Research in Astronomy and Astrophysics, 24:115019 (14pp), 2024 November https://doi.org/10.1088/1674-4527/ad86a6

Ensemble Learning for Stellar Classification and Radius Estimation from Multimodal Data

Zhi-Jie Deng[1], Sheng-Yuan Yu[2], A-Li Luo[3,4], Xiao Kong[3], and Xiang-Ru Li[1]
[1] School of Computer Science, South China Normal University, Guangzhou 510631, China; lixiangru@scnu.edu.cn
[2] School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China
[3] CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China; lal@nao.cas.cn
[4] School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 101408, China
Received 2024 April 30; revised 2024 August 14; accepted 2024 August 15; published 2024 November 14

Abstract

Stellar classification and radius estimation are crucial for understanding the structure of the Universe and stellar evolution. With the advent of the era of astronomical big data, multimodal data are available and theoretically effective for stellar classification and radius estimation. A problem is how to improve the performance of this task by jointly using the multimodal data. However, existing research primarily focuses on using single-modal data. To this end, this paper proposes a model, Multi-Modal SCNet, and its ensemble model Multimodal Ensemble for Stellar Classification and Regression (MESCR) for improving stellar classification and radius estimation performance by fusing two modality data. In this problem, a typical phenomenon is that the sample numbers of some types of stars are evidently more than others. This imbalance has negative effects on model performance. Therefore, this work utilizes a weighted sampling strategy to deal with the imbalance issues in MESCR. Some evaluation experiments are conducted on a test set for MESCR and the classification accuracy is 96.1%, and the radius estimation performance Mean of Absolute Error and σ are 0.084 dex and 0.149 $R_{\odot}$, respectively. Moreover, we assessed the uncertainty of model predictions, confirming good consistency within a reasonable deviation range. Finally, we applied our model to 50,871,534 SDSS stars without spectra and published a new catalog.

Key words: methods: data analysis – techniques: image processing – methods: statistical

1. Introduction

In recent years, various large-scale sky survey projects have been continuously carried out, such as the Sloan Digital Sky Survey (SDSS; York et al. 2000; Yanny et al. 2009), Gaia (Gilmore et al. 2012), LAMOST (Luo et al. 2015), and the Large Synoptic Survey Telescope (LSST; LSST Dark Energy Science Collaboration 2012). These large sky survey projects have provided us with an unprecedented amount of data. The big data enriches our understanding of the distribution of stars in the Universe, their physical properties and evolutionary process. To conduct astronomy research based on these big data, two essential tasks are to classify the observed star data and estimate their parameters.

Based on the Morgan-Keenan classification method (Morgan & Keenan 1973) and surface temperature, stars are classified into O, B, A, F, G, K, M from hot to cold. Each category is further classified into ten subclasses from 0 to 9 in case of surface temperature decrease. Historically, the classification task was often conducted by astronomers through visual inspection. However, it is infeasible in the era of big data. Therefore, it is necessary to develop some automated classification methods, such as template matching (e.g., Duan et al. 2009) and machine learning methods (e.g., Sharma et al. 2020).

The stellar radius is an important parameter for understanding stellar morphology. Methods for measuring the stellar radius can be divided into direct and indirect types. Direct methods include interferometry or eclipsing binary fitting, etc. Interferometry allows astronomers to directly measure the angular diameter of stars and calculate their actual radius by combining known distances (Monnier 2003; van Belle & von Braun 2009). The eclipsing binary fitting method obtains physical information about the star by analyzing the light curve of the eclipsing binary (Blay & Lovekin 2015; Kallrath 2022). Indirect methods include deriving the radius by analyzing the brightness, temperature, and distance of stars or using empirical relations based on observational data for fitting (e.g., Moya et al. 2018). However, the accuracy of these methods is limited by the models and assumptions that they rely on, and the accumulation of errors may increase the uncertainty of the results. In recent years, machine learning methods have become increasingly popular for automatically estimating stellar parameters (e.g., Flores et al. 2023).

The machine learning methods mentioned above typically use data from a single modality for classification or parameter estimation. With large-scale sky survey projects continuing to develop, we can now obtain data from different bands and

types of instruments. This diversity of data provides more comprehensive and richer information. However, how to integrate data from multiple modalities, making the information from different data types complementary to achieve higher accuracy has become a key topic. Fortunately, some research has focused on this area (e.g., Hong et al. 2022; Kang et al. 2023; Zhang et al. 2023). However, there are not any effects until now for stellar classification and stellar radius estimation. Therefore, this work focuses on establishing a machine learning model to improve the stellar classification and stellar radius estimation performance by multimodal fusing.

The structure of this paper is as follows: Section 2 discusses the data acquisition process used in this study, Section 3 discusses the principles of the MESCR solution proposed in this study, Section 4 presents the performance of our model and related experiments, Section 5 applies MESCR to the SDSS photometric images to generate a new catalog, and Section 6 concludes the paper.

2. Data

In this study, the neural network accepts parallax and proper motion along with photometric images as inputs, simultaneously performing stellar classification and radius estimation. The proper motion, parallax, and radius of stars can be obtained from Gaia, while photometric images and their spectral types can be obtained from SDSS. Therefore, the reference set was obtained by cross-matching the SDSS Data Release (DR) 17 catalog and the Gaia DR3 catalog. Subsequent sections will detail the process of establishing the reference data set and the relevant data preprocessing steps.

2.1. Reference Dataset Based on Common Observations from SDSS DR17 and Gaia DR3

The Sloan Digital Sky Survey (SDSS) is one of the largest optical sky survey projects to date, dedicated to creating a detailed 3D map that covers about one-third of the sky. SDSS uses CCD filters to collect photometric images in five optical bands: u, g, r, i, and z. The photometric images and spectra from SDSS can be accessed online via the Casjobs server.[5] Shi et al. (2023) selected a subset of objects from SDSS DR17 by making the sample numbers of various types of stars as equal as possible, and established a novel catalog (the Shi catalog). The type balance in the Shi catalog is helpful for the machine learning model establishing. Actually, Shi et al. (2023) established their catalog by cross-matching the SDSS DR17 SpecPhoto and SpecObj catalogs. To ensure the quality of the data, they further filtered out targets fainter than the limiting magnitudes, compiling a data set of 46,245 stars. They used the catalog for the task of stellar classification, and achieved an accuracy of 86.1%. The Shi catalog, however, does not include parallax, proper motion or radius, all of which are necessary for this study. Therefore, this work established a reference set of some common observations from SDSS DR17 and Gaia DR3 by cross-matching the Shi catalog and Gaia DR3.

The Gaia sky survey provides astrometric, photometric, and spectroscopic samples of nearly 2 billion stars in the Milky Way, as well as important samples of extragalactic and solar system objects. The Gaia Archive offers a rich data set that includes calculated positions, parallaxes, proper motions, radial velocities, and brightness measurements of stars and other celestial bodies, accessible online via the Gaia Archive server.[6] Our data set was constructed from Gaia DR3 (Gaia Collaboration et al. 2023). The parallax (*parallax*) and proper motion (*pm*) used in this paper are from the Gaia source catalog, whereas the stellar radius (*radius_flame*) is from the astrophysical parameters catalog. For the stellar radius, we choose to use the stellar radius calculated from the Stefan–Boltzmann law, derived from the star's temperature and luminosity. Luminosity of the star from Final Luminosity Age Mass Estimator (FLAME) using G band magnitude, extinction, distance, and a bolometric correction (Fouesneau et al. 2023). The extinction calculation is performed by the GSP-Phot module, which analyzes BP and RP spectral data in conjunction with the star's apparent G magnitude and parallax. This analysis is based on model fitting from the Aeneas best library that achieves the highest goodness-of-fit value. The GSP-Phot module uses the Markov Chain Monte Carlo (MCMC) method to obtain the median value of multiple samples to calculate the extinction value in the G band (Ulla et al. 2022).

Due to some bright sources from Gaia that may saturate SDSS, we used the *flags* from SDSS DR17 to filter out the saturated sources in the Shi catalog before performing the cross-match. Then, we use the Shi catalog to cross-match with the Gaia source and astrophysical parameters catalogs based on R.A. and decl. through the Gaia Archive server, with a cross-match radius of $3''$. However, upon examining the radius distribution after cross-match, we found that the radii of O and B-type stars did not match the sample. Despite further cross-matching with the SIMBAD astronomical database (Wenger et al. 2000) to refine the data set, the errors remained significant. To reduce data errors, we decided to exclude O and B-type stars from the data set and limit our research to A, F, G, K, and M-type stars.

