning Representations (ICLR 2015). p. 1
Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, ApJS, 214, 15
Stensbo-Smidt K., Gieseke F., Igel C., Zirm A., Steenstrup Pedersen K., 2016, MNRAS, 464, 2577
Štrumbelj E., Kononenko I., 2013, Knowl. Inform. Syst., 41, 647
Surana S., Wadadekar Y., Bait O., Bhosale H., 2020, MNRAS, 493, 4808
Tagliaferri R., Longo G., Andreon S., Capozziello S., Donalek C., Giordano G., 2003, Neural Nets. Lecture Notes in Computer Science. Vol. 2859. Springer, Berlin, p. 226
Tanaka M. et al., 2018, PASJ, 70, S9
Tohill C., Ferreira L., Conselice C. J., Bamford S. P., Ferrari F., 2021, ApJ, 916, 4
Wolpert D. H., 1992, Neural Netw., 5, 241

SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

Euclid_XXIII_appendix

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

[1]*Dipartimento di Fisica e Astronomia 'G.Galilei', Universitá di Padova, Via Marzolo 8, I-35131 Padova, Italy*
[2]*INAF-Osservatorio di Astrofisica e Scienza dello Spazio di Bologna, Via Piero Gobetti 93/3, I-40129 Bologna, Italy*
[3]*School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK*
[4]*Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK*
[5]*Sterrenkundig Observatorium, Universiteit Gent, Krijgslaan 281 S9, B-9000 Gent, Belgium*
[6]*INFN Section of Naples, Via Cinthia 6, I-80126 Napoli, Italy*
[7]*INAF-Osservatorio Astronomico di Capodimonte, Via Moiariello 16, I-80131 Napoli, Italy*
[8]*Department of Physics 'E. Pancini', University Federico II, Via Cinthia 6, I-80126 Napoli, Italy*
[9]*Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto, CAUP, Rua das Estrelas, P-4150-762 Porto, Portugal*
[10]*INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy*
[11]*Institute of Cosmology and Gravitation, University of Portsmouth, Portsmouth PO1 3FX, UK*
[12]*Istituto Nazionale di Astrofisica (INAF) - Osservatorio di Astrofisica e Scienza dello Spazio (OAS), Via Gobetti 93/3, I-40127 Bologna, Italy*
[13]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, NL-9700 AV Groningen, The Netherlands*
[14]*Université Paris-Saclay, CNRS, Institut d'Astrophysique Spatiale, F-91405 Orsay, France*
[15]*Dipartimento di Fisica e Astronomia, Universitá di Bologna, Via Gobetti 93/2, I-40129 Bologna, Italy*
[16]*INFN-Sezione di Bologna, Viale Berti Pichat 6/2, I-40127 Bologna, Italy*
[17]*Universitäts-Sternwarte München, Fakultät für Physik, Ludwig-Maximilians-Universität München, Scheinerstrasse 1, D-81679 München, Germany*
[18]*Max Planck Institute for Extraterrestrial Physics, Giessenbachstr. 1, D-85748 Garching, Germany*
[19]*INAF-Osservatorio Astrofisico di Torino, Via Osservatorio 20, I-10025 Pino Torinese (TO), Italy*
[20]*Dipartimento di Fisica, Universitá degli Studi di Genova, INFN-Sezione di Genova, via Dodecaneso 33, I-16146 Genova, Italy*
[21]*INFN-Sezione di Roma Tre, Via della Vasca Navale 84, I-00146 Roma, Italy*
[22]*Dipartimento di Fisica, Universitá degli Studi di Torino, Via P. Giuria 1, I-10125 Torino, Italy*
[23]*INFN-Sezione di Torino, Via P. Giuria 1, I-10125 Torino, Italy*
[24]*INAF-IASF Milano, Via Alfonso Corti 12, I-20133 Milano, Italy*
[25]*Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, Campus UAB, 08193 Bellaterra (Barcelona), Spain*
[26]*Port d'Informació Científica, Campus UAB, C. Albareda s/n, E-08193 Bellaterra (Barcelona), Spain*
[27]*Institut d'Estudis Espacials de Catalunya (IEEC), Carrer Gran Capitá 2-4, E-08034 Barcelona, Spain*

