The Astrophysical Journal Supplement Series, 277:19 (11pp), 2025 March https://doi.org/10.3847/1538-4365/adae86

The Stellar Abundances and Galactic Evolution Survey (SAGES). II. Machine Learning–based Stellar Parameters for 21 Million Stars from the First Data Release

Hongrui Gu[1,2], Zhou Fan[1,2], Gang Zhao[1,2], Huang Yang[1,2], Timothy C. Beers[3,4], Wei Wang[1], Jie Zheng[1], Jingkun Zhao[1], Chun Li[1], Yuqin Chen[1], Haibo Yuan[5], Haining Li[1], Kefeng Tan[1], Yihan Song[1], Ali Luo[1], Nan Song[6], and Yujuan Liu[1]

[1] CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, People's Republic of China; zfan@nao.cas.cn
[2] School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing, People's Republic of China
[3] Department of Physics and Astronomy, University of Notre Dame, Notre Dame, IN 46556, USA
[4] Joint Institute for Nuclear Astrophysics—Center for the Evolution of the Elements (JINA-CEE), USA
[5] Department of Astronomy, Beijing Normal University, Beijing 100875, People's Republic of China
[6] China Science and Technology Museum, Beijing 100101, People's Republic of China
Received 2024 June 9; revised 2024 December 1; accepted 2025 January 24; published 2025 February 25

Abstract

Stellar parameters for large samples of stars play a crucial role in constraining the nature of stars and stellar populations in the Galaxy. An increasing number of medium-band photometric surveys are presently used in estimating stellar parameters. In this study, we present a machine learning approach to derive estimates of stellar parameters, including [Fe/H], log g, and $T_{\rm eff}$, based on a combination of medium-band and broadband photometric observations. Our analysis employs data primarily sourced from the Stellar Abundances and Galactic Evolution Survey (SAGES), which aims to observe much of the Northern Hemisphere. We combine the *uv*-band data from SAGES DR1 with photometric and astrometric data from Gaia EDR3, and apply the random forest method to estimate stellar parameters for approximately 21 million stars. We are able to obtain precisions of 0.09 dex for [Fe/H], 0.12 dex for log g, and 70 K for $T_{\rm eff}$. Furthermore, by incorporating Two Micron All Sky Survey and Wide-field Infrared Survey Explorer infrared photometric and Galaxy Evolution Explorer ultraviolet data, we are able to achieve even higher precision estimates for over 2.2 million stars. These results are applicable to both giant and dwarf stars. Building upon this mapping, we construct a foundational data set for research on metal-poor stars, the structure of the Milky Way, and beyond. With the forthcoming release of additional bands from SAGES such DDO51 and Hα, this versatile machine learning approach is poised to play an important role in upcoming surveys featuring expanded filter sets.

Unified Astronomy Thesaurus concepts: Stellar abundances (1577); Fundamental parameters of stars (555); Astronomy data analysis (1858); Photometry (1234)

1. Introduction

Accurate determination of stellar parameters for large samples of stars in our Galaxy plays an important role in many fields, including stellar physics and the structure and evolution of the Milky Way. Spectroscopic observations can accurately determine stellar parameters from spectral lines. As early as 2009, the Sloan Extension for Galactic Understanding and Exploration (B. Yanny et al. 2009) project, part of the Sloan Digital Sky Survey (SDSS; D. G. York et al. 2000), used low-resolution ($R \sim 1800$) spectra to obtain stellar parameters for some 240,000 stars. Results for an additional 118,958 unique stars were reported on by C. M. Rockosi et al. (2022). Subsequently, the SDSS APOGEE project (S. R. Majewski et al. 2017) obtained stellar parameters for over 700,000 stars using high-resolution ($R \sim 22{,}500$) near-infrared spectra. With the advent of the Large Sky Area Multi-Object Fiber Spectroscopy Telescope (LAMOST, X.-Q. Cui et al. 2012; G. Zhao et al. 2012), multifiber spectroscopic survey telescopes have greatly improved the efficiency of acquiring spectra. To date, LAMOST has observed more than 11 million stars, with over 7 million stars with available parameters, increasing the sample size by an order of magnitude. However, compared to photometric survey projects with over a billion stars at limiting magnitudes of around $G = 20$, such as Gaia DR3 (Gaia Collaboration et al. 2021), there is still a 1–2 order of magnitude gap between the sample sizes of photometric surveys and spectroscopic surveys, underscoring the importance of obtaining stellar-parameter estimates through the use of photometry.

Stellar-parameter estimates based on multiband photometric surveys have expanded over the past several decades. The Hauck–Mermilliod sky survey (HM; B. Hauck & M. Mermilliod 1998) and the Geneva–Copenhagen Survey (GCS; B. Nordström et al. 2004), contemporaneous with the SDSS, included medium-band filters, but had limiting magnitudes of only $v \sim 8$, far from meeting current research needs. Ž. Ivezić et al. (2008) used spectral and photometric data from SDSS to estimate metallicities for some 200,000 FGK dwarfs, with a precision of about 0.2–0.3 dex down to [Fe/H] ~ -2.0. With improvements in photometric accuracy and precision, H. Yuan et al. (2015) estimated metallicities for some 500,000 FGK dwarfs to a level of 0.1–0.2 dex in SDSS/Stripe 82 down to [Fe/H] ~ -2. R.-Y. Zhang et al. (2021) demonstrated that estimates for giant stars can also achieve 0.2 dex precision down to a similar limit, with differing model parameters

compared to FGK stars. However, estimating parameters of giant stars using dwarf-star models could introduce a certain degree of systematic bias (D. An & T. C. Beers 2020). Space-based telescopes offer higher photometric accuracy and more uniform survey coverage. S. Xu et al. (2022) used Gaia EDR3 data to provide metallicity estimates for 27 million FGK stars across the sky with 0.2 dex precision. Due to the bandwidth limitations of wide-band photometric survey filters, it is difficult to significantly improve upon metallicity and other stellar-parameter estimates further using such filters alone.

The Skymapper Southern Sky Survey (SMSS; S. Keller et al. 2007), led by the Australian National University, has adopted the Strömgren–Crawford filter system (B. Strömgren 1964; D. L. Crawford et al. 1970), similar to the GCS and HM surveys in some of its filters, covering as faint as a 20th of a magnitude (at 5σ) in the Southern Hemisphere over 21,000 square degrees. Y. Huang et al. (2019) used polynomial fitting to establish empirical relationships between stellar parameters and photometric colors, deriving accurate atmospheric parameters for about one million red giants from SMSS DR1.1. A. Chiti et al. (2021) used a grid-based synthetic-photometry method to obtain over 250,000 photometric metallicities for giants from SMSS DR2. Subsequently, using recalibrated SMSS DR2 data and Gaia EDR3 data, Y. Huang et al. (2022) employed polynomial-fitting methods to produce a catalog of stellar parameters, including metallicity, for about 24 million stars with a precision of 0.1–0.3 dex, with carefully selected training data sets enabling metallicity estimates down to [Fe/H] ~ -3.5.

By installing a single narrowband filter with a response curve centered around the Ca II H and K lines on the CFHT, the Pristine Survey (E. Starkenburg et al. 2017) is capable of observing thousands of square degrees to a depth of about a 20th of a magnitude, while maintaining a 0.1–0.2 dex precision in metallicity estimates down to [Fe/H] ~ -3.5 in their DR1 (N. F. Martin et al. 2024).

With the advent of multiple narrowband photometric surveys, such as J-PLUS (A. J. Cenarro et al. 2019), S-PLUS (C. Mendes de Oliveira et al. 2019), and J-PAS (S. Bonoli et al. 2021), traditional grid-based, parameter-fitting, and other empirically based algorithms become increasingly difficult to handle in such high-dimensional scenarios. Therefore, various machine learning algorithms have been employed in the measurement of stellar parameters. Examples include neural networks (V. F. Ksoll et al. 2020; D. D. Whitten et al. 2021; L. Yang et al. 2022), and random forest algorithms (Y. Bai et al. 2019; C. A. Galarza et al. 2022). These studies have demonstrated that machine learning algorithms are highly effective methods for extracting stellar parameters from high-dimensional data.

The Stellar Abundance and Galaxy Evolution Survey (SAGES, W. Wang et al. 2014; J. Zheng et al. 2018, 2019; K.-F. Tan et al. 2022; Z. Fan et al. 2023) is a photometric survey in the Northern Hemisphere that includes narrow- and medium-band filters. Its v_s-band central wavelength coincides with the location of the Ca II H and K lines, and its u band is similar to SMSS, providing the opportunity to obtain more accurate atmospheric parameters, and reducing the impact of molecular carbon bands. In addition to the u- and v-band data used in this work, broadband photometric data, such as SDSS-like g, r, i, and medium-band data, such as DDO51 and Hα, will continue to be released. In this approach, an algorithm must be chosen that can be applied to both low-dimensional data and complex high-dimensional data. This is necessary, since the current techniques (e.g., Y. Huang et al. 2023, hereafter H23) of measuring stellar parameters based on DR1 data cannot be iterated with the release of additional filter bands in SAGES. Robust statistical methods have played a significant role in handling high-dimensional astronomical data sets, and it is logical to expect that they would perform similarly well or even better in low dimensions. By applying such statistical methods to low-dimensional data sets, such as multiband photometric data from SAGES, it is possible to simultaneously attain high-precision stellar parameters, and gain experience for the future exploration of large multidimensional photometric data sets.