To further obtain a more reliable catalog, we applied the following criteria to the cross-matched catalog to balance data quality and quantity: (1) $parallax\_over\_error > 2$ to select more accurate parallax data. (2) $pmra\_error < 0.2$ and $pmdec\_error < 0.2$ to exclude proper motion data with large errors. (3) $ruwe < 1.4$ to exclude potentially problematic observational data, such as binaries or other complex background interferences.

[5] http://casjobs.sdss.org/casjobs/

[6] https://gea.esac.esa.int/archive/

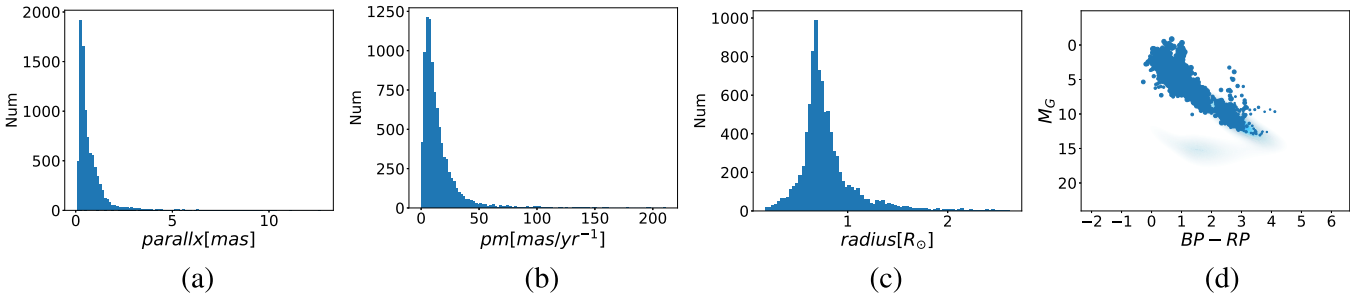


Figure 1. The distributions of the used samples in (a) the *parallax* space, (b) the *pm* space, (c) the *radius_flame* space, and (d) the color–magnitude diagram (CMD). The CMD indicates that the data primarily consists of main sequence stars. In (d), Gaia stars are used as a background, with the size of the points in the diagram representing the signal-to-noise ratio, and larger points indicating a higher signal-to-noise ratio.

Our final data set consists of 9180 stars: 3368 K-type stars, 2918 F-type stars, 1350 G-type stars, 918 A-type stars and 626 M-type stars. Their distributions are presented in Figure 1. It is worth noting that there are relatively more K and F-type stars than G, A and M-type stars, resulting in an imbalanced data set. This poses some challenges for model learning, with detailed solutions to be provided in Section 3.2.2. Moreover, we will further process the photometric images.

2.2. Photometric Image Preprocessing

The photometric images obtained from SDSS DR16 are typically uncropped, high-resolution images that contain a large amount of irrelevant information, such as background noise, other celestial bodies, or blank areas. Therefore, we first use the World Coordinate System (WCS) to locate the target star, and preliminarily crop the image around the star as the center to a size of 64×64. This greatly reduces the presence of irrelevant information and ensures that the target star part of the image is not cropped, preventing information loss. The photometric images include five bands: *g*, *r*, *i*, *u*, *z*. To maximize the utilization of each band's imagery, we repetitively use the *r*-band images to divide the five bands into two groups, *gri* and *urz* (Shi et al. 2023). Following the method described by Lupton et al. (2004), the cropped images from the *gri* and *urz* bands are converted into RGB images, as shown in Figure A1. They are subsequently referred to as *gri* and *urz* images respectively.

To reduce irrelevant information in the images, we need to perform more refined cropping on the *gri* and *urz* images. The algorithm starts expanding the ROI from the center of the image to check if it contains all information of the star. It calculates the position and size of each star in the *r*-band photometric image and then crops the *gri* and *urz* images accordingly. To complete this algorithm, first define the detection area side length l, pixel intensity threshold α, and the invalid pixel ratio threshold β. When the pixel intensity of a point in the image is less than or equal to α, we call this pixel "invalid;" otherwise, it is "valid." For each 64×64 size *r*-band image, we initially set a detection ROI of size $l \times l$ at the center of the image, as shown in the left of Figure A2, and calculated the ratio of invalid pixels to total pixels within this area. If this ratio is higher than β, we expand the side length and continue detecting. Otherwise, the current ROI area is the target area, as shown on the right of Figure A2. Based on the results of multiple experiments, we set the initial ROI side length $l = 6$, pixel intensity threshold $\alpha = 0.05$, and invalid pixel ratio threshold $\beta = 0.2$. To improve the efficiency of image preprocessing, the algorithm only needs to calculate the newly added annular detection area in each iteration. After refined cropping, most image pixels fall within the 10–20 pixel range. To standardize the image size for neural network learning, we resize the cropped images to 64×64 pixels using bilinear interpolation. We also recognize that scaling from such a small size to 64×64 may introduce some degree of detail loss and artifacts. This issue will be further discussed in Section 4.3.2.

3. Method

A neural network is a hierarchical computational model consisting of a series of computing units. This computing unit is referred to as a neural cell and the characteristic of a neural network is the ability of automatic feature learning (Ma et al. 2022; Kang et al. 2023). This work developed a model, Multi-Modal SCNet (MMSCNet), and its ensemble model Multi-modal Ensemble for Stellar Classification and Regression (MESCR) to improve the accuracy of stellar classification and radius prediction by fusing multimodal data.

The core of this scheme is to improve overall performance by integrating the outputs of multiple sub-models as the final result through weighted sampling and ensemble learning. The following sections will detail the network architecture of MMSCNet, the design motivation and principles of the MESCR, and how to use the MESCR to achieve the final prediction objectives.

3.1. MMSCNet

The structure of MMSCNet is shown in Figure 2. The MMSCNet is composed of three main parts: image feature

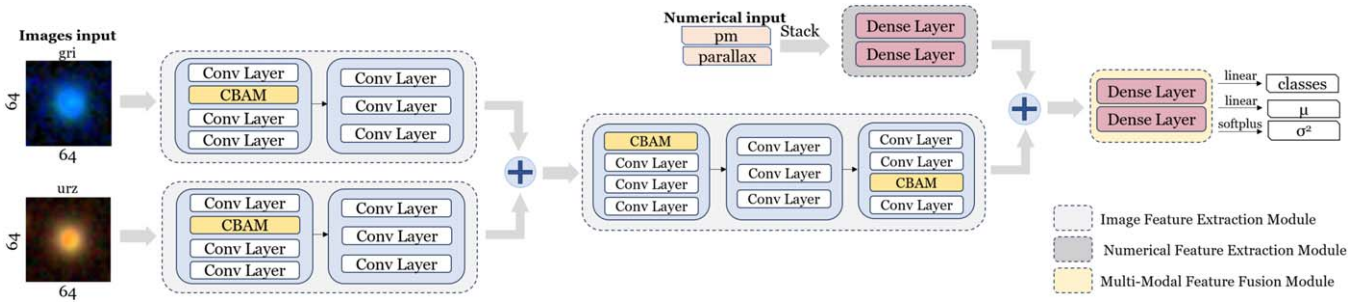


Figure 2. MMSCNet Structure Diagram. This network is used for stellar classification, star radius prediction, and outputting the uncertainty associated with the star radius. It has three input branches: one for receiving numerical data such as *parallax* and *pm*, and two for receiving RGB images synthesized from the *gri* and *urz* bands, respectively. Different background colors represent three distinct modules: image feature extraction module, numerical feature extraction module, and multimodal fusion module.

extraction module, numerical feature extraction module, and multimodal feature fusion module. Each of these modules will be introduced in turn.

In the image feature extraction module, we selected SCNet (Shi et al. 2023) as the backbone network due to its superior performance in stellar classification. SCNet integrates the CBAM attention mechanism (Woo et al. 2018) and is divided into two feature extraction stages. The first stage simultaneously accepts images from the *gri* and *urz* bands as inputs, while the second stage stacks the feature maps from both branches for further feature extraction. The numerical data inputs consist only of parallax and proper motion. To avoid overfitting while extracting features from this data, we used two dense layers for the numerical data feature extraction module. In the multimodal feature fusion module, the model concatenates the processed image feature vectors with the numerical feature vectors, forming a cross-modal feature vector. After the concatenation, dense layers with softplus and linear outputs provide the scores for each class, stellar radius, and the corresponding uncertainty.

