Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **540,** 128–142 (2025) https://doi.org/10.1093/mnras/staf690
Advance Access publication 2025 April 28

Efficient search for extremely metal-poor galaxies in the local universe using convolutional neural networks

Ting-Yun Cheng ★ and Ryan J. Cooke

Centre for Extragalactic Astronomy, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

Accepted 2025 April 25. Received 2025 April 25; in original form 2025 February 1

ABSTRACT

Nearby extremely metal-poor galaxies (XMPs) allow us to study primitive galaxy formation and evolution in greater detail than is possible at high redshift. This work promotes the use of convolutional neural networks (CNNs) to efficiently search for XMPs in multiband imaging data based on their predicted N2 index (N2 $\equiv \log\{[\mathrm{N\,II}]\,\lambda6585/\mathrm{H}\,\alpha\}$). We developed a sequential characterization pipeline, composed of three CNN procedures: (i) a classifier for metal-poor galaxies, (ii) a classifier for XMPs, and (iii) an N2 predictor. The pipeline is applied to over 7.7 million Sloan Digital Sky Survey (SDSS) DR17 imaging data without SDSS spectroscopy. The predicted N2 values are used to select promising candidates for observations. This approach was validated by new observations of 45 candidates with redshifts less than 0.065 using the 2.54 m Isaac Newton Telescope and the 4.1 m Southern Astrophysical Research Telescope between 2023 and 2024. All 45 candidates are confirmed to be metal poor, including 28 new discoveries. There are 18/45 galaxies lacking detectable [N II] λ6585 lines (S/N < 2); for these, we report 2σ upper limits on their oxygen abundance. Our XMPs have estimated oxygen abundances of $7.1 \leq 12 + \log(\mathrm{O/H}) \leq 8.7$ (2σ upper limit), based on the N2 index, and 21 of them with estimated metallicity $< 0.1\,\mathrm{Z}_\odot$. Additionally, we identified 4 potential candidates of low-metallicity AGNs at $\lesssim 0.1\,\mathrm{Z}_\odot$. Finally, we found that our observed samples are mostly brighter in the g band compared to other filters, similar to blueberry galaxies, resembling green pea galaxies and high-redshift Ly α emitters.

Key words: methods: data analysis – galaxies: abundances – galaxies: dwarf.

1 INTRODUCTION

Extremely metal-poor galaxies (hereafter XMP) are commonly defined to have a gas-phase metallicity ten times lower than the Sun (Kunth & Östlin 2000). Some of the well-known examples include I Zwicky 18 (Sargent & Searle 1970) and SBS 0335–052 (Izotov et al. 1997), each having a metallicity of $\simeq 1/30\,Z_\odot$. Due to selection effects, the XMPs are mostly star-forming dwarf galaxies such as blue compact dwarf galaxies, that are characterized by prominent hydrogen emission lines. They tend to be less massive (10^6–$10^8\,\mathrm{M}_\odot$) but contain massive stars, harbour near-pristine gas, and appear to be at the early stage of galaxy evolution. These characteristics are analogous to some of the primeval galaxies which were formed in a primordial gas environment during the early stages of cosmic history and provide an excellent laboratory in the local Universe for the studies of galaxy evolution and the formation of massive stars (Bromm et al. 2009; Bromm &Yoshida 2011; Wise et al. 2012; Fukushima et al. 2024).

Additionally, XMPs have been used to study big bang nucleosynthesis such as determining the primordial ^{4}He abundance (Fukugita & Kawasaki 2006; Izotov, Thuan & Guseva 2014; Peimbert, Peimbert & Luridiana 2016; Fernández et al. 2019; Hsyu et al. 2020; Aver et al. 2021; Matsumoto et al. 2022), because they have not experienced much chemical evolution. The relative abundances of the light elements (such as H, He, and Li) that were formed shortly after the big bang allow us to study the properties of the early Universe, as well as search for possible extensions to the standard model (Steigman 2007).

Given the great interest and wide applications of XMPs, there have been many attempts to search for new candidates. With great effort, the number of XMPs has been increased from 31 samples listed in the review of Kunth & Östlin (2000) to a few hundreds of XMPs reported in literature (e.g. van Zee 2000; Thuan & Izotov 2005; Izotov et al. 2006, 2009; Guseva et al. 2007; Izotov & Thuan 2007, 2009; Pustilnik et al. 2010; Izotov, Thuan & Guseva 2012; Skillman et al. 2013; Hirschauer et al. 2016; Guseva et al. 2017; Hsyu et al. 2017; James et al. 2017; Yang et al. 2017; Hsyu et al. 2018; Ruiz-Escobedo et al. 2018; Kojima et al. 2020; Nakajima et al. 2022; Nishigaki et al. 2023). The modern searches of XMP samples include the following approaches: (1) search for low redshift H I 21 cm emission associated with blue optical colours (e.g. Skillman et al. 2013; Hirschauer et al. 2016; Karachentsev et al. 2023); (2) human inspection and colour selection of multiband imaging to identify galaxies that look similar to known XMP (e.g Hsyu et al. 2018; Grossi et al. 2025); (3) trawling survey spectra to identify galaxies with weak metal emission lines (e.g Guseva et al. 2017; Zou et al. 2024); and (4) applying machine learning algorithms to photometric properties to create a list of XMP candidates, combined with follow-up longslit spectroscopy (Kojima et al. 2020). Without spectroscopy, the identification process of (1) and (2) is generally based on colour–colour selections. The selection criteria usually also require human inspection. This not only leads

★ E-mail: tycheng.sunny@gmail.com

to a prejudiced decision boundary, but one also cannot easily assess more than three dimensions of colour–colour diagrams with human inspection alone.

Approach (4), by Kojima et al. (2020), is the first attempt with a machine learning approach using a neural network classifier that can provide numerical decision boundaries in a multidimensional space. Their goal is to separate XMPs from stars, QSOs, and other galaxies using the photometric magnitudes in different bands such as the Hyper Suprime-Cam (HSC) $griz$ bands and the Sloan Digital Sky Survey (SDSS) $ugriz$ bands. The photometric magnitudes are generated from the spectral energy distribution (SED) models of star, QSOs, XMPs, and non-XMPs for the HSC and SDSS, respectively. Hence, their training XMP samples are a set of emulated magnitudes from the SED models covering the physical properties of typical known XMPs. The samples for validation are selected at $z < 0.03$ with $12 + \log(\mathrm{O/H}) < 7.69$ ($\sim$0.1 $\mathrm{Z}_{\odot}$) from literatures. Their classifier accomplishes a completeness of 86 per cent and a purity of 46 per cent XMP classifications. This indicates that their classifier could possibly misidentify over a half of the samples as XMPs, given only the photometric magnitudes in different bands.

In this work, we promote three main improvements with our deep learning (DL) pipeline composed of three individual models, made of convolutional neural networks (CNN). During the visual inspection process, the focus typically lies in identifying key characteristics, such as compact, ball-like structures and notably bluer colours. Therefore, by employing CNN, we directly work with multiband imaging data, which provide pixel-wise visual analysis and yields more comprehensive information, including both morphology and colour of the target objects. In addition to providing more information, the morphological and colour features contained in images are directly extracted by the CNNs without human interference. Our approach could be more inclusive towards different types of XMPs than applying a specific SED model. Secondly, due to the scarcity of the XMPs in the local Universe, it can be challenging to look for this tiny needle from a haystack of galaxies. Our DL pipeline performs a sequential classification procedure to simplify the task for each CNN model that helps to purify the XMP classification. Finally, in addition to classification, the pipeline predicts a proxy of metallicity that can be used to select the most promising XMP candidates.

Our paper is outlined as follows: The data sets used in this work are described in Section 2, while the details of our methodology are introduced in Section 3. With the DL prediction, the selection of XMP candidates and the observations of these candidates are described in Section 4. The analysis and discussion of the new observations are carried out in Section 5. We summarize the validation of the proposed methodology and the analyses of our new observations in Section 6.

2 DATA SET

All DL approaches rely on reliable training data set and labels in order to perform their task. In this section, we describe the data set and labels that we have adopted to train our algorithm. For the labels, we used the N2 index ($\mathrm{N2} \equiv \log\{[\mathrm{N\,II}]\,\lambda6585/\mathrm{H}\,\alpha\}$) as the proxy for metallicity, probed by the oxygen abundance (Storchi-Bergmann, Calzetti & Kinney 1994; Raimann et al. 2000; Denicoló, Terlevich & Terlevich 2002; Pettini & Pagel 2004; Yin et al. 2007), to select XMP candidates. The N2 index is a well-established strong line diagnostic that exhibits a monotonic relationship with oxygen abundance, making it a valuable tool for metallicity estimation. For practical reasons, the [N II] λ6585 and H α lines are located close to each other, making the N2 index much less sensitive to dust extinction and flux calibration errors. Additionally, the H α line is typically very strong, offering an advantage in deriving more robust measurements even under poor observing conditions. Nevertheless, in the future, we plan to accommodate other lines such as [O III], [O II], and H β lines to improve the pipeline.

The photometric data (including five bands, u, g, r, i, z) of the SDSS Data Release 17 (DR17; Abdurro'uf et al. 2022) were used as the input images to the DL pipeline. As mentioned in Section 1, one of the key characteristics of XMPs is their notably bluer visual colours, primarily driven by the continuua of the residing OB stars. Thus, colour information plays a critical role in this search. To preserve colour information in the CNN models, we converted the fluxes (pixel values) of galaxy cutouts in each band into the relative fluxes to the ones of the r-band image by

$$f_{jk,\mathrm{norm}} = \frac{f_{jk} - n_k}{f_{r,\mathrm{max}} - n_r}, \tag{1}$$

where the subscript j represents the pixel index from 0 to $N-1$ (with N being the number of pixels) and the subscript k indicates five filters: u, g, r, i, and z. The numerator measures the intrinsic flux by subtracting the raw flux of each pixel in each band (f_{jk}) by the background level of each band (n_k) which is determined by the average flux of the most common pixel values in a galaxy cutout of the filter. To determine the most common value, we construct a histogram of the pixel fluxes, and set the background level to be the peak of the distribution. The denominator represents the difference between the maximum value of flux in r band ($f_{r,\mathrm{max}}$) and the background level measured in r band (n_r). The scaled flux of each pixel in each band, $f_{jk,\mathrm{norm}}$, therefore provides an analogy of 'colour' per pixel for each band relative to the r band.

2.1 Known samples: training the CNN models

With the SDSS spectroscopic observation, we select samples that not only have a spectroscopic measurement of the N2 index but also have appropriate imaging coverage across all five SDSS bands. The N2 index is provided by the MPA-JHU measurements – `galSpec` – using SDSS DR12 data[1] (Kauffmann et al. 2003a; Brinchmann et al. 2004; Tremonti et al. 2004). This gives the N2 index for 180 369 galaxies. Additionally, we collect other known low-metallicity samples and their N2 values from the series of the EMPRESS project (Kojima et al. 2020; Isobe et al. 2022; Xu et al. 2022; Nakajima et al. 2022) and the following literature: Zee (2000); Thuan & Izotov (2005), Izotov et al. (2006, 2009), Guseva et al. (2007), Izotov & Thuan (2007, 2009), Pustilnik et al. (2010), Skillman et al. (2013), Hirschauer et al. (2016), Guseva et al. (2017), James et al. (2017), Hsyu et al. (2018), Ruiz-Escobedo et al. (2018). This provides the N2 index of an additional 108 nearby galaxies at redshift $\lessapprox 0.05$.

In this work, we define a metal-poor galaxy (hereafter MP) as a galaxy with $\mathrm{N2} \leq -1.0$, and an XMP is defined to have $\mathrm{N2} \leq -1.5$ [approximately $12 + \log(\mathrm{O/H}) \lessapprox 8.0$; estimated using equation 9 in Yin et al. (2007, hereafter Y07)]. With the collection of the SDSS spectroscopic samples and additional MP samples from the literature, the initial training sample contains 180 477 galaxies in total with 5097 MP ($\mathrm{N2} \leq -1.0$; $\sim$2.82 per cent of total samples) and 384 XMP ($\mathrm{N2} \leq -1.5$; $\sim$0.2 per cent of total samples). The redshift distributions of the initial training sample is $\lessapprox$0.1 with the average value of $\langle z \rangle \sim 0.05$.