This paper is organized as follows. Section 2 describes the data and methods used in this work. Section 3 introduces the results for the training data. Section 4 describes the stellar parameters derived for SAGES stars, followed by a summary in Section 5.

2. Data and Methods

In order to train the random forest algorithm, we constructed two sets of training data, including the colors and filter magnitudes provided by SAGES and Gaia and the estimated stellar parameters provided by spectroscopic surveys, as described below.

2.1. Photometric Data

The first training set uses photometric data from SAGES DR1 and Gaia EDR3 (Gaia Collaboration et al. 2021) as input; the second set additionally includes the photometric data in the infrared from Two Micron All Sky Survey (2MASS; R. M. Cutri et al. 2003) and ALLWISE (R. M. Cutri et al. 2021) and ultraviolet from Galaxy Evolution Explorer (GALEX; L. Bianchi et al. 2014). SAGES is a multiband photometric survey focused on estimation of stellar parameters and interstellar extinction. Its narrow v-band filter is quite sensitive to [Fe/H]. Its u band and v band are located on either side of the Balmer jump, providing good sensitivity to surface gravity (hereafter, log g). Gaia EDR3 provides a large amount of accurate photometric information (the ultrawide Bp and Rp bands) for the stars in SAGES, so it can provide effective temperature estimates (hereafter, $T_{\rm eff}$). In addition, infrared photometric information such as from ALLWISE and 2MASS (K. C. Schlaufman & A. R. Casey 2014), and ultraviolet photometric information such as from GALEX (X. Lu et al. 2024), can also be used to identify candidate metal-poor stars, taking advantage of the sensitivity of the infrared and ultraviolet bands to metallicity .

In this work, two sets of training data are constructed. Data in the first set are chosen to apply to as large a sample of SAGES stars as possible, while providing parameter estimates with reasonable accuracy and precision. Data in the second set sacrifices sample size in order to improve the accuracy and precision as much as possible for a subset of the SAGES stars.

For the extinction in the various photometric bands, we employ the reddening coefficients provided by K.-F. Tan et al. (2022) to correct the SAGES u- and v-band data, and the coefficients provided in Table 2 of R. Zhang & H. Yuan (2023) to correct the rest of data. All values of reddening coefficient

Table 1
Reddening Coefficients for Photometry Data Used in this Work

Band	Survey	Reddening Coefficient	Citation
NUV	GALEX	7.294	R. Zhang & H. Yuan (2023)
u	SAGES	4.324	K.-F. Tan et al. (2022)
v	SAGES	3.917	K.-F. Tan et al. (2022)
G	Gaia	2.364	R. Zhang & H. Yuan (2023)
Bp	Gaia	2.998	R. Zhang & H. Yuan (2023)
Rp	Gaia	1.737	R. Zhang & H. Yuan (2023)
J	2MASS	0.748	R. Zhang & H. Yuan (2023)
H	2MASS	0.453	R. Zhang & H. Yuan (2023)
K	2MASS	0.306	R. Zhang & H. Yuan (2023)
*W*1	WISE	0.194	R. Zhang & H. Yuan (2023)
*W*2	WISE	0.138	R. Zhang & H. Yuan (2023)

we used are shown in Table 1. The extinction map we used is SFD98 (D. J. Schlegel et al. 1998).

Figure 1 is the color–magnitude diagram (x: $(Bp - Rp)_0$, y: M_{G0}) for SAGES stars, after application of the reddening and extinction corrections. We have divided the targets into dwarf stars (yellow dots), giant stars (red dots), hot subdwarfs (blue dots), and white dwarfs (green dots). Targets with Gaia parallax errors larger than the parallax itself are considered to be likely giants. Table 2 lists the constraints we imposed on the input training data and the output prediction data. The total sample size of the first set of data, where we emphasize retaining as large a sample as feasible, is 21,071,305, including 19,663,040 dwarfs and 1,408,265 giants. The total sample size of the second set of data, emphasizing accuracy and precision of the derived stellar parameters, is 2,191,452, including 2,037,528 dwarf stars and 153,924 giants.

2.2. Stellar Parameters from Spectroscopic Surveys

Spectroscopic observation is the most direct way to obtain stellar parameters. Even relatively low-resolution spectra can be used to estimate stellar parameters with higher accuracy and precision than current photometric methods, in particular if they are of a high signal-to-noise ratio. LAMOST (X.-Q. Cui et al. 2012), as the largest spectroscopic survey project in the Northern Hemisphere at present, can provide the three stellar parameters required in this work: $T_{\rm eff}$, log g, and [Fe/H]. However, through DR10, LAMOST spectra were only low resolution ($R \sim 1800$). The accuracy and precision of the stellar parameters in DR10 is lower than that of the medium- or high-resolution spectra, and the most metal-poor star reliably estimated by most pipelines in DR10 is only be about [Fe/H] ~ -2.5. In order to obtain training stars with lower metal abundance, we considered using samples from multiple spectral surveys. We employ samples from another three spectroscopic sky surveys: (1) APOGEE DR17 (Abdurro'uf et al. 2022), a high-resolution near-infrared survey with wavelengths ranging from 1.51 to 1.70 μm and $R \sim 22{,}500$, (2) RAVE DR5 (A. Kunder et al. 2017), a spectroscopic survey aiming to measure radial velocity with wavelengths centered on the Ca I triplet (8410–8795 Å) and resolving power of $R \sim 7500$, and (3) PASTEL (C. Soubiran et al. 2016), a catalog of collected stellar parameters from the literature obtained by various high-resolution spectroscopic observations. (4) A number of additional papers have reported on results based on high-resolution follow-up observations for very metal-poor (VMP; [Fe/H] $\leqslant -2.0$) and extremely metal-poor (EMP; [Fe/H] $\leqslant -3.0$) stars identified in previous surveys that are not included in the PASTEL catalog. We collected a total of 15,628 stars from D. K. Lai et al. (2004), P. Bonifacio et al. (2009), D. Yong et al. (2013), W. Aoki et al. (2013), J. Yoon et al. (2016), P. François et al. (2018), H. Li et al. (2018), R. Ezzeddine et al. (2020), and G. Limberg et al. (2021). These stars have metallicities in the range [−0.1, −7.8]; 93.1% are VMP stars and 9.8% are EMP stars.

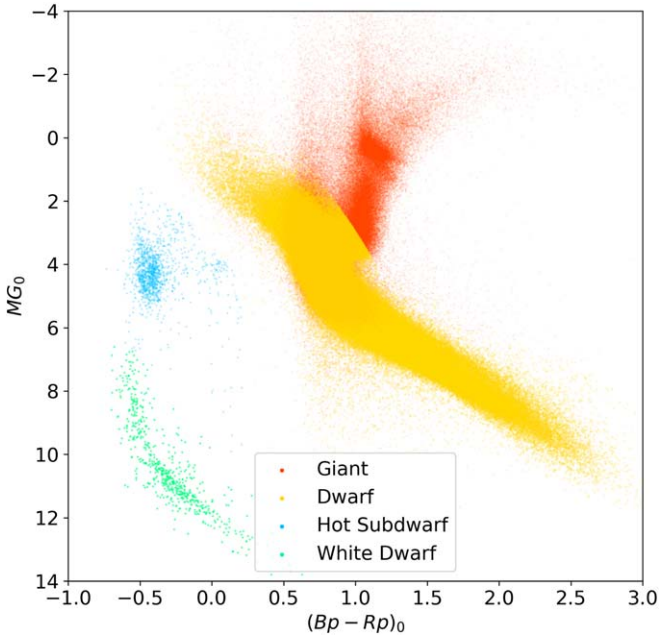

Figure 1. Distribution of the data used in this work in the Gaia color–absolute magnitude diagram. The colors distinguish four different types of stars. The data used in this work are the dwarf (yellow) and giant (red) stars.

