doi:** 10.3969/j.issn.1000-8349.2026.02.02

2026-07-07

ChinaRxiv

Abstract

At present, the vast majority of imaging-dominated galaxy and quasar survey projects rely heavily on photometric redshifts to obtain redshift information for tens of millions to billions of celestial objects, thereby supporting frontier scientific research on the large-scale structure of the universe, the properties of dark energy, and related topics. In this context, machine learning methods, owing to their efficiency and scalability, have become mainstream tools for obtaining high-precision photometric redshifts. This paper systematically reviews recent advances in the application of machine learning to photometric redshifts, including algorithm classifications, model optimization strategies, and typical application scenarios, and, in conjunction with practical case studies, analyzes the technical characteristics and performance differences of different machine learning architectures.

Full Text

doi: 10.3969/j.issn.1000-8349.2026.02.02

Application of Machine Learning in Photometric Redshift Estimation

Junhao Lu1, Zhijian Luo1, Jianzhen Chen1, Lujin Zeng1, Chenggang Shu1

(1 Shanghai Key Laboratory for Semi-Analytical Research on Galaxies and Cosmology, Shanghai Normal University, Shanghai 200234)

Abstract: At present, the vast majority of imaging-based galaxy and quasar survey projects rely heavily on photometric redshifts to obtain redshift information for tens of millions to several billions of celestial objects, thereby supporting frontier scientific research on the large-scale structure of the Universe, dark-energy properties, and related topics. Against this background, machine-learning methods, owing to their efficiency and scalability, have become mainstream tools for obtaining high-precision photometric redshifts. This paper systematically reviews recent research progress in the application of machine learning to photometric redshifts, including algorithm classification, model-optimization strategies, and typical application scenarios. In combination with practical cases, it analyzes the technical characteristics and performance differences of different machine-learning architectures.

Keywords: photometric redshift; machine learning; data analysis

Chinese Library Classification Number: P157.2 **Document Code:** A

1 Introduction

Major scientific questions in current extragalactic astronomy include the formation and evolutionary pathways of galaxies, the spatial distribution and physical properties of dark matter, and the formation and evolutionary laws of the large-scale structure of the Universe. Breakthroughs in these areas all depend on accurate distance measurements for large samples of galaxies. Traditionally, such distance measurements have mainly been achieved by observing the spectroscopic redshift caused by cosmic expansion. Limited by actual observing conditions—especially for extremely faint objects or large-scale galaxy samples—spectroscopic observations remain time-consuming and highly inefficient, even though advanced multi-object spectroscopy and slitless-spectroscopy techniques have developed rapidly. This limitation directly promoted the emergence of photometric redshift as an alternative estimation method. The theoretical prototype of this method was first proposed by Baum[1], then systematically improved by Butchins[2], and a complete theoretical framework was established by Connolly et al.[3]. Photometric-redshift methods are based on the statistical association between the intrinsic energy distribution and redshift derived from multi-band photometric data of galaxies; they can estimate celestial distances in the absence of spectroscopic observations, thereby providing key methodological support for galaxy-cosmology studies using large-scale multicolor survey data.

In the history of photometric-redshift research, the true milestone was the Sloan Digital Sky Survey (SDSS). As the first large-scale digital survey project to realize coordinated multi-band photometric and spectroscopic observations[4], SDSS not only revolutionized the mode of astronomical data acquisition, but also provided an unprecedented experimental platform for the diversified development of galaxy-distance measurement methods. This pioneering project enabled astronomers to systematically explore the effectiveness of various photometric-redshift methods. SDSS' s successful expe[[unclear: text continues beyond the visible page]]

Received: 2025-05-14; **Revised:** 2025-09-04

Funding: National Natural Science Foundation of China (12141302, 12573009); China Manned Space Engineering Survey Space Telescope Special Scientific Research Program (CMS-CSST-2025-A05, CMS-CSST-2025-A07)

Corresponding author: Zhijian Luo, zjluo@shnu.edu.cn

experiments directly promoted the implementation of a subsequent series of major survey projects, including the Cosmic Evolution Survey (COSMOS)[5], the Dark Energy Survey (DES)[6], the Kilo Degree Survey (KiDS)[7], the Hyper Suprime-Cam Survey (HSC)[8], the James Webb Space Telescope (JWST)[9], the Euclid Space Telescope[10], the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST)[11], as well as the soon-to-be-operated Nancy Grace Roman Space Telescope (Roman)[12] and the Chinese Space Station Survey

Telescope (CSST)[13]. Against this background, by integrating high-precision multi-band photometric data for massive galaxy samples with prior spectral libraries, astronomers have successfully carried out studies in precision cosmology and achieved remarkable results, marking the formal entry of photometric redshift into the stage of large-scale engineering application.

The principle of photometric redshift methods is that, under the influence of cosmological expansion, galaxy spectra shift toward the red end, causing galaxies of the same type but at different distances to exhibit distinguishable photometric features in their multi-band magnitudes and colors within a given photometric system. A typical example is the Lyman-α galaxy (Lyα emitter, LAE)[14], which at high redshift can be observed because cosmic expansion redshifts the Lyα emission line into the visible or near-infrared bands. However, practical application is far from so simple: the mapping function from photometric-parameter space to redshift space is highly nonlinear, and its complexity arises from the coupled action of multiple factors, including different galaxy morphologies, large-scale-structure environments, and evolutionary stages.

Current photometric-redshift estimation methods can be grouped into two major categories:

(1) Spectral energy distribution (SED) template fitting[15,16]. This method estimates redshift values by comparing the degree of match between observational data and theoretical or empirical templates. Theoretical templates[17] synthesize theoretical spectra based on stellar-evolution models, while empirical templates[18] directly use observed galaxy spectra to construct template libraries. The advantage of this method lies in its strong physical interpretability, but it depends heavily on the completeness of the template library and is readily affected by uncertainties such as interstellar extinction effects and instrumental response functions.

(2) Machine-learning (ML)-driven empirical methods[19–22]. Such data-driven methods make predictions by establishing nonlinear mappings from photometric-parameter space to redshift space. Supervised learning trains regression models on spectroscopically verified samples; unsupervised learning uses methods such as self-organizing mapping (SOM) to mine potential associations between photometric features and redshifts in unlabeled data; and hybrid enhancement strategies combine transfer learning with semi-supervised learning to alleviate the scarcity of spectroscopic data.

Compared with the former category, machine-learning methods stand out in terms of large-data processing efficiency and the ability to capture complex patterns, but they face challenges such as weak physical interpretability and limited extrapolative generalization capability. Recent research trends show that the two categories of methods are gradually moving toward integration. For example, SED templates, galaxy images, galaxy morphology, galaxy mass, and other quantities can be embedded as physical constraint terms into neural-network

loss functions, forming a new framework of physics-informed ML[23]. It should be noted that SED template-fitting methods likewise have diverse variants; various hybrid methods, such as those combining Bayesian inference with nested-sampling techniques[24], have further expanded the application dimensions of this class of methods.

Accurate determination of the basic physical parameters of galaxies is one of the scientific goals of large-scale surveys. In theory, machine-learning methods can not only estimate redshifts, but also further infer key physical quantities such as the star formation rate (SFR) and galaxy stellar mass[25]. For example, Bonjean et al.[26] trained a random forest (RF) model based on SDSS Data Release 8 (DR8) spectroscopic data, successfully applying it to the Wide-field Infrared Survey Explorer,

WISE)[27], for the simultaneous estimation of the SFR and galaxy stellar mass of near-infrared source samples; Mucesh et al.[28] applied this method to DES survey data, where it still exhibited good performance despite limited photometric parameters. The Euclid team[29] investigated the performance of template fitting and machine-learning methods (CatBoost, deep neural networks, etc.) in recovering redshift, galaxy stellar mass, SFR, and other physical parameters from simulated Euclid satellite observations. In view of the data characteristics of the Euclid wide survey (EWS) and the Euclid deep fields (EDF), they compared the advantages and disadvantages of training strategies based on matched labels and mixed labels, finding that, for wide-field surveys, the best results can be obtained by using labels derived from template-fitting training on deep-field data together with wide-field features[29–31]. These results fully demonstrate that machine learning can effectively uncover the complex correlations between photometric redshift and the physical properties of galaxies, laying a key foundation for the scientific output of next-generation large-scale survey projects such as Euclid and CSST.

This paper focuses on machine-learning estimation methods for photometric redshifts and systematically analyzes the advantages and limitations of various algorithms. Through a comprehensive discussion of the key issues in this field, it aims to provide astronomers with theoretical support and practical guidance for constructing a more efficient and accurate redshift-estimation system, thereby serving the scientific exploration of next-generation survey projects such as Euclid and CSST and enhancing their scientific value.

2 Machine-Learning Methods

The observing speed of current galaxy photometric surveys has far exceeded the capacity of spectroscopic follow-up observations. For example, LSST is expected to observe 37 million galaxies per year, whereas spectroscopic observations can cover only about 0.1%. Therefore, photometric-redshift methods have become an indispensable research tool in extragalactic astronomy. Studies have shown that, within the coverage range of the spectroscopic training set, machine-

learning methods can achieve a prediction accuracy of $|\Delta z|/(1+z_{\rm spec}) < 0.01$, where $\Delta z = z_{\rm phot} - z_{\rm spec}$ is the difference between the predicted value and the true value. However, when this method is extrapolated to regions of parameter space not covered by the training set (e.g., $z > 2.5$), its predictive accuracy shows a systematic decline[32]. Even so, machine-learning methods have also made major progress in studies of special types of objects. For example, in quasar redshift estimation, random forest algorithms have successfully reduced the outlier rate to one-third that of traditional methods[33]; generalized additive models have successfully predicted the redshifts of 276 long-duration gamma-ray bursts[34]. These innovative methods have promoted the transformation of photometric redshift from a purely "statistical tool" into a "physical diagnostic instrument." The representative methods and basic principles in this field are summarized below.

First are methods dominated by classical machine learning:

(1) Self-organizing maps[35–37]. SOM projects high-dimensional photometric data onto a two-dimensional grid through unsupervised learning, so that objects with similar photometric features are distributed near one another in the grid. By further combining a small number of labeled nodes with known redshifts, or subsequent regression methods, the redshifts of objects at unlabeled nodes can be inferred. This method is particularly suitable for batch calibration of photometric redshifts in large-scale surveys, but it is relatively sensitive to samples at node boundaries. Because SOM can perform local calibration based on known-redshift samples within nodes (such as mean-value or regression correction)[38], and because it uses topology-preserving discretized mapping, it can effectively identify the data distribution to correct biases, thereby reducing systematic errors. This method is suitable for visualization analysis of massive data sets and preliminary estimation of photometric redshifts.

(2) Supervised feed-forward neural networks[22,39–46]. Supervised feed-forward neural networks use multiband photometric data of celestial objects (such as magnitudes and colors) as input and, using a labeled training set (with known spectroscopic redshifts), learn nonlinear mapping relations. After feature extraction by hidden layers, they output predicted redshift values, and then optimize parameters through backpropagation to minimize prediction errors, thereby realizing regression modeling from photometric data to redshift.