[28]*Institute of Space Sciences (ICE, CSIC), Campus UAB, Carrer de Can Magrans, s/n, E-08193 Barcelona, Spain*
[29]*INAF-Osservatorio Astronomico di Roma, Via Frascati 33, I-00078 Monteporzio Catone, Italy*
[30]*Dipartimento di Fisica e Astronomia 'Augusto Righi' - Alma Mater Studiorum Universitá di Bologna, Viale Berti Pichat 6/2, I-40127 Bologna, Italy*
[31]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[32]*ESAC/ESA, Camino Bajo del Castillo, s/n., Urb. Villafranca del Castillo, E-28692 Villanueva de la Cañada, Madrid, Spain*
[33]*European Space Agency/ESRIN, Largo Galileo Galilei 1, I-00044 Frascati, Roma, Italy*
[34]*Université de Lyon, Université Claude Bernard Lyon 1, CNRS/IN2P3, IP2I Lyon, UMR 5822, F-69622 Villeurbanne, France*
[35]*Observatoire de Sauverny, Ecole Polytechnique Fédérale de Lau- sanne, CH-1290 Versoix, Switzerland*
[36]*Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK*
[37]*Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Edifício C8, Campo Grande, P-1749-016 Lisboa, Portugal*
[38]*Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, P-1749-016 Lisboa, Portugal*
[39]*Department of Astronomy, University of Geneva, ch. dÉcogia 16, CH-1290 Versoix, Switzerland*
[40]*Department of Physics, Oxford University, Keble Road, Oxford OX1 3RH, UK*
[41]*INFN-Padova, Via Marzolo 8, I-35131 Padova, Italy*
[42]*Université Paris-Saclay, Université Paris Cité, CEA, CNRS, Astrophysique, Instrumentation et Modélisation Paris-Saclay, F-91191 Gif-sur-Yvette, France*
[43]*INAF-Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, I-34143 Trieste, Italy*
[44]*Aix-Marseille Université, CNRS/IN2P3, CPPM, F-13007 Marseille, France*
[45]*Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Via Irnerio 46, I-40126 Bologna, Italy*
[46]*INAF-Osservatorio Astronomico di Padova, Via dell'Osservatorio 5, I-35122 Padova, Italy*
[47]*Dipartimento di Fisica 'Aldo Pontremoli', Universitá degli Studi di Milano, Via Celoria 16, I-20133 Milano, Italy*
[48]*INAF-Osservatorio Astronomico di Brera, Via Brera 28, I-20122 Milano, Italy*
[49]*INFN-Sezione di Milano, Via Celoria 16, I-20133 Milano, Italy*
[50]*Institute of Theoretical Astrophysics, University of Oslo, P.O. Box 1029 Blindern, N-0315 Oslo, Norway*
[51]*Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA*
[52]*von Hoerner & Sulger GmbH, SchloßPlatz 8, D-68723 Schwetzingen, Germany*
[53]*Technical University of Denmark, Elektrovej 327, DK-2800 Kgs. Lyngby, Denmark*
[54]*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[55]*Université de Genève, Département de Physique Théorique and Centre for Astroparticle Physics, 24 quai Ernest-Ansermet, CH-1211 Genève 4, Switzerland*
[56]*Department of Physics and Helsinki Institute of Physics, Gustaf Hällströmin katu 2, FI-00014 University of Helsinki, Finland*
[57]*NOVA Optical Infrared Instrumentation Group at ASTRON, Oude Hoogeveensedijk 4, NL-7991 PD Dwingeloo, the Netherlands*
[58]*Argelander-Institut für Astronomie, Universität Bonn, Auf dem Hügel 71, D-53121 Bonn, Germany*
[59]*Dipartimento di Fisica e Astronomia 'Augusto Righi' - Alma Mater Studiorum Università di Bologna, via Piero Gobetti 93/2, I-40129 Bologna, Italy*
[60]*Department of Physics, Institute for Computational Cosmology, Durham University, South Road, DH1 3LE, UK*
[61]*Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Boulevard de l'Observatoire, CS 34229, F-06304 Nice cedex 4, France*
[62]*Institute of Physics, Laboratory of Astrophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Observatoire de Sauverny, CH-1290 Versoix, Switzerland*
[63]*European Space Agency/ESTEC, Keplerlaan 1, NL-2201 AZ Noordwijk, The Netherlands*
[64]*Department of Physics and Astronomy, University of Aarhus, Ny Munkegade 120, DK-8000 Aarhus C, Denmark*
[65]*Space Science Data Center, Italian Space Agency, via del Politecnico snc, I-00133 Roma, Italy*
[66]*Centre National d'Etudes Spatiales, F-31400 Toulouse, France*
[67]*Institute of Space Science, Bucharest, Ro-077125, Romania*
[68]*Université Paris Cité, CNRS, Astroparticule et Cosmologie, F-75013 Paris, France*
[69]*Departamento de Física, FCFM, Universidad de Chile, Blanco Encalada 2008, Santiago, Chile*
[70]*Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Avenida Complutense 40, E-28040 Madrid, Spain*
[71]*Instituto de Astrofísica e Ciências do Espaço, Faculdade de Ciências, Universidade de Lisboa, Tapada da Ajuda, PT-1349-018 Lisboa, Portugal*
[72]*Departamento de Electrónica y Tecnología de Computadoras, Universidad Politécnica de Cartagena, E-30202 Cartagena, Spain*
[73]*Infrared Processing and Analysis Center, California Institute of Technology, Pasadena, CA 91125, USA*
[74]*Instituto de Astrofísica de Canarias, Calle Vía Láctea s/n, E-38204 San Cristóbal de La Laguna, Tenerife, Spain*
[75]*Institut d'Astrophysique de Paris, UMR 7095, CNRS, Sorbonne Université, 98 bis boulevard Arago, 75014 Paris, France*
[76]*AIM, CEA, CNRS, Université Paris-Saclay, Université de Paris, F-91191 Gif-sur-Yvette, France*
[77]*NASA Ames Research Center, Moffett Field, CA 94035, USA*
[78]*INFN-Bologna, Via Irnerio 46, I-40126 Bologna, Italy*
[79]*IFPU, Institute for Fundamental Physics of the Universe, via Beirut 2, I-34151 Trieste, Italy*
[80]*Dipartimento di Fisica e Scienze della Terra, Universitá degli Studi di Ferrara, Via Giuseppe Saragat 1, I-44122 Ferrara, Italy*
[81]*INAF, Istituto di Radioastronomia, Via Piero Gobetti 101, I-40129 Bologna, Italy*
[82]*Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, F-31400 Toulouse, France*
[83]*Institute for Theoretical Particle Physics and Cosmology (TTK), RWTH Aachen University, D-52056 Aachen, Germany*
[84]*Department of Physics & Astronomy, University of California Irvine, Irvine, CA 92697, USA*
[85]*University of Lyon, UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, F-69100 Lyon, France*
[86]*Aix-Marseille Université, CNRS, CNES, LAM, F-13013 Marseille, France*
[87]*INFN-Sezione di Genova, Via Dodecaneso 33, I-16146, Genova, Italy*
[88]*INAF-Istituto di Astrofisica e Planetologia Spaziali, via del Fosso del Cavaliere, 100, I-00100 Roma, Italy*
[89]*School of Physics, HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK*
[90]*Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, E-28049 Madrid, Spain*
[91]*Department of Physics, University of Helsinki, P.O. Box 64, FI-00014 University of Helsinki, Finland*
[92]*Astronomical Institute (AIRUB), Faculty of Physics and Astronomy, Ruhr University Bochum, German Centre for Cosmological Lensing (GCCL), D-44780 Bochum, Germany*
[93]*Department of Physics, Lancaster University, Lancaster LA1 4YB, UK*
[94]*Instituto de Astrofísica de Canarias (IAC), Departamento de Astrofísica, Universidad de La Laguna (ULL), E-38200 La Laguna, Tenerife, Spain*
[95]*Université de Paris, LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Université, F-75013/4 Paris, France*