Table 2
Adopted Constraints on the Data Sets for Training and Prediction

Parameter	Training Cut	Prediction Cut
Parallax over error	1	1
$E\,(B-V)$	0.05	0.5
$T_{\rm efferr}$	200	…
log $g_{\rm err}$	0.3	…
$[{\rm Fe/H}]_{\rm err}$	0.2	…
$u_{\rm err}$	0.1	0.2
$v_{\rm err}$	0.1	0.2
$G_{\rm err}$	0.01	0.1
$Bp_{\rm err}$	0.01	0.1
$Rp_{\rm err}$	0.01	0.1
$J_{\rm err}$	0.04	0.1
$H_{\rm err}$	0.05	0.1
$K_{\rm err}$	0.05	0.2
$W1_{\rm err}$	0.03	0.1
$W2_{\rm err}$	0.05	0.1
$\rm NUV_{err}$	0.1	0.2

2.3. Training Samples

To obtain stellar parameters from the SAGES photometric data, it is necessary to crossmatch the stars in common between both spectroscopic surveys and SAGES (constraints are as shown in Table 2). During the training step, the apparent magnitudes, reddening corrections, and other parameters are used as input data; the stellar parameters provided by spectral

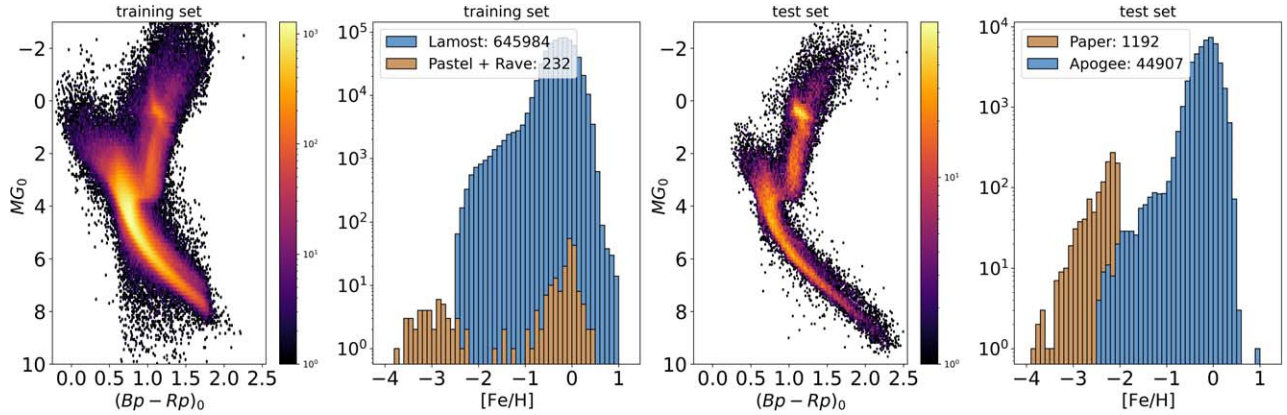

Figure 2. Left two panels: number-density distribution of the training samples for Data Set 1 in the color–absolute magnitude diagram, and the histogram of [Fe/H] from different spectral survey catalogs. Right two panels: same as the left two panels but from the test set.

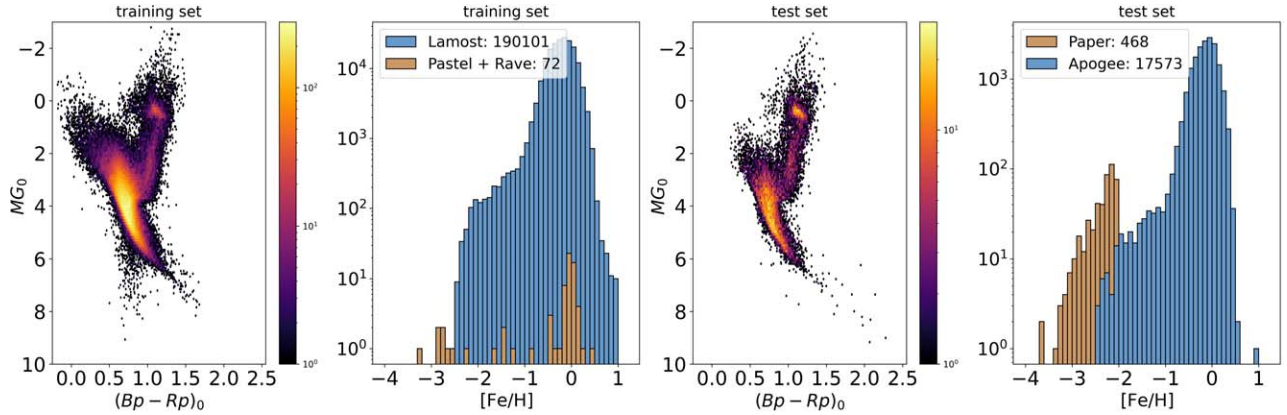

Figure 3. Similar to Figure 2, but using data for Data Set 2.

data are used to constrain output predictions. To ensure the independence and the coverage of different parameters in the parameter space, this work uses LAMOST data as the training set and APOGEE data as the test set. To extend the parameter range at low metallicity, data from PASTEL and RAVE are utilized. Additionally, data from another nine papers are used to test the result at low metallicity. Considering the systematic differences in the parameters provided by different data sets, a correction of 0.427 in [Fe/H] is applied to the data from those nine papers to account for these systematic differences. Figures 2 and 3 show the number-density distribution of these two samples in the color–magnitude diagram and the number-density distribution of metallicity in training data and testing data.

2.4. Method: Random Forest

After comparing the results of a simple three-layer fully connected neural network under different loss functions, polynomial fitting, and the random forest method, we find out that the random forest algorithm not only achieved the fastest speed but also the best performance, slightly outperforming polynomial fitting and significantly outperforming the simple fully connected neural network. Therefore, this work chose to use the random forest algorithm (L. Breiman 2001) for subsequent work. Random forest is an ensemble-learning approach, assembled from multiple independent decision-tree algorithms. Each decision tree divides the space through a series of discriminants that seek to minimize the "entropy," defined as the Gini index. The decision trees are trained through sets of partial training samples; the final result is obtained by weighting the results of the multiple decision trees. The random forest has the advantage that is can prevent overfitting, and thus is more generally applicable to a variety of input data.

Moreover, since each sample is drawn with replacement from the data set to serve as the training set for each tree, the samples that are not selected can be used as a validation set, known as the out-of-bag error. Therefore, when using a random forest, there is no need to establish an additional validation set to perform real-time evaluation of the model's performance.

Random forest, as a machine learning algorithm, has several hyperparameters, and selecting an appropriate set of these hyperparameters is crucial. In this work, two very important hyperparameters were chosen: n_estimators and max_features. n_estimators refers to the number of decision trees in the random forest, while max_features is the number of features considered in each decision tree. Ultimately, we selected n_estimators = 300 and max_features = "None," which means all input features are used. After determining and fixing the most important two parameters, another hyperparameter, max_depth, which represents the maximum depth of each individual decision tree, was tested. After testing different depths, the final value for max_depth was determined to be 17. Details of the above testing are shown in Figure 4.

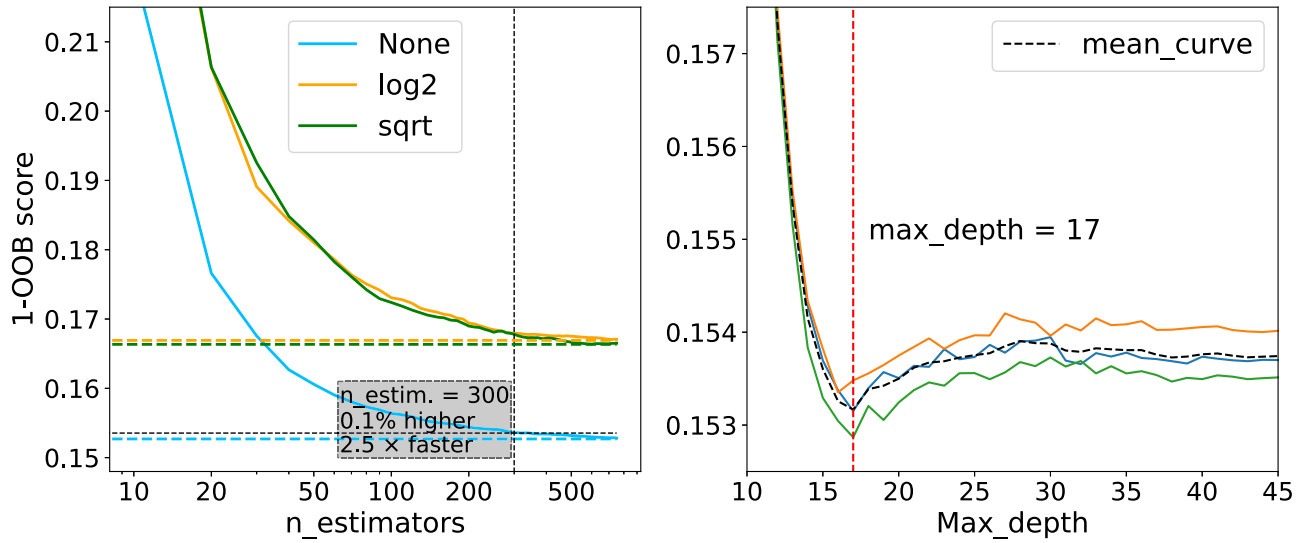

Figure 4. Left panel: The out-of-bag (OOB) scores for different combinations of hyperparameters are shown, with different colors representing different max_features values, and the *x*-axis representing different n_estimators values. The *y*-axis represents 1 − OOB score, so a smaller value on the *y*-axis indicates better performance. For the selected n_estimators = 300, the accuracy is only 0.1% lower than when n_estimators = 750, but the computational speed is 2.5 times faster. Right panel: For n_estimators = 300 and max_features = "None," the 1 − OOB score varies with the max_depth parameter. A smaller value on the *y*-axis indicates better performance. By testing each max_depth three times and averaging the results to obtain the mean curve (represented by the black dashed line), it can be seen that the optimal performance is achieved when max_depth = 17.