(3) Methods for adaptively detecting and removing anomalous photometric or spectroscopic data[47–50]. Adaptive neural networks (self-adaptive neural networks) dynamically adjust network weights or structures and combine them with anomaly-scoring mechanisms—such as reconstruction error and prediction bias—to learn online the distributional characteristics of normal photometric data, identify anomalous points that deviate from the pattern or are affected by noise, and adaptively optimize the removal of data above a threshold, thereby

achieving highly robust outlier filtering. For example, Cavuoti et al.[51] introduced a redshift–color prior relation into the loss function, improving prediction accuracy by embedding constraints from physical terms.

(4) Support vector machines[19, 52]. A support vector machine (SVM) searches for the optimal hyperplane in a high-dimensional feature space and maximizes the separation between different redshift classes, thereby enabling classification or regression prediction of photometric redshifts. Its kernel functions, such as the Gaussian kernel, can nonlinearly map photometric-band data and reduce the complex degeneracy between redshift and photometric features, making the method particularly suitable for small-sample, high-dimensional data. SVMs are often combined with template-fitting methods or spectral features to improve robustness for low signal-to-noise data. The choice of kernel function has a significant impact on performance, and the computational complexity grows superlinearly with sample size.

(5) Tree-based methods[21, 53–58]. These include a series of variants such as random forests, gradient boosting decision trees (GBDT), and CatBoost, which offer both interpretability and the ability to handle missing data. Among them, RF[21, 59] constructs multiple decision trees to vote on or average predictions from the photometric data of celestial objects, using bootstrap aggregation (bagging) to reduce variance and enhance generalization; each tree randomly selects features and samples to reduce overfitting, thereby stably handling high-dimensional and nonlinear photometric data. Its inherent feature-importance assessment can also help select key bands and improve prediction accuracy. GBDT[57, 60], by contrast, iteratively trains multiple weak decision trees, with each tree focusing on correcting the prediction residuals of the preceding tree, thereby progressively optimizing the nonlinear relationship between photometric data and redshift. By combining a boosting strategy with regularization techniques, such as learning-rate and tree-depth control, it effectively reduces bias and suppresses overfitting, enabling high-precision redshift prediction from noisy, multi-feature photometric data; feature-importance analysis can further reveal the contribution of key bands to redshift estimation. CatBoost[54] uses oblivious trees and ordered boosting to process categorical features, such as filter bands, and numerical features in astronomical photometric data, effectively reducing overfitting. Its built-in feature permutation and adaptive learning-rate mechanisms can automatically optimize the redshift-prediction model, making it particularly suitable for high-dimensional, noisy photometric data. At the same time, it efficiently encodes categorical variables through target-variable statistics, improving the accuracy and robustness of redshift regression.

(6) k-nearest-neighbour algorithms[20, 61–64]. The k-nearest-neighbours (kNN) method computes the similarity, such as Euclidean distance, between the photometric data of a target celestial object and objects

in the training set, selects the nearest k samples, and uses their mean redshift or a weighted value as the prediction result. This method requires no explicit modeling and relies directly on the local data distribution, making it suitable for nonlinear relationships; however, it is sensitive to feature scaling and sample density, and its computational complexity increases substantially as the data volume grows.

(7) Gaussian processes[65, 66]. A Gaussian process (GP) assumes that the mapping between photometric data and redshift follows a Gaussian random field, and uses a covariance function, such as a radial basis kernel, to characterize the similarity among the photometric features of different celestial objects, thereby predicting redshifts and their uncertainties. Its nonparametric nature allows it to flexibly fit complex nonlinear relationships and naturally provide prediction confidence intervals, making it especially suitable for small-sample, high-precision redshift estimation, although it has relatively high computational-resource requirements.

(8) Mixture density networks[67, 68]. A mixture density network (MDN) combines neural networks with a Gaussian mixture model, learning the multimodal conditional probability distribution from photometric data to redshift rather than producing a single predicted value. Its outputs are the parameters of multiple Gaussian distributions—weights, means, and variances—which can capture complex nonlinear relationships and uncertainties in redshift estimation and provide comprehensive probabilistic prediction results, making it particularly suitable for data with multiple solutions or noise interference.

With the development of computer performance and the increasingly stringent requirements for redshift accuracy, deep learning has begun to be widely applied to photometric-redshift estimation. Representative methods include:

(9) Neural networks for combined images[69–73]. Many neural-network models are suitable for combined images, including deep neural networks (DNNs), convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory networks (LSTMs). These models directly learn, end to end, the complex mapping between multi-band photometric images of celestial objects and redshift. Through convolutional layers, they automatically extract local pixel patterns (such as galaxy morphology and color gradients) and global distributional features. Combined with multi-task learning or attention mechanisms, they can simultaneously model the nonlinear relationship between images and redshift, significantly improving prediction accuracy. They are especially suitable for high-dimensional, structured photometric image data and surpass traditional methods based on tabular features.

(10) Hybrid applications of machine-learning methods[50, 74–76]. Hybrid machine-learning methods combine the advantages of different algorithms (such as SVM kernel techniques, feature selection in random forests,

and nonlinear fitting by neural networks) and collaboratively model the complex relationship between photometric data and redshift through cascading, stacking, or weighted fusion. As a result, they surpass single models in accuracy, robustness, and uncertainty quantification, and are particularly suitable for astronomical data with multimodal distributions; they can be used in cosmological studies that require high photometric-redshift accuracy. For example, Lin et al.[50] used supervised contrastive learning (SCL) and the kNN algorithm to construct and calibrate estimates of the original redshift probability distribution function (PDF).

The machine-learning methods above can be broadly divided into unsupervised learning (method (1)) and supervised learning (methods (2)–(10)). Among them, the premise for applying supervised learning methods to photometric-redshift prediction is that a complex mapping model must be established between multi-band photometric data and the distances to celestial objects. If the redshift distribution of the training set can continuously cover the target interval, the sampling density in color–magnitude space is sufficient, and the proportions of special object types are representative, then machine-learning models can exhibit excellent redshift-prediction performance[30, 32]. However, as noted above, this high-precision predictive capability is strictly limited by the boundaries of the feature space defined by the training set. When the object to be predicted exceeds the parameter coverage of the training samples—such as having a higher redshift or more extreme color features—the model performance will show systematic degradation. It is worth noting that, compared with traditional machine learning (methods (1)–(8)), deep learning (method (9)) appears on the surface to be a tool for solving problems, but in essence it represents a major leap. Mathematically, it can be shown that a neural network containing at least one hidden layer can fit arbitrarily complex functions, whereas traditional machine learning is constrained by shallow models and can only approximate simple functions; by increasing the number of network layers, deep learning can achieve an exponential improvement in model-fitting capability. The most fundamental distinction between the two in photometric-redshift estimation lies in how input features are processed. The advantage of deep learning is its end-to-end learning paradigm and its capacity for "automatic feature-representation learning" : through multiple nonlinear transformations, it can automatically learn and extract optimal hierarchical features for redshift prediction directly from more primitive data forms, thereby bypassing or greatly simplifying the steps of manual feature engineering and shifting the primary task from designing the best features for the algorithm to designing the best model architecture.

In addition to dependence on the coverage of the training set, machine-learning methods also face another challenge: handling missing data. When a large amount of data is missing from the training set, the results of most models often produce systematic biases. The mechanism behind such biases is that the model needs to define an effective "metric distance," and for this distance to operate correctly, it must be based on input spaces with the same dimensionality. The complexity of this problem is especially pronounced in astronomy, because

"missing data" is not a single concept. It can be divided into two categories: first, "data voids" caused by observational limitations, such as detector bad pixels or an object lying outside the observed region; such missingness carries no information about the object itself. Second, the observed magnitude has reached the limit. Although the latter is often marked as a missing value in data tables, it clearly indicates that the object' s brightness in that band is below a known threshold, which is itself important physical constraint information. A classic example is the "Lyman break" or "dropout" phenomenon used to discover high-redshift galaxies. For example, in a galaxy with redshift $z \approx 4$, radiation in its spectrum below the Lyα line (rest-frame 121.6 nm) is completely absorbed by neutral hydrogen, causing it to "disappear" or become undetectable in blue bands (such as the u and g bands), but in red bands (such as

r and i bands) remain visible. Such "non-detections" in specific bands are precisely the key features for identifying high-redshift objects. Therefore, effectively using non-detection information, rather than simply discarding or filling it in, is crucial for constructing robust photometric-redshift models. When the proportion of missing data is low, common approaches are to remove incomplete data or apply various data-imputation techniques[77]; for datasets with high missing rates, however, a more reliable solution is to adopt algorithms that are more robust to missing data, such as probabilistic random forest[49] and generative adversarial imputation network (GAIN)[78].

In the actual model-training process, partitioning the sample set is also a key step. For randomly dividing samples into training, validation, and test sets, there is currently neither a clearly defined standard for the relative proportions nor a unified specification for the sampling mechanism; methods such as random sampling and interval sampling may be used. Although trial and error can be used to seek the optimal partitioning strategy and sampling method, as a rule of thumb, when the data volume is sufficient—at least on the order of 10^3—a random data split in proportions of 60%, 20%, and 20% is a conventional choice.

To overcome the bottlenecks of any single method, researchers have begun exploring innovative schemes that integrate different approaches. Traditional SED-fitting methods, while theoretically having no upper redshift limit, usually achieve lower predictive accuracy than supervised machine-learning methods when sufficient data are available; conversely, machine-learning methods, though more accurate, are constrained by the parameter space of the training samples. It is precisely this pronounced complementarity that has stimulated extensive recent research on hybrid methods combining the strengths of both, with the aim of breaking through their respective inherent limitations. A relatively direct model is the classification-aided photometric redshift (z) estimation (CP_z) method[79], which combines template fitting with RF classifiers to identify objects such as stars, normal galaxies, AGN, and quasars (QSO), and to estimate their redshifts. This method first uses LePhare[15] for SED template fitting, then employs three RF classifiers respectively for star identi-

fication, optimal redshift-template-library matching, and outlier detection, and finally integrates the results through probability thresholds. Experiments on SDSS multi-band data[80] show that the CP_z method achieves a photometric-redshift accuracy of $\sigma = 0.039$ for normal galaxies and AGN, with an outlier rate of only 2.3%; moreover, it can effectively distinguish QSOs from stars without requiring X-ray data, thereby providing a new tool for high-precision cosmological studies. Another scheme adopts a hierarchical Bayesian fusion strategy; by integrating the PDFs of different models—including PDFs obtained from template-fitting and machine-learning methods—this approach significantly improves the accuracy of single-model performance evaluation[81,82]. Another mainstream direction for fusion focuses on optimizing the SED templates themselves. For example, perturbation algorithms can be used to dynamically adjust the shapes of SED templates according to observed photometric data, after which the adjusted templates are applied to template fitting, thereby obtaining more accurate redshift predictions[83]. This method has been validated on a CSST simulated catalog, successfully reducing the outlier rate from 3.86% to 2.55%[84]. In addition, some schemes use machine learning to directly construct large-scale, more representative SED template libraries from photometric data and combine them with multi-band data for outlier analysis. This not only optimizes existing photometric-redshift software that combines template fitting and machine learning, such as Delight[85], but also reveals a new phenomenon in which narrow-band data may introduce outliers[35,65].