[1] https://www.sdss4.org/dr17/spectro/galaxy_mpajhu/

2.2 Training samples and labels

With the scarce number of XMPs, the prediction of the N2 index could be driven by the majority of samples with higher metallicities. Hence, we carry out a sequential approach with multiple CNN models focusing on different tasks: (i) classifying MP candidates from the total sample; (ii) classifying XMP candidates from the predicted MP samples; and (iii) predicting the N2 index from the predicted XMP samples (see details in Section 3.1). To perform these different tasks, the corresponding training samples and labels are different, as follows:

(i) MP classifier: applied to all samples

  (a) MP: N2 ≤ -1.0
  (b) non-MP: N2 > -1.0

(ii) XMP classifier: applied to the subset with N2 ≤ -0.5

  (a) XMP: N2 ≤ -1.5
  (b) non-XMP: $-0.5 \geq$ N2 > -1.5

(iii) N2 predictor: applied to the MP samples

Amongst these samples, we randomly select 1000 testing data, following the distribution of the total samples (Section 2.1). This contains approximately 28 MPs including two XMPs. These testing data are removed from the training procedures. To reduce the impact of data selection on training a CNN model, we also create three different training and testing sets for each procedure [(i), (ii), and (iii)].

Finally, to avoid any training bias caused by the number differences between the target outputs in each task (examined in Cheng et al. 2020), we balance the number of samples between the target outputs by rotating galaxy cutouts. Each galaxy is rotated by 90, 180, and 270 degrees due to the fixed cutout frame. This provides only three times more extra data in training. To further increase our training data set, an additional negligible Gaussian noise, generated with a dispersion equal to 1percnt of the standard deviation of the pixel values in a cutout, is then added to each cutout after rotation. With these tiny perturbations in inputs, the models are further improved even if using the repeated rotated images (check the discussion of relevant approaches in Goodfellow, Shlens & Szegedy 2014). Note that in this work the utilisation of these additive noises is not to change the visual appearance of the cutout nor help regularisation of the model training, but just provide a nominal difference in pixel values. This step of data augmentation ensures that our training is performed on a balanced data set.

For the classifiers at the procedure (i) and (ii), the balancing is carried out between two target classes. For the N2 predictor at the procedure (iii), the target output is a floating value. Hence, we augment the data across several bins of the N2 index to ensure that each bin has an equal number of samples. We divide the range between -1.0 and -2.1 into 11 bins with an interval of 0.1. For the samples with N2 ≤ -2.1, we form one bin due to the scarce population in this range. There are 2064 galaxies in the first bin of $(-1.0, -1.1)$, and the last bin of $(-2.1,)$ contains 15 galaxies. The number of data in each bin is augmented to equal the number of data in the first bin.

2.3 Working samples: SDSS images without spectroscopy

In this subsection, we describe the SDSS imaging sample that we use, in combination with our trained algorithm, to discover new XMP candidates that do not currently have spectroscopic confirmation from SDSS. The query for the SDSS DR17 is based on the physical properties such as colours and brightness of the known MP samples in the initial training samples. We applied query criteria as shown in Table 1 when retrieving SDSS DR17 imaging data. We select sources classified as galaxies by the SDSS pipeline, without applying any additional SDSS flags. The query criteria covers greater than 99 per cent of the MP samples in our data set. Note that the colour criteria was applied to prevent wasting computational resources on unlikely samples, rather than determining the final list of candidates. We therefore allowed a broader coverage in these criteria, such that only the upper limits of the colour distributions were used, to avoid excluding desired samples that may have physical properties that differ from the currently known MPs.[2] Additionally, as CNN models are capable of extrapolating beyond the training distributions (Cheng et al. 2021, 2023), we considered galaxies that are up to 2 mag fainter in r band than the training set. The limit of $\mathrm{mag}_r \leq 22$ is chosen as it becomes challenging to observe any candidates with a continuum fainter than this limit using 4-m telescopes (see Section 4). Additionally, we further exclude samples that are missing one or more of the five filters or are positioned near the image edge resulting in incomplete cutouts. By excluding data with SDSS spectroscopic observation and our samples, the number of working samples is 7763 821.

Table 1. The selection criteria for SDSS DR17 data query.

$16 \leq \mathrm{mag}_r \leq 22$		$(u-g) \leq 1.7$	$(u-r) \leq 2.1$
$(u-i) \leq 2.2$	$(u-z) \leq 2.5$	$(g-r) \leq 0.6$	$(g-i) \leq 0.9$
$(g-z) \leq 1.2$	$(r-i) \leq 0.7$	$(r-z) \leq 0.9$	$(i-z) \leq 1.0$

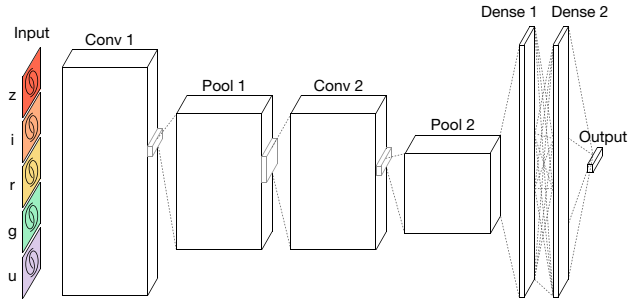

Figure 1. Schematic diagram of the CNN architecture used in this work. The input is a galaxy image of five different filters (u, g, r, i, z). The 'Conv 1' and 'Conv 2' represent convolutional layers, and each layer is followed by a pooling layer (Pool 1 and Pool 2), respectively. Finally, two dense layers (Dense 1 and Dense 2) are used before the output layer.

3 DEEP LEARNING APPROACH

We use multiband imaging data as the input of a DL pipeline to carry out sequential predictions of classifying XMP candidates and estimating their N2 index. Due to the scarce population of XMPs, the DL pipeline is composed of three CNN algorithms: (i) MP classifier; (ii) XMP classifier; and (iii) N2 predictor. The architecture of each CNN model is the same (see Fig. 1)for simplicity. The input cutouts have a dimension of 32 by 32 pixels and contain five bands (u, g, r, i, z). The architecture contains two convolutional layers (Conv 1 and

[2]We note that our final XMP candidates are all well-inside our selection box. This suggests our selection box has not significantly impacted the selection of XMP candidates.

Table 2. The hyper-parameters used for each CNN model. The 'conv_1' and 'kernel_1' are the channel and kernel size for the first convolutional layer (Conv 1), and the 'conv_2' and 'kernel_2' are for the second convolutional layer (Conv 2). The 'neuron_1' and 'neuron_2' are the number of neurons used in the dense layers (Dense 1 and Dense 2), respectively.

	Learning rate	l2	Dropout	conv_1	conv_2	kernel_1	kernel_2	neuron_1	neuron_2
MP classifier	0.0001	0.0	0.5	16	256	3	7	64	156
XMP classifier	0.0001	0.0	0.0	256	256	3	3	512	16
N2 predictor	0.0004	0.0	0.0	128	128	7	3	128	256

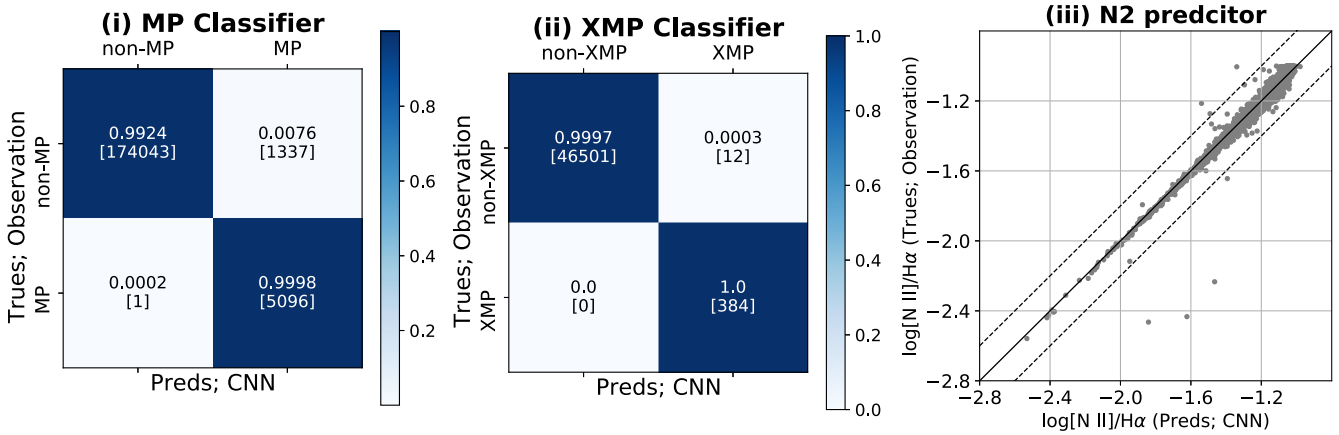

Figure 2. Each panel presents the result applying the nine trained CNN models to the specific data set (including training and testing samples) for each procedure (Section 2.2). The left and middle panels are the confusion matrices of the MP and XMP classifiers. The classification probability thresholds for assigning classes are >0.5. The value in each quadrant indicates the fraction (number) of the samples predicted by CNN in each true (observed) class. The right panel shows the comparison between the predicted N2 index by CNN models and the observed N2 values from literature (listed in Section 2.1). The solid line shows a one-to-one relation and the dashed lines indicate a scatter of 0.2 dex.

Conv 2) followed by a pooling layer (Pool 1 and Pool 2) for each convolutional layer, as well as two dense layers (Dense 1 and Dense 2). Two dropout layers were implemented after the second pooling layer and before the output layer, to reduce the number of parameters inside the networks. The dropout rate is one of the hyper-parameters and a constant for both dropout layers.

The hyper-parameters for each CNN model, however, are optimized individually for each model using the Bayesian optimization method (Frazier 2018) due to the use of different data sets (see Section 2.2) in training different models. Table 2 shows the hyper-parameters used for the MP classifier, the XMP classifier, and the N2 predictor, respectively. We applied the `Adam` optimizer (Kingma & Ba 2015), and the learning rates are also hyper-parameters optimized independently for each model. The maximum number of iterations for each training is 20 epochs, but only the model with the minimum validation loss within the 20 epochs is saved.

3.1 Training XMP classification and N2 prediction

As mentioned in Section 2.2, we created three different training and testing sets for each procedure, (i), (ii), (iii) to account for the impact of the quality of randomly selected data sets. Furthermore, we train three independent CNN models for each procedure to account for the variation caused by having a random initial state when training a new model. Therefore, there are 3×3, i.e. 9 CNN models trained for each procedure, and each model is assessed by its corresponding testing set.

In this section, we describe the details of how each algorithm is trained and evaluated independently. For classifiers, the median values of the output probabilities from the nine CNN models, trained for each classifier, are used to assign classes. Similarly, for the N2 predictor, the median value of the predicted N2 indices from the nine models is used for the selection of the candidates. The validation of each procedure is carried out separately using their whole assigned samples (including training and testing sets) for each procedure, as stated in Section 2.2.

In detail, the MP classifier is trained with all available samples (excluding their testing sets) separated into two classes: MP (N2 ≤ -1.0) and non-MP (N2 > -1.0). The numbers of MP and non-MP samples in the training set are equal after data augmentation for training. The median value of the predicted probabilities from the nine models is used to assign classes. The trained models are applied to all samples in this work including 180 477 galaxies, and the result is shown at the left panel of the Fig. 2. This confusion matrix of the MP classifier uses a threshold of 0.5 applied to the median predicted probabilities. The classification accuracy is about 99.26 per cent, which indicates the fraction between the number of correctly classified samples and the total number of samples. The trained MP classifier correctly identifies over 99.98 per cent MP galaxies (i.e. the recall is 99.98 per cent). The false positive, which the model identifies as a MP but with N2 > -1.0, occupies about 20.78 per cent of the predicted MP samples, containing mostly (over a fraction of 0.9686) galaxies with N2 ≤ -0.5.