3.2. Multimodal Ensemble for Stellar Classification and Regression

As previously mentioned, this study considers a comprehensive task involving both classification and regression. In the regression tasks, to derive the uncertainty of predictions and accurately estimate the probability density functions of physical parameters, deep ensemble methods are often required (Xiong et al. 2022; Li et al. 2023). In the classification tasks of this study, there exists an imbalance in the data set. The number of K and F-type stars significantly exceeds that of G, A and M-type stars, presenting a challenging task in machine learning (YANG & WU 2006). Numerous studies have been proposed to address such imbalanced classification tasks, including adjusting algorithm mechanisms (e.g., Li et al. 2018), resampling techniques, cost-sensitive methods (Wan & Yang 2020), and ensemble methods.

Resampling techniques are widely used due to their universality and effectiveness. Ensemble methods not only have significant effects in solving data imbalance problems (e.g., Krawczyk et al. 2016; Li et al. 2022) but also match the requirements of the regression tasks in this study. Therefore, we use methods based on the resampling techniques and ensemble methods as the basis for the MESCR. Specifically, we first perform repeatable weighted sampling on the imbalanced data set to construct multiple sub-data sets, then adopt an ensemble learning method similar to bagging (Breiman 1996). In the following two subsections, we will introduce the process of constructing the data set and the overall idea of MESCR, respectively.

3.2.1. Weighed Sampling and Data Augmentation

Weighted sampling is a probability sampling technique in which each sample is assigned a weight that reflects its probability of being selected. Typically, the inverse of the sample frequency is used as the weight, meaning that minority class samples will receive more attention. We first replicate the data set six times, corresponding to the data sets of five different classes of stars D_a, D_f,..,D_m, and one data set that does not favor any class D_{base}. The initial weight of a sample is set as the inverse of its class frequency. The weights in the D_{base} data set remain unchanged, while for the data set D_i, which targets a specific class, the weight of samples within that class is squared to further emphasize: $w_i' = 1/f_i^2$.

Subsequently, the weights of all samples are normalized to form the final sampling probability distribution: $P(x_i) = w_i'/\sum_{j=1}^{n} w_j'$. w_i' is the adjusted weight of the sample, and the denominator is the sum of all adjusted sample weights. This method gives additional attention to rare classes in an imbalanced data set, increasing the probability of these class samples being selected. During sampling, the weight of minority classes is further increased due to the squaring effect, while the weight of majority classes, which are already small,

does not lead to more severe data imbalance after being squared.

Ultimately, we obtain five data sets biased toward their respective classes D_a, D_f,..,D_m and one relatively balanced data set $D_{\rm base}$. However, this repeatable sampling operation will produce duplicate samples in our data set, potentially leading to issues such as overfitting. Therefore, we perform five types of data augmentation on randomly selected data, including rotation, flipping, adding Gaussian noise, cutout, and erase, with actual effects as shown in Figure A3. During training, data that has been sampled repeatedly will undergo data augmentation at least once, while other data will be randomly augmented with a certain probability. This method can increase the diversity of the data.

To validate the model's performance under real-world imbalanced data distribution, we performed repeatable weighted sampling only on the training and validation sets, without any processing on the test set. We first divided the original data set into a training set and a test set in an 8:2 ratio, resulting in a training set with 7344 samples and a test set with 1836 samples. Subsequently, we replicated the training set six times. For each copy, we applied the aforementioned weighted sampling and at least one data augmentation method for each image. This increased the number of samples in each training set to 9000. Finally, we divided this enhanced training set again in an 8:2 ratio to obtain a training set with 7200 samples and a validation set with 1800 samples, which were used for subsequent model training. The distribution of stellar spectral types in the resulting training, validation, and test sets is shown in Figure A4, while the distribution of *pm*, *parallax*, and *radius_flame* is shown in Figure A5.

3.2.2. MESCR Overall Framework and Training

The proposed MESCR scheme is an ensemble learning strategy akin to bagging, consisting of six identical sub-models. These models use different weights in the sampling methods to construct the data sets. During training, we apply the five data augmentation methods mentioned in Section 3.2.1 to the training samples. The final output is divided into two parts: classification results and regression results. The classification results are obtained by averaging the outputs of the six models, with weights determined by the models' performance on the validation set. Through experimentation, we set the weights as follows: {Model Base:0.1, Model A:0.2, Model F:0.1, Model G:0.2, Model K:0.15, Model M:0.25}. The ensemble stellar radius is determined by the average prediction of the six models, and the ensemble uncertainty $\hat{\sigma}^2(x)$ is calculated using the formula:

$$\hat{\sigma}^2(x) = \frac{1}{6}\sum_{i=1}^{6}(\sigma_{a,i}^2(x) + \mu_{a,i}^2(x)) - \hat{\mu}^2(x), \tag{1}$$

where $\mu_{a,i}^2(x)$ and $\sigma_{a,i}^2(x)$ represent the predicted stellar radius and its corresponding uncertainty for the ith MMSCNet, respectively, and $\hat{\mu}^2(x)$ is the mean of the stellar radii obtained by the six sub-models.

For the training of each sub-model, we adopted the same strategy. We initialized each model with random weights to increase diversity among the models. We use two different loss functions for stellar classification and radius estimation, respectively. For stellar classification, we use the cross-entropy loss function. For stellar radius, to enable the model to predict the probability density distribution, we use the negative log-likelihood of the normal distribution (Bialek et al. 2020), as the model's loss function:

$$\mathrm{Loss}(x|\theta) = \frac{\log \sigma_\theta^2(x)}{2} + \frac{(y - \mu_\theta(x))^2}{2\sigma_\theta^2(x)}, \tag{2}$$

where x and y represent the input data and the corresponding reference labels, respectively. $\mu_\theta(x)$ and $\sigma_\theta^2(x)$ represent the mean and variance of the Gaussian distribution predicted by the model, θ are the parameters of the MMSCNet model to be optimized.

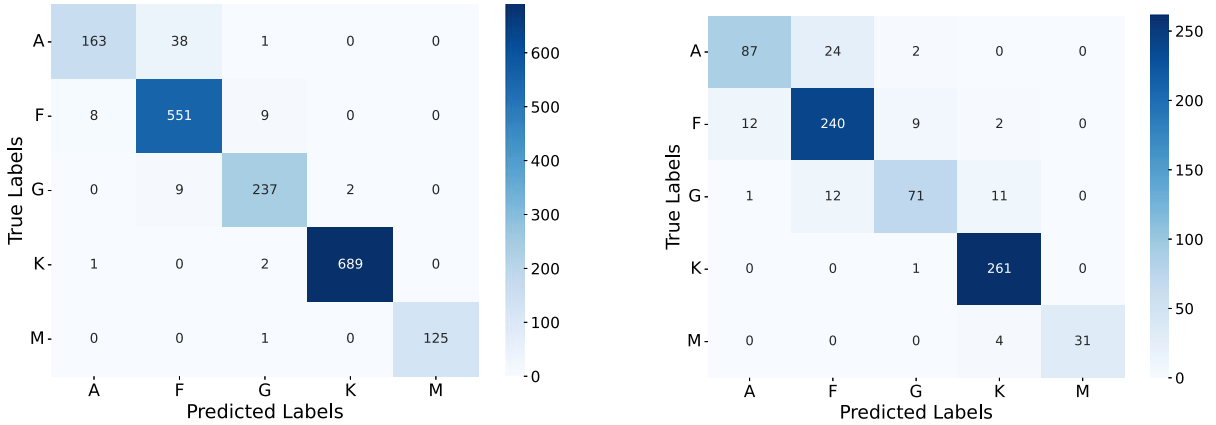
Then, we adopted the following hyperparameters: batch size set to 256; total training epochs set to 200; baseline learning rate set at 0.001, with learning rate updated using the cosine annealing algorithm, and performing a warm-up for the first three epochs, with the final learning rate set to 1e-5; using the SGD optimizer to optimize model parameters. The deep learning framework used was PYTORCH, with an RTX 4060ti GPU. Training each sub-model took approximately 1 hr, while training the entire MESCR took about 6 hr in total. Figure A6 shows the changes in loss values for each model of MESCR during the training process. As can be seen from the figure, the loss values for each model on both the training and validation sets gradually decrease with the increase in iterations. This indicates that the models' fitting and generalization capabilities improve progressively as the training progresses.