Recently, a hybrid algorithm named HAYATE (hybrid algorithm for wi(Y)de-range photo-z estimation with artificial neural networks and template fitting) has further demonstrated the potential of this field. This method innovatively combines artificial neural networks with template fitting, using the SEDs of low-redshift galaxies to generate simulated data covering a broader redshift range while simultaneously optimizing redshift estimation and probability distributions, thereby achieving dual improvements in accuracy and robustness. Tests show that HAYATE yields smaller errors than the traditional template-fitting method EAZY in the low-redshift range, performs comparably in the high-redshift range, and is about 100 times faster computationally. Although this method provides an efficient, high-precision redshift-estimation solution for large-scale survey projects such as JWST, the study also points out that, for rare objects such as active galactic nuclei, its template library still needs to be expanded[86].

3 Methods and Metrics for Evaluating Model Performance

Given the growing diversity of photometric-redshift estimation methods, it is essential to establish a unified evaluation framework so that different methods can be compared fairly; large survey projects have further stimulated a series of standardized assessments. Many studies have attempted to systematically compare the strengths and weaknesses of different methods by formulating unified data protocols and evaluation standards[32,87]. The Photometric Redshift

Accuracy Testing (photo-z accuracy testing, PHAT)[88,89] pioneered this type of assessment, and similar competitions were subsequently launched by the Euclid and LSST projects[30,32]. These tests all adopt a blind-testing mechanism: part of the data set is used as an independent test sample, and participants cannot access these data during model construction, thereby ensuring the objectivity and fairness of the evaluation[90]. Evaluation is usually based on standardized statistical metrics, including standard deviation, bias, normalized median absolute deviation, root-mean-square error, and outlier fraction. Assessing the quality of photometric-redshift estimation from the perspective of point estimates is a widely recognized evaluation system with sufficient credibility in the community.

Because relying solely on point estimates to represent photometric-redshift quality introduces systematic deficiencies and may lead to significant biases, the field is gradually shifting toward using PDFs to construct a more comprehensive confidence-interval evaluation system, thereby effectively improving the reliability of high-precision applications by quantifying the uncertainty distribution of redshift estimates. From the perspective of information content, a PDF can provide rich information that point estimates cannot capture, making it a superior evaluation carrier; in particular, when degeneracies exist in parameter space, the PDF can reflect secondary solutions caused by those degeneracies. In practical applications, PDFs can not only improve the precision of cosmological and weak-lensing measurements, but also enable a full analysis of uncertainties in various cosmological parameters ranging from weak lensing tomography to baryon acoustic oscillation (BAO)[91]. At present, survey projects such as KiDS, Euclid, and LSST have comprehensively upgraded the standards of their data products: their catalogs contain not only point estimates, but also complete PDF statistical features[30].

Although the community has reached a consensus on adopting PDFs, no unified standard has yet been formed for how to quantitatively evaluate the quality of the PDFs themselves. This divergence is particularly evident in the two major survey projects Euclid and LSST. The Euclid project first assigns sources to different photometric-redshift bins according to their point estimates, obtains $\mathrm{PDF}(z-z_{\mathrm{spec}})$, and then uses as optimization parameters the ratios of the PDF area within the intervals $\pm 0.05(1+z_{\mathrm{spec}})$ (denoted $f_{0.05}$) and $\pm 0.15(1+z_{\mathrm{spec}})$ (denoted $f_{0.15}$) around the peak of the stacked PDF to the total PDF area[30]. Amaro et al.[92] pointed out that such metrics have significant limitations: if the designed PDF is a unimodal distribution around the point estimate, the prediction results can easily meet the optimization criterion, thereby leading to distorted evaluation results. Tests based on the KiDS DR3 data set ($z_{\mathrm{spec}} < 1$) show that, when a mock PDF scheme is adopted, its $f_{0.05}$ and $f_{0.15}$ metrics can reach 93.1% and 99.0%, respectively; whereas the corresponding metrics obtained by mature algorithms (METAPHOR, ANNz2, BPZ) are 65.6%, 76.9%, and 46.9% ($f_{0.05}$), and 91.0%, 97.7%, and 92.6% ($f_{0.15}$), respectively. This finding indicates that using $f_{0.05}$ and $f_{0.15}$ as comprehensive evaluation metrics

for photometric-redshift PDFs has significant limitations.

LSST, by contrast, adopts commonly used metrics related to the cumulative distribution function (CDF) of the PDF, including the probability integral transform (PIT)

$$PIT(x_{\mathrm{val}}, y_{\mathrm{val}}) = \int_{-\infty}^{y_{\mathrm{val}}} \hat{p}(y|x_{\mathrm{val}})\, dy, \tag{1}$$

and the highest probability density (HPD)

$$HPD(x_{\mathrm{val}}, y_{\mathrm{val}}) = \int_{y:\hat{p}(y|x_{\mathrm{val}})\geq\hat{p}(y_{\mathrm{val}}|x_{\mathrm{val}})} \hat{p}(y|x_{\mathrm{val}})\, dy, \tag{2}$$

where x_{val} denotes the photometric data, $\hat{p}$ denotes the PDF, y is the redshift, and y_{val} is the true redshift used for validation; schematic illustrations of the two calculations are shown in Figure 1. The PIT and HPD metrics are directly related to the p-value and e-value in Bayesian diagnostics. It should be noted that, although each source

The PDFs all have PIT/HPD values, but model performance is discussed through the PIT/HPD distributions formed from a large number of samples. In general, this is carried out by visually analyzing quantile-quantile (QQ) plots of the overall PIT and HPD distributions. The study by Harrison et al. [93] showed that, even in the multivariate case, if the model is well calibrated at the overall level, HPD, like PIT, will follow a uniform distribution. A QQ plot compares the distribution of the statistic (PIT/HPD) with the hypothesized uniform distribution. In an ideal QQ plot, all points should lie tightly along the diagonal, indicating complete agreement between the empirical and theoretical distributions. If the probability distribution of PIT/HPD lies toward the two ends, it indicates that the model predictions are overconfident; if it lies in the middle region, the model is underconfident. Some other indicators are also commonly used to quantitatively determine the uniformity of the PIT/HPD distribution, such as relative entropy (Kullback-Leibler divergence, KL), the KS (Kolmogorov-Smirnov) test, the CvM (Cramér-von Mises) test, and the AD (Anderson-Darling) statistic.

Probability Integral Transform (PIT) Highest Probability Density (HPD)

y_{val} y_{val}

Note: The solid line is the assumed redshift PDF, and y_{val} denotes the true redshift value. The shaded area in the figure represents the corresponding PIT and HPD values.

Figure 1 Construction principles of PIT and HPD [87]

Although PIT and HPD indicators are widely used, subsequent studies have pointed out that they are not perfect and may still give misleading results

under certain circumstances. Schmidt et al. [32] noted that even if the PDF is not accurately estimated, its values may still exhibit the characteristics of a uniform distribution. In the absence of the true photometric-redshift posterior distribution, at present only the conditional density estimation (CDE) loss can be used to evaluate the performance of photometric-redshift PDFs; it is defined as

$$\mathcal{L}(f,\hat{f}) = \iint \left[\hat{f}(z|x) - f(z|x)\right]^2 \mathrm{d}z\,\mathrm{d}P(x), \tag{3}$$

where $f(z|x)$ represents the unknown true redshift PDF, and $\hat{f}(z|x)$ is the PDF given by the model based on the photometric data x. This loss function can be regarded as the counterpart of the mean-square error (MSE) in standard regression. Since the CDE loss depends on the true PDF, its value cannot be calculated directly, but it can be estimated in the following way:

$$\hat{\mathcal{L}} = E_x\left[\int \hat{f}(z|x)^2\,\mathrm{d}z\right] - 2E_{x,z}\left[\hat{f}(z|x)\right] + K_f, \tag{4}$$

where the first term denotes the expectation of the photometric-redshift posterior with respect to the marginal distribution of the photometric data x; the second term denotes the expectation with respect to the joint distribution of x and all possible redshift space z; and the third term, K_f, is a constant that depends only on the true PDF. The most significant advantage of the CDE loss function is that, even when the true PDF distribution is not known, it can still estimate a loss function relatively close to the true one.

functions, differing at most by a constant term (for the detailed derivation, see Eq. (7) and the related discussion in Ref. [94]). This property makes it possible, even in the current absence of true PDF data, to quantitatively compare the estimation methods for existing datasets.

In summary, whether using traditional point-estimation metrics or emerging PDF evaluation functions, each method has its own applicable scenarios and limitations. Table 1 summarizes the principal statistics currently used for point estimation and PDF optimization. Although optimizing any single one of these indicators may lead to the generation of a meaningless PDF, considering these indicators in combination can still yield relatively reliable criteria for judgment. Selecting appropriate optimization methods and metrics for evaluating PDF reliability remains an open research question.

Table 1 Statistics used for photometric-redshift point estimation and PDF optimization [50]

Statistic	Calculation method
Mean point estimate	$\int_0^{z_{\max}} z \times p(z \mid x)\,\mathrm{d}z$
Peak point estimate	$\{z \mid \max(p(z \mid x))\}$
Cumulative distribution function $C(z \mid x)$	$\int_0^{z} p(z \mid x)\,\mathrm{d}z$
Relative residual	$(z_{\mathrm{phot}} - z_{\mathrm{spec}})/(1 + z_{\mathrm{spec}})$
Normalized median absolute deviation (σ_{NMAD})	$1.482\,6 \times M\left(\left\lvert\frac{\Delta z - M(\Delta z)}{1 + z_{\mathrm{spec}}}\right\rvert\right)$
Probability integral transform [32]	$\int_0^{z_{\mathrm{spec}}} p(z \mid x)\,\mathrm{d}z$
Odds [68]	$\int_{z_{\mathrm{phot}}-\xi_z}^{z_{\mathrm{phot}}+\xi_z} p(z \mid x)\,\mathrm{d}z$
Highest probability density [87]	$\int_{y:\,p(z\mid x)\geq p(z_{\mathrm{spec}}\mid x)} p(z \mid x)\,\mathrm{d}z$
First-order optimal transport distance	$\int_0^{z_{\mathrm{spec}}} C(z \mid x)\,\mathrm{d}z + \int_{z_{\mathrm{spec}}}^{z_{\max}} \lvert 1 - C(z \mid x)\rvert\,\mathrm{d}z$
Continuous ranked probability score	$\int_0^{z_{\mathrm{spec}}} C(z \mid x)^2\,\mathrm{d}z + \int_{z_{\mathrm{spec}}}^{z_{\max}} (1 - C(z \mid x))^2\,\mathrm{d}z$
Cross entropy	$-\lg(p(z_{\mathrm{spec}} \mid x))$
Entropy	$-\int_0^{z_{\max}} p(z \mid x)\lg(p(z \mid x))\,\mathrm{d}z$
Standard deviation (σ)	$\sqrt{\int_0^{z_{\max}} (z - z_{\mathrm{phot}})^2 p(z \mid x)\,\mathrm{d}z}$
Skewness	$\int_0^{z_{\max}} (z - z_{\mathrm{phot}})^3 p(z \mid x)\,\mathrm{d}z/\sigma^3$
Kurtosis	$\int_0^{z_{\max}} (z - z_{\mathrm{phot}})^4 p(z \mid x)\,\mathrm{d}z/\sigma^4$
Conditional density loss ($\widehat{\mathcal{L}}$) [32]	$E_x\left[\int p(z \mid x)^2\,\mathrm{d}z\right] - 2E_{x,z}[p(z \mid x)] + K_f$
$f_{0.05/0.15}$ [30]	$\int_{z_{\mathrm{peak}}-0.05/0.15}^{z_{\mathrm{peak}}+0.05/0.15} p\big((z - z_{\mathrm{spec}})/(1 + z_{\mathrm{spec}}) \mid x\big)\,\mathrm{d}z$

Note: M denotes the median function.