With the high fraction of false positives with N2 ≤ -0.5, the XMP classifier is trained on the subset of samples with N2 ≤ -0.5. The negative label, 'non-XMP', is therefore for the galaxies with N2 between -0.5 and -1.5 (see also, Section 2.2). The numbers of XMP and non-XMP training samples are balanced with data augmentation for training. Again, the median value of the predicted probabilities is used to assign classes, and the middle panel of Fig. 2 shows the

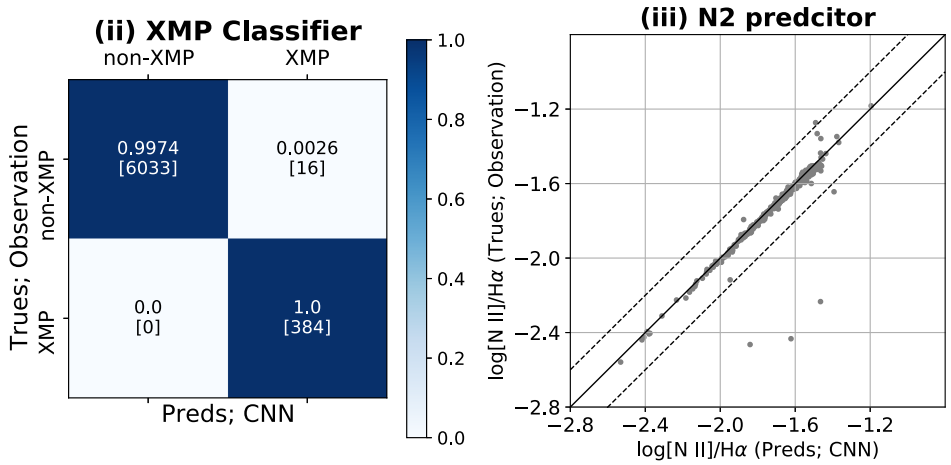

Figure 3. Unlike Fig. 2, this figure presents the evaluation of the sequential process. The left panel shows the confusion matrix of the XMP classifier for the predicted MP candidates from Fig. 2. The classification probability threshold is >0.5. The value in each quadrant indicates the fraction (number) of the samples predicted by CNN in each true (observed) class. The right panel shows the comparison between the predicted N2 index of the predicted XMP candidates by CNN models and their observed N2 values from literature. The solid line shows a one-to-one relation and the dashed lines indicate a scatter of 0.2 dex.

result applying the trained models of the XMP classifier to all subset samples with N2 ≤ -0.5. The XMP classifier also reaches a high classification accuracy of 99.97 per cent. There are around 3 per cent of false positives, which the model identifies as a XMP but with N2 > -1.5. In this test, all of the false positives are in fact MP galaxies with N2 ≤ -1.0.

We therefore anticipate that the vast majority of the classified XMPs after the sequential classification of two classifiers shall satisfy the definition of MP. The N2 predictor is trained with only the MP samples with N2 ≤ -1.0 to focus on effectively predicting the N2 index at the lowest range. As stated in Section 2.2, we separate the MP samples into 12 bins based on their N2 index. The first 11 bins have an interval of 0.1, and the last bin covers the remaining samples with a broader range of N2 values, $(-2.6, -2.1)$, where -2.6 is the lowest N2 value in our sample. We augment the number of data in each bin to match the number of data in the first bin, $(-1.1, -1.0]$. Since the primary goal of this work is to identify XMP candidates with the lowest possible N2 index, we introduce a loss weighting factor ($3\times$) for the systems that have a true N2 index ≤ -2.1. This ensures that the network is more severely penalized when it incorrectly predicts the N2 index of the most metal-poor XMPs.

The right panel of Fig. 2 shows the comparison of all MP samples (5097 galaxies) between the predicted N2 index by CNN and the observed values collected from the literature (Section 2.1). The prediction of the N2 index is accurate with the root-mean-squared deviation (RMSD) of 0.031 dex and the median absolute deviation (MAD) of 0.015 dex.

3.2 Evaluation of the sequential process

When a new set of galaxy cutouts is fed to the sequential process, only the predicted MP candidates with $P_{\rm MP} > 0.5$ from the MP classifier proceed to the XMP classifier; similarly, only the predicted XMP candidates with $P_{\rm XMP} > 0.5$ from the XMP classifier advance to the N2 predictor. Therefore, the sequential process constructs a list of XMP candidates with their MP predicted probability ($P_{\rm MP} > 0.5$), XMP predicted probability ($P_{\rm XMP} > 0.5$), and the predicted N2 index. The output N2 index is used to select the most promising candidate for observation, which we discuss further in Section 4.1.

The assessment of the sequential process with all available samples (i.e. 180 477 galaxies) is shown in Fig. 3. Unlike Fig. 2, only the predicted MP candidates in the left panel of Fig. 2 (i.e. 6433 samples) proceed to the XMP classifier (the left panel of Fig. 3), and only the predicted XMP candidates (containing 400 samples) continue to the N2 predictor (the right panel of Fig. 3). With the two classifiers carrying out a sequential classification, the fraction of 0.96 and 0.99 of predicted XMP candidates are indeed XMP galaxies and MP galaxies, respectively (i.e. the precision is 96 per cent and 99 per cent). The predicted N2 index has the RMSD of 0.13 dex and a MAD of 0.012 dex. With such small statistics, the RMSD is skewed towards the outliers; while the MAD, which is less affected by outliers, is consistent with the individual test in Fig. 2.

4 OBSERVATIONS AND DATA REDUCTION

4.1 Sample selection

The trained CNN models are applied to the working samples (Section 2.3) – SDSS DR17 multiband imaging without SDSS spectroscopy. This gives 232 954 XMP candidates with $P_{\rm MP} > 0.5$ and $P_{\rm XMP} > 0.5$ and the predictions of their N2 values from over 7 million SDSS galaxy cutouts.

We select only the most promising XMP candidates with the lowest range of metallicities for observation by applying $P_{\rm MP} > 0.99$, $P_{\rm XMP} > 0.99$, and N2 < -1.8. This contains approximately 550 candidates. We then performed a fast visual inspection to exclude apparent artefacts and faulty images, leaving 390 highly possible XMP candidates with lowest ranges of N2 index (< -1.8) predicted by the DL pipeline. A few subtle artefacts remain among the samples, requiring further assessment before they can be selected as final observational candidates.

4.2 Observations

To validate the effectiveness of the DL pipeline as well as discover new XMP galaxies, we conducted observational programmes using the 2.54 m Isaac Newton Telescope (INT) and the 4.1 m Southern Astrophysical Research (SOAR) Telescope between 2023 and 2024. Table 3 lists 45 targets for which we acquired spectra as part of

Table 3. The list of observed XMP samples from INT and SOAR telescope. The order is sorted based on the predicted N2 values by our CNN algorithms. The 'ObsDate' is the date that the observations were conducted. The 'ExpTime' provides the exposure time used for a single exposure, and the 'ExpN' is the number of exposures for an object. The 'ref.' column provides references to XMP galaxies that have been investigated in the literature.

Name[a]	RA (J2000.0)	DEC (J2000.0)	N2 (CNN)	Instrument	ObsDate	ExpTime (s)	ExpN	ref.[b]
XMP0013+1354	00 13 42.62	+13 54 17.1	−1.92±0.07	INT/IDS	21-Aug-2023	1000	3	–
XMP0124+0838	01 24 06.58	+08 38 06.9	−2.17±0.10	SOAR/Goodman	8-Nov-2023	800	3	–
XMP0219−0059	02 19 30.34	−00 59 14.3	−1.81±0.04	SOAR/Goodman	6-Oct-2023	800	3	1,7
XMP0429+0026	04 29 51.38	+00 26 52.0	−1.88±0.03	INT/IDS	4-Mar-2024	1200	3	–
XMP0742+1103	07 42 18.07	+11 03 30.6	−1.86±0.04	INT/IDS	5-Mar-2024	1800	3	–
XMP0752+2340	07 52 11.06	+23 40 16.7	−2.00±0.09	INT/IDS	3-Mar-2024	1000	3	–
XMP0801+2640	08 01 03.92	+26 40 54.3	−1.89±0.05	INT/IDS	5-Mar-2024	600	3	2
XMP0803+1635	08 03 16.04	+16 35 44.6	−1.81±0.06	INT/IDS	3-Mar-2024	1500	3	–
XMP0827+1059	08 27 46.65	+10 59 11.1	−1.97±0.06	INT/IDS	6-Mar-2024	1500	3	12
XMP0850+1150	08 50 57.57	+11 50 45.6	−2.06±0.07	INT/IDS	28-Feb-2024	1000	3	–
XMP0856+2414	08 56 01.07	+24 14 21.5	−1.88±0.04	INT/IDS	6-Mar-2024	1400	3	6
XMP0916+5002	09 16 06.66	+50 02 30.6	−2.11±0.07	INT/IDS	5-Mar-2024	1800	3	–
XMP0916+0257	09 16 25.09	+02 57 43.2	−2.13±0.02	INT/IDS	4-Mar-2024	1800	3	–
XMP0922+6324	09 22 23.86	+63 24 36.9	−2.12±0.06	INT/IDS	4-Mar-2024	2000	3	12
XMP0928+3601	09 28 44.73	+36 01 04.2	−1.91±0.08	INT/IDS	3-Mar-2024	1200	3	4
XMP0930+4934	09 30 04.97	+49 34 29.7	−1.95±0.06	INT/IDS	28-Feb-2024	1000	3	–
XMP0931+2617	09 31 14.14	+26 17 27.4	−1.98±0.15	INT/IDS	6-Mar-2024	1800	3	–
XMP1003+2746	10 03 10.80	+27 46 31.7	−1.92±0.13	INT/IDS	6-Mar-2024	1600	3	–
XMP1030+3151	10 30 44.81	+31 51 24.0	−2.04±0.09	INT/IDS	3-Mar-2024	1100	3	–
XMP1032+5035	10 32 00.39	+50 35 07.7	−2.08±0.08	INT/IDS	5-Mar-2024	1000	3	–
XMP1035+3814	10 35 07.20	+38 14 30.4	−1.89±0.02	INT/IDS	28-Feb-2024	1500	3	6
XMP1139+0040	11 39 00.41	+00 40 42.6	−2.02±0.07	INT/IDS	4-Mar-2024	1500	3	12
XMP1140+5037	11 40 45.72	+50 37 07.6	−1.95±0.06	INT/IDS	5-Mar-2024	1400	3	–
XMP1214+1245	12 14 33.11	+12 45 49.2	−2.18±0.02	INT/IDS	3-Mar-2024	1100	3	3
XMP1228+4313*	12 28 48.09	+43 13 48.9	−2.00±0.04	INT/IDS	6-Mar-2024	600	3	–
XMP1230+0544	12 30 11.99	+05 44 50.7	−1.96±0.02	INT/IDS	6-Mar-2024	1000	3	–
XMP1238+3246*	12 38 40.25	+32 46 00.9	−2.02±0.02	INT/IDS	28-Feb-2024	2200	3	10
XMP1322+2251	13 22 01.75	+22 51 31.5	−2.03±0.03	INT/IDS	5-Mar-2024	1800	2	–
XMP1329+2237	13 29 24.31	+22 37 12.3	−2.00±0.05	INT/IDS	28-Feb-2024	1000	3	–
XMP1344+0621	13 44 57.48	+06 21 46.3	−1.96±0.02	INT/IDS	6-Mar-2024	1000	3	–
XMP1347+0755	13 47 56.00	+07 55 32.1	−2.31±0.07	INT/IDS	4-Mar-2024	1500	3	12
XMP1408+1753	14 08 16.16	+17 53 50.9	−1.86±0.04	INT/IDS	3-Mar-2024	1000	3	–
XMP1422+5414	14 22 38.85	+54 14 09.2	−1.81±0.07	INT/IDS	3-Mar-2024	600	3	8
XMP1631+4426	16 31 14.25	+44 26 04.7	−2.42±0.02	INT/IDS	4-Mar-2024	1800	3	5,9
XMP1638+2421	16 38 58.01	+24 21 39.2	−2.07±0.08	INT/IDS	5-Mar-2024	1200	3	–
XMP1655+6337	16 55 39.20	+63 37 03.3	−2.10±0.02	INT/IDS	6-Mar-2024	1500	1	3
XMP2048−0559	20 48 34.22	−05 59 01.4	−2.07±0.08	INT/IDS	21-Aug-2023	2300	1	–
XMP2136−0307	21 36 09.38	−03 07 30.7	−1.86±0.05	SOAR/Goodman	6-Oct-2023	800	3	–
XMP2149−0535	21 49 12.62	−05 35 05.6	−2.16±0.08	INT/IDS	21-Aug-2023	3000	1	–
XMP2156+0856	21 56 33.58	+08 56 36.6	−1.86±0.06	INT/IDS	20-Aug-2023	800	2	6
XMP2212+2205	22 12 59.31	+22 05 05.5	−1.89±0.03	INT/IDS	21-Aug-2023	1000	3	–
XMP2325+2008	23 25 37.75	+20 08 17.2	−1.91±0.14	INT/IDS	22-Aug-2023	1000	3	–
XMP2329+0226	23 29 26.60	+02 26 28.1	−1.85±0.05	INT/IDS	21-Aug-2023	1000	3	11
XMP2331+2226	23 31 00.91	+22 26 34.8	−1.81±0.04	INT/IDS	21-Aug-2023	2000	3	–
XMP2336−0404	23 36 31.64	−04 04 36.6	−1.96±0.08	SOAR/Goodman	8-Nov-2023	800	3	–

[a] A '∗' indicates a nearby H II region ($z < 0.002$).