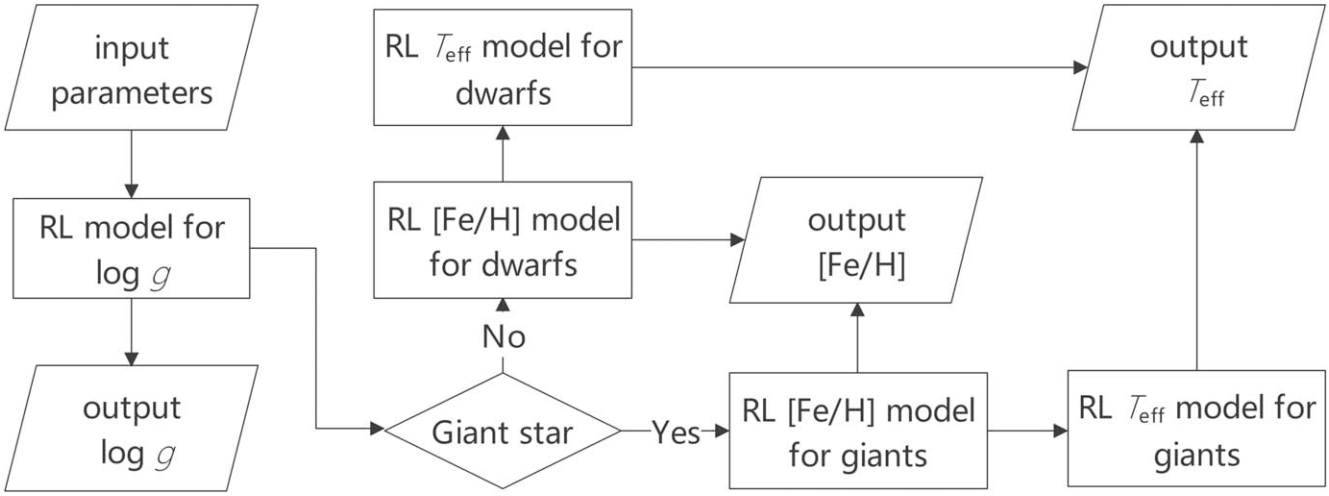

Figure 5. Flow chart of this work.

2.5. Model Establishment

Since the prediction accuracy and precision of a single model to simultaneously predict multiple parameters is generally less than that obtained by individual models to predict single parameters, we constructed multiple random forest models to predict the [Fe/H], $T_{\rm eff}$, and log g for dwarf and giant stars, respectively. Surface gravity is one of the primary criteria to distinguish dwarf and giant stars; if we were to predict the log g of dwarfs and giants stars separately, it would introduce some misclassification error for the model. Therefore, we employ a unified model to predict the log g of dwarf and giant stars together. Five random forest models are trained for the two data sets in this work, which are models for (1) [Fe/H] of dwarf stars, (2) [Fe/H] of giant stars, (3) $T_{\rm eff}$ of dwarf stars, (4) $T_{\rm eff}$ of giant stars, and (5) log g for both dwarfs and giants. The flow chart is shown in Figure 5.

The input data for these models include the magnitude after extinction correction, the colors after reddening correction, the parallax, the G-band absolute magnitude, and $E(B-V)$. The magnitudes and $E(B-V)$ can provide the model with extinction, reddening, and effective temperature, and the G-band absolute magnitude and colors can provide the model with information similar to the color–magnitude diagram, which not only constrain the model $T_{\rm eff}$ and stellar-classification information, but also provides constraints on the derived log g. The u- and v-band photometry and related colors mainly provide information for the models to constrain the estimated [Fe/H]. Because the metallic absorption lines cover both the visible, ultraviolet, and near-infrared spectra, visible photometry from Gaia (S. Xu et al. 2022), ultraviolet photometry from GALEX (X. Lu et al. 2024), and infrared photometry from Wide-field Infrared Survey Explorer (WISE) and 2MASS (K. C. Schlaufman & A. R. Casey 2014) can also provide information on the metallicity of stars. During the training, we used parameters from LAMOST as the training set and APOGEE as the test set. Additionally, we supplemented the

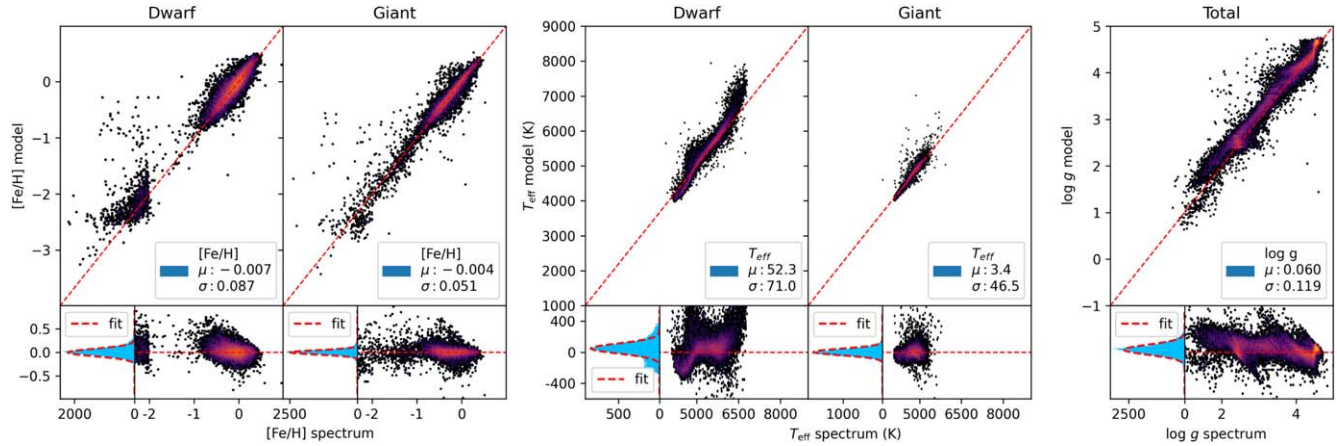

Figure 6. Comparisons between the spectroscopic and photometric stellar parameters for the test set from Data Set 1, for five models (dwarf [Fe/H], giant [Fe/H], dwarf $T_{\rm eff}$, giant $T_{\rm eff}$, log g). The lower portion of each panel shows the number-density distribution of the difference between two spectroscopic and photometric stellar parameters. The mean and standard deviation are obtained from fitting the histograms, and their values are shown in the legends.

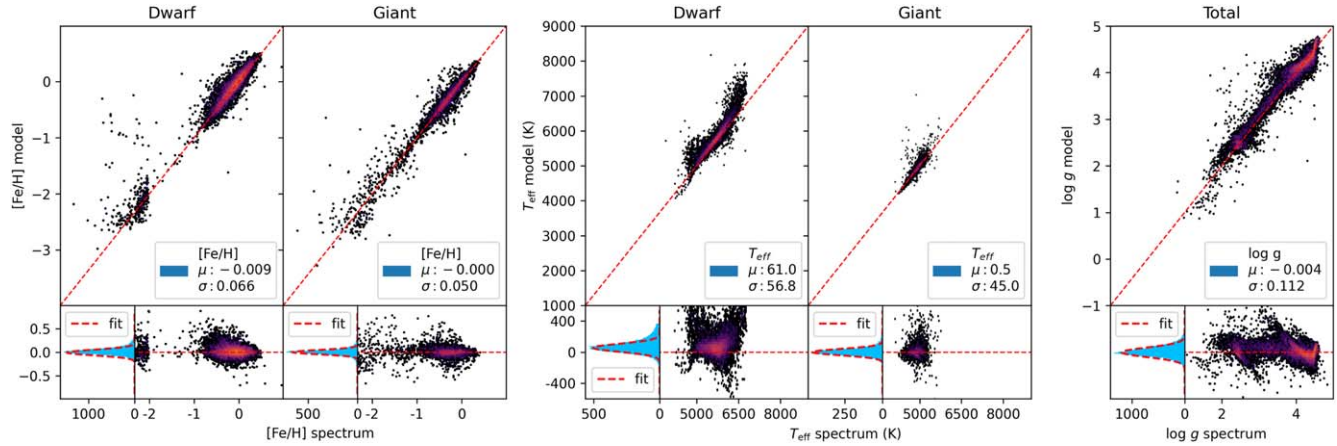

Figure 7. Similar to Figure 6, but using the test set from Data Set 2.

training set with metallicity data from PASTEL and RAVE for metal-poor stars, and we used the metallicity data from the nine papers to supplement the metal-poor stars in the test set. Due to the systematic difference between the nine papers and LAMOST, this work applied a systematic correction of 0.427 to the metallicity values from the nine papers.