It should be noted that, in practical applications, machine-learning models often face complex and variable scenarios, and the stability and reliability of their performance still require further investigation; for example, not all photometric data are suitable for machine learning. The next chapter will focus on describing the uncertainties that may arise for such models under different circumstances, in order to understand the characteristics of the models and the potential impact of their applications.

4 Uncertainty in Photometric-Redshift Estimation by Machine Learning

Modern galaxy cosmology research imposes extremely stringent requirements on reducing various errors. For example, in tomographic imaging studies on cosmological scales, in order to control the influence of noise in dark-energy estimation,

it is necessary to ensure that the bias and dispersion within each redshift bin are approximately 0.003 or lower [95]. Photometric-redshift errors blur large-scale structure and affect the measurement precision of clustering algorithms, but this effect is mainly confined to relatively small scales. In the context of large-area photometric surveys, a certain level of error can also provide valuable structural information, which is of great significance for cosmological research [96].

Different types of galaxies exhibit different behavior in photometric-redshift estimation. For example, machine-learning studies based on SDSS data show that the redshift scatter obtained for luminous red galaxy samples ($\sigma \approx 0.028$) is reduced by about half compared with that for blue galaxies ($\sigma \approx 0.05$)[97]. For galaxies with redshift $z > 0.4$, redshifting moves key spectral features and the radiation peak of galaxies into the near-infrared band, making sufficiently deep near-infrared imaging particularly important[98]. By combining optical and infrared-band data, the accuracy of photometric redshifts for galaxy samples can be significantly improved. To ensure that photometric-redshift accuracy meets the requirements of cosmological measurements, a certain optical depth must be reached ($r \approx 24$ mag)[96]. The main factors affecting photometric-redshift estimation are briefly listed below.

4.1 Data-level Uncertainty

The most direct factors affecting the accuracy of photometric-redshift estimation arise from the input data themselves. Such uncertainties mainly include the errors, coverage, and quantity of the input data, as well as the reliability of the labels. Machine-learning models rely to a large extent on high-quality input data for training. If the input photometric data themselves contain large measurement errors, or if the wavelength coverage is insufficient, the model will have difficulty capturing accurate galaxy properties. Similarly, an insufficient number of training samples or low accuracy in their redshift labels will directly limit the model' s learning capacity and generalization performance, leading to increased uncertainty in the final redshift estimates.

4.1.1 Photometric Errors

The accuracy of photometric redshifts depends strongly on the quality of the input photometric data, among which photometric errors are a key factor. When simulating photometric catalogs, photometric errors are usually assumed to follow a Gaussian distribution; however, real photometric-error distributions show significant non-Gaussian characteristics. The causes of this "heavy-tailed distribution" are multifaceted. In addition to idealized thermal noise and readout noise, they also include a series of systematic effects, such as imperfectly subtracted sky-background fluctuations, cosmic-ray hits, bad or saturated detector pixels, and photometric contamination caused by the overlap of objects in crowded sky regions (blending). This non-Gaussianity is especially pronounced for faint sources whose signal-to-noise ratios are close to the detection limit.

When about 10% of the sources in a sample lie in the tails, their impact on the predicted scatter increases by more than 2σ, even exceeding the effect of halving the signal-to-noise ratio[99]. This means that tail errors in magnitudes and colors need to be minimized as much as possible, especially in the noise-sensitive U band, because this band not only has relatively low instrumental and atmospheric throughput, but is also critical for identifying the Balmer break in low-redshift galaxies or the Lyman break in intermediate-redshift galaxies. Inaccurate U-band photometry can very easily lead to a large fraction of outliers.

In machine-learning applications, the conventional approach for handling such non-Gaussian errors is to use the photometric errors provided in the catalog as part of the input features, and to train the model by feeding them together with information such as fluxes and colors[76]. The internal logic of this method is that the model is expected to learn the complex patterns of the noise automatically. However, for models such as standard feed-forward networks or random forests, they may treat the errors merely as another independent numerical feature, making it difficult for them to truly understand their physical meaning as "confidence" or "probability-distribution width." This is especially the case in high-dimensional feature spaces, where the information carried by the error features may be diluted. To make deeper use of error information, an effective method is error-based, weighted data augmentation. This scheme does not directly use the error features; instead, it treats them as standard deviations and draws perturbations from a Gaussian distribution to generate an "augmented" simulated catalog. The model can then be trained on this reasonably perturbed "infinite" dataset, thereby learning robustness at different signal-to-noise ratios. In tests on CSST simulated catalogs, compared with a model that directly inputs error features, this method successfully reduced the redshift outlier rate from 2.3% to 2.0%[21]. In essence, this method allows the model to learn the various forms that the data may take within a given error range, thereby forcing it to learn the more fundamental physical relationship between magnitude and redshift. More advanced probabilistic models, such as Bayesian neural networks and mixture density networks, can naturally incorporate the uncertainty of the input data within their mathematical frameworks[68]. In such models, one input feature corresponds to a probability distribution rather than to a single value, supporting the learning of one-to-many mappings. These models not only predict redshift values, but can also directly infer the probability distribution of the prediction results from inputs with errors, thereby providing more complete uncertainty

···uncertainty quantification, which is crucial for describing situations in which multiple redshift solutions may exist because of color degeneracies. Although machine-learning methods can automatically learn noise features, they require an appropriate parameter space and a consistent data set. By effectively controlling and modeling the error distribution during the model-training stage, the model' s dependence on a complete and enormous spectroscopic training sample can be significantly reduced[100]. In practice, obtaining spectroscopic confirmations for all types of faint objects across the entire parameter space is

unrealistic. Therefore, a model that can learn robustly from photometric data with real noise will be key to maximizing the scientific output of these surveys. Figure 2 shows the study by Blake and Bridle[96], giving the variation of the confidence boundary for photometric redshifts with the error threshold over a survey area of 10 000 $\deg^2$, and providing a reference for the design of deep surveys.

In-figure labels: confidence (10 000 $\deg^2$); photometric-redshift error $\sigma = \delta(z)/(1+z)$; r-band magnitude / mag; 2σ, 3σ, 4σ, 6σ.

Figure 2. Variation of photometric-redshift confidence with error and r-band magnitude[96]

4.1.2 Number of Observed Bands and Bandpass Coverage

The accuracy and reliability of photometric redshifts are closely related to the number of bands in the input photometric data and the wavelength coverage. By sampling the SED of an astronomical object over a broader wavelength range, its SED shape can be constrained more accurately, thereby breaking the "degeneracy" between different redshifts and object types. A widely verified conclusion is that broader bandpass coverage usually yields better redshift estimates; the physical principle behind this differs with object type. For normal galaxies, the key lies in precisely locating spectral-break features. One of the most important features in galaxy spectra is the break at 4 000 Å, which is formed by the dense superposition of absorption lines in stellar atmospheres. In the rest frame, this break lies in the blue optical part of the spectrum; as redshift increases, the feature gradually shifts toward the red end and even into the near-infrared bands. The study by Fu et al.[101] shows that data in the near-infrared bands (Y, J, H, K_s) make a positive contribution to the accuracy of photometric-redshift prediction. They compared results using only four optical bands with those after adding near-infrared bands, and found that adding near-infrared data enabled more accurate redshift prediction for high-redshift galaxies and reduced the misclassification rate by 15%, significantly improving the precision of constraints on the cosmological parameters σ_8 and Ω_M. This indicates that increasing bandpass coverage can effectively alleviate degeneracies in parameter space; especially at redshifts $z > 0.4$, near-infrared data are crucial for improving accuracy[98].

For objects with special SEDs, broad bandpass coverage is even more indispensable. Taking active galactic nuclei (active galactic nucleus, AGN) and quasars as examples, their radiation does not originate from a single stellar population, but is instead a complex continuum composed of multiple components, including the accretion disk around the central supermassive black hole (radiating in the ultraviolet and optical bands) and the surrounding dust torus (re-radiating in the infrared bands). In eROSITA (extended ROentgen Survey with an Imaging Telescope Array)

In sky surveys[^102], Brescia et al.[^39] studied the contribution of the pho-

tometry corresponding to X-ray sources to the quality of photometric-redshift measurements for AGN sources. The results showed that adding bands can improve photometric-redshift estimation. For example, using only optical bands leads to overestimated redshift values for low-redshift sources; adding the WISE mid-infrared bands can reduce this effect, and further adding the IRAC bands can improve statistical precision and reduce the outlier rate. At the observational depth of eROSITA, machine-learning models and SED-fitting methods show similar performance in terms of the outlier fraction[^39].

It should be noted that more input-feature bands are not necessarily better. Carrasco and Brunner[^59] found that, after using RF to remove several of the least important input features, the prediction accuracy instead improved. This is because unimportant band features may contain substantial noise or measurement errors. For example, observations in some bands may be strongly affected by instrumental limitations or environmental factors (such as atmospheric absorption), resulting in low signal-to-noise data. If a model attempts to learn from such noise, errors may instead be introduced; especially when the amount of data is limited, the model may overfit the noise rather than the true signal, affecting its generalization ability. On the other hand, in high-dimensional space the sparsity of the data increases, and the model needs more samples in order to learn effectively. After irrelevant features are removed, the data distribution becomes more compact, and the model can more easily capture the true relationship between redshift and effective bands in a low-dimensional space, improving computational efficiency and accuracy. Moreover, as noted above, strong correlations may exist between different bands; for example, photometric data in adjacent bands may be similar, causing unstable model-parameter estimates. Removing redundant features can reduce collinearity, enhance model robustness, and make weight allocation more reasonable.

Large-scale sky-survey projects themselves must preferentially select combinations of photometric bands that minimize parameter degeneracies and extend wavelength coverage in order to achieve their scientific goals. To strike a balance between "quantity" and "coverage," machine-learning methods themselves also provide tools for evaluating band importance. Taking the out-of-bag (OOB) sampling technique of RF models as an example, when constructing decision trees, randomly drawn data samples that do not participate in training can be used to verify the information contribution of photometric features, and to identify and remove information-redundant or misleading features. It should be noted that the degree of importance given by machine-learning methods is determined by the model itself and the algorithm, and correlations among the data may mislead the judgment, causing redundant information to appear in the importance ranking. The Euclid team[^103] found that when the VIS-u feature was duplicated, the feature importance assigned by RF changed. Lu et al.[^21] used different feature-importance calculation methods and found that, in the simulated CSST catalog, the importance of features changed before and after adding error weights. Then, after applying a Spearman rank-correlation test to classify the input data, they found that the z band and y band had a

very strong correlation.

4.1.3 Reliability of spectroscopic-redshift labels Although many photometric-redshift estimation methods are currently available, none has yet achieved the precision of spectroscopic-redshift measurements (about 10^{-3}). For example, for photometric redshifts obtained from broadband photometry, the error of the best precision is about 0.02, which still differs substantially from the precision of spectroscopic redshifts[^104].