[b] Reference: (1) Ann, Seo & Ha (2015), (2) Griffith et al. (2011), (3) Hsyu et al. (2018), (4) James et al. (2017), (5) Kojima et al. (2020), (6) Liu et al. (2023), (7) Micheva et al. (2013), (8) Thuan, Izotov & Lipovetsky (1995), (9) Thuan, Guseva & Izotov (2022), (10) van Zee & Haynes (2006) (HII region), (11) Wang et al. (2018), (12) Yang et al. (2017).

these observations.[3] Although some targets have been reported in the literature, they either lack N2 measurements or only have values with upper limits or large uncertainties. The order of this table is sorted according to their right ascension. If a galaxy has previously been observed in the literature, we also list its corresponding references.

[3] Due to time allocations and weather conditions, we were unable to observe all candidates during this run. A follow-up observation request has been submitted for the remaining targets.

The details of each of these observing programmes are provided in the following subsections.

4.2.1 INT observations

We collected 41 new optical spectra using the Intermediate Dispersion Spectrograph (IDS) with the `Red + 2` detector and `R632V` grating on the INT during 2023 August 19–22 (2023B) and 2024 February 28–March 6 (2024A). The `Red + 2` detector provides a

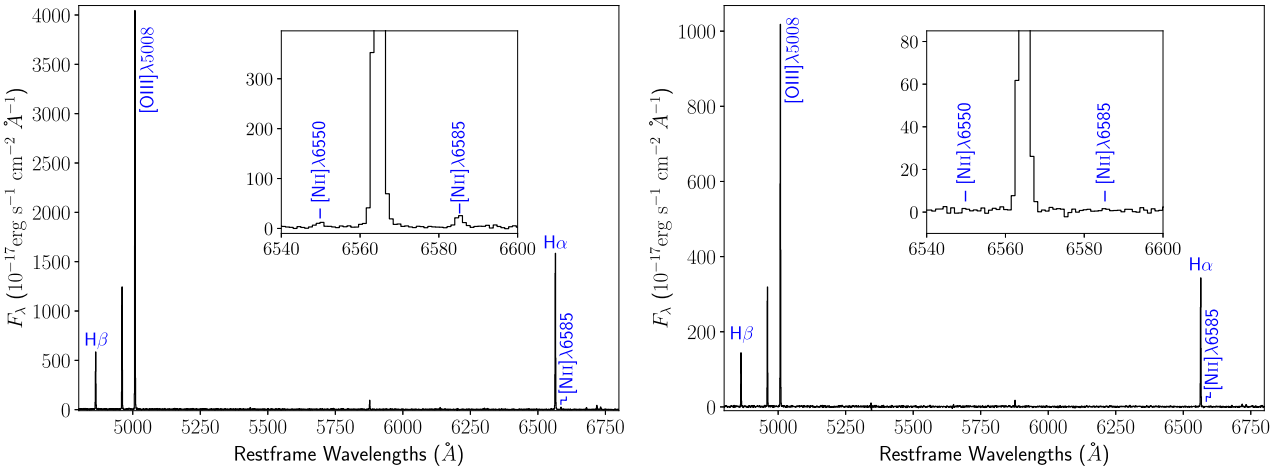

Figure 4. Examples of the reduced spectra with (XMP0801+2640: left; $>10\sigma$) and without (XMP1347+0755: right; $<2\sigma$) clear detection of the [N II] λ6585 lines.

spatial scale of 0.44 arcsec per pixel. The detector binning was set to be 1 × 1 (i.e. unbinned). The `R632V` grating has a total wavelength coverage of 2178 Å and the spectral resolution of $R \sim 2270$ at 4500 Å, with a nominal sampling of ∼1 Å/pixel. This grating was chosen to resolve the weak [N II] line from the much stronger H α line, as well as its better efficiency (>60 per cent) within the target range of wavelengths. All observations were made using a 1 arcsec slit and the slit angle is at the approximate parallactic angle during the observations. We use observations of G191-B2B[4] to flux calibrate the data taken in both observation periods.

The 2023B observation serves as a preliminary assessment for validating the DL pipeline. Since the redshifts of our selected samples were unknown, a central wavelength of 6460.5 Å was chosen to cover a wavelength range from 5370 to 7540 Å for good-quality observation of the [N II] λ6585 line at $z < 0.15$. Spectroscopic observation of eight XMP candidates were made, and their redshifts were all less than 0.055. This observation provides an important validation of the DL pipeline. For the 2024A observation, we adjust the central wavelength to 5940 Å with the prior knowledge from previous observation to ensure coverage on [O III] doublet at λλ4960, 5008 Å, H β, [N II] doublet at λλ6550, 6585 Å, and H α lines. We observed 33 XMP candidates total, and their redshifts are all less than 0.065.

4.2.2 SOAR observations

We collect four spectra using the Goodman High Throughput Spectrograph (Goodman; Clemens, Crain & Anderson 2004) with the `SOAR_GHTS_BLUECAM` camera on the SOAR telescope on 2023 October 6 and 2023 November 8. The spatial scale is 0.15 arcsec per pixel, and the binning for the detector is set to be 2×2. We use the `SYZY_400` grating, which has a spectral resolution of $R \sim 850$ at 5500 Å (assuming a 1 arcsec slit). This setup provides optimal throughout for the wavelength range covering the [O III] doublet at λλ4960, 5008 Å, H β, [N II] doublet at λλ6550, 6585 Å, and H α lines. The slit size is set to 1 arcsec and the slit angle is at the approximate parallactic angle during the observations. The standard star for flux calibration is LTT 3864.[5] Three observations of 800 s each are made for each object.

[4]https://www.eso.org/sci/observing/tools/standards/spectra/g191b2b.html
[5]https://www.eso.org/sci/observing/tools/standards/spectra/ltt3864.html

4.3 Data reduction

The data reduction is carried out using the PYPEIT data reduction pipeline (Prochaska et al. 2020a, 2020b). The reduction process includes the subtraction of bias frames, the correction of the flat-field using dome flats, identification and masking of cosmic rays, sky subtraction, wavelength calibration using arc frames, 1D boxcar extraction and flux calibration using the chosen standard stars. When available, three exposures of a single candidate are combined using the PYPEIT coaddition tools. The number of exposures for each target is also listed in Table 3. Examples of the reduced and combined spectra with clear (left) and unclear (right) detection of [N II] λ6585 lines are shown in Fig. 4.

5 ANALYSIS AND DISCUSSION

Using data from INT and SOAR, we have collected 45 spectroscopically confirmed XMPs, including 28 new discoveries. Most of these samples exhibit typical characteristics of XMPs, such as bluer colours, and compact, ball-like structures. However, some of our observed samples appear to display diffuse, irregular or tidal structures. This suggests a diversity of structural morphologies among dwarf galaxies, even within the lower metallicity regime.

In this study, our primary focus is the N2 index for validating the CNN pipeline and deriving oxygen abundances via strong line diagnostics. This work provides 29 first spectra and 36 new N2 measurements. Where possible, we also measured the [O III] λ5008/H β flux ratio (i.e. the O3 index; $\mathrm{O3} \equiv \log\{[\mathrm{O\,III}]\,\lambda5008/\mathrm{H}\,\beta\}$), to assess the existence of AGN activity in our samples. The O3 index will be used to improve our pipeline in future work. However, note that eight XMPs observed during INT 2023B do not have the necessary coverage for O3 index measurement.

5.1 The fluxes and ratios of emission lines

The fluxes of emission lines, [N II] λ6585 and H α, are initially fit with Gaussian models using a χ^2 minimization Absorption LIne Software (ALIS; see more details in Cooke et al. 2014). We use this fitting method to assess whether there is a clear detection of [N II] λ6585. If the [N II] λ6585 emission line is confidently detected (S/N ≥ 2), we then measure the integrated fluxes of [N II] λ6585 and H α lines by summing the pixel values associated with the target line. Conversely,

Table 4. The measurements of H α emission line flux, the N2 index, and the O3 index. The missing values are indicated with three dots. For samples with significance $\geq$2, we provide the N2 values measured with the integrated flux approach; for other samples, we provide the 2σ values.

Name	Redshift	F (H α)[a]	Significance[b]	N2	O3
XMP0013+1354	0.0524	2921±48	2.41	−1.99±0.21	–
XMP0124+0838	0.0487	912.7±4.4	9.47	−2.01±0.05	0.828±0.005
XMP0219−0059	0.0085	1947.6±6.6	17.28	−2.24±0.04	0.716±0.004
XMP0429+0026	0.0119	519.1±8.0	3.88	−2.04±0.18	0.801±0.015
XMP0742+1103	0.0438	136.5±4.3	1.84	<−1.50	0.726±0.026
XMP0752+2340	0.0474	1048±11	5.37	−1.93±0.09	0.910±0.007
XMP0801+2640	0.0265	3963±27	12.98	−1.82±0.04	0.805±0.007
XMP0803+1635	0.0211	440.4±6.0	2.71	−2.13±0.19	0.842±0.013
XMP0827+1059	0.0436	327.4±7.6	2.27	−2.48±0.89	0.723±0.022
XMP0850+1150	0.0293	620.9±8.9	2.37	−1.68±0.09	0.820±0.014
XMP0856+2414	0.0511	295.9±8.3	2.81	−1.72±0.25	0.897±0.020
XMP0916+5002	0.0497	206.2±4.7	2.48	−2.12±0.41	0.918±0.018
XMP0916+0257	0.0385	421.2±6.0	1.75	<−2.03	0.786±0.011
XMP0922+6324	0.0395	457.0±5.7	2.63	−2.00±0.16	0.728±0.014
XMP0928+3601	0.0312	244.1±6.1	0.26	<−1.89	0.890±0.018
XMP0930+4934	0.0247	507.0±8.5	3.70	−1.79±0.12	0.895±0.014
XMP0931+2617	0.0638	164.1±5.5	−1.60	<−1.70	0.804±0.033
XMP1003+2746	0.0398	219.1±5.9	0.27	<−1.73	0.783±0.031
XMP1030+3151	0.0436	1617±12	4.74	−2.41±0.13	0.782±0.007
XMP1032+5035	0.0318	123.9±6.2	1.03	<−1.29	0.948±0.035
XMP1035+3814	0.0254	553.1±6.9	3.67	−1.94±0.14	0.900±0.014
XMP1139+0040	0.0418	281.5±5.5	0.61	<−1.87	0.784±0.019
XMP1140+5037	0.0278	270.8±5.7	2.07	−1.93±0.25	0.812±0.020
XMP1214+1245	0.0192	154.3±6.1	1.00	<−1.59	0.314±0.031
XMP1228+4313	0.0017	2001±21	9.32	−1.86±0.08	0.649±0.013
XMP1230+0544	0.0397	421.8±9.6	1.58	<−1.69	0.869±0.019
XMP1238+3246	0.0011	360.4±5.6	1.56	<−1.94	0.165±0.019
XMP1322+2251	0.0373	131.6±4.5	0.14	<−1.55	0.820±0.029
XMP1329+2237	0.0247	2246±16	11.65	−1.82±0.04	0.836±0.007
XMP1344+0621	0.0229	661±11	2.69	−2.26±0.29	0.799±0.018
XMP1347+0755	0.0438	848.5±8.0	0.65	<−2.36	0.844±0.009
XMP1408+1753	0.0238	5116±23	13.13	−2.26±0.04	0.823±0.004
XMP1422+5414	0.0212	2904±22	4.27	−2.52±0.18	0.844±0.007
XMP1631+4426	0.0313	133.5±4.0	1.43	<−1.54	0.264±0.031
XMP1638+2421	0.0344	630.2±8.0	1.10	<−2.08	0.928±0.013
XMP1655+6337	0.0211	522±12	0.92	<−1.81	0.409±0.023
XMP2048−0559	0.0480	719±28	1.15	<−1.46	–
XMP2136−0307	0.0536	1736.6±5.8	17.26	−2.01±0.03	0.779±0.003
XMP2149−0535	0.0542	237±20	0.02	<−1.26	–
XMP2156+0856	0.0118	3506±58	2.76	−1.89±0.16	–
XMP2212+2205	0.0288	3320±50	3.95	−1.90±0.16	–
XMP2325+2008	0.0391	1992±41	2.46	−1.87±0.22	–
XMP2329+0226	0.0293	2327±45	2.39	−2.11±0.33	–
XMP2331+2226	0.0231	609±20	1.98	<−1.41	–
XMP2336−0404	0.0303	1190.6±5.1	4.67	−2.22±0.06	0.703±0.004

[a]The integrated flux of the H α line. The unit is 10^{-17} erg s^{-1} cm^{-2}.
[b]The significance of the N2 measurements. The [N II] λ6585 and H α lines are fitted with Gaussian profiles using χ^2 minimization, where the signal represents the amplitude ratio of the fits and the noise denotes the uncertainty.

if the [N II] λ6585 line is not confidently detected (S/N $<$ 2), we use the 2σ value from the Gaussian profile fitting to provide an upper limit on the N2 index.