3. Results

We now compare the results for the two sets of training samples. The first data set includes the photometric information provided by SAGES DR1 and Gaia EDR3, and the second data set uses information from SAGES DR1, Gaia EDR3, WISE, 2MASS, and GALEX.

3.1. Data Set 1: Photometry from SAGES DR1 and Gaia DR3

The training data for this input data contains 579,185 dwarf stars and 67,031 giant stars. Figure 6 shows the accuracy and precision of the five models on the test set; each panel represents the training result for one model. The upper part of each panel compares the known stellar parameters derived from spectroscopy (x-axis) with the parameters predicted by the model (y-axis). The left portion of the lower part in each panel represents the Gaussian fitting results for the residuals. The right portion of the lower part in each panel shows the residual distribution of the predicted values minus the spectroscopic values, as a function of the spectroscopic values. The fitting parameters are shown in the legend in the upper part of each panel.

3.2. Data Set 2: Addition of Two Infrared Photometric Surveys

After crossmatching with 2MASS J, H, and K bands, WISE $W1$ and $W2$ bands, as well as the GALEX near-UV band, a second set of training sample is assembled containing 178,542 dwarf stars and 11,559 giant stars. Figure 7 illustrates the accuracy and precision of the five models on this test set, with each subplot representing the training result of one model. When contrasting Figure 7 with Figure 6, it is clear that the statistical accuracy and precision of the [Fe/H] and log g for the dwarf stars has been improved somewhat after adding the infrared photometric data. However, in subsequent studies, we believe that the higher precision observed here may be an artifact due to differences in sample distribution.

3.3. Final Accuracy and Precision

Figure 8 illustrates the performance of the mean and standard deviation for the five models, with the three columns of panels representing the [Fe/H], $T_{\rm eff}$, and log g, respectively. The two rows display the variations of the standard deviation (first row) and mean (second row) with respect to the known stellar parameters. In the figure, different colors represent distinct

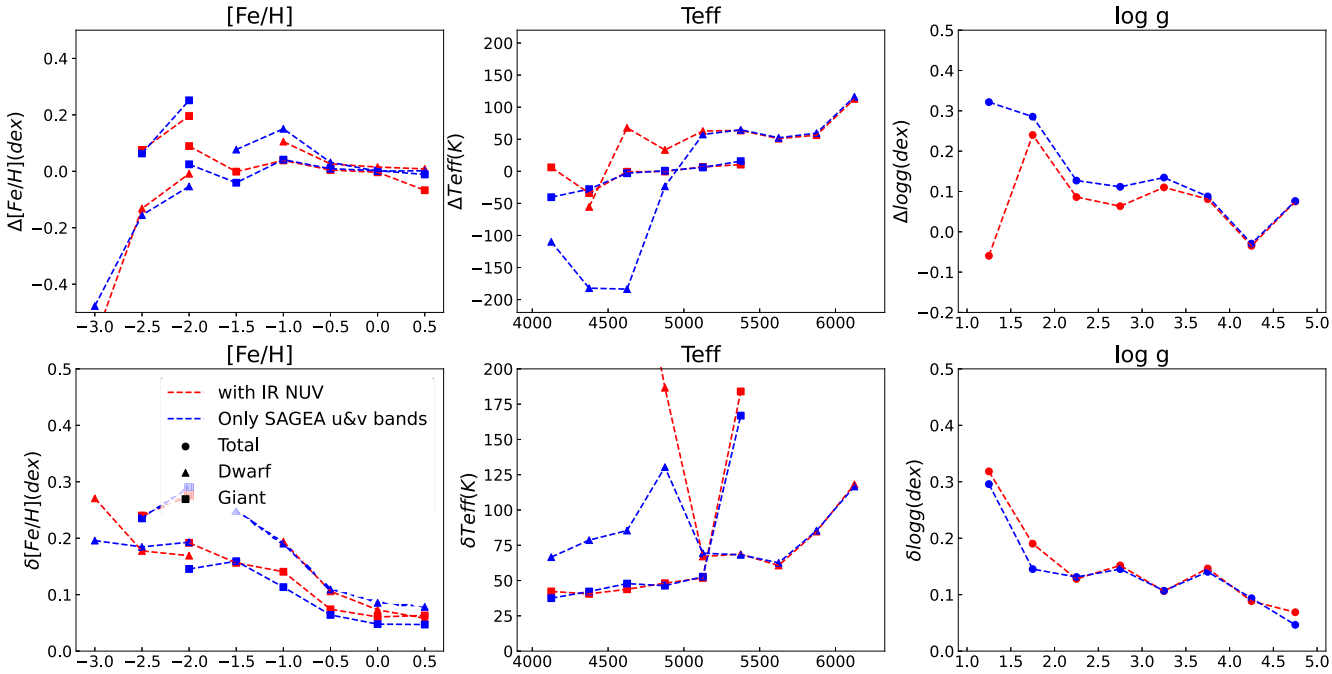

Figure 8. Systematic errors and standard deviations for three stellar parameters ([Fe/H], $T_{\rm eff}$, log g), as a function of their values. Different colors represent different test sets and the different shapes of the points represent different types of stars (triangles for dwarfs, squares for giants, circles for giants and dwarfs).

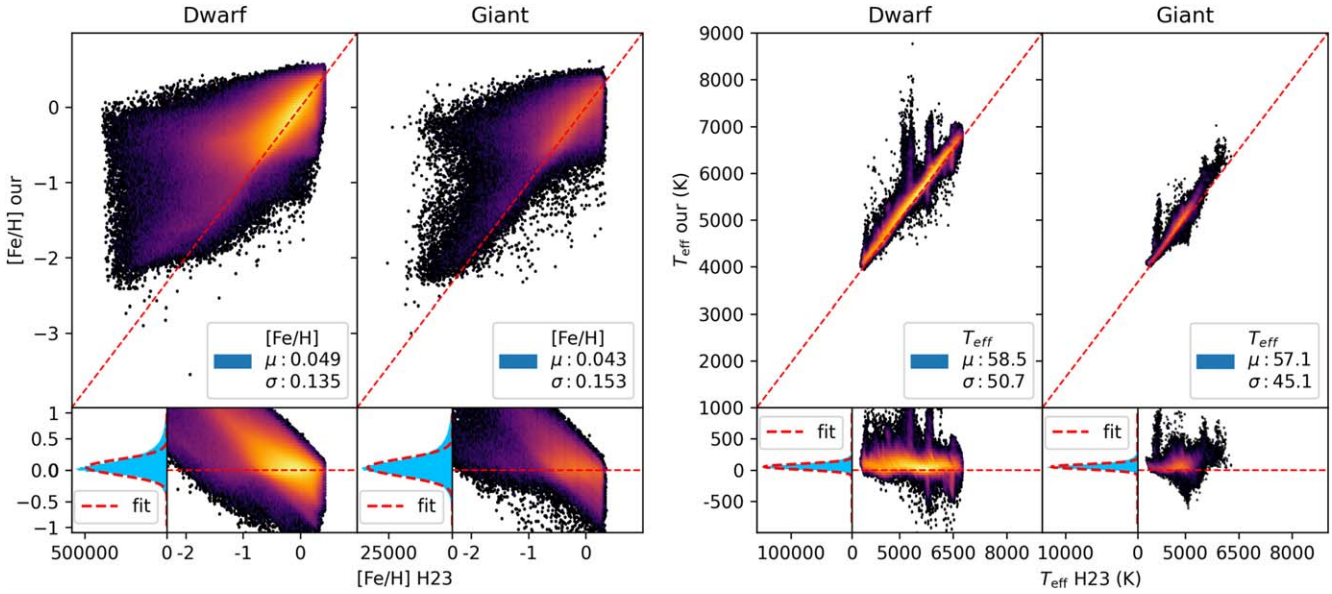

Figure 9. A comparison of the performance of the H23 catalog and our catalog for four models with two parameters. The lower half of each panel shows the distribution of the differences between our model and the H23 model, based on the parameters provided by the H23 model, along with a histogram of the distribution of the differences. The mean and standard deviation were obtained through Gaussian fitting of the histogram, and the fitted values are displayed in the legend of each panel. The red-dashed line in the upper panels is the one-to-one line, and is the zero-residual line in the lower panels.

training sets, while the various shapes represent different types of stars.