Photometric redshifts obtained through supervised learning depend to a large extent on the completeness and quality of the spectroscopic catalog used as the reference. Incompleteness of spectroscopic samples is usually more pronounced in the faint part of the photometry, which may trigger selection effects and thereby affect the entire parameter space. At the same time, residual errors in spectroscopic redshifts also affect the reliability of the validation metrics used in machine-learning model training, and thus directly affect the quality of photometric redshifts. In general, the reliability of spectroscopic redshifts lies between 95% and 99%, which means that 1%–5% of the training samples are unreliable[^35]. At present, it is not possible to effectively distinguish the effects of spectroscopic and photometric uncertainties in the sample. The catalogs obtained by large sky-survey projects are enormous, and manual analysis based on visual inspection is difficult to implement; therefore, automatic-search machine-learning algorithms need to be explored to distinguish sources of uncertainty in the data.

Razim et al.[^35] proposed a method for identifying unreliable spectroscopic-redshift samples, using the Deep Imaging Multi-Object Spectrograph (DEIMOS) catalog and the COSMOS2015 catalog to improve measur[[unclear: text continues off page]].

photometric-redshift estimation: by combining SOM with the MLPQNA neural network and introducing the spectral quality coefficient K_{spec}, unreliable spectroscopic-redshift samples in the COSMOS data can be effectively removed, reducing the scatter of photometric-redshift prediction by about one half. In addition to K_{spec}, that study also used DEIMOS spectroscopic redshifts as a validation set and employed the so-called galaxy occupation map concept, so that the photometric-redshift sample and the validation-set sample occupied the same regions in the SOM clustering grid. This ensured an accurate correspondence between the photometric data and the spectroscopic data, and the outlier rate was thereby significantly reduced from about 11% to about 2%. It can thus be seen that unreliable spectroscopic redshifts introduce biases into machine learning that are difficult to handle. Future large survey projects need to take into account both the quantity and the quality of spectroscopic training samples, and to develop automated validation techniques, so as to fully unleash the potential of photometric redshifts in studies of galaxy cosmology.

4.1.4 Influence of photometric images The two major challenges facing photometric-redshift estimation are the decline in accuracy at high redshift and the degeneracy between photometric colors and spectroscopic redshifts; that is, sources at different redshifts may have similar color features, making it difficult for supervised-learning methods that rely only on magnitudes and colors to distinguish them. Traditional photometric data, such as magnitudes and colors, use only a small fraction of the image information. This is mainly because they are affected by factors such as the point-spread function and aperture selection, and therefore cannot fully capture the morphological features of galaxies. These morphological features are closely related to physical parameters such as a galaxy' s star-formation history and stellar mass, and thus affect SED and photometric-redshift estimation. In addition, image quality—including signal-to-noise ratio, pixel resolution, and background noise—determines whether a model can accurately extract effective morphological and color information from images. In images with low signal-to-noise ratio, the faint features of galaxies may be submerged by noise, causing the model to be unable to effectively distinguish different types of galaxies or to estimate their accurate redshifts. Therefore, when machine-learning methods are used for photometric-redshift estimation, in-depth analysis of the quality of photometric images and of galaxy morphological features is likewise crucial. Here one should note the related but different concept of the surface brightness fluctuations (SBF)[105] distance-measurement method. SBF is a classical method that uses high-signal-to-noise images to measure the distances of nearby early-type galaxies (about 100 Mpc). Its principle is that the farther away a galaxy is, the "smoother" its image appears on the detector, and the smaller the fluctuations in the number of stars contained within a unit pixel. The SBF method is applicable only to old early-type galaxies with little gas and dust, such as elliptical galaxies, S0 galaxies, and the bulges of spiral galaxies, because it relies on a smooth, stable red-giant-branch population. For spiral-galaxy disks filled with young bright stars, star clusters, and dust lanes, this method becomes unusable. Although SBF also uses image information, it is an independent, high-precision distance-calibration technique and is part of the cosmic distance ladder; its physical principles and application scenarios are entirely different from the photometric-redshift estimation methods discussed in this paper, which are applicable to large-scale, distant-galaxy samples. Photometric redshift essentially estimates the shape of the SED, whereas SBF provides a direct, physical distance estimate. In photometric-redshift applications, SBF can serve as a strong-distance prior for calibrating and constraining photometric redshifts. For more advanced models such as CNNs, SBF information can be used indirectly: CNNs can automatically identify texture information in images, and the model can learn to associate a "grainy" elliptical galaxy with a nearer distance and a "smooth" elliptical galaxy with a farther distance. Moreover, multi-branch neural networks can simultaneously process multi-band photometric data and high-resolution image cutouts, extracting multi-level features such as color gradients, disk inclination angles, and special morphologies, thereby avoiding biases from manual feature selection, and finally fusing the two types of information to make the final redshift judgment. D' Isanto and Polsterer[100]

combined CNN with MDN to simultaneously extract color and morphological information from SDSS images, controlling the prediction bias of quasar photometric redshifts to 0.004 and reducing the standard deviation to 0.069. Pasquet et al.[106] used CNN to analyze images in the five SDSS bands; the prediction bias was only ± 0.02 within the 1σ range, significantly lower than the ± 0.05 of the kNN method, and no systematic bias directly correlated with redshift was found. Zhou et al.[69] used a hybrid transfer algorithm to incorporate CSST simulated images into training; compared with using fluxes alone, the outlier rate decreased from 1.43% to 0.9%.

Although deep learning has shown promise in mining image features, its application to photometric redshift estimation still requires validation. Pasquet et al.[106] found that CNN predictions exhibit a slight bias of $\leq 5\%$ in redshift intervals densely populated in the training sample, and further cross-validation with photometric data is needed. Future large-scale survey projects are expected to use deep learning to extract information directly from pixel-level images, avoiding the limitations of traditional choices of photometric parameters and providing a more comprehensive solution to the problems of high redshift and degeneracy.

4.2 Uncertainties at the Level of Training Samples and Methods

After the issue of data quality has been addressed, how to use these data for effective training introduces another layer of uncertainty. This is mainly reflected in the representativeness of the samples and the way the data are processed. The training samples must adequately represent the distribution of the entire survey data set; otherwise, the model may perform poorly in redshift or type intervals it has not encountered. In addition, data preprocessing methods—such as feature engineering, normalization, or outlier treatment—also directly affect model performance. Improper data processing may lose important physical information or introduce new biases, thereby exacerbating the uncertainty in photometric redshift estimation.

4.2.1 Insufficient Training Sets and Outliers

In photometric redshift prediction, identifying and analyzing outliers in the training set is extremely important, because these outliers may lead to errors in distance estimation. Identifying outliers can not only provide effective guidance for subsequent spectroscopic observations, optimize the training data set, and supplement data in undersampled regions, but also statistically predict redshift quality and evaluate the effectiveness of different combinations of photometric features[107]. For example, both the supervised-learning RF model[59] and the unsupervised-learning SOM[108] can optimize parameter space; the latter identifies anomalous regions without spectroscopic priors by constructing a Kohonen map.

This problem is particularly evident in redshift prediction for galaxies containing AGN, because the fraction of the total radiation contributed by the AGN nuclear region is unknown and varies from source to source. Since X-ray selection has low reliability at $z < 1$, at present only Seyfert galaxies, when narrow-/medium-band photometric data are available (e.g., in the COSMOS survey), can achieve photometric redshift quality comparable to that of normal galaxies[109]. Although hybrid methods (machine learning + SED fitting) have made progress[79], it remains difficult to choose appropriate models when performing SED fitting[110], and their photometric redshift estimation efficiency is still inferior to that for normal galaxies. Similarly, supervised machine-learning models are constrained by whether sufficiently large and complete spectroscopic training samples are available, while most spectroscopic samples are usually extracted from catalogs of sources selected in optical bands. This inevitably leads to an imbalanced distribution of AGN or other special objects in the sample. The impact of this bias on future radio surveys such as the Square Kilometre Array (SKA) can be found in the study by Norris et al.[111].

To address the challenges posed by the scarcity and bias of AGN spectroscopic training samples, recent studies have begun to use a new generation of deep imaging surveys, combined with more refined methods for extracting physical features, to construct large-scale, high-quality training sets dedicated to AGN. The idea behind this solution is that, rather than struggling to handle AGN outliers within samples dominated by normal galaxies, one should actively establish a dedicated and representative training benchmark for this class of "outliers," namely AGN. In a recent study, Saxena et al.[112] used the tenth data release of the Dark Energy Spectroscopic Instrument (DESI) imaging Legacy Survey (LS10) to estimate the photometric redshifts of X-ray-detected AGN. This study used 14,000 X-ray-detected AGN as the training sample. This selection method fundamentally bypassed the bias of optical samples against AGN and alleviated the previous problem of insufficient AGN spectroscopic sample size. To tackle the difficulty that the radiation contributions from the AGN nucleus and the host galaxy are unknown, the study did not use simple single-aperture magnitudes, but instead employed a refined multi-aperture photometric technique. This method aims to better distinguish and quantify the light from the bright nucleus and the extended host galaxy by analyzing changes in object brightness under different apertures, thereby providing machine-learning models with input features of greater physical significance. Based on the above high-quality training set and features, they employed a fully connec-

trained using the neural-network algorithm CIRCLEZ. When tested on an independent test set consisting of 2,913 AGNs, the model achieved a prediction scatter of 0.067, with the outlier fraction controlled at 11.6%. This result is even better than those of previous studies, demonstrating that by constructing a dedicated, unbiased training set and supplementing it with refined physical-quantity inputs, efficient redshift estimation can be fully achieved for special objects such as AGNs. This work, together with the updated AGN photometric-redshift catalog released for the eROSITA/eFEDS field[112,113], has provided an important

approach to solving the long-standing problem of AGN redshift estimation.

4.2.2 Processing of Input Features

For complete samples with reliable redshifts and photometric data, different ways of processing input features will affect the results. This type of processing is generally called feature engineering: the process of transforming raw data into features that better represent the essence of the problem, serving as a bridge between the raw data and machine-learning algorithms.

Constructing new features from existing data is also a common way to improve prediction accuracy. In photometric-redshift estimation, the most basic feature engineering is the use of fluxes in different bands. Although fluxes themselves are direct observables, the "colors"constructed from them—i.e., magnitude differences between two bands—are often more physically meaningful features. This is because color directly reflects the slope of an object' s SED, and the shape of the SED is closely related to key physical properties such as the object' s redshift, age, metallicity, and dust content. The contribution of different bands to redshift prediction essentially depends on whether they can effectively capture the key spectral features that shift with redshift. Below, the specific roles and contributions of commonly used photometric bands are analyzed one by one within this physical framework.

(1) Ultraviolet and blue bands (UV, about 200 nm; u, about 350 nm; g, about 480 nm). In the low-redshift Universe ($z < 0.5$), the main task of these bands is to accurately locate the 4,000 Å break. Studies show that if a survey lacks u-band data, the accuracy of redshift estimates for galaxies at $z < 0.5$ deteriorates significantly, and may even fail to meet the precision requirements of projects such as LSST for cosmological studies[114]. In the high-redshift Universe ($z > 2.5$), the roles of the u and g bands shift to target selection using the Lyman break. The Lyα line at a rest-frame wavelength of 1,216 Å and the Lyman limit at 912 Å produce a sharp spectral drop due to absorption by neutral hydrogen in the intergalactic medium. Therefore, the absence of UV and blue bands can lead to confusion between high-redshift objects and low-redshift, dust-reddened galaxies.