When calculating the integrated fluxes of emission lines, we fit the continua of observed spectra with a quartic polynomial function. Since the H α and [N II] λ6585 lines may blend together, we first integrate the flux across both lines (as a measurement of the H α+[N II] λ6585 flux. We then refit the curve between the H α and [N II] λ6585 lines with quartic polynomial function to obtain an accurate measurement of the [N II] λ6585 line flux alone. By subtracting the two fluxes, we obtain the H α flux measurement for calculating the N2 index. When calculating the O3 index, the fluxes of the [O III] λ5008 and H β lines are integrated separately. The fitting results are compiled in Table 4.

For the samples with detectable [N II] λ6585 line (S/N $\geq$ 2), the measurements using Gaussian fitting and integrated fluxes are consistent with each other, with a MAE of $\sim$0.11 dex.

From the literature, there are overlapping N2 measurements for nine objects. If the literature reported the flux ratio between [N II] λ6585 + [N II] λ6550 and H α lines, we converted their quoted ([N II] λ6550 + [N II] λ6585)/H α ratio to [N II] λ6585Ha line by multiplying their line ratio by a constant of 2.96/3.96 (i.e. we

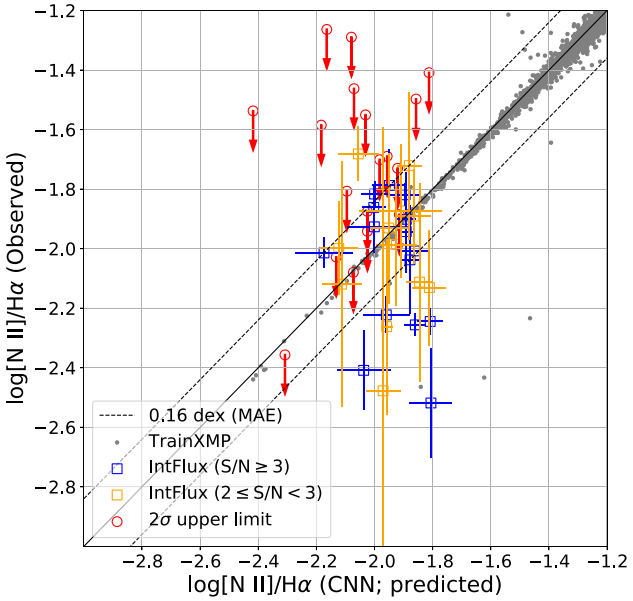

Figure 5. Comparison of the N2 index between CNN predictions and the observed values. The grey dots show the values of our training samples. The dashed lines indicate the MAE measured using the samples marked as squares. The blue and orange squares represent the values measured using integrated flux with good (S/N $\geq$ 3) and fair (2 $\leq$ S/N $<$ 3) detection of [N II] λ6585 line, respectively. The red circles are 2σ upper limits, due to the lack of a detectable [N II] λ6585 line.

assume the [N II] λ6585/[N II] λ6550 = 2.96; Tachiev & Froese Fischer 2001). The comparison between literature values and our measurements ($S/N \geq 2$) are consistent with each other to within a MAE of 0.086 dex.

5.2 The N2 index: CNN versus observation

Fig. 5 shows the comparison of the N2 index between our CNN predictions and our observational measurements. The MAE of the values with detectable [N II] λ6585 line (S/N $\geq$ 2; squares in Fig. 5) is around 0.16 dex. We use this value to distinguish outliers. The outliers above the upper dashed line mostly lack robust detection of [N II] λ6585 lines. On the other hand, we found that a subset of samples tends to have lower N2 values than those predicted by the CNN model, falling below the bottom dashed line (hereafter, lower outliers). We do not find any evident differences in the visual morphology and colour distributions between the lower outliers and non-outliers. This indicates that our CNN pipeline behaves correctly by assigning the values corresponding to their visual appearances. However, these galaxies somehow have lower N2 values than their visual appearances suggest. This also means that some of our new observations behave differently to the majority of our training sets. Interestingly, we also identify three similar outliers within the training set itself (see grey dots in Fig. 5), for which our CNN exhibits similar predictive behaviour. While we can artificially balance the data across different ranges of N2 values, the diversity of XMPs present in the training sets is fixed. Therefore, we propose that our XMP sample shows some differences to the majority of the training set.

If this difference has a physical origin, there are couple possible causes: (1) these galaxies may have lost their nitrogen gas during their evolutionary history without altering their colour or morphology; (2) their nitrogen gas may have a distinct origin compared to most of the galaxies in our training set (Chiappini, Romano & Matteucci 2003; Pilyugin, Thuan & Vílchez 2003; Roy et al. 2021); or (3) these galaxies might have high-ionization parameters, leading to enhanced [N III] and [N IV] lines rather than [N II]. The first reason would require substantial outflows from these galaxies or gas stripping effects from the environments in which they reside. Investigating this possibility requires more extensive and deeper spectroscopic observations to detect any features indicating the flows, as well as understanding their environments. As for the second possibility, one could examine the N/O and O/H relationship. Secondary nitrogen production in stars occurs via the CNO cycle, catalysed by the carbon that was already present in the interstellar medium before the star was born (ISM; e.g. Meynet & Maeder 2002). In contrast, primary production happens when the carbon catalyst is derived directly from the helium-burning core rather than from the ISM (Marigo 2001). If primary production dominates, the N/O ratio should be independent of O/H, whereas a correlation between N/O and O/H would indicate dominant secondary production. Currently, we cannot check this hypothesis because the [O II] $\lambda\lambda$3727, 3730 doublet, which is necessary to determine the N/O ratio, is not covered by our observations. Additionally, this doublet can be used to probe the ionization parameter, helping to address the third proposed possible cause. These possibilities will be investigated in the future with our follow-up observations.

5.3 Derived physical properties

We have derived several physical properties of the XMP galaxies observed with our programme, including the oxygen abundance, star formation rate (SFR), and stellar mass; these values are summarized in Table 5.

5.3.1 Oxygen abundance

The oxygen abundance is estimated using the observed N2 values together with the empirical relation from Y07:

$$12 + \log(\mathrm{O/H}) = 9.263 + 0.836 \times N2. \quad (2)$$

The relation is based on a linear least-squares fit to the data collected from SDSS and various literature sources, providing a better description of the data compared to the empirical relation of Pettini & Pagel (2004). The uncertainty of this linear fit to the data used in Y07 is 0.159 dex. If a galaxy's [N II] λ6585 line is not detected, we report a 2σ upper limit on the oxygen abundance. The estimated oxygen abundance of our samples ranges between 7.1 $\leq 12 + \log(\mathrm{O/H}) \leq$ 8.7 (2σ upper limit). We have 21 samples with estimated oxygen abundances of $12 + \log(\mathrm{O/H}) \leq 7.7$ ($\sim$0.1 $Z_\odot$), and 18 samples are reported with 2σ upper limits. Planned follow-up observations of these targets will firmly pin down the chemistry of these near-pristine galaxies. The distribution of oxygen abundance estimates is shown in the leftmost panel of Fig. 6.

5.3.2 Distance, Hα luminosity, and star formation rate

For distance-derived properties, such as luminosity and SFR, we use the luminosity distance to estimate these quantities. The luminosity distance (D_L) is calculated using spectroscopic redshifts from our observations in combination with ASTROPY's cosmology package (Astropy Collaboration 2022). We assume a flat ΛCDM cosmology with $H_0 = 67.4 \pm 0.5$ km s^{-1}Mpc^{-1} and $\Omega_m = 0.315 \pm 0.007$

Table 5. Derived physical properties of the XMP galaxies reported in this work. Oxygen abundances are derived by the N2 index (Y07). For samples with significance S/N < 2, we report a 2σ upper limit on the oxygen abundance. Calculations of luminosity distance ($D_{\rm L}$), Hα luminosities (L (Hα)), SFR, and stellar mass (M_*) are described in Section 5.3.

Name	$12+\log$(O/H)	$D_{\rm L}$ (Mpc)	$D_{\rm L}$ (corrected) (Mpc)	L (Hα) ($\times10^{39}$ erg s^{-1})	SFR ($\times10^{-3}$M$_\odot$ yr^{-1})	M_* ($\times10^{6}$ M$_\odot$)
XMP0013+1354	7.60±0.24	241.99	245.2±1.8	210.2±4.7	922±20	7100±1600
XMP0124+0838	7.58±0.16	224.39	228.3±1.7	56.91±0.89	249.8±3.9	1510±350
XMP0219−0059	7.39±0.16	38.26	40.66±0.32	3.853±0.063	16.91±0.27	0.57±0.13
XMP0429+0026	7.56±0.22	53.37	51.93±0.62	1.675±0.048	7.35±0.21	1.19±0.28
XMP0742+1103	<8.47	201.47	200.7±1.5	6.58±0.23	28.9±1.0	198±46
XMP0752+2340	7.65±0.17	218.53	221.5±1.7	61.5±1.1	270.0±4.9	2990±690
XMP0801+2640	7.74±0.16	120.11	123.36±0.94	72.2±1.2	316.7±5.2	120±28
XMP0803+1635	7.48±0.23	95.48	100.8±1.6	5.36±0.19	23.51±0.83	39.6±9.2
XMP0827+1059	7.19±0.76	200.26	203.7±1.5	16.25±0.45	71.3±2.0	500±110
XMP0850+1150	7.86±0.18	133.03	133.2±1.1	13.17±0.29	57.8±1.3	86±20
XMP0856+2414	7.82±0.26	235.91	237.2±1.8	19.91±0.63	87.4±2.8	1170±270
XMP0916+5002	7.49±0.38	229.22	232.4±1.7	13.32±0.36	58.5±1.6	403±93
XMP0916+0257	<8.07	176.27	176.5±1.7	15.70±0.37	68.9±1.6	157±36
XMP0922+6324	7.59±0.21	180.99	180.4±1.4	17.79±0.35	78.1±1.5	125±29
XMP0928+3601	<8.17	142.17	145.2±1.2	6.16±0.18	27.03±0.81	25.4±5.9
XMP0930+4934	7.77±0.19	111.97	106.41±0.89	6.87±0.16	30.15±0.71	21.1±4.9
XMP0931+2617	<8.28	297.41	297.2±3.5	17.34±0.71	76.1±3.1	1500±350
XMP1003+2746	<8.27	182.32	187.2±1.4	9.19±0.29	40.3±1.3	164±38
XMP1030+3151	7.25±0.19	200.11	203.5±1.5	80.2±1.4	351.8±6.0	840±190
XMP1032+5035	<8.65	145.04	151.9±1.2	3.42±0.18	15.02±0.79	324±75
XMP1035+3814	7.64±0.20	115.08	118.27±0.88	9.26±0.18	40.63±0.79	30.8±7.1
XMP1139+0040	<8.16	191.86	186.5±1.4	11.72±0.29	51.4±1.3	301±69
XMP1140+5037	7.65±0.26	126.21	122.38±0.98	4.85±0.13	21.30±0.56	78±18
XMP1214+1245	<8.44	86.82	88.07±0.68	1.432±0.061	6.28±0.27	70±16
XMP1228+4313	7.71±0.17	7.69	4.87±0.13	0.0568±0.0030	0.249±0.013	0.0251±0.0059
XMP1230+0544	<8.33	181.72	183.4±1.4	16.98±0.46	74.5±2.0	325±75
XMP1238+3246	<8.19	4.87	8.14±0.66	0.0286±0.0046	0.125±0.020	0.0196±0.0055
XMP1322+2251	<8.41	170.74	171.2±1.3	4.62±0.17	20.26±0.75	74±17
XMP1329+2237	7.75±0.16	111.80	113.95±0.87	34.89±0.59	153.1±2.6	151±35
XMP1344+0621	7.37±0.29	103.60	103.1±3.1	8.41±0.53	36.9±2.3	18.8±4.5
XMP1347+0755	<7.77	201.43	201.4±1.5	41.19±0.72	180.8±3.2	640±150
XMP1408+1753	7.38±0.16	107.82	109.86±0.97	73.9±1.3	324.3±5.9	14.2±3.3
XMP1422+5414	7.16±0.22	95.79	98.45±0.73	33.68±0.56	147.8±2.5	38.0±8.8
XMP1631+4426	<8.43	142.33	146.4±1.1	3.42±0.11	15.03±0.50	16.5±3.8
XMP1638+2421	<7.98	156.82	159.6±1.2	19.20±0.39	84.3±1.7	82±19
XMP1655+6337	<8.21	95.46	100.88±0.75	6.36±0.17	27.90±0.75	5.2±1.2
XMP2048−0559	<8.49	220.98	217.7±1.8	40.8±1.7	179.0±7.6	1190±270
XMP2136−0307	7.58±0.16	248.06	247.8±1.8	127.6±1.9	559.9±8.5	4010±930
XMP2149−0535	<8.64	251.08	253.0±1.9	18.1±1.5	79.6±6.7	1280±290
XMP2156+0856	7.68±0.21	52.83	51.94±0.39	11.31±0.25	49.7±1.1	3.47±0.80
XMP2212+2205	7.67±0.21	130.70	133.74±1.00	71.1±1.5	311.8±6.6	262±61
XMP2325+2008	7.70±0.25	179.25	174.6±1.4	72.7±1.9	318.9±8.3	1370±320
XMP2329+0226	7.50±0.32	133.18	134.65±0.99	50.5±1.2	221.6±5.4	369±85
XMP2331+2226	<8.55	104.64	97.8±1.2	6.97±0.28	30.6±1.2	26.5±6.1
XMP2336−0404	7.40±0.17	137.63	138.8±1.1	27.46±0.44	120.5±1.9	94±22