[Fe/H] basically has no systematic error, and the deviation is within 0.1 dex, but the dispersion is large at the metal-poor end. When [Fe/H] < -2.0, it is about 0.2–0.3 dex, and when [Fe/H] > -0.5, it can achieve accuracy better than 0.1 dex. The systematic error in $T_{\rm eff}$ is approximately 20 K for giant stars and around 50 K for dwarf stars. Outliers are primarily due to the larger intrinsic dispersion caused by small sample sizes and the lower luminosity of cooler stars. For giant stars with temperatures below 5250 K, the temperature dispersion is better than 50 K. For most dwarf stars, the temperature dispersion is around 100 K. The systematic error of log g is also relatively small as a whole and at about 0.1 dex. For most stars, the log g accuracy can be better than 0.15 dex. The dispersion of this parameter becomes larger with decreasing log g.

In Figure 8, red-dashed lines represent the systematic error and standard deviation of the corresponding parameters in the test set of Data Set 2. When comparing the results of the first data set (blue-dashed line), it can be seen that the second data set exhibits similar precision and accuracy across various

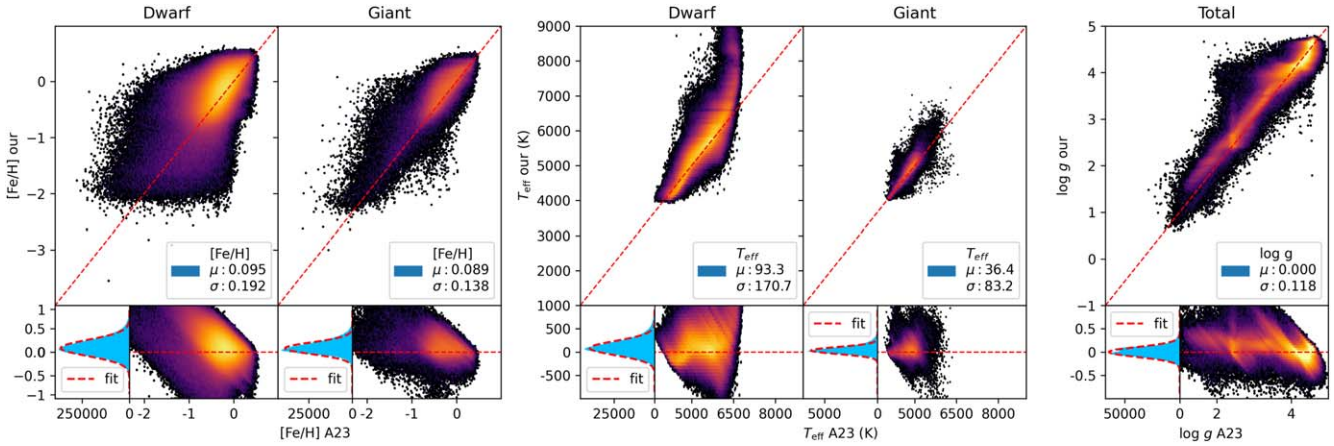

Figure 10. Similar to Figure 9, but for a comparison with the A23 catalog.

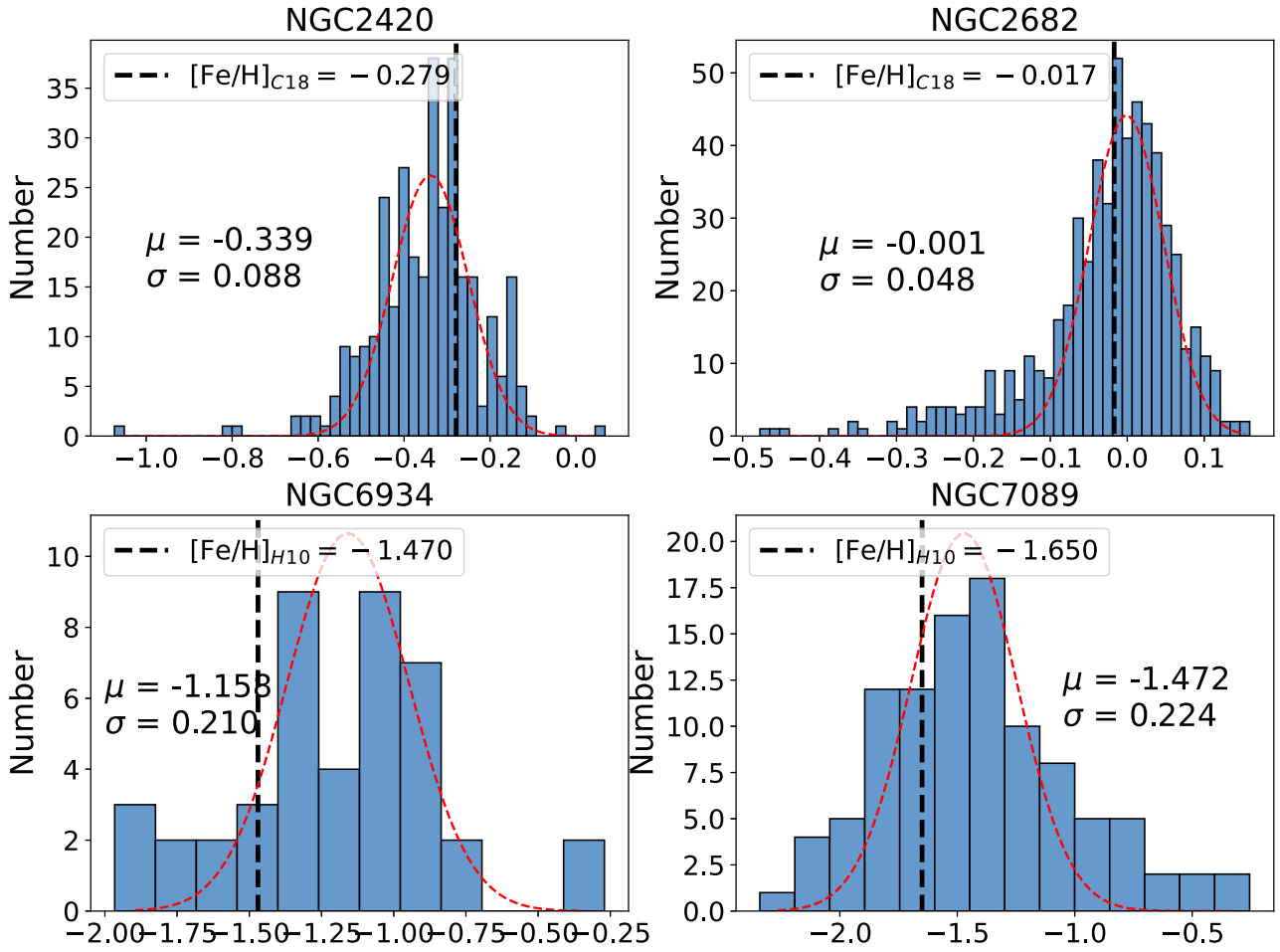

Figure 11. This figure presents the metallicity histograms from this work. The top two panels show the open clusters NGC 2420 and NGC 2682, and the bottom two panels display the globular clusters NGC 6934 and NGC 7089. The red-dashed lines represent the best-fit Gaussian distributions to the histograms, with the mean and standard deviation displayed in the panels. The black dashed lines indicate the cluster metallicities from literature, with their values shown in legends.

parameters. However, the second data set, which includes several different survey data with different limiting magnitudes and a wide range of wavelength, suffers from significant selection effects for both low effective temperature and high effective temperature stars. Additionally, the sample size of the second data set is only one-tenth that of the first data set therefore has a smaller coverage range of stellar parameters. In consequence, subsequent tests will utilize only the first data set.

3.4. External Tests with Other Catalogs

In this study, a substantial portion of publicly available spectroscopic survey data was collected for model training and testing. To conduct external tests, we crossmatch the stellar parameters obtained from the XP spectra provided by Gaia DR3 (R. Andrae et al. 2023, hereafter A23) with those obtained by previous studies utilizing SAGES (H23).