(2) Middle optical bands (r, about 620 nm; i, about 750 nm; z, about 900 nm). These three red optical bands are mainly used to study the intermediate-redshift Universe ($0.4 < z < 1.6$). Taking over from the u and g bands, they continue to track the 4,000 Å break as it shifts continuously toward longer wavelengths with increasing redshift.

(3) Near-infrared (NIR) bands (Y, about 1 μm; J, about 1.25 μm; H, about 1.65 μm; K, about 2.2 μm). When the key 4,000 Å break is redshifted beyond 1 μm, outside the detection range of traditional optical CCDs, the NIR bands take over the task of detecting it. NIR bands are crucial for breaking the age–dust–redshift degeneracy. Observations in the NIR

can effectively distinguish these two cases: old galaxies exhibit a sharp drop at the 4,000 Å break, whereas the SEDs of dust-obscured galaxies are usually smooth, reddened continua. In addition, NIR bands mainly detect radiation from old stellar populations such as red giants; therefore, they correlate well with a galaxy' s total stellar mass and help determine the nature of its stellar populations more accurately.

(4) Mid-infrared (MIR) bands (W1, about 3.4 μm; W2, about 4.6 μm). These bands mainly detect not the direct radiation from stellar photospheres, but thermal radiation re-emitted by dust after it absorbs starlight. For example, the center of an AGN is surrounded by a dust torus, whose thermal radiation forms a distinctive "red tail" in the MIR bands; this becomes a decisive feature for distinguishing AGNs from normal galaxies. Without MIR data, the redshifts of many AGNs can be severely misidentified[115].

In photometric-redshift estimation, early studies mostly adopted trial-and-error methods or combined searches using machine learning[100], which were computationally expensive and did not necessarily yield the optimal solution. In recent years, random forests have been widely applied because of their ability to evaluate feature importance effectively. For example, the ϕLAB (parameter handling investigation laboratory, PhiLAB) method developed by Brescia et al.[39] combines random forests with L1 regularization and introduces "shadow features" to quantify feature contributions, thereby identifying key features in multiband photometric data. They found that when only magnitudes were used as input features, the K band was the most important because it corresponds to the bend in the galaxy SED; after color information was added, the total importance of color features exceeded 60%[39]. When RF was used on a CSST mock galaxy catalog to assess the importance of features, it was found that although the importances of the i- and z-band fluxes were extremely low (0.006 and 0.009, respectively), the importance of the color $i - z$ formed by the two bands increased sharply (0.398), because it directly corresponds to the location of the 4 000 Å break near $z \approx 1$[21]. Therefore, great caution is required when removing apparently useless features during preprocessing.

In essence, colors do not provide additional information beyond what is already contained in the fluxes. The fact that traditional machine-learning models need colors as input parameters indicates that the model architectures cannot capture the correlations among the input photometric data. Luo et al.[116] recently proposed a new RNN model based on LSTM for estimating photometric redshifts. This model avoids the input-feature selection problem faced by most traditional machine-learning models: it only needs the fluxes in each band to be arranged in wavelength order to form an input sequence, and can then automatically and effectively learn their dependencies and capture the patterns and correlations in the flux sequence. Owing to the inherent sequence-processing capability of the LSTM model, it can effectively capture correlations among data points in the sequence, thereby reducing the need for feature engineering. That is, the model no longer needs manually combined input features to generate colors, nor does

it require complex feature-selection engineering, thus avoiding human error or subjective bias.

4.3 Uncertainty at the Level of Surveys and Redshift Intervals

In addition to the data and training process, the characteristics of redshift intervals and survey strategies also pose challenges to the performance of machine-learning models. Within specific redshift intervals, even excellent models may face intrinsic difficulties. The observing strategy adopted by different survey projects determines the quality of the photometric data that can be obtained.

4.3.1 Effects in Different Redshift Intervals

The final accuracy and reliability of photometric redshifts are the result of the combined action of multiple factors. In addition to the aforementioned error handling, band selection, and feature engineering, machine-learning methods show different predictive performance in different redshift intervals. The low-redshift interval ($z < 0.5$) is the most mature and accurate regime for photometric-redshift estimation. In this range, objects are usually brighter, and the 4 000 Å break lies in optical bands with favorable observing conditions, making the redshift signal very clear. Benefiting from large-scale spectroscopic surveys such as SDSS, the training samples in this interval are not only enormous in number, but also highly complete and representative. Therefore, both template-fitting methods and machine-learning methods can usually achieve very high accuracy. The main challenges come from low-redshift galaxies whose spectral features are not prominent and from stars in the Milky Way, as well as outliers caused by a small amount of dust or AGN activity. For example, a hybrid-input CNN model recently used on SDSS data achieved an extremely low root-mean-square error of only 0.000 7 in the range $z < 0.3$[117].

The intermediate-redshift range ($0.5 < z \lesssim 1.5$) is crucial for cosmological studies such as weak gravitational lensing and baryon acoustic oscillations. Although the predictive performance of various photometric-redshift methods remains good, it begins to decline relative to the low-redshift regime, and the scatter and outlier rate begin to increase[118]. In this redshift interval, the 4 000 Å break gradually shifts into the red-end optical bands (i, z) and even into the near-infrared Y band; at this point, redshift precision depends strongly on the photometric quality in these red-end bands. The problem of degeneracy in color space begins to stand out: galaxies of different types or at different redshifts may exhibit similar colors, leading to model misclassification and an increased outlier rate. In this redshift range, the completeness of spectroscopic surveys gradually declines; they usually tend to observe brighter galaxies or specific types of luminous red galaxies, which

means that the training set is less representative of the complete galaxy population observed by photometric surveys[119].

The high-redshift regime ($z > 1.5$) is a key window for studying the early Universe and galaxy formation, and it is also where the performance of most standard, purely data-driven machine-learning models declines sharply, with substantially increased scatter and outlier rates. Because the spectroscopic samples available in this redshift range are sparse and are strongly biased toward the brightest and rarest objects, such as quasars, these objects cannot represent the typical galaxy population. This forces machine-learning models to extrapolate far beyond the range of their training data, and poor extrapolation ability is precisely a well-known weakness of machine-learning methods. For example, the ANNZ+ algorithm already struggles to maintain its accuracy above redshift $z \approx 1.3$[46]; at this point, the main objective shifts to identifying high-redshift candidates rather than precisely estimating their redshift values[56]. In addition, high-redshift objects have very low intrinsic luminosities, resulting in low signal-to-noise ratios in photometric measurements. This noise further blurs the already degenerate color information and increases the uncertainty, especially in the special redshift interval known as the "redshift desert" ($1.4 < z < 2.5$), where the Lyman break has not yet entered the optical bands, while the 4 000 Å break has completely shifted into the near-infrared. For surveys lacking ultraviolet (u) and near-infrared (J, H, K) coverage, galaxies have almost no strong spectral features in the optical bands, leading to extremely large uncertainties in redshift estimation and making this a region with a high incidence of catastrophic outliers.

The decline in machine-learning predictive performance in the high-redshift regime is not a sign that a particular algorithm is "bad," but rather a fundamental and expected behavior of supervised-learning models when they are forced to extrapolate into a data domain—high-redshift, faint, and different types of galaxies—that is fundamentally different from the training data: low-redshift, bright, and specific types of galaxies. High-redshift galaxies are not merely farther away; their intrinsic properties may differ, for example in having different star-formation histories and metallicities, and their observed properties are affected by absorption from the intergalactic medium, an effect for which there are no analogous samples at low redshift. Therefore, requiring a standard machine-learning model trained on galaxy data at $z < 1.5$ to predict the redshift of a galaxy at $z \approx 4$ is an extrapolation into a physically completely different domain. This also explains why methods that incorporate physical knowledge or attempt to simulate these missing data—such as hybrid models like HAYATE[86] —are crucial for tackling high-redshift prediction.

4.3.2 Impact of Photometric Survey Strategies

The quality of photometric redshifts is directly tied to the depth, area, and wavelength coverage of the photometric database on which they rely. Different survey projects serve different scientific goals, and the characteristics of their databases also determine the range of application of their photometric redshifts. As a well-known shallow, ultra-wide-field optical survey, SDSS used five

broadband filters (u, g, r, i, z) to image more than 14 000 deg^2 of the northern sky. Although its depth is relatively shallow compared with modern surveys ($r \approx 22.2$ mag), its enormous spectroscopic observations provided spectroscopic redshifts for millions of galaxies, forming the main training and validation samples on which the development of machine-learning photometric redshifts relied for many years, and becoming the cornerstone of photometric-redshift studies of the low-redshift Universe ($z < 0.7$)[120]. However, its relatively shallow photometric depth and only five optical bands make it almost unusable for precise studies at higher redshifts. The subsequent Pan-STARRS1 survey (PS1) used five filters (g, r, i, z, y) to conduct observations deeper than SDSS ($r \approx 23.2$ mag) over the entire northern sky, about 30 000 deg^2, providing larger samples and deeper, more homogeneous multi-band data.

Another survey strategy adopts deep, multi-wavelength survey modes, such as COSMOS and CANDELS. Surveys of this type cover extremely small areas ($0.22 \sim 2$ deg^2), but within these sky regions they assemble ultra-deep, full-band data from X-rays to radio, obtained from the Hubble and Spitzer space telescopes as well as major ground-based observatories. These deep, panchromatic data sets provide the most reliable photometric redshifts currently available for faint galaxies, and almost all of the most advanced high-redshift photometric-redshift algorithms have been validated on these data sets.

Modern deep, wide-area surveys are a compromise between the first two types, such as DES, KiDS, and HSC-SSP. Such surveys aim to unify the large area required for cosmological applications with high-precision photometric redshifts. They often cover sky areas of several thousand square degrees, with depths

far exceeding SDSS, with wavelength coverage also extending to the near-infrared Y band. These surveys are currently the main force in measuring the properties of dark energy by means such as weak gravitational lensing. Their photometric-redshift precision is required to reach an unprecedented level of 1%–2% within the "cosmological sample" ($0.2 < z < 1.3$). Table 2 summarizes a comparison of the main characteristics and photometric-redshift precision of several major current photometric survey projects.