(Planck Collaboration VI 2020). The distance is corrected for peculiar velocity based on the results from Carrick et al. (2015). Both the original ($D_{\rm L}$) and corrected [$D_{\rm L}$ (corrected)] values are listed in Table 5. The quoted uncertainties include the uncertainties on the redshift, peculiar velocity, and cosmological parameters. The Hα luminosity, L (Hα), is then calculated by the following conversion using the corrected luminosity distances and the measured Hα fluxes:

$$L\,(\mathrm{H}\alpha) = F\,(\mathrm{H}\alpha)\,4\pi\,{D_{\rm L}}^2. \quad (3)$$

The SFR can be estimated using hydrogen emission lines such as Hα line. This line is produced by the recombination of ionized hydrogen in the H II regions. Thus, the Hα luminosity is linked with the number of ionizing photons and traces the formation of young (<20 Myr), massive (>10 M$_\odot$) stars. The following conversion between SFR and Hα luminosity is provided in Kennicutt (1998) derived by Kennicutt, Tamblyn & Congdon (1994) and Madau, Pozzetti & Dickinson (1998).

$$\mathrm{SFR} = 7.9\times10^{-42}\ \times (L\,(\mathrm{H}\alpha)\,/\mathrm{erg\ s^{-1}})\ \mathrm{M_\odot\ yr^{-1}}. \quad (4)$$

Note that this relationship assumes a Salpeter initial mass function (IMF; Salpeter 1955). We apply a correction factor of 1.8 to convert this relationship to a Chabrier (2003) IMF. The resulting values of $\log_{10}$ (SFR) are between −3.9 and −0.035, with the median value of -1.16 ± 0.60, which includes ∼69 per cent of the samples. The SFR distribution of our sample is shown in the middle panel of Fig. 6.

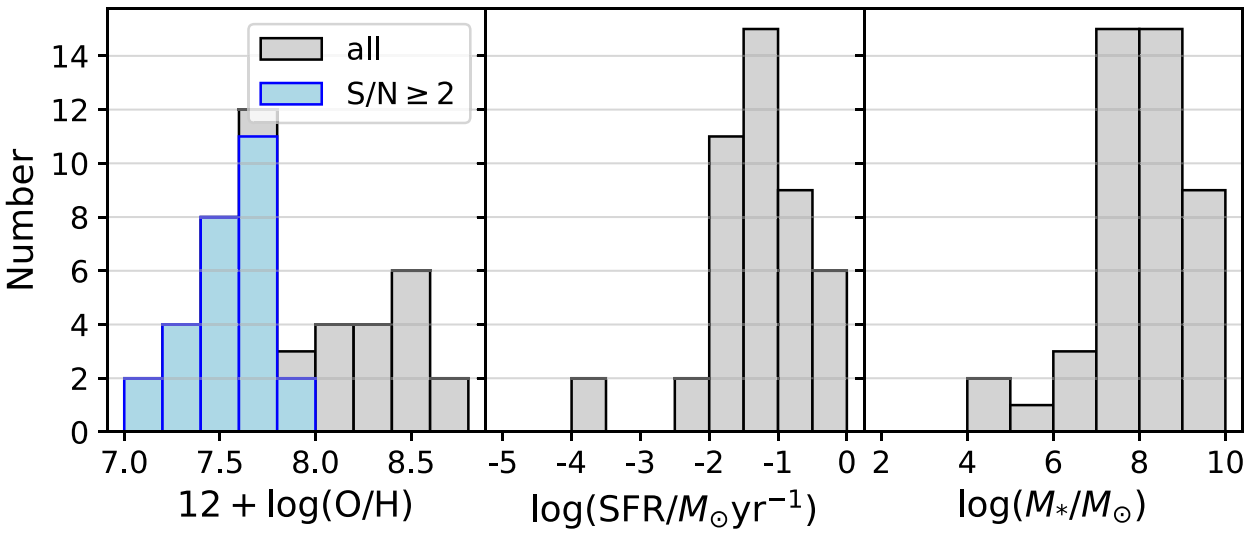

Figure 6. The distributions of oxygen abundance, SFR, and stellar mass of our XMP galaxy sample. The grey histograms show the values of all samples, while the blue histograms in the first panel represent the values of the samples with significantly detected [N II] λ6585 lines (S/N ≥ 2).

5.3.3 Stellar mass

The stellar mass is estimated using stellar mass-to-light ratio, obtained with the broad-band luminosity in SDSS i-band, in combination with the $r - i$ colour (Bell et al. 2003, B03). The conversion of the solar absolute magnitude to SDSS i-band filter is from Blanton & Roweis (2007). The filter choices are to reduce contamination from strong emission lines. However, we note that this kind of relationship is not robust for metal-poor dwarf galaxies. The results presented in this work just serve as an indicative measure of the stellar mass of the XMP galaxies in our sample. We adopt the following relationship:

$$\log_{10}\left(\frac{M_*}{L}\right) = 0.006 + 1.114 \times (r - i)\,, \tag{5}$$

where the value of M_*/L is expressed in solar units. Another relationship, using more complex stellar models, was introduced in Zibetti, Charlot & Rix (2009, Z09). To assess which of these aforementioned stellar mass models are more suitable for the XMP galaxies of our sample, we compared the stellar mass of Leo P estimated using both B03 and Z09 relations to the robust stellar mass measurement derived from *Hubble Space Telescope* imaging by McQuinn et al. (2015). Among the two, the B03 relation yielded a closer estimate to the robust value. Therefore, for stellar mass estimation in this work, we adopt equation (5) from B03, and divided the value by a factor of correct the 'diet' Salpeter IMF used in B03 to Chabrier (2003) IMF. The resulting values of $\log_{10}(M_*/\mathrm{M}_\odot)$ are between 4.3 and 9.8, with the median value of 8.1 ± 1.0 including about 71 per cent of the sample. The distribution is shown in the rightmost panel of Fig. 6.

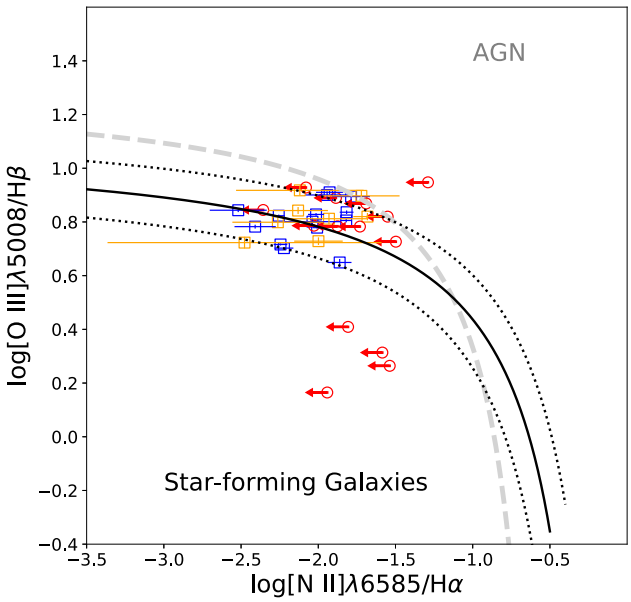

Figure 7. Diagnostic diagram of [N II] λ6585/H α versus [O III] λ5008/H β. The grey dashed line shows equation 1 from Kauffmann et al. (2003b), which delineates star-forming galaxies and AGN. The black solid line represents the mean of local star-forming sequences for SDSS galaxies analysed in Kewley et al. (2006), while the dotted black lines show the error of 0.1 dex from their models (Kewley et al. 2013). The symbols show our sample of XMP galaxies, with the same colour-coding as used in Fig. 5.

5.4 AGN activity

Fig. 7 shows the BPT diagram (Baldwin, Phillips & Terlevich 1981) – [N II] λ6585/H α versus [O III] λ5008/H β diagnostic diagram – for assessing AGN activity in our sample of XMP galaxies. The grey dashed line represents the relation provided in equation 1 of Kauffmann et al. (2003b) to separate the population of star-forming galaxies and AGN. This indicates that our XMP galaxies do not show any indication of AGN activity. The black solid line and dotted lines show the mean of SDSS star-forming galaxies at redshift $0.04 < z < 0.1$ (Kewley et al. 2006, 2013). This range contains 91 per cent of the SDSS star-forming galaxies from Kewley et al. (2006). Four of our objects are outliers to this trend with lower O3 values. This could be due to their low-metallicity nature. Groves, Heckman & Kauffmann (2006) studied the evolution of diagnostic diagrams for low-metallicity AGN (lower than 1 $\mathrm{Z}_\odot$). At metallicity of 0.1 $\mathrm{Z}_\odot$, their simulations predicted a decrease in the [O III] λ5008/H β ratio (lower than 0.5). This may explain the outliers' behaviour on the

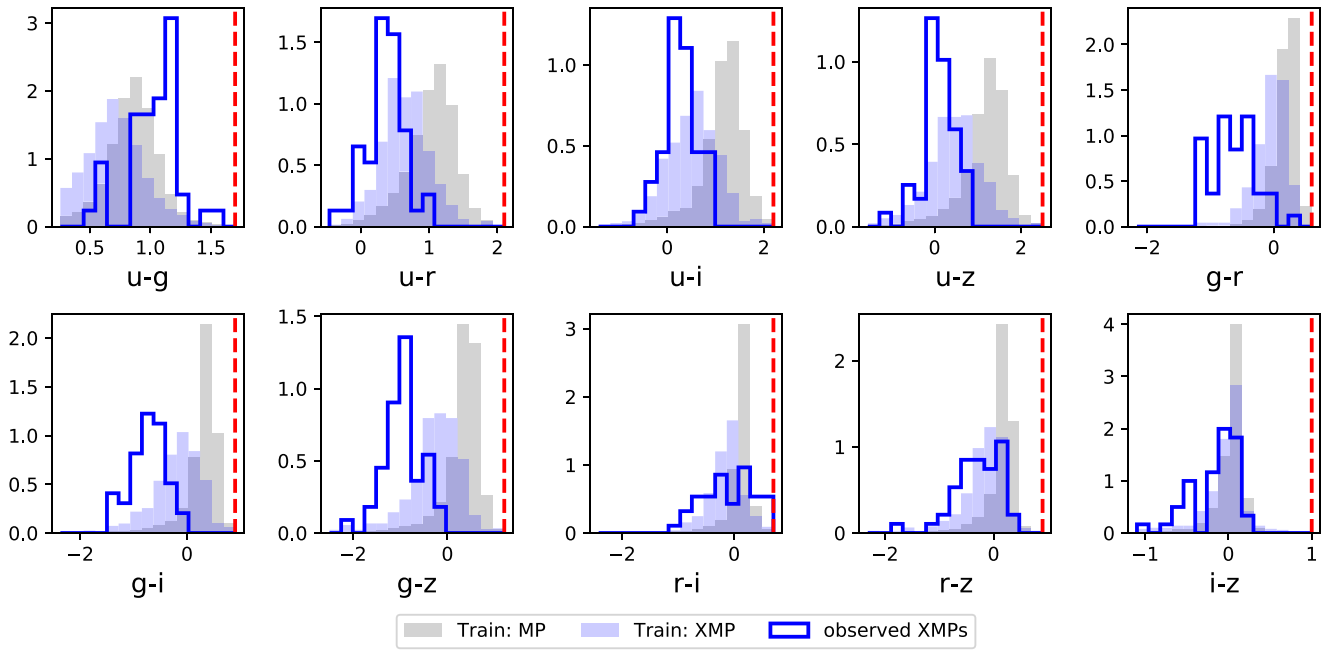

Figure 8. Colour distributions of training MP samples (grey shadings), training XMP samples (blue shadings) and the observed XMPs (unfilled blue histogram). The red dashed lines indicate the upper limit applied to each colour for querying the working samples (see Table 1 and the discussion in Section 2.3).

diagnostic diagram, indicating that they may contain low-metallicity AGN. To test this possibility, future observations will target the [O II], [Ne III], and [Ne V] emission lines.