4. Results and Discussions

This section is to evaluate the performance of the proposed model (Section 4.1), and investigates the stellar radius estimation uncertainty (Section 4.2). Furthermore, we also studied the impact of image band combinations and resolutions (Section 4.3), and assessed the improvements by incorporating *parallax* and *pm* (Section 4.4).

4.1. Model Evaluation

After selecting the optimal hyperparameters for the model and completing the training, this section uses the test set to comprehensively evaluate the predictive performance of MESCR. To explore the model's classification and regression performance separately, we will use different evaluation metrics to assess the model.



(a) Comparing MESCR's predictions with the SDSS DR17 catalog (b) Comparing MESCR's predictions with the LAMOST DR9 catalog

Figure 3. The confusion matrix obtained from MESCR predictions. Each cell displays the number of predictions made by the model, where rows correspond to true labels and columns correspond to predicted labels.

Table 1
MESCR Classification Performance Evaluation by Calculating the Consistency of its Predictions with SDSS DR17 and LAMOST DR9

	SDSS DR17 Results					LAMOST DR9 Results				
	A	F	G	K	M	A	F	G	K	M
F1-score	87.1%	94.5%	95.1%	99.6%	99.6%	81.7%	89.0%	79.8%	96.7%	93.9%
Recall	80.7%	97.0%	95.6%	99.6%	99.2%	77.0%	91.3%	74.7%	99.6%	88.6%
Precision	94.7%	92.1%	94.8%	99.7%	99.2%	87.0%	87.0%	85.5%	93.9%	99.8%
Count	202	568	248	692	126	113	263	95	262	35

4.1.1. Classification Result Evaluation

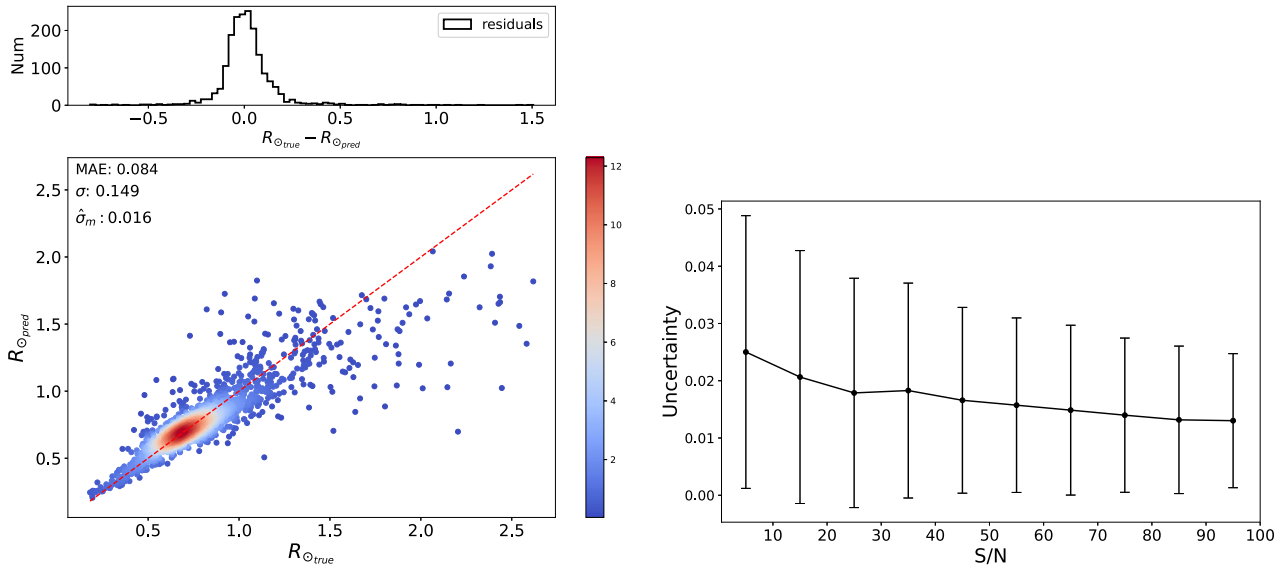
In this section, we evaluate the overall performance of the model using accuracy and measure the model's ability to predict different categories using recall, precision, and F1 scores. Detailed definitions can be found in Zeng et al. (2020).

MESCR achieved an overall accuracy of 96.1% on the test set. Table 1 (left) shows the F1 scores, recall, precision and corresponding counts for MESCR's predictions on the test set. Figure 3(a) shows the confusion matrix of the MESCR's predictions on the test set. The numbers on the main diagonal represent correct predictions by the model, while other non-zero positions indicate incorrect predictions. From Figure 3(a), it can be seen that most predictions on the main diagonal, indicating the model's good classification ability. As for the misclassified samples, they are mainly distributed in adjacent classes. For example, in the samples where F and G-type stars were misclassified, all of them were classified into adjacent categories. We used the samples from SDSS described in Section 2.1 for a homologous match with LAMOST DR9, obtaining a test set containing 768 stars. We found that this phenomenon also occurs between SDSS and LAMOST, where an A-type star in SDSS may be classified into an adjacent category (B or F) in LAMOST. We compared the MESCR prediction results with the LAMOST spectral classification results, and the accuracy was 89.7%. The corresponding confusion matrix is shown in Figure 3(b), and the F1-score, recall, and precision are presented in Table 1 (right). Although our training samples are from SDSS, the accuracy of each metric remains within an acceptable range when using LAMOST spectral classification results as the reference. Additionally, as shown in Figure 3(b), most of the misclassified results are categorized into adjacent classes, indicating that it is normal for models to classify into adjacent classes.

4.1.2. Regression Result Evaluation

In this section, we use the standard deviation (σ) between MESCR predictions and reference values and the Mean of Absolute Error (MAE) as the main evaluation metrics. Detailed definitions can be found in Zhong et al. (2024).

The resulting MAE is 0.084 dex, σ is 0.149 $R_{\odot}$, and the mean predicted 1σ uncertainty is 0.016 $R_{\odot}$. The lower subplot in Figure 4(a) shows a 1:1 scatter plot of the distribution of predicted and reference values, while the upper subplot shows a histogram of the distribution of differences between predicted and reference values. From the upper subplot, it is evident that most of the differences are distributed in the range of $[-0.5, 0.5]$, indicating that the deviations between the model's predictions and the actual values are small. In the scatter plot



(a) Consistency of MESCR predictions with references

(b) Dependencies of stellar radius estimation on S/N

Figure 4. MESCR performance evaluation. (a) The dashed line represents the theoretical consistency between predictions and references, the color indicates the density of samples; The subplot above presents a histogram of the differences between predictions and references. (b) The dots in the graph represent the mean value of 1σ uncertainty (predictive uncertainty $\sigma_{\rm pred}$) provided by MESCR for each signal-to-noise (S/N) ratio interval; The lines indicate the mean value of the standard deviation of the prediction results of the eight sub-models within each signal-to-noise ratio range, and is referred to as model's uncertainty $\sigma_{\rm net}$; The S/N represents the r-band signal-to-noise ratio.

below, the red dashed line represents the ideal prediction result, where the predicted values are exactly equal to the true values. From the plot, it can be seen that most of the data points are concentrated in the lower left corner and close to the main diagonal, indicating that the model performs well in most cases. Overall, the model's predictions closely match the reference values, and the high accuracy in reconstructing star properties indicates our model performs well.

4.2. Uncertainty Analysis of Stellar Radius

This study employs a deep integration approach for predicting stellar radii and is capable of providing corresponding predictive uncertainties $\sigma_{\rm pred}$. Moreover, the MESCR consists of six sub-models, each trained with a different training set and random initial weights, ultimately yielding six distinct prediction results $\{y_1, y_2,...,y_8\}$. We calculate the standard deviation among these six prediction results as the model's uncertainty, $\sigma_{\rm net}$.