Table 2 Main characteristics and photometric-redshift precision of selected photometric survey projects

Survey project	Survey area / deg^2	Filter system	5σ limiting magnitude / mag	Typical machine-learning photometric-redshift precision σ
SDSS	about 14 500	u, g, r, i, z	$r \approx 22.2$	about 0.02[121]

Survey project	Survey area / deg^2	Filter system	5σ limiting magnitude / mag	Typical machine-learning photometric-redshift precision σ
Pan-STARRS1	about 30 000	g, r, i, z, y	$r \approx 23.2$	about 0.03[122]
COSMOS	about 2	More than 30 bands(ultraviolet to radio)	$i \approx 26+$	about 0.007–0.015[123]
CANDELS	about 0.22	Near-infrared + optical	$H \approx 27$	about 0.01–0.02[124]
DES	about 5 000	g, r, i, z, Y	$r \approx 24.4$	about 0.03–0.04[125]
KiDS	about 1 350	u, g, r, i	$r \approx 24.8$	about 0.019–0.022(bright galaxies)[90]
HSC-SSP (wide-field survey)	about 1 400	g, r, i, z, y	$r \approx 26.0$	about 0.04–0.05[126]
DESI Legacy	about 14 000	g, r, z, (+WISE)	$r \approx 23.9$	about 0.014–0.026[73,127]
S-PLUS	about 9 300	5 broad bands + 7 narrow bands	$r \approx 21$ (broad band)	about 0.019–0.023[128]
LSST	about 18 000	u, g, r, i, z, y	$r \approx 27.5$ (stacked)	required to be less than 0.02
Euclid (wide-field survey)	about 14 000	1 broad band (VIS) +3 near-infrared bands (Y, J, H) +ground-based u, g, r, i, z	$I \approx 24.5$	required to be less than 0.05[129]

In addition to survey sky area, the precision of photometric redshifts is also limited by the spectral resolution of the photometry. Broad-band systems—

such as the u, g, r, i, z bands of SDSS and the g, r, i, z, Y bands of DES—have 4–6 filters, high observing efficiency, and can collect large numbers of photons, thereby enabling deep surveys over large sky areas. However, their sampling of the SED is very coarse, which limits the precision to $\sigma_z/(1+z) \approx 0.03$–$0.1$. Narrow-band systems—such as the 12 filters of S-PLUS and the 40 filters of PAUS—provide much higher spectral resolution ($R \approx 20$–50). This enables them to resolve features such as the 4 000 Å break and even detect strong emission lines, as if each object were equipped with an extremely low-resolution spectrograph. In this case, the precision can be significantly improved to $\sigma_z/(1+z) \approx 0.003$–$0.02$. The price is that, for a fixed exposure time, narrow-band filters have much lower signal-to-noise ratios, thereby limiting survey depth[130,131].

Even when the spectral resolution is sufficiently high, galaxy blending also affects photometric precision[132]. As surveys such as HSC and LSST reach unprecedented depths, the projected surface density of galaxies on the sky increases sharply, so that as many as 50% or even more objects physically overlap in two-dimensional images. When galaxies overlap, standard photometric algorithms have difficulty separating their light, leading to erroneous measurements of the fluxes, colors, and shapes of individual components. The resulting photometry may be a mixture of multiple objects at completely different redshifts, which can lead to an entirely erroneous photometric-redshift estimate and significantly increase the number of outliers. This has been recognized as a major source of systematic error in LSST weak-lensing cosmology studies[133,134]. Image quality directly affects the blending problem, especially the atmospheric "seeing," which determines the size of the point-spread function (PSF). Better seeing means that galaxies appear sharper and are less likely to overlap, making their photometric results more reliable. Ground-based survey projects located at sites with excellent seeing—such as HSC' s $0.6''$—have a significant advantage in mitigating blending compared with surveys at poorer sites. However,

Moreover, even the superior apparent magnitude depth of HSC cannot completely eliminate this problem; this highlights the advantage of the spatially diffraction-limited imaging from HST/CANDELS in providing uncontaminated, real samples.

These factors that affect photometric-redshift performance do not exist in isolation; rather, they are coupled to one another and sometimes amplify each other' s negative effects. This is especially true of observing depth, which creates a contradiction in observing strategy. The main driver of modern surveys is the pursuit of deeper observations to increase the number of galaxies and thereby improve the statistical power of cosmological measurements. However, increasing depth directly leads to two negative consequences: first, a larger fraction of galaxy samples has a lower signal-to-noise ratio; second, the projected density of galaxies increases, causing the fraction of image blending to rise sharply. Image blending contaminates the photometric data, while a low signal-to-noise ratio also means that the photometry itself is full of noise. Thus, the very act of pursuing deeper observations to improve cosmological measurements simultane-

ously reduces the basic data quality of newly added objects. These degraded photometric data ultimately lead to poorer photometric redshifts and introduce systematic errors, which may offset the statistical gains initially obtained by increasing the number of galaxies. This feedback loop is a problem for current surveys, but at the same time it is driving the development of new algorithms that attempt to model photometry, blending, and photometric redshifts simultaneously.

At present, no single survey can provide all the data required for precision cosmology, and different survey programs are used to cross-calibrate one another. For example, deep fields (such as COSMOS) are used to calibrate photometric redshifts for hundreds of millions of faint galaxies without spectroscopic redshifts in wide-area surveys (such as DES). This method usually involves SOMs. First, all available filters are used, and a multidimensional "color space" is defined by combining the wide-area survey with deep near-infrared data. The SOM is then used to map this high-dimensional color space onto a two-dimensional grid; galaxies with similar colors are placed in the same "cell," and deep multi-band photometric data from fields such as COSMOS are used to fill this mapping. Because these fields have highly reliable photometric or spectroscopic redshifts, each cell is associated with a known redshift distribution. A galaxy in the wide-area survey is then placed into the corresponding cell according to its observed colors, and its redshift distribution is assumed to be that cell' s distribution as determined from the COSMOS data. This process effectively "transfers" high-quality redshift information from small-area deep fields to large-area wide surveys, thereby calibrating the overall redshift distribution required for weak-lensing tomographic imaging. However, this process is not perfect: photometry differs systematically among surveys, and the deep fields themselves may not be fully representative. These effects must be modeled, usually by injecting simulated galaxies into real images to measure selection effects and photometric biases accurately—for example, the Balrog code in DES[^135].

On the other hand, one cannot simply train a machine-learning model on the data from one survey and then apply it to another. Even if the filters have similar nominal definitions, subtle differences in filter transmission curves, photometric calibration, and data-processing pipelines can produce systematic biases in the measured colors. Machine-learning models learn these survey-specific systematic errors and fail when applied to data with different systematics. This highlights the need for extremely careful cross-calibration among different surveys and for adopting consistent data-processing pipelines. To make photometric-redshift calibration robust, one must either process all data using exactly the same pipeline or accurately model the bias-transfer functions between different data sets[^55].

5 Summary and Outlook

Astronomy lies at the frontier of data-intensive science. Forthcoming large photometric survey projects are expected to generate data at the PB or even EB

scale, creating an urgent need for machine-learning methods to achieve efficient data processing and analysis. At present, deep learning has demonstrated the potential to predict photometric redshifts directly from pixel-level data in calibrated images; this strategy can not only avoid traditional photometric parameter

selection biases, but can also mine implicit morphological information in images—such as surface brightness and color gradients—thereby providing a new pathway for addressing the decline in prediction accuracy and parameter degeneracies at high redshift[^106].

The advantages of data-driven machine-learning methods are also reflected in many other aspects. Unsupervised models can identify the optimal applicable regions in multidimensional feature space, helping to assess the balance of data distributions between photometric and spectroscopic spaces; by combining Bayesian statistics with mixture models, joint prediction of redshift, galaxy stellar mass, and SFR can be achieved[^26][^28]; and techniques such as SOM can effectively detect outliers and unreliable spectroscopic samples in the training data[^35].

By combining empirical models, SED template fitting, and Bayesian inference in hybrid approaches, future development will focus on cross-method integration and standardized evaluation. The accuracy of redshift estimation will be improved and extended to high-redshift regimes. Meanwhile, large-scale photometric surveys, through unified data and evaluation metrics, have promoted fair comparisons among different methods; for example, in LSST, PIT and QQ plots have been used to verify the effectiveness of deep-learning models in PDF estimation[^32].

The deep integration of computational science and astronomy will be an important strategy for coping with the massive astronomical data sets of the future. Its goal is not only to improve the accuracy of photometric redshifts, but also to provide reliable foundational tools for in-depth frontier scientific research, including cosmological-parameter measurements and probes of the dark-matter distribution[^91].

References:

[1] Baum W A. IAU Symposium, 1962, 15: 390

[2] Butchins S A. A&A, 1981, 97(2): 407

[3] Connolly A J, Csabai I, Szalay A S, et al. AJ, 1995, 110: 2655

[4] York D G, Adelman J, Anderson J, John E, et al. AJ, 2000, 120(3): 1579

[5] Scoville N, Aussel H, Brusa M, et al. ApJS, 2007, 172(1): 1

[6] Dark Energy Survey Collaboration, Abbott T, Abdalla F B, et al. MNRAS, 2016, 460(2): 1270

[7] de Jong J T A, Kuijken K, Applegate D, et al. The Messenger, 2013, 154: 44

[8] Aihara H, Arimoto N, Armstrong R, et al. PASJ, 2018, 70: S4

[9] Kauffmann O B, Le Fèvre O, Ilbert O, et al. A&A, 2020, 640: A67

[10] Laureijs R, Amiaux J, Arduini S, et al. https://arxiv.org/abs/1110.3193, 2011

[11] Abell P A, Allison J, Anderson S F, et al. https://arxiv.org/abs/0912.0201, 2009

[12] Green J, Schechter P, Baltay C, et al. https://arxiv.org/abs/1208.4012, 2012

[13] Gong Y, Liu X, Cao Y, et al. ApJ, 2019, 883(2): 203

[14] Witten C, Laporte N, Martin-Alvarez S, et al. Nature Astronomy, 2024, 8: 384

[15] Ilbert O, Arnouts S, McCracken H J, et al. A&A, 2006, 457(3): 841

[16] Brammer G B, van Dokkum P G, Coppi P. ApJ, 2008, 686(2): 1503

[17] Bruzual G, Charlot S. MNRAS, 2003, 344(4): 1000

[18] Assef R J, Kochanek C S, Brodwin M, et al. ApJ, 2010, 713(2): 970

[19] Retana-Montenegro E, Röttgering H J A. A&A, 2020, 636: A12

[20] Holwerda B W, Hsu C C, Hathi N, et al. MNRAS, 2024, 529(2): 1067

[21] Lu J, Luo Z, Chen Z, et al. MNRAS, 2024, 527(4): 12140

[22] Dennis M T, Hu E M, Cowie L L. ApJ, 2025, 983(2): 173

[23] Dey B, Andrews B H, Newman J A, et al. MNRAS, 2022, 515(4): 5285

[24] Feroz F, Hobson M P, Cameron E, et al. The Open Journal of Astrophysics, 2019, 2(1): 10