5.5 Colour distribution

In Fig. 8, we compare the colour distributions of the observed galaxies and the training samples (MP and XMP). Note that the majority of the training XMP samples have an N2 value between −1.5 and −1.8 (>77 per cent). Thus, we expect there to be a slight difference in the colour distributions between the training XMP samples and our observed XMP galaxies with predicted N2 < −1.8.

However, a notable discrepancy exists in the colour distributions of those calculated particularly using the *g* band between the training samples and the observed XMPs. These samples tend to be brighter in *g* band compared to the majority of training XMP samples. The brighter *g*-band magnitudes are likely due to significant emission from the [O III] doublet, which dominates the flux in the SDSS *g* band given the redshifts of our observed XMPs. This characteristic is similar to that of blueberry (BB) galaxies (Yang et al. 2017, hereafter Y17), which are low-redshift counterparts of green pea galaxies (Cardamone et al. 2009) and high redshift Ly α emitting galaxies.

Fig. 9 shows the colour distributions of the observed XMP (blue contours), BB galaxies from Y17 (black dots), and the training XMP samples (grey dots) in this work. The red dashed lines in each panel represent the colour criteria used in Y17 to select green pea galaxies at $z \lessapprox 0.05$. This figure demonstrates that the observed XMPs are mostly greener[6] than the majority of the training samples. About 38 per cent (17 XMPs) of the observed objects fall into the colour regions of BB galaxies defined by Y17. We found that the N2 values of the training samples satisfying the colour criteria of BB galaxies are lower than those outside these criteria. Specifically, the median N2 value of BB-like training XMP samples (within the red-dashed lines in Fig. 9) is −2.02, while those outside these regions have a median N2 value of −1.64. This indicates that the phenomenon shown here are simply due to the blind selection of samples with lower N2 index.

5.6 Comparison with Blueberry Galaxies

Since BB galaxies are considered promising analogues to high redshift star-forming galaxies, we compare the derived physical properties of our XMPs with those of BB galaxies reported in Y17 in this section. In Fig. 10, we compare derived physical properties, such as stellar mass and SFR, as well as the measured O3 index, between our observed XMPs and the BB galaxies from Y17. We find a distinct difference in the physical properties between the two samples. In terms of stellar mass, Y17 estimates their values using Starburst99 models (Leitherer et al. 1999) based on photometric data, whereas our estimates rely on the photometric relation from B03, which is likely less reliable in the low metallicity regime (see Section 5.3). Given the differing estimation methods, the observed discrepancy in stellar mass between the two samples is therefore reasonable. In terms of SFRs, although Y17 estimates are based on the H β line while this work uses the H α line, both methods yield fairly robust estimates. Nonetheless, the BB galaxies from Y17 exhibit higher SFRs than our observed XMP samples, even among those that satisfy the colour selection criteria defined for BB galaxies in Y17 (blue dots/shadings). One of the key characteristics of BB galaxies is their high-ionization parameters, which may be a primary factor contributing to the observed differences in SFRs. However, in the right panel of Fig. 10, we do not observe a significant difference in the distributions of their O3 index. It is important to

[6]We describe our samples as 'greener' rather than 'bluer' because only *g* band appears brighter, while the *u* band does not exhibit a similar increase in brightness.

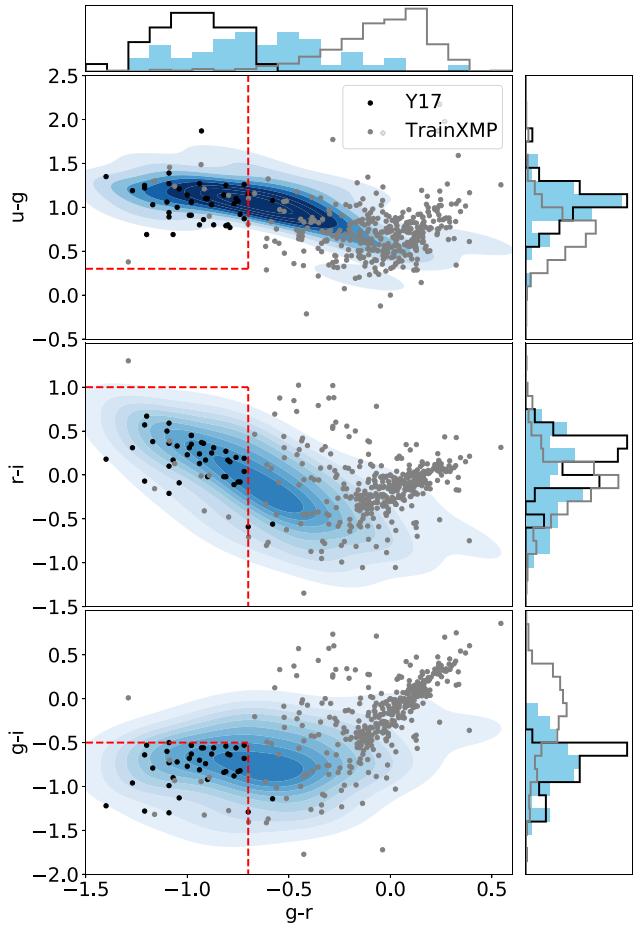

Figure 9. The comparison of the colour–colour diagram and histograms between the observed XMP (blue contours and histograms), samples of BB galaxies from Y17 (black dots and histogram), and training XMP samples (TrainXMP) from this work (grey dots and histogram). The red dashed lines are the colour criteria used in Y17 to select green pea galaxies at $z \lessapprox 0.05$.

note that a more reliable probe of ionization parameters requires the [O II] λλ3727, 3730 doublet, which is not covered by our current observations. As mentioned in Section 5, we will investigate this further with follow-up observations.

6 SUMMARY

To advance the discovery of new XMP galaxies, we promote the use of CNNs to accelerate the process of XMP identification and characterization for current and upcoming wide-area multiband sky surveys. A primary advantage of this approach is to efficiently consider both morphology and colour information simultaneously from broad-band images. Our DL pipeline is built from three individual CNN procedures: (i) MP classifier, (ii) XMP classifier, and (iii) N2 predictor, to conduct sequential classification and predictions of the N2 index ($\mathrm{N2} \equiv \log\{[\mathrm{N\,II}]\,\lambda6585/\mathrm{H}\,\alpha\}$) for MP galaxies. The N2 index is then used to select the most promising XMP galaxies. This design is to ensure an effective and efficient training and identification of the extremely sparse population of XMPs. Each CNN procedure contains nine CNN models (i.e. 3 different initializations × 3 training data sets) to account for the variation in each training run and the impact of quality of the selected training subsets. The median values of these nine CNNs are used for each classifier and the N2 predictor.

The trained DL pipeline is applied to the multiband imaging data without spectroscopy from the SDSS DR17. There are 232 954 XMP candidates selected with the criteria of $P_{\mathrm{MP}} > 0.5$ and $P_{\mathrm{XMP}} > 0.5$ from over 7 million SDSS galaxies. For observational candidates, we further select 390 promising candidates with $P_{\mathrm{MP}} > 0.99$, $P_{\mathrm{XMP}} > 0.99$, and $\mathrm{N2} < -1.8$. Among them, we successfully observed 45 XMP candidates with redshifts less than 0.065 using the 2.54 m INT and the 4.1 m SOAR Telescope between 2023 and 2024. All 45 observed XMP candidates are spectroscopically confirmed to be metal poor, including 28 new discoveries. Additionally, our observations provide the first N2 measurements for 36 XMPs. These N2 measurements are found to be consistent with the CNN predictions with a MAE of 0.16 dex. However, we found a set of samples with differences between predicted and observed N2 values greater than 0.16. These objects do not show distinct morphologies and colours from those within the MAE threshold, indicating that these objects somehow possess lower N2 values than their morphologies and colours suggest. These galaxies may have experienced significant outflows or gas stripping in their evolutionary history. Another hypothesis is that their nitrogen gas may have originated from a different route compared to the majority of our training XMPs. Or, they may have a high-ionization parameter, leading to enhanced ionized nitrogen gas, such as [N III] and [N IV]. Nevertheless, as this work aims to report a new methodology and the discovery of new XMP galaxies, our observations do not have sufficient wavelength coverage and depth to validate this hypothesis. We will address this with forthcoming observations.

The reported samples have estimated oxygen abundances of $7.1 \lesssim 12 + \log(\mathrm{O/H}) \lesssim 8.7$ (2σ upper limit), based on the N2 index. There are 21/45 galaxies with estimated oxygen abundances below 7.7, and 18/45 galaxies lack of detectable [N II] λ6585 lines ($\mathrm{S/N} < 2$). These samples offer an exciting opportunity to identify a record-breaking XMP, providing valuable insights into chemical abundances and evolutionary processes within galaxies in extremely metal-deficient environments. The SFRs of our XMPs are between $10^{-3.9}$–$10^{-0.035}$ $\mathrm{M_\odot\,yr^{-1}}$, and their stellar masses are in the range $10^{4.3}$–$10^{9.8}$ $\mathrm{M_\odot}$ based on the B03 calibration, with a correction to the Chabrier IMF. The BPT diagram of our XMPs shows that our objects are mostly star-forming galaxies without AGN activity. However, four XMPs without detectable [N II] λ6585 lines deviate from the typical trend of star-forming galaxies on the BPT diagram, suggesting that they may potentially be low-metallicity AGN at metallicity $\lesssim 0.1\,\mathrm{Z_\odot}$. Future observations that cover the emission lines [O II], [Ne III], and [Ne V] at wavelengths <4000 Å are required to test this possibility.

Finally, we examined the colour distributions of our observed galaxies, and found that they tend to be brighter in the SDSS g band than the training samples. This is likely to be caused by significant emission of the [O III] λλ4960, 5008 doublet that coincide with the bandpass of this filter. This characteristic is reminiscent of green pea galaxies and high-redshift Ly α emitters, but at lower redshifts ($z \lessapprox 0.05$), similar to BB galaxies. By applying the Y17 BB colour criteria, 38 per cent (17 XMPs) of the observed samples are categorized as BB galaxies. We found that our training samples, which share similar colour characteristics to the BB galaxies, tend to have lower N2 values. This leads to a skew of our observed samples, selected based on low predicted N2 values, towards the colour regions associated with BB galaxies. However, when comparing physical properties such as stellar mass and SFR, we found a discrepancy between our observed XMPs and the Y17 samples, even for those that satisfy the colour criteria. To resolve this discrepancy, observations covering additional emission lines, such as [O II], are required.