Figure 4(b) shows the variation in uncertainty within different r-band SNR intervals. The dots in the graph represent the average uncertainty provided by MESCR for each SNR interval, serving as the predictive uncertainty $\sigma_{\rm pred}$. The line segments centered on the circles denote the average standard deviation among the results of the six sub-models within each SNR interval, serving as the model's uncertainty $\sigma_{\rm net}$. Overall, the uncertainty of MESCR is generally at a low level. As the S/N increases, both $\sigma_{\rm net}$ and $\sigma_{\rm pred}$ decrease. When S/N > 30, $\sigma_{\rm pred}$ stabilizes around 0.015 $R_\odot$. This indicates that the model performs better with higher S/N data, and its predictive ability is more stable for data with medium to high S/N.

4.3. The Impact of Image Bands and Resolution on Model Performance

In this section, we tested the model's performance using different image resolutions and various band combinations separately. In this section's experiments, we tested each band combination and resolution using a single MMSCNet, retraining with the $D_{\rm base}$ data set from Section 3.2.1.

4.3.1. The Influence of Image Bands on Model Performance

In this work, we employ the r-band image reuse strategy, inputting *gri* and *urz* bands into the model. However, we aim to explore if there are more optimal combination schemes to assess the performance of various band combinations for specific tasks. Therefore, we have selected the following band combinations for detailed analysis: [*gri* + *ruz*, *gri* + *iuz*, *gri* + *guz*, *gru* + *giz*, *gru* + *riz*, *gru* + *iuz*, *giu* + *grz*, *giu* + *riz*, *giu* + *ruz*, *giz* + *riu*, *giz* + *ruz*, *guz* + *riu*, *guz* + *riz*, *riu* + *grz*, *iuz* + *grz*]. These combinations all involve reusing one of the bands to form a new combination. In previous work, Wu & Boada (2019) created three-channel images by averaging two channels (for example, for *gr*-bands, the blue and red channels correspond to g and r-bands, respectively, while the green channel is the average of the two). Hence, we also adopted a

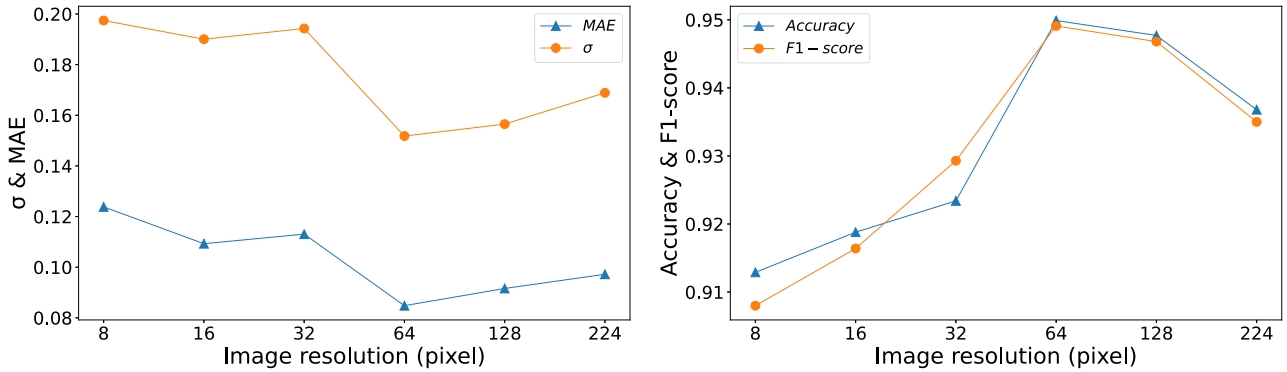


Figure 5. Impact of different image resolutions as input on the classification and regression performance of the model. Classification results are evaluated using accuracy and F1-score, while regression results are assessed using MAE and σ.

Table 2
When Using Data from Different Modalities as Input, the F1 Scores for Each Star Type, as well as the Overall Classification Accuracy, MAE, and σ, with the Best Values Highlighted in Bold

Input	A	F	G	K	M	Accuracy	MAE	σ
Photometric image	83.6%	91.9%	93.2%	99.6%	**99.6%**	94.5%	0.090	0.159
Photometric image + *parallax*	84.1%	93.7%	94.8%	**99.7%**	**99.6%**	95.6%	0.085	0.156
Photometric image + *pm*	83.7%	93.6%	94.0%	99.5%	**99.6%**	95.2%	0.088	0.152
Photometric image + *parallax* + *pm*	**87.1%**	**94.5%**	**95.1%**	99.6%	**99.6%**	**96.1%**	**0.084**	**0.149**

similar approach to test the following band combinations: [*gri* + *uz*, *gru* + *iz*, *grz* + *iu*, *giu* + *rz*, *giz* + *ru*, *guz* + *ri*, *riu* + *gz*, *riz* + *gu*, *ruz* + *gi*, *iuz* + *gr*].

The experimental result in Table A1 clearly shows the significant impact of different band combinations on model performance. In terms of stellar classification, the accuracy fluctuated from a low of 92.7% (with the *giz* + *ru* combination) to a high of 95.7% (with the *gru* + *iuz* combination), indicating a potential performance improvement of up to 2.4%. Similarly, the MAE in regression tasks also showed significant variability, from the best 0.077 (with the *guz* + *ri* combination) to the worst 0.087 (with the *guz* + *riu* combination), a difference of nearly 11.5%. These results confirm the decisive role of band selection for model efficacy.

4.3.2. The Influence of Image Resolution on Model Performance

Next, we fixed the band combination to *gri* + *urz* to explore the impact of image resolution on model performance. The experiment investigated various resolution settings ranging from 8×8 pixels to 224×224 pixels. We used bilinear interpolation to enlarge images and area interpolation to reduce the size of images, and retrained the model at different resolutions.

Figure 5 shows the impact of different resolutions as inputs on regression and classification results. From the figures, it can be seen that within the 8–128 range, as the resolution increases, the model's accuracy improves to some extent. This suggests that within a certain range higher resolution images can provide more detailed information, ultimately aiding the model in better learning and recognizing different features. However, when the image resolution is scaled up to 224×224 pixels, the accuracy of both classification and regression decreases. This may be due to the bilinear interpolation method introducing excessive blur or artifacts while enlarging the image, leading to a reduction in accuracy. In summary, to balance accuracy and computational efficiency, a resolution of 64×64 pixels is considered the best choice for this study.

4.4. The Impact of Excluding Parallax and Proper Motion Data

In this section, we keep the photometric images as fixed input and compare the impact of using different inputs on the model's performance. For different inputs, we retrained using MESCR, and the results are shown in Table 2.

For the classification results, photometric images provide unique color and brightness characteristics of different stars. Therefore, the model can achieve good classification results even when using only photometric images as input. After introducing parallax, the F1 scores for A, F, and G-type stars have all improved. This is because parallax indirectly infers the absolute brightness of stars and determines their actual distances, thereby improving the model's classification accuracy. As for proper motion, although it introduces dynamic information about the star's movement, the improvement in accuracy is smaller compared to the addition of parallax. When photometric images, parallax, and proper motion are used together as input, the model shows optimal performance in

overall classification accuracy. This indicates that integrating data from multiple modalities can significantly enhance the predictive accuracy of the model.

For radius estimation, our experimental results show that different inputs have a slight impact on MAE and σ. This may be due to the limited information provided by the additional parallax and proper motion for predicting stellar radii, and the model has a certain stability in predicting stellar radii. Considering that the stellar radius can be derived from temperature using the Stefan–Boltzmann law, we used the *Teff_gspphot* provided by Gaia DR3 as an input to the model to check whether temperature can constrain the radius. However, the results showed that the accuracy of the radius estimation did not improve significantly after including temperature. In future work, we will continue to explore other factors that may affect the accuracy of radius estimation and further optimize the model to improve the accuracy of radius estimation.

5. Application To Spectral-Free Data

In previous work, Clarke et al. (2020) used a random forest machine learning model to classify target sources into three categories: stars, galaxies, and quasars, ultimately creating a catalog containing a total of 111,395,468 objects (hereinafter referred to as the Clarke Catalog). Of these, 58,840,082 are stars, and none of these targets had corresponding spectra in SDSS. Shi et al. (2023) further cross-matched the Clarke Catalog with LAMOST, using LAMOST's classification results as reference values. They found that when the *class_prob_star* (the probability of being predicted as a star) was greater than 0.8, the accuracy was 0.99. To ensure data reliability, we further filtered the Clarke Catalog to include only stellar samples with *class_prob_star* greater than 0.8. We then cross-matched these with Gaia DR3 using TOPCAT (Taylor 2017) by R.A. and decl. within a radius of three arcseconds to obtain corresponding parallax and proper motion. The resulting catalog contains 50,871,534 stars without spectra but with associated parallax and proper motion. We applied MESCR to this catalog, resulting in a new catalog, the columns of which are shown in Table A2. It is important to note that none of the stars processed by MESCR exceed the limiting magnitude. The entire catalog is available online via the China-VO PaperData repository.