[25] de Diego J A, Nadolny J, Bongiovanni Á, et al. A&A, 2021, 655: A56

[26] Bonjean V, Aghanim N, Salomé P, et al. A&A, 2019, 622: A137

[27] Wright E L, Eisenhardt P R M, Mainzer A K, et al. AJ, 2010, 140(6): 1868

[28] Mucesh S, Hartley W G, Palmese A, et al. MNRAS, 2021, 502(2): 2770

[29] Enia A, Bolzonella M, Pozzetti L, et al. A&A, 2024, 691: A175

[30] Desprez G, Paltani S, Coupon J, et al. A&A, 2020, 644: A31

[31] Bisigello L, Kuchner U, Conselice C J, et al. MNRAS, 2020, 494(2): 2337

[32] Schmidt S J, Malz A I, Soo J Y H, et al. MNRAS, 2020, 499(2): 1587

[33] Baron D. https://arxiv.org/abs/1904.07248, 2019

[34] Narendra A, Dainotti M G, Sarkar M, et al. A&A, 2025, 698: A92

[35] Razim O, Cavuoti S, Brescia M, et al. MNRAS, 2021, 507(4): 5034

[36] Stölzner B, Joachimi B, Korn A, et al. MNRAS, 2023, 519(2): 2438

[37] Jalan P, Bilicki M, Hellwing W A, et al. A&A, 2024, 692: A177

[38] Wright A H, Hildebrandt H, van den Busch J L, et al. A&A, 2020, 637: A100

[39] Brescia M, Salvato M, Cavuoti S, et al. MNRAS, 2019, 489(1): 663

[40] Broussard A, Gawiser E. ApJ, 2021, 922(2): 153

[41] Beck R, Dodds S C, Szapudi I. MNRAS, 2022, 515(4): 4711

[42] Sen S, Singh K P, Chakraborty P. New Astronomy, 2023, 99: 101959

[43] Qu H, Sako M. ApJ, 2023, 954(2): 201

[44] Crenshaw J F, Kalmbach J B, Gagliano A, et al. AJ, 2024, 168(2): 80

[45] Zhou X, Gong Y, Zhang X, et al. ApJ, 2024, 977(1): 69

[46] Pathi I M, Soo J Y H, Wee M J, et al. JCAP, 2025, 2025(1): 097

[47] Hoyle B, Rau M M, Zitlau R, et al. MNRAS, 2015, 449(2): 1275

[48] Baron D, Poznanski D. MNRAS, 2017, 465(4): 4530

[49] Reis I, Baron D, Shahaf S. AJ, 2019, 157(1): 16

[50] Lin Q, Ruan H, Fouchez D, et al. A&A, 2024, 691: A331

[51] Cavuoti S, Brescia M, D' Abrusco R, et al. MNRAS, 2014, 437(1): 968

[52] Jones E, Singal J. A&A, 2017, 600: A113

[53] Humphrey A, Cunha P A C, Paulino-Afonso A, et al. MNRAS, 2023, 520(1): 305

[54] Li C, Zhang Y, Cui C, et al. AJ, 2024, 168(6): 233

[55] Janiurek L, Hendry M A, Speirits F C. MNRAS, 2024, 533(3): 2786

[56] Ye G, Zhang H, Wu Q. ApJS, 2024, 275(1): 19

[57] Huang J, Luo B, Brandt W N, et al. ApJ, 2025, 979(2): 107

[58] Kim T, Sohn J, Hwang H S, et al. ApJS, 2025, 277(2): 41

[59] Carrasco K M, Brunner R J. MNRAS, 2013, 432(2): 1483

[60] Reza M. New Astronomy, 2025, 115: 102316

[61] Han B, Qiao L N, Chen J L, et al. Research in Astronomy and Astrophysics, 2021, 21(1): 017

[62] Curran S J, Moss J P, Perrott Y C. MNRAS, 2021, 503(2): 2639

[63] Luken K J, Norris R P, Park L A F, et al. Astronomy and Computing, 2022, 39: 100557

[64] Luken K J, Norris R P, Wang X R, et al. Publications of the Astronomical Society of Australia, 2023, 40: e039

[65] Soo J Y H, Joachimi B, Eriksen M, et al. MNRAS, 2021, 503(3): 4118

[66] Naidoo K, Johnston H, Joachimi B, et al. A&A, 2023, 670: A149

[67] Ansari Z, Agnello A, Gall C. A&A, 2021, 650: A90

[68] Teixeira G, Bom C R, Santana-Silva L, et al. Astronomy and Computing, 2024, 49: 100886

[69] Zhou X, Gong Y, Meng X M, et al. MNRAS, 2022, 512(3): 4593

[70] Yao L, Qiu B, Luo A L, et al. MNRAS, 2023, 523(4): 5799

[71] Ait O R, Arnouts S, Pasquet J, et al. A&A, 2024, 683: A26

[72] Jones E, Do T, Li Y Q, et al. ApJ, 2024, 974(2): 159

[73] Zhou X, Li N, Zou H, et al. MNRAS, 2025, 536(3): 2260

[74] Mu Y H, Qiu B, Zhang J N, et al. Research in Astronomy and Astrophysics, 2020, 20(6): 089

[75] Li M, Gao Z, Qiu B, et al. MNRAS, 2021, 506(4): 5923

[76] Zhou X, Gong Y, Meng X M, et al. Research in Astronomy and Astrophysics, 2022, 22(11): 115017

[77] Ejaz Awan S, Bennamoun M, Sohel F, et al. Neurocomputing, 2021, 453: 164

[78] Luo Z, Tang Z, Chen Z, et al. MNRAS, 2024, 531(3): 3539

[79] Fotopoulou S, Paltani S. A&A, 2018, 619: A14

[80] Alam S, Albareti F D, Allende P C, et al. ApJS, 2015, 219(1): 12

[81] Carrasco K M, Brunner R J. MNRAS, 2014, 442(4): 3380

[82] Duncan K J, Brown M J I, Williams W L, et al. MNRAS, 2018, 473(2): 2655

[83] Crenshaw J F, Connolly A J. AJ, 2020, 160(4): 191

[84] Li Y, Fu L, Chen Z, et al. Research in Astronomy and Astrophysics, 2025, 25(5): 055021

[85] Leistedt B, Hogg D W. ApJ, 2017, 838(1): 5

[86] Tanigawa S, Glazebrook K, Jacobs C, et al. MNRAS, 2024, 530(2): 2012

[87] Dalmasso N, Pospisil T, Lee A B, et al. Astronomy and Computing, 2020, 30: 100362

[88] Hildebrandt H, Arnouts S, Capak P, et al. A&A, 2010, 523: A31

[89] Cavuoti S, Brescia M, Longo G, et al. A&A, 2012, 546: A13

[90] Bilicki M, Hoekstra H, Brown M J I, et al. A&A, 2018, 616: A69

[91] Mandelbaum R. ARA&A, 2018, 56: 393

[92] Amaro V, Cavuoti S, Brescia M, et al. MNRAS, 2019, 482(3): 3116

[93] Harrison D, Sutton D, Carvalho P, et al. MNRAS, 2015, 451(3): 2610

[94] Izbicki R, Lee A B. Electronic Journal of Statistics, 2017, 11(2): 2800

[95] Ma Z, Hu W, Huterer D. ApJ, 2006, 636(1): 21

[96] Blake C, Bridle S. MNRAS, 2005, 363(4): 1329

[97] Csabai I, Budavári T, Connolly A J, et al. AJ, 2003, 125(2): 580

[98] Bolzonella M, Miralles J M, Pelló R. A&A, 2000, 363: 476

[99] Wittman D, Riechers P, Margoniner V E. ApJL, 2007, 671(2): L109

[100] D' Isanto A, Polsterer K L. A&A, 2018, 609: A111

[101] Fu L, Liu D, Radovich M, et al. MNRAS, 2018, 479(3): 3858

[102] Cappelluti N, Predehl P, Böhringer H, et al. Memorie della Societa Astronomica Italiana Supplementi, 2011, 17: 159

[103] Humphrey A, Bisigello L, Cunha P A C, et al. A&A, 2023, 671: A99

[104] Salvato M, Ilbert O, Hoyle B. Nature Astronomy, 2019, 3: 212

[105] Cantiello M, Blakeslee J P. https://arxiv.org/abs/2307.03116, 2023

[106] Pasquet J, Bertin E, Treyer M, et al. A&A, 2019, 621: A26

[107] Breiman L. Machine Learning, 2001, 45(1): 5

[108] Carrasco K M, Brunner R J. MNRAS, 2014, 438(4): 3409

[109] Salvato M, Hasinger G, Ilbert O, et al. ApJ, 2009, 690(2): 1250

[110] Ananna T T, Salvato M, LaMassa S, et al. ApJ, 2017, 850(1): 66

[111] Norris R P, Salvato M, Longo G, et al. PASP, 2019, 131(1004): 108004

[112] Saxena A, Salvato M, Roster W, et al. A&A, 2024, 690: A365

[113] Salvato M, Wolf J, Saxena A, et al. EAS2024, European Astronomical Society Annual Meeting, Padova, Italy: European Astronomical Society, 2024: 2540

[114] Crenshaw J F, Leistedt B, Graham M L, et al. https://arxiv.org/abs/2503.06016, 2025

[115] Kunsági-Máté S, Beck R, Szapudi I, et al. MNRAS, 2022, 516(2): 2662

[116] Luo Z, Li Y, Lu J, et al. MNRAS, 2024, 535(2): 1844

[117] Henghes B, Thiyagalingam J, Pettitt C, et al. MNRAS, 2022, 512(2): 1696

[118] Beck R, Lin C A, Ishida E E O, et al. MNRAS, 2017, 468(4): 4323

[119] Sánchez C, Alarcon A, Bernstein G M, et al. MNRAS, 2023, 525(3): 3896

[120] Soo J Y H, Shuaili I Y K A, Pathi I M. American Institute of Physics Conference Series, 2023, 2756: 040001

[121] Beck R, Dobos L, Budavári T, et al. MNRAS, 2016, 460(2): 1371

[122] Tarrío P, Zarattini S. A&A, 2020, 642: A102

[123] Eriksen M, Alarcon A, Cabayol L, et al. MNRAS, 2020, 497(4): 4565

[124] Kodra D, Andrews B H, Newman J A, et al. ApJ, 2023, 942(1): 36

[125] Toribio S C L, de Vicente J, Sevilla-Noarbe I, et al. A&A, 2024, 686: A38

[126] Tanaka M, Coupon J, Hsieh B C, et al. PASJ, 2018, 70: S9

[127] Li C, Zhang Y, Cui C, et al. MNRAS, 2023, 518(1): 513

[128] Lima E V R, Sodré L, Bom C R, et al. Astronomy and Computing, 2022, 38: 100510

[129] Stanford S A, Masters D, Darvish B, et al. ApJS, 2021, 256(1): 9

[130] Laur J, Tempel E, Tamm A, et al. A&A, 2022, 668: A8

[131] Navarro-Gironés D, Gaztañaga E, Crocce M, et al. MNRAS, 2024, 534(2): 1504

[132] Melchior P, Joseph R, Sanchez J, et al. Nature Reviews Physics, 2021, 3(10): 712

[133] Boucaud A, Huertas-Company M, Heneka C, et al. MNRAS, 2020, 491(2): 2481

[134] Nourbakhsh E, Tyson J A, Schmidt S J, et al. MNRAS, 2022, 514(4): 5905

[135] Everett S, Yanny B, Kuropatkin N, et al. ApJS, 2022, 258(1): 15

Machine Learning for Photometric Redshift Estimation

LU Junhao1, LUO Zhijian1, CHEN Jianzhen1, ZENG Lujin1, SHU Chenggang1

(1. Shanghai Key Lab for Astrophysics, Shanghai Normal University, Shanghai 200234, China)

Abstract: Consequently, most imaging-based galaxy and quasar surveys rely heavily on photometric redshifts to provide redshift information for tens of millions to billions of celestial objects, enabling forefront research on the large-scale structure of the universe and the nature of dark energy. Against this background, machine learning (ML) methods have emerged as mainstream tools for obtaining high-precision photometric redshifts due to their efficiency and scalability. In the present paper, a comprehensive review of recent advances in the application of ML in photo-z estimation, including algorithmic classifications, model optimization strategies, and typical application scenarios, together with examples to show the technical characteristics and performance differences of various ML architectures, are summarized.

Key words: photometric redshift; machine learning; data analysis

Note: Figure translations are in progress. See original paper for figures.