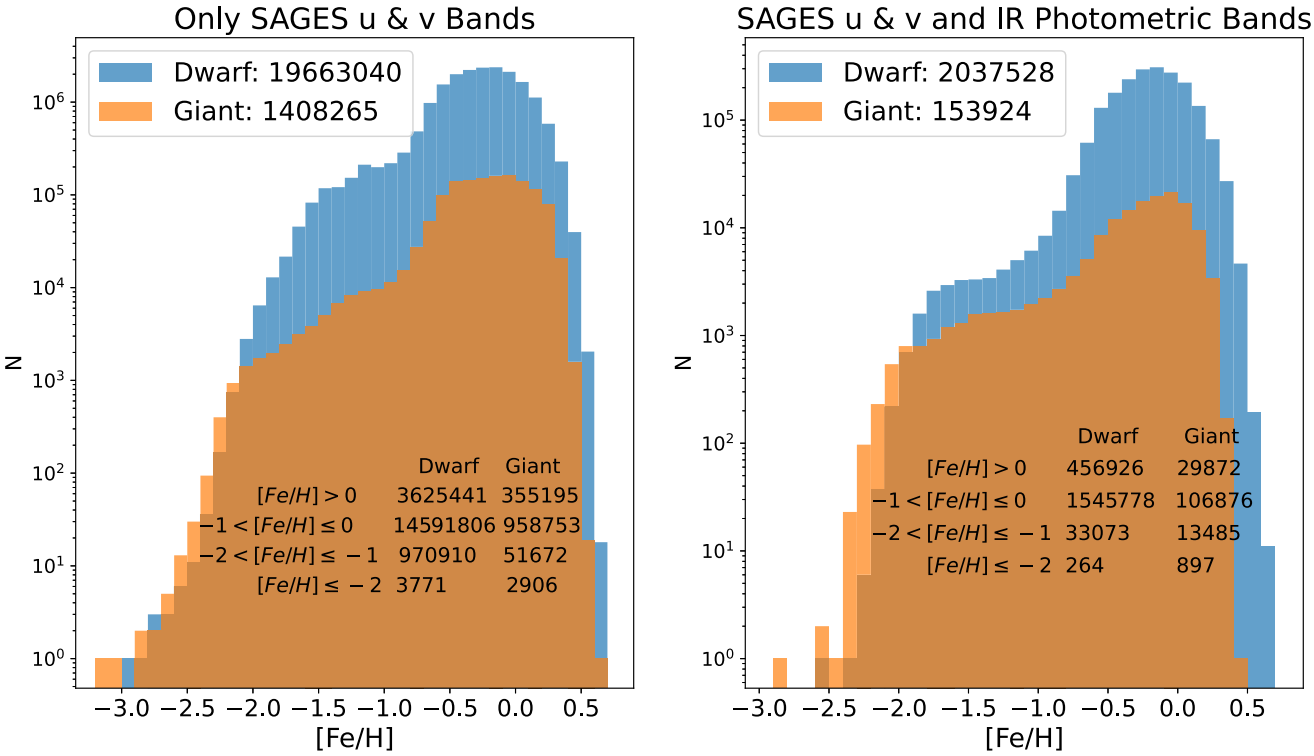

Figure 12. Left panel: The photometric-metallicity distributions of dwarf (blue) and giant (yellow) stars in Data Set 1. The number of dwarf and giant stars in each bin of metallicity are also given. Right panel: similar to the left panel, but using data from Data Set 2.

Table 3
Description of the Final Sample

Parameter	Description	Unit
R.A.	R.A. from SAGES DR1 (J2000)	degrees
Decl.	Decl. from SAGES DR1 (J2000)	degrees
$T_{\rm eff}$	Effective temperature (unreliable data are marked with −9999)	kelvin
log g	Logarithm surface gravity (unreliable data are marked with −9999)	cgs
[Fe/H]	Photometric metallicity (unreliable data are marked with −9999)	…
ebv_sfd	Value of $E\,(B-V)$ from the extinction map of SFD98	…
u/v	Magnitude for the SAGES u/v bands from DR1	mag
u_err/v_err	Uncertainty of magnitude for the SAGES u/v bands from DR1	mag
$G/Bp/Rp$	Magnitudes for three Gaia bands from EDR3	mag
G_err/Bp_err/Rp_err	Uncertainty of magnitudes for three Gaia bands from EDR3	mag
$J/H/K$	Magnitudes for three 2MASS bands	mag
J_err/H_err/K_err	Uncertainty of magnitudes for three 2MASS bands	mag
$W1/W2$	Magnitude for the WISE $W1/W2$ bands	mag
$W1$_err/$W2$_err	Uncertainty of magnitude for the WISE $W1/W2$ bands	mag
NUV	Magnitude for the GALEX NUV bands	mag
NUV_err	Uncertainty of magnitude for the GALEX NUV bands	mag
parallax	Parallax from Gaia EDR3	mas
err_parallax	Uncertainty of parallax from Gaia EDR3	mas
pm/pmra/pmdec	Proper motion in total/R.A./decl. from Gaia EDR3	mas yr^{-1}
err_pmra/err_pmdec	Uncertainty of proper motion in R.A./decl. from Gaia EDR3	mas yr^{-1}
type	Flag to indicate classifications of stars, 1 means dwarf, 0 means giant	…

After crossmatching with the H23 catalog, we identified 13,455,865 sources in common in Data Set 1. The error distribution of the three parameters provided by the five models is depicted in Figure 9. Since the H23 catalog lacks log g information, a direct performance comparison with this model was not possible. The mean and standard deviation of the models were determined by Gaussian fitting of the residual values of stellar parameters measured by the two models (values in our catalog minus values in H23 catalog). The mean offsets for both models in metallicity is close to zero, and the standard deviation is slightly higher than 0.1 dex, indicating that, in terms of metallicity, the parameters generated by our models are in agreement with the values given in the H23 paper. Assuming equal errors in H23 and our model, the errors for both are approximately 0.1 dex.

However, in the effective temperature models, there is a significant systematic difference of about 50 K between the two. This substantial systematic difference may be attributed to

the training set. Our training set used the parameters directly provided by LAMOST DR10, while H23 used stellar parameters they calculated themselves from LAMOST spectra. In both the giant- and dwarf-star temperatures, there are some vertical structures, indicating that a group of stars with similar temperatures in the H23 catalog are considered to have different effective temperature in this work. Given that this structure does not appear in comparisons with other studies, we believe it may be caused by some systematic errors inherent in the SAGES DR1. The polynomial fitting, due to its limited number of parameters, is unable to correct the complex systematic error in input data therefore cause large systematic error in output data.

Similarly, after crossmatching with the A23 parameter catalog, there are 11,367,335 sources in common. The error distribution of the three parameters provided by the five models is illustrated in Figure 10. The generation method of the figure is consistent with the previous discussion. It can be seen that the overall correlation of the five models is quite good. An unusual structure in the surface gravity and effective temperature models is evident. This may be due to the different training data sets used. In the crossmatched catalog between LAMOST and APOGEE, the log g and $T_{\rm eff}$ also exhibit similar anomalous structures. Because the temperature range of the training set used by A23 only extends up to 7000 K, most stars with temperatures greater than 7000 K are considered as stars with temperatures of 6500 to 7000 K. Consequently, a larger standard deviation is exhibited when comparing our model to A23 for dwarf stars.

3.5. External Tests with Star Clusters

This section presents the results of the tests using star clusters. We crossmatched our catalog with the proper motion from Gaia DR3, and performed cuts in R.A. and decl. near the centers of the clusters, restricting the proper-motion errors in both directions to be less than 0.3 mas yr^{-1}. We then selected a stringent sample of cluster members by cutting the overdense regions in the proper-motion space. By examining the color–magnitude diagram of this sample, we confirmed that it is virtually free of field-star contamination. Using this method for member selection, we identified high-purity member stars for two open clusters, NGC 2420 and NGC 2682, and two globular clusters, NGC 6934 and NGC 7089. After applying a signal-to-noise ratio $>$ 20 criterion in the SAGES u and v bands, we conducted metallicity precision tests on the selected member stars.

Figure 11 shows the metallicity distributions of the four clusters, fitted with Gaussian functions, and the resulting mean and standard deviation. The standard deviations are consistent with the metallicity dispersions presented in Section 3.3. By comparing the metallicity measurements of the globular and open clusters with those provided by W. E. Harris (2010; hereafter H10) and T. Cantat-Gaudin et al. (2018; hereafter C18) (indicated by the black dashed lines in Figure 11), it can be seen that the systematic errors in metallicity at the metal-poor end are also consistent with the results from the test set in Section 3.3.

4. The Final Sample

Through the application of the trained models to all photometric data, two sets of value-added catalogs are obtained. The first data set contains 21,071,305 stars, including 19,663,040 dwarf stars and 1,408,265 giant stars; the second data set has 2,191,452 stars, including 2,037,528 dwarf stars and 153,924 giant stars.

In view of the poor extrapolation ability of the random forest algorithm, the results obtained from extrapolating $T_{\rm eff}$ were unsatisfactory. To remove this low-temperature star sample, we performed a polynomial fit to the temperature using $(Bp - Rp)_0$ from the training data and excluded the three atmospheric parameters of the low-temperature stars with $(Bp - Rp)_0 > 1.598$. Figure 12 shows the distribution of metal abundance predicted by the two models. For the first data set, the fractions of stars with [Fe/H] $\leqslant 0$, $\leqslant -1$, and $\leqslant -2$ are 81.11%, 4.88%, and 0.03%, respectively. For the second data set, the fractions of stars with [Fe/H] $\leqslant 0$, $\leqslant -1$, and $\leqslant -2$ are 77.59%, 2.18%, and 0.05%, respectively. The numbers of stars in each interval of [Fe/H] and stellar types are shown in the histogram. Table 3 lists the parameters provided in the final data product.