6. Conclusions

In this paper, we introduce an ensemble learning scheme named MESCR based on the MMSCNet. MESCR integrates multimodal data to simultaneously perform star classification and radius estimation. To obtain the final training samples, we first perform cross-matching to acquire the necessary data, then crop the original images to reduce irrelevant information. Subsequently, we weighted sampled the data set to obtain 6 different data sets and trained 6 sub-models separately, finally integrating the predictions of all sub-models to produce the final output. Our model achieved a classification accuracy of 96.1% on the test set, with the MAE for stellar radius at 0.084 dex, and σ at 0.149 $R_{\odot}$. Through uncertainty analysis, it further proves that our MESCR solution has good accuracy and robustness.

Next, we tested the impact of different band combinations and resolutions on model performance. We found that specific band combinations and appropriate image resolutions can significantly optimize model performance. Additionally, to quantitatively assess the impact of using data from different modalities as inputs on model accuracy, further testing was conducted. The results indicate that integrating various modalities of data can enhance the model's ability to distinguish between different types of stars. Ultimately, we applied MESCR to 50,871,534 stars without spectra and released a new catalog.

In our future work, we will further improve the precision of the model and explore more survey data, while also enhancing the accuracy of our catalog, thereby providing more reference value for astronomers and data analysis researchers.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (12261141689, 12273075, and 12373108), the National Key R&D Program of China No. 2019YFA0405502, and the science research grants from the China Manned Space Project with No. CMS-CSST-2021-B05.

Data Availability

The complete code of this study and its instructions for use will be published at https://github.com/yuu250/MESCR after this paper is accepted and finalized, for reference by astronomers and data processing personnel.

Appendix
Supplementary Figures and Tables

This appendix contains supplementary images and tables that are discussed and analyzed in the article. These materials are intended to provide more detailed visual evidence and data support to enhance the reader's understanding.

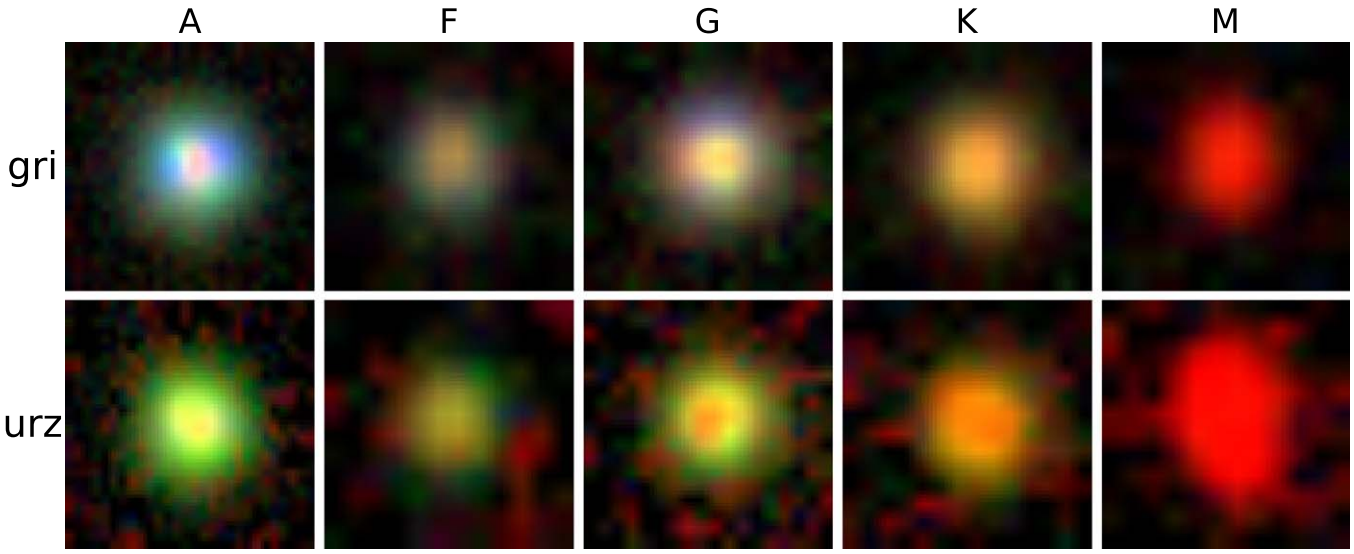


Figure A1. An example of synthesizing photometric images. The first row represents the RGB images synthesized from the *gri* bands, while the second row represents the RGB images synthesized from the *urz* bands. The *urz* band images exhibit more noise compared to the *gri* band images.

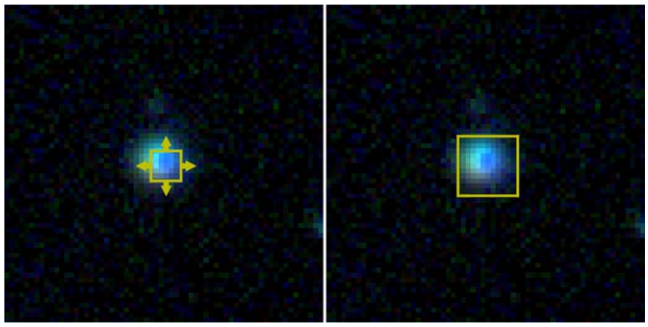


Figure A2. Image cropping schematic, the size of the image before cropping is 64 × 64. The left image shows the initial Region of Interest (ROI), with a size of 6 × 6. The right image shows the final cropped target ROI, with most ROIs sized between 10 × 10 and 20 × 20.

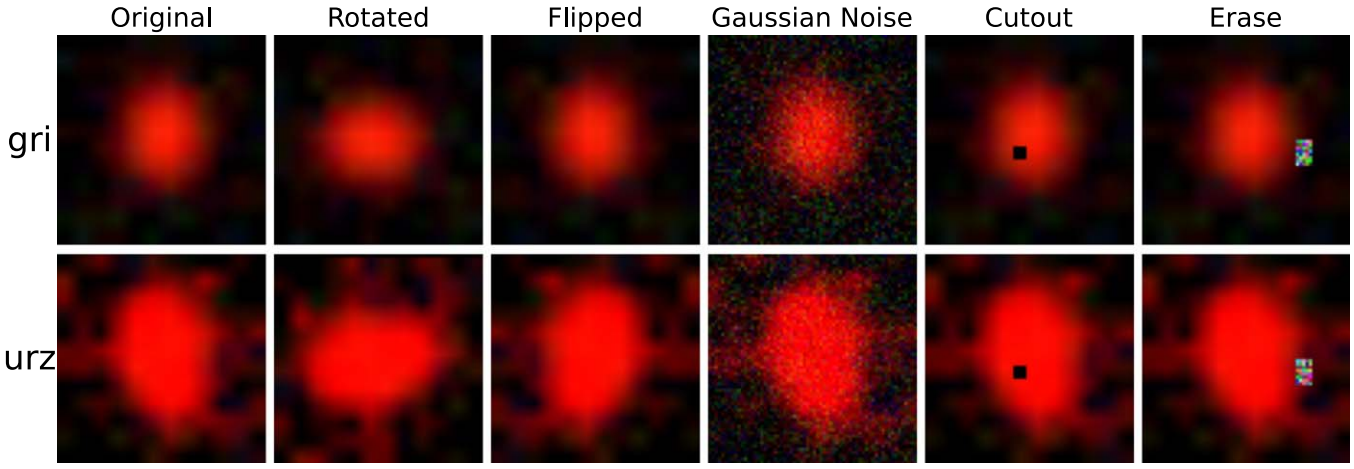


Figure A3. Randomly selected M-type star showing the effect of data augmentation. Titles display the corresponding data augmentation methods. In this study, both *gri* images and *urz* images undergo the same data augmentation methods. The first row shows the effects of data augmentation on *gri* images, while the second row displays the effects on *urz* images.