[96]*Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK*
[97]*Astrophysics Group, Blackett Laboratory, Imperial College London, London SW7 2AZ, UK*
[98]*Université Grenoble Alpes, CNRS, Grenoble INP, LPSC-IN2P3, 53, Avenue des Martyrs, F-38000 Grenoble, France*
[99]*Centre de Calcul de l'IN2P3, 21 avenue Pierre de Coubertin, F-69627 Villeurbanne Cedex, France*
[100]*School of Engineering, University of Applied Sciences and Arts of Northwestern Switzerland, CH-5210 Windisch, Switzerland*
[101]*Dipartimento di Fisica - Sezione di Astronomia, Universitá di Trieste, Via Tiepolo 11, I-34131 Trieste, Italy*
[102]*INFN, Sezione di Trieste, Via Valerio 2, I-34127 TS Trieste, Italy*
[103]*Department of Mathematics and Physics E. De Giorgi, University of Salento, Via per Arnesano, CP-I93, I-73100 Lecce, Italy*
[104]*INFN – Sezione di Lecce, Via per Arnesano, CP-193, I-73100 Lecce, Italy*
[105]*Institute for Computational Science, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland*
[106]*Higgs Centre for Theoretical Physics, School of Physics and Astronomy, The University of Edinburgh, Edinburgh EH9 3FD, UK*
[107]*Institut d'Astrophysique de Paris, 98bis Boulevard Arago, F-75014, Paris, France*
[108]*Institut für Theoretische Physik, University of Heidelberg, Philosophenweg 16, D-69120 Heidelberg, Germany*
[109]*Faculty of Science, Université St Joseph, Beirut, Lebanon*
[110]*Department of Astrophysical Sciences, Peyton Hall, Princeton University, Princeton, NJ 08544, USA*
[111]*Helsinki Institute of Physics, University of Helsinki, Gustaf Hällströmin katu 2, FI-00560 Helsinki, Finland*
[112]*SISSA, International School for Advanced Studies, Via Bonomea 265, I-34136 Trieste TS, Italy*

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