5. Summary

In this work, we combine stellar-parameter estimates from spectral surveys and high-resolution spectral data collected from the literature with photometric data from SAGES DR1, Gaia DR3, WISE, 2MASS, and GALEX. Using the random forest algorithm, we derive high-precision stellar parameters for a total of 21,071,305 stars, including 19,663,040 dwarf stars and 1,408,265 giant stars. The overall precision is about 0.1 dex in [Fe/H] and log g, and the $T_{\rm eff}$ precision is better than 100 K.

Currently, the log g of stars primarily relies on the input parallax from Gaia DR3. Ongoing SAGES observations in the DDO51 band, which is sensitive to log g, will provide more accurate log g parameter information once the observations are completed. Subsequently, the large number of stars with available stellar parameters can be used as the basis for a variety of follow-up efforts, such as using metal-abundance information provided by this catalog to identity and study stellar streams with Gaia kinematic information, or to find candidate targets for future spectroscopic observation. At present, two catalogs of this work are available at doi:10.12149/101557.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (NSFC) under grant Nos. 11988101, 12261141689, 12090044, and 12090040. T.C.B. acknowledges partial support for this work from grant PHY 14-30152; Physics Frontier Center/JINA Center for the Evolution of the Elements (JINA-CEE), and OISE-1927130: The International Research Network for Nuclear Astrophysics (IReNA), awarded by the US National Science Foundation. We thank the staff of the University of Arizona and the mountain operation team of the Steward Observatory, including Bill Wood, Michael Lesser, Ed Olszewski, Joe Hoscheidt, Gary Rosenbaum, Jeff Rill, and Richard Green, for assistance with the observations. Sponsored by the Xinjiang Uygur Autonomous Region "Tianchi Talent" Introduction Plan. The global common challenge project of the Chinese Academy of Sciences grant No. 178GJHZ2022040GC has partially supported this work.

ORCID iDs

Hongrui Gu https://orcid.org/0009-0007-5610-6495
Zhou Fan https://orcid.org/0000-0003-3067-3540
Gang Zhao https://orcid.org/0000-0002-8980-945X
Huang Yang https://orcid.org/0000-0003-3250-2876
Timothy C. Beers https://orcid.org/0000-0003-4573-6233
Wei Wang https://orcid.org/0000-0002-9702-4441
Jie Zheng https://orcid.org/0000-0001-6637-6973
Jingkun Zhao https://orcid.org/0000-0003-2868-8276
Chun Li https://orcid.org/0009-0000-4835-7525
Yuqin Chen https://orcid.org/0000-0002-8442-901X
Haibo Yuan https://orcid.org/0000-0003-2471-2363
Haining Li https://orcid.org/0000-0002-0389-9264
Kefeng Tan https://orcid.org/0000-0003-0173-6397
Yihan Song https://orcid.org/0000-0001-7255-5003
Ali Luo https://orcid.org/0000-0001-7865-2648

References

Abdurro'uf, Acetta, K., Aerts, C., et al. 2022, ApJS, 259, 35
An, D., & Beers, T. C. 2020, ApJ, 897, 39
Andrae, R., Rix, H.-W., & Chandra, V. 2023, ApJS, 267, 8
Aoki, W., Beers, T. C., Lee, Y. S., et al. 2013, AJ, 145, 13
Bai, Y., Liu, J., Bai, Z., Wang, S., & Fan, D. 2019, AJ, 158, 93
Bianchi, L., Conti, A., & Shiao, B. 2014, yCat, II/335
Bonifacio, P., Spite, M., Cayrel, R., et al. 2009, A&A, 501, 519
Bonoli, S., Marìn-Franch, A., Varela, J., et al. 2021, A&A, 653, A31
Breiman, L. 2001, Mach. Learn., 45, 5
Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, A&A, 618, A93
Cenarro, A. J., Moles, M., Cristóbal-Hornillos, D., et al. 2019, A&A, 622, A176
Chiti, A., Frebel, A., Mardini, M. K., et al. 2021, ApJS, 254, 31
Crawford, D. L., Barnes, J. V., & Golson, J. C. 1970, AJ, 75, 624
Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, RAA, 12, 1197
Cutri, R. M., Skrutskie, M. F., van Dyk, S., et al. 2003, 2MASS All Sky Catalog of Point Sources, http://irsa.ipac.caltech.edu/applications/Gator/
Cutri, R. M., Wright, E. L., Conrow, T., et al. 2021, yCat, II/328
Ezzeddine, R., Rasmussen, K., Frebel, A., et al. 2020, ApJ, 898, 150
Fan, Z., Zhao, G., Wang, W., et al. 2023, ApJS, 268, 9
François, P., Caffau, E., Bonifacio, P., et al. 2018, A&A, 620, A187
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1
Galarza, C. A., Daflon, S., Placco, V. M., et al. 2022, A&A, 657, A35
Harris, W. E. 2010, arXiv:1012.3224
Hauck, B., & Mermilliod, M. 1998, A&AS, 129, 431
Huang, Y., Beers, T. C., Wolf, C., et al. 2022, ApJ, 925, 164
Huang, Y., Beers, T. C., Yuan, H., et al. 2023, ApJ, 957, 65
Huang, Y., Chen, B. Q., Yuan, H. B., et al. 2019, ApJS, 243, 7
Ivezić, Ž., Sesar, B., Jurić, M., et al. 2008, ApJ, 684, 287
Keller, S., Bessell, M., Schmidt, B., & Francis, P. 2007, in ASP Conf. Ser. 364, The Future of Photometric, Spectrophotometric and Polarimetric Standardization, ed. C. Sterken (San Francisco, CA: ASP), 177
Ksoll, V. F., Ardizzone, L., Klessen, R., et al. 2020, MNRAS, 499, 5447
Kunder, A., Kordopatis, G., Steinmetz, M., et al. 2017, AJ, 153, 75
Lai, D. K., Bolte, M., Johnson, J. A., & Lucatello, S. 2004, AAS Meeting, 205, 52.10
Li, H., Tan, K., & Zhao, G. 2018, ApJS, 238, 16
Limberg, G., Rossi, S., Beers, T. C., et al. 2021, ApJ, 907, 10
Lu, X., Yuan, H., Xu, S., et al. 2024, ApJS, 271, 26
Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, AJ, 154, 94
Martin, N. F., Starkenburg, E., Yuan, Z., et al. 2024, A&A, 692, A115
Mendes de Oliveira, C., Ribeiro, T., Schoenell, W., et al. 2019, MNRAS, 489, 241
Nordström, B., Andersen, J., Holmberg, J., et al. 2004, PASA, 21, 129
Rockosi, C. M., Lee, Y. S., Morrison, H. L., et al. 2022, ApJS, 259, 60
Schlaufman, K. C., & Casey, A. R. 2014, ApJ, 797, 13
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Soubiran, C., Le Campion, J.-F., Brouillet, N., & Chemin, L. 2016, A&A, 591, A118
Starkenburg, E., Martin, N., Youakim, K., et al. 2017, MNRAS, 471, 2587
Strömgren, B. 1964, ApNr, 9, 333
Tan, K.-F., Zhao, G., Fan, Z., et al. 2022, RAA, 22, 105004
Wang, W., Zhao, G., Chen, Y., & Liu, Y. 2014, in IAU Symp. 298, Setting the Scene for Gaia and LAMOST, ed. S. Feltzing et al. (Cambridge: Cambridge Univ. Press), 326
Whitten, D. D., Placco, V. M., Beers, T. C., et al. 2021, ApJ, 912, 147
Xu, S., Yuan, H., Zhang, R., et al. 2022, ApJS, 263, 29
Yang, L., Yuan, H., Xiang, M., et al. 2022, A&A, 659, A181
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
Yong, D., Norris, J. E., Bessell, M. S., et al. 2013, ApJ, 762, 26
Yoon, J., Beers, T. C., Placco, V. M., et al. 2016, ApJ, 833, 20
York, D. G., Adelman, J., Anderson, J. E. J., et al. 2000, AJ, 120, 1579
Yuan, H., Liu, X., Xiang, M., Huang, Y., & Chen, B. 2015, ApJ, 803, 13
Zhang, R., & Yuan, H. 2023, ApJS, 264, 14
Zhang, R.-Y., Yuan, H.-B., Liu, X.-W., et al. 2021, RAA, 21, 319
Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, RAA, 12, 723
Zheng, J., Zhao, G., Wang, W., et al. 2018, RAA, 18, 147
Zheng, J., Zhao, G., Wang, W., et al. 2019, RAA, 19, 003