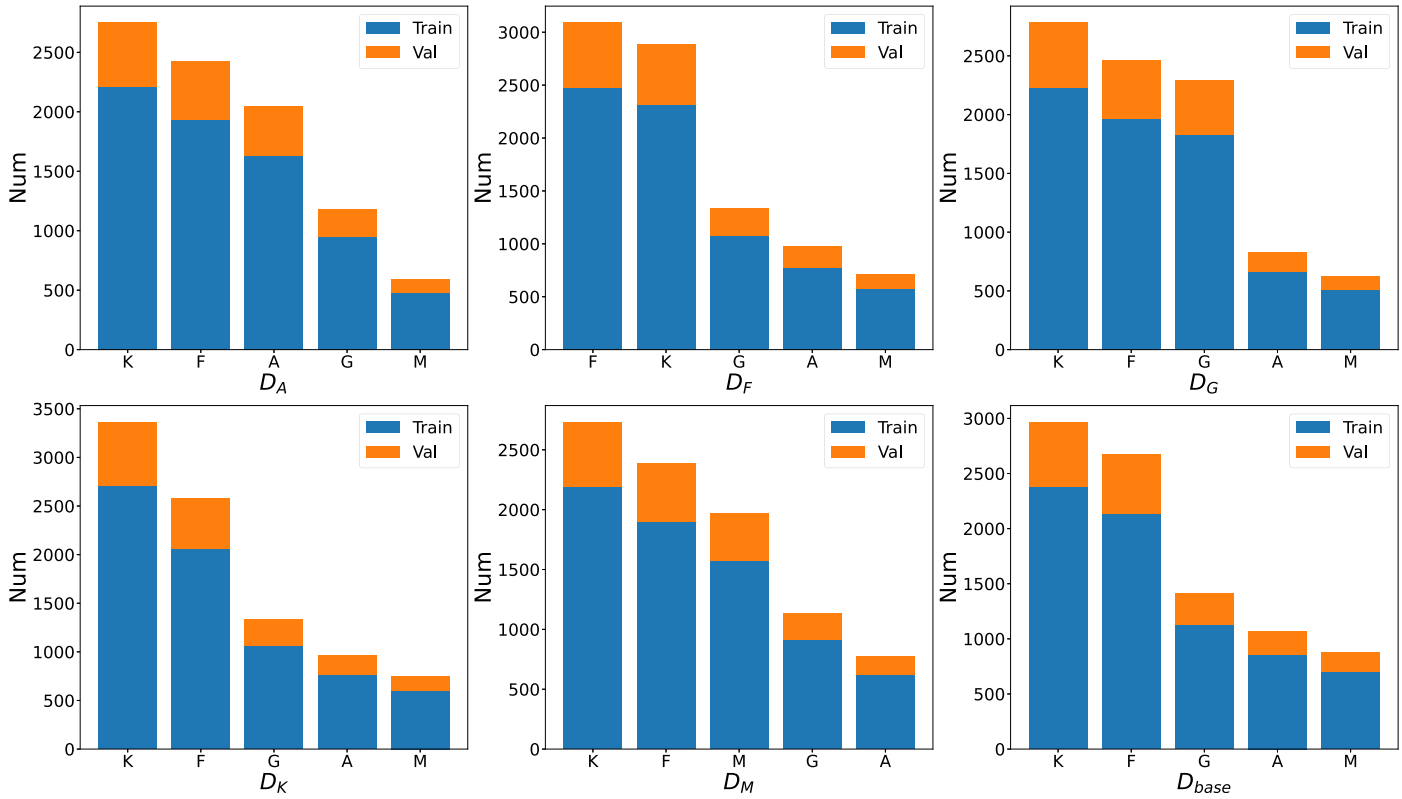


Figure A4. The six subfigures represent the distribution of stellar spectral classes for six data sets after weighted sampling and division into training and validation sets.

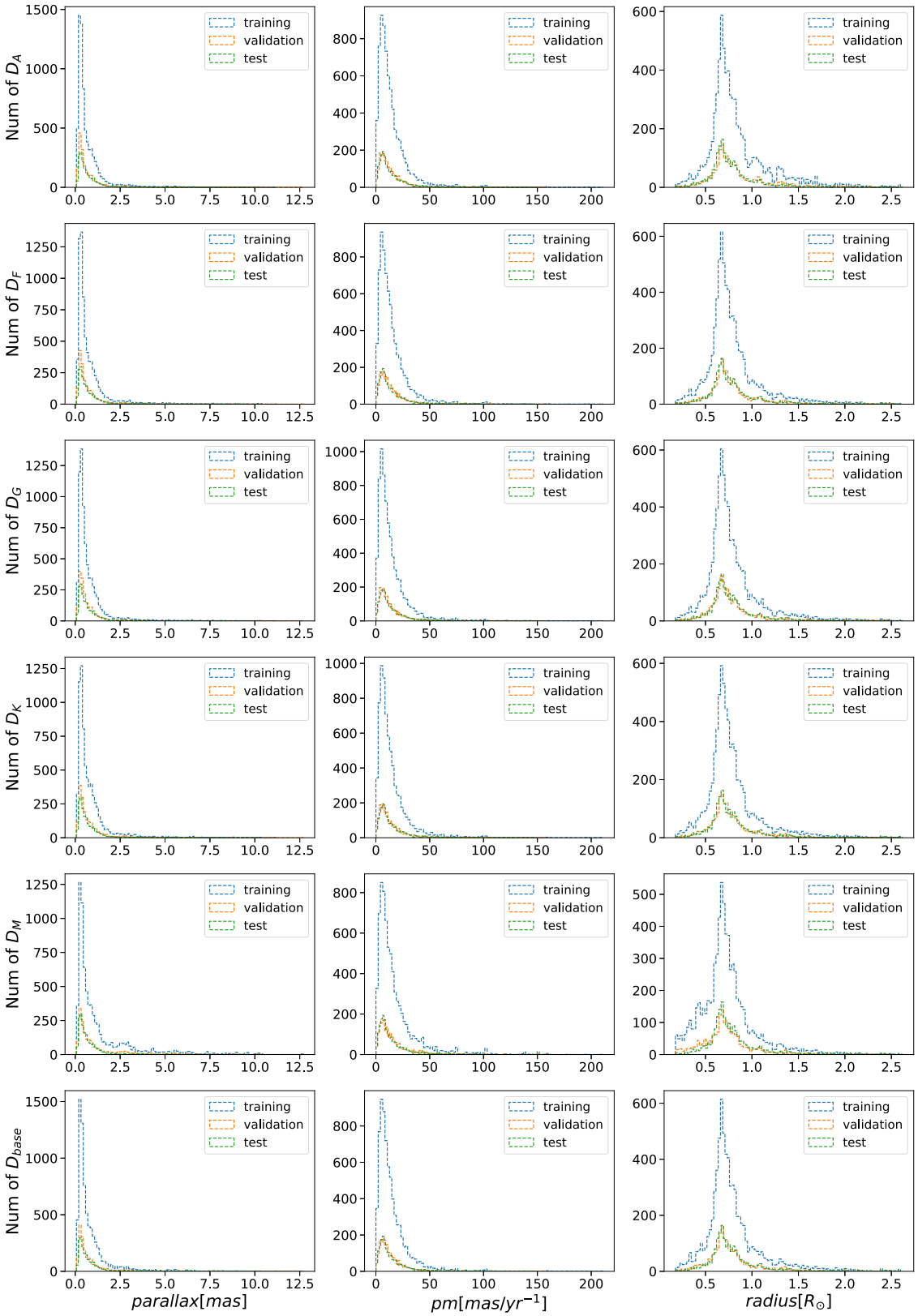


Figure A5. After weighted sampling and splitting into training, validation, and test sets for six data sets, the corresponding distribution of *parallax*, *pm*, and *radius_flame*. It is important to note that the six data sets use the same test set. Therefore, the distribution of the test set remains consistent.