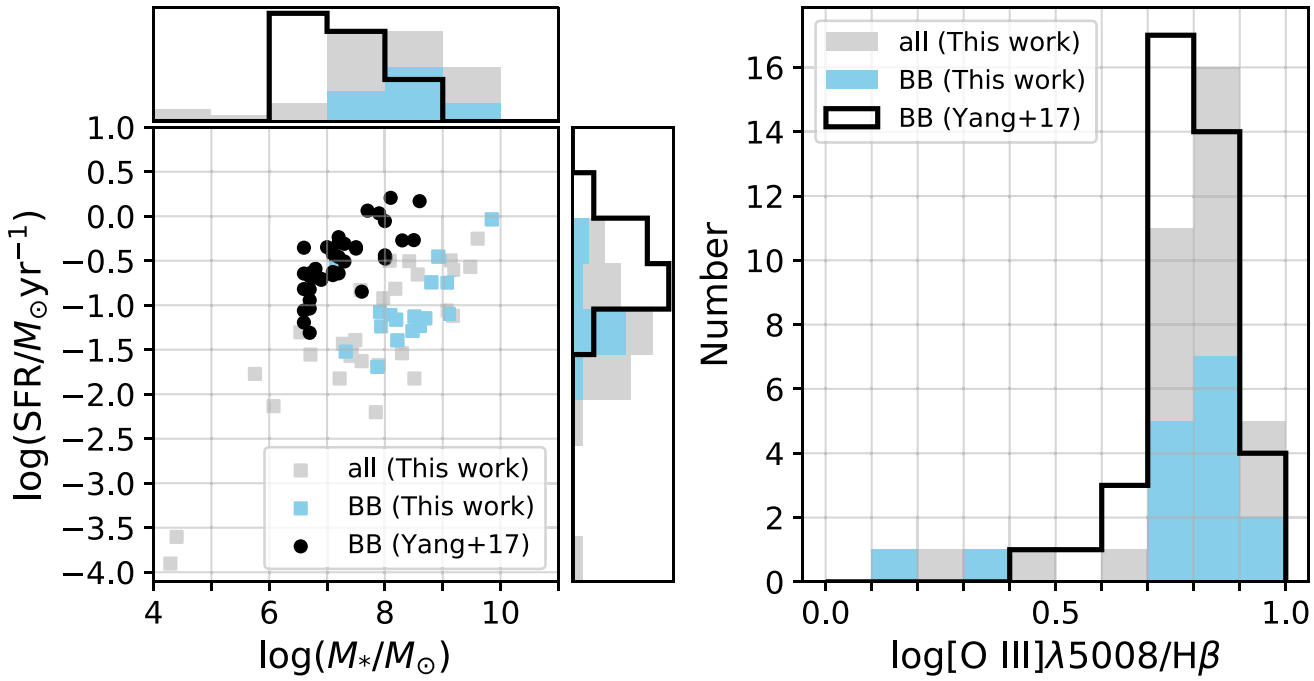

Figure 10. Comparisons of stellar mass, SFR, and the O3 index between all observed XMPs (grey squares/shadings), BB-like observed XMPs (blue squares/shadings), and BB galaxies from Y17 (dots/unfilled black histogram).

In this work, we developed a DL pipeline and validated its effectiveness using new observations. In the near-future, we will conduct follow-up spectroscopic observations that will cover additional key emission lines to: (1) enable direct measurements of oxygen abundances; (2) measure the primordial ^{4}He abundance; and (3) address questions that were raised in our analysis, including: (a) the origin of outliers in our pipeline – potentially due to outflows, environmental impact, or nitrogen produced through different channels; (b) test if the 4 galaxies with low [O III] λ5008/H β and low [N II] λ6585/H α contain a low-metallicity AGN; and (c) develop the connection between our discoveries, and high-redshift galaxies.d

ACKNOWLEDGEMENTS

During this work, RJC was funded by a Royal Society University Research Fellowship. TYC and RJC acknowledge support from STFC (ST/T000244/1). The Issac Newton Telescope is operated on the island of La Palma by the Isaac Newton Group of Telescopes in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. The IDS spectroscopy was obtained as part of P4 programme. Based on observations obtained at the SOAR telescope, which is a joint project of the Ministério da Ciência, Tecnologia e Inovações (MCTI/LNA) do Brasil, the US National Science Foundation's NOIRLab, the University of North Carolina at Chapel Hill (UNC), and Michigan State University (MSU).

DATA AVAILABILITY

The observational data via open-time programme can be shared upon request. The catalogues for the measured quantities presented in this work will be available upon the publication of this manuscript.

REFERENCES

Abdurro'uf et al., 2022, ApJS, 259, 35
Ann H. B., Seo M., Ha D. K., 2015, ApJS, 217, 27
Astropy Collaboration, 2022, ApJ, 935, 167
Aver E., Berg D. A., Olive K. A., Pogge R. W., Salzer J. J., Skillman E. D., 2021, J. Cosmol. Astropart. Phys., 2021, 027
Baldwin J. A., Phillips M. M., Terlevich R., 1981, PASP, 93, 5
Bell E. F., McIntosh D. H., Katz N., Weinberg M. D., 2003, ApJS, 149, 289
Blanton M. R., Roweis S., 2007, AJ, 133, 734
Brinchmann J., Charlot S., White S. D. M., Tremonti C., Kauffmann G., Heckman T., Brinkmann J., 2004, MNRAS, 351, 1151
Bromm V., Yoshida N., 2011, ARA&A, 49, 373
Bromm V., Yoshida N., Hernquist L., McKee C. F., 2009, Nature, 459, 49
Cardamone C. et al., 2009, MNRAS, 399, 1191
Carrick J., Turnbull S. J., Lavaux G., Hudson M. J., 2015, MNRAS, 450, 317
Chabrier G., 2003, PASP, 115, 763
Cheng T.-Y. et al., 2020, MNRAS, 493, 4209
Cheng T.-Y. et al., 2021, MNRAS, 507, 4425
Cheng T. Y. et al., 2023, MNRAS, 518, 2794
Chiappini C., Romano D., Matteucci F., 2003, MNRAS, 339, 63
Clemens J. C., Crain J. A., Anderson R., 2004, in Moorwood A. F. M., Iye M., eds, Proc. SPIE Conf. Ser. Vol. 5492, Ground-based Instrumentation for Astronomy. SPIE, Bellingham, p. 331
Cooke R. J., Pettini M., Jorgenson R. A., Murphy M. T., Steidel C. C., 2014, ApJ, 781, 31
Denicoló G., Terlevich R., Terlevich E., 2002, MNRAS, 330, 69
Fernández V., Terlevich E., Díaz A. I., Terlevich R., 2019, MNRAS, 487, 3221
Frazier P. I., 2018, preprint (arXiv:1807.02811)
Fukugita M., Kawasaki M., 2006, ApJ, 646, 691
Fukushima K., Nagamine K., Matsumoto A., Isobe Y., Ouchi M., Saitoh T., Hirai Y., 2024, preprint (arXiv:2401.06450)
Goodfellow I. J., Shlens J., Szegedy C., 2014, CoRR. Explaining and Harnessing Adversarial Examples, preprint (arXiv:1412.6572)
Griffith R. L. et al., 2011, ApJ, 736, L22
Grossi M. et al., 2025, preprint (arXiv:2501.18498)

Groves B. A., Heckman T. M., Kauffmann G., 2006, MNRAS, 371, 1559
Guseva N. G., Izotov Y. I., Papaderos P., Fricke K. J., 2007, A&A, 464, 885
Guseva N. G., Izotov Y. I., Fricke K. J., Henkel C., 2017, A&A, 599, A65
Hirschauer A. S. et al., 2016, ApJ, 822, 108
Hsyu T., Cooke R. J., Prochaska J. X., Bolte M., 2017, ApJ, 845, L22
Hsyu T., Cooke R. J., Prochaska J. X., Bolte M., 2018, ApJ, 863, 134
Hsyu T., Cooke R. J., Prochaska J. X., Bolte M., 2020, ApJ, 896, 77
Isobe Y. et al., 2022, ApJ, 925, 111
Izotov Y. I., Thuan T. X., 2007, ApJ, 665, 1115
Izotov Y. I., Thuan T. X., 2009, ApJ, 690, 1797
Izotov Y. I., Lipovetsky V. A., Chaffee F. H., Foltz C. B., Guseva N. G., Kniazev A. Y., 1997, ApJ, 476, 698
Izotov Y. I., Papaderos P., Guseva N. G., Fricke K. J., Thuan T. X., 2006, A&A, 454, 137
Izotov Y. I., Guseva N. G., Fricke K. J., Papaderos P., 2009, A&A, 503, 61
Izotov Y. I., Thuan T. X., Guseva N. G., 2012, A&A, 546, A122
Izotov Y. I., Thuan T. X., Guseva N. G., 2014, MNRAS, 445, 778
James B. L., Koposov S. E., Stark D. P., Belokurov V., Pettini M., Olszewski E. W., McQuinn K. B. W., 2017, MNRAS, 465, 3977
Karachentsev I. D., Makarova L. N., Koribalski B. S., Anand G. S., Tully R. B., Kniazev A. Y., 2023, MNRAS, 518, 5893
Kauffmann G. et al., 2003a, MNRAS, 341, 33
Kauffmann G. et al., 2003b, MNRAS, 346, 1055
Kennicutt Robert C. J., 1998, ARA&A, 36, 189
Kennicutt Robert C. J., Tamblyn P., Congdon C. E., 1994, ApJ, 435, 22
Kewley L. J., Groves B., Kauffmann G., Heckman T., 2006, MNRAS, 372, 961
Kewley L. J., Dopita M. A., Leitherer C., Davé R., Yuan T., Allen M., Groves B., Sutherland R., 2013, ApJ, 774, 100
Kingma D. P., Ba J., 2015, Adam: A Method for Stochastic Optimization, 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Conference Track Proceedings. preprint (arXiv:1412.6980)
Kojima T. et al., 2020, ApJ, 898, 142
Kunth D., Östlin G., 2000, A&AR, 10, 1
Leitherer C. et al., 1999, ApJS, 123, 3
Liu S., Luo A. L., Zhang W., Kong X., Zhang Y.-X., Shen S.-Y., Zhao Y.-H., 2023, ApJS, 267, 16
Madau P., Pozzetti L., Dickinson M., 1998, ApJ, 498, 106
Marigo P., 2001, A&A, 370, 194
Matsumoto A. et al., 2022, ApJ, 941, 167
McQuinn K. B. W. et al., 2015, ApJ, 812, 158
Meynet G., Maeder A., 2002, A&A, 390, 561
Micheva G., Östlin G., Bergvall N., Zackrisson E., Masegosa J., Marquez I., Marquart T., Durret F., 2013, MNRAS, 431, 102
Nakajima K. et al., 2022, ApJS, 262, 3
Nishigaki M. et al., 2023, ApJ, 952, 11
Peimbert A., Peimbert M., Luridiana V., 2016, Rev. Mex. Astron. Astrofis., 52, 419
Pettini M., Pagel B. E. J., 2004, MNRAS, 348, L59
Pilyugin L. S., Thuan T. X., Vílchez J. M., 2003, A&A, 397, 487
Planck Collaboration VI, 2020, A&A, 641, A6
Prochaska J. X. et al., 2020a, pypeit/PypeIt: Release 1.0.0
Prochaska J. X. et al., 2020b, J. Open Source Softw., 5, 2308
Pustilnik S. A., Tepliakova A. L., Kniazev A. Y., Martin J. M., Burenkov A. N., 2010, MNRAS, 401, 333
Raimann D., Bica E., Storchi-Bergmann T., Melnick J., Schmitt H., 2000, MNRAS, 314, 295
Roy A., Dopita M. A., Krumholz M. R., Kewley L. J., Sutherland R. S., Heger A., 2021, MNRAS, 502, 4359
Ruiz-Escobedo F., Peña M., Hernández-Martínez L., García-Rojas J., 2018, MNRAS, 481, 396
Salpeter E. E., 1955, ApJ, 121, 161
Sargent W. L. W., Searle L., 1970, ApJ, 162, L155
Skillman E. D. et al., 2013, AJ, 146, 3
Steigman G., 2007, Annu. Rev. Nucl. Part. Sci., 57, 463
Storchi-Bergmann T., Calzetti D., Kinney A. L., 1994, ApJ, 429, 572
Tachiev G., Froese Fischer C., 2001, Can. J. Phys., 79, 955
Thuan T. X., Izotov Y. I., 2005, ApJS, 161, 240
Thuan T. X., Izotov Y. I., Lipovetsky V. A., 1995, ApJ, 445, 108
Thuan T. X., Guseva N. G., Izotov Y. I., 2022, MNRAS, 516, L81
Tremonti C. A. et al., 2004, ApJ, 613, 898
Wang L.-L. et al., 2018, MNRAS, 474, 1873
Wise J. H., Turk M. J., Norman M. L., Abel T., 2012, ApJ, 745, 50
Xu Y. et al., 2022, ApJ, 929, 134
Yang H., Malhotra S., Rhoads J. E., Wang J., 2017, ApJ, 847, 38 (Y17)
Yin S. Y., Liang Y. C., Hammer F., Brinchmann J., Zhang B., Deng L. C., Flores H., 2007, A&A, 462, 535 (Y07)
van Zee L., 2000, ApJ, 543, L31
van Zee L., Haynes M. P., 2006, ApJ, 636, 214
Zibetti S., Charlot S., Rix H.-W., 2009, MNRAS, 400, 1181
Zou H. et al., 2024, ApJ, 961, 173

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