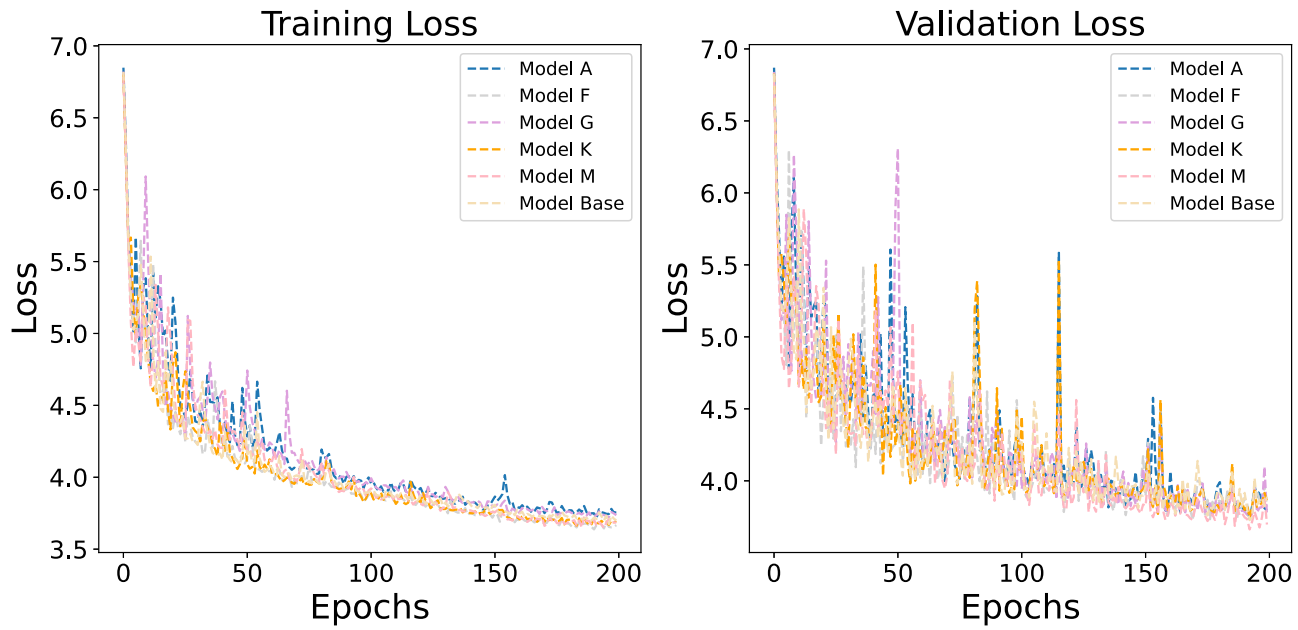


Figure A6. The loss curves for each model during the training process, on both the training and validation sets. The left graph represents the changes in loss values for the training set, while the right graph shows the changes for the validation set. Loss curves for different models are distinguished using different colors.

Table A1
Evaluation Results on the Test Set After Retraining the Model with Different Band Combinations

Bound	Accuracy	F1-score	MAE	σ
gri+urz	94.9%	94.9%	0.085	0.151
giu+grz	95.6%	95.5%	0.083	0.150
giu+riz	94.7%	94.6%	0.082	0.148
giu+ruz	95.2%	95.2%	0.086	0.153
giz+riu	95.3%	95.3%	0.083	0.152
giz+ruz	95.7%	95.7%	0.082	0.149
gri+guz	95.2%	95.1%	0.081	0.146
gri+iuz	95.5%	95.4%	0.081	0.149
gri+ruz	95.5%	95.4%	0.082	0.150
gru+giz	95.1%	95.1%	0.081	0.149
gru+iuz	**95.7%**	**95.6%**	0.082	0.148
gru+riz	95.5%	95.4%	0.079	0.148
guz+riu	94.3%	94.2%	0.087	0.154
guz+riz	94.5%	94.3%	0.081	0.151
iuz+grz	95.6%	95.5%	0.086	0.152
riu+grz	95.2%	95.2%	0.082	0.150
gri+uz	94.9%	94.9%	0.084	0.149
giu+rz	94.8%	94.7%	0.081	0.150
giz+ru	92.7%	95.2%	0.083	0.152
gru+iz	95.2%	95.1%	0.078	**0.145**
grz+iu	95.2%	95.1%	0.083	0.151
guz+ri	94.5%	94.4%	**0.077**	**0.145**
iuz+gr	94.9%	94.8%	0.084	0.152
riu+gz	94.7%	94.6%	0.083	0.151
riz+gu	93.3%	93.1%	0.079	**0.145**
ruz+gi	95.2%	95.2%	0.083	0.150

Note. Bold indicates the best performance.

Table A2
Description of Our Catalog Column

Column Name	Description
objid	Unique SDSS object identifier
R.A.	J2000 R.A. (*r*-band)
decl.	J2000 decl. (*r*-band)
psfMag_u (*g*, *r*, *i*, *z*)	SDSS PSF magnitude in the *u*(*g*, *r*, *i*, *z*) band
spectral_class	Class of model predictions
prob_O(B, A, F, G, K, M)	Probability of prediction as an O(B, A, F, G, K, M)-type star
radius	Radius of model predictions
radius_err	The uncertainty provided by the model when predicting the radius

References

Bialek, S., Fabbro, S., Venn, K. A., et al. 2020, MNRAS, 498, 3817
Blay, G., & Lovekin, C. 2015, AIP Conf. Proc. 1697 (Melville, NY: AIP), 110001
Breiman, L. 1996, Mach. Learn., 24, 123
Clarke, A. O., Scaife, A. M. M., Greenhalgh, R., & Griguta, V. 2020, A&A, 639, A84
Duan, F.-Q., Liu, R., Guo, P., Zhou, M.-Q., & Wu, F.-C. 2009, RAA, 9, 341
Flores, M., Corral, R., Fierro-Santillán, L. J., & Navarro, S. G, C. R. 2023, A&C, 45, 100760
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2023, A&A, 674, A28
Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1
Gilmore, G., Randich, S., Asplund, M., et al. 2012, Msngr, 147, 25
Hong, S., Zou, Z., Luo, A.-L., et al. 2022, MNRAS, 518, 5049
Kallrath, J. 2022, Galax, 10, 17
Kang, Z., Zhang, Y., Zhang, J., et al. 2023, PASP, 135, 094501
Krawczyk, B., Galar, M., Jeleń, & Herrera, F. 2016, Appl. Soft Comput., 38, 714

Li, G., Li, M., Liu, Y., et al. 2022, Journ. Compute. App., 42, 3307
Li, J., Wang, Y., Song, X., & Xiao, H. 2018, CoPhC, 100, 1
Li, X., Zhang, X., Xiong, S., Zheng, Y., & Li, H. 2023, MNRAS, 523, 5230
LSST Dark Energy Science Collaboration 2012, arXiv:1211.0310
Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, RAA, 15, 1095
Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, PASP, 116, 133
Ma, X., Li, X., Luo, A., Zhang, J., & Li, H. 2022, MNRAS, 519, 4765
Monnier, J. D. 2003, RPPh, 66, 789
Morgan, W. W., & Keenan, P. C. 1973, ARA&A, 11, 29
Moya, A., Zuccarino, F., Chaplin, W. J., & Davies, G. R. 2018, ApJS, 237, 21
Sharma, K., Kembhavi, A., Kembhavi, A., et al. 2020, MNRAS, 491, 2280
Shi, J.-H., Qiu, B., Luo, A.-L., et al. 2023, MNRAS, 520, 2269
Taylor, M. 2017, Informatics, 4, 3
Ulla, A., Creevey, O. L., Álvarez, M. A., et al. 2022, Gaia DR3 documentation Chapter 11: Astrophysical parameters 11, European Space Agency
van Belle, G. T., & von Braun, K. 2009, ApJ, 694, 1085
Wan, W., & Yang, M. 2020, JoS, 31, 113
Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9
Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. 2018, Proc. of the European Conf. on Computer Vision (ECCV) (Cham: Springer),
Wu, J. F., & Boada, S. 2019, MNRAS, 484, 4683
Xiong, S., Li, X., & Liao, C. 2022, ApJS, 261, 36
YANG, Q., & WU, X. 2006, Int. J. Inf. Technol. & Decision Making, 5, 597
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
York, D. G., Adelman, J. K., Anderson, J. M., et al. 2000, AJ, 120, 1579
Zeng, Q., Li, X., & Lin, H. 2020, MNRAS, 494, 3110
Zhang, M., Gao, J., Luo, A. L., et al. 2023, RASTI, 2, 408
Zhong, J., Deng, Z., Li, X., et al. 2024, MNRAS, 271, 2011