



A Catalog of 12,766 Carbon-enhanced Metal-poor Stars from LAMOST Data Release 8

Ziyu Fang¹, Xiangru Li¹ , and Haining Li² ¹ School of Computer Science, South China Normal University, 510631 Guangzhou, People's Republic of China; lixiangru@scnu.edu.cn² Key Lab of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, A20 Datun Road, Chaoyang, 100102 Beijing, People's Republic of China; lihn@nao.cas.cn

Received 2024 August 29; revised 2024 December 4; accepted 2024 December 6; published 2025 March 6

Abstract

Metal-poor stars are a rare and ancient type of star; carbon-enhanced metal-poor (CEMP) stars are a subset of these celestial bodies that show an enrichment of carbon relative to iron. They are believed to be formed from gas polluted by the first generation of stars after the Big Bang and are important objects for studying the early Universe, galaxy evolution, and nucleosynthesis. Due to their rarity, the search for metal-poor stars and CEMP stars is a valuable task. This study investigates the search for CEMP stars based on the low-resolution stellar spectra from the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) Data Release 8 (DR8), and proposes a deep learning scheme. From the LAMOST DR8 spectral library, this work discovered 12,766 CEMP star candidates. For ease of reference and use, we provide the estimated parameters T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, and $[\text{C}/\text{H}]$ for them.

Unified Astronomy Thesaurus concepts: [Astronomy data analysis \(1858\)](#); [CEMP stars \(2105\)](#); [Chemical abundances \(224\)](#); [Catalogs \(205\)](#)

1. Introduction

Stars that were born in the early Universe usually tend to have lower metallicities (e.g., with $[\text{Fe}/\text{H}] < -2.0$),³ and are referred to as very metal-poor (VMP) stars (T. C. Beers & N. Christlieb 2005). They preserve important clues to the first generation of stars in the Universe, and are regarded as fossil records of the early evolution of the Galaxy. Detailed chemical abundance patterns of these objects can be compared to theoretical models and constrain the nucleosynthesis of early generations of supernovae (K. Nomoto et al. 2013), while their observed abundance trends along with metallicities provide essential information about the chemical history of the Milky Way (A. Frebel & J. E. Norris 2015).

In the past decades, significant effort has been devoted to searching for VMP stars and exploring their properties through survey projects and high-resolution follow-up spectroscopy (W. Aoki et al. 2012; J. E. Norris et al. 2012; J. Yoon et al. 2016, 2018; Y. S. Lee et al. 2017, 2019; K. A. Venn et al. 2019; D. D. Whitten et al. 2021; H. Li et al. 2022; M. Lucey et al. 2023). One of the important discoveries about VMP stars is that a large proportion of these objects exhibit a significant excess of carbon, and they are referred to as carbon-enhanced metal-poor (CEMP) stars (T. C. Beers & N. Christlieb 2005; W. Aoki et al. 2007). Moreover, the fraction of CEMP stars becomes higher at lower metallicities. The CEMP star occurrence fraction could reach over 20% at $[\text{Fe}/\text{H}] < -2.5$ (V. M. Placco et al. 2014), and their high occurrence makes CEMP stars the most pristine objects in the Universe. Therefore, the abundance patterns and origins of CEMP stars are very important tracers to understand early chemical evolution (P. Bonifacio et al. 2012).

Based on their overabundance characteristics for neutron-capture elements, CEMP stars can be divided into different categories such as CEMP-no, CEMP-s, CEMP-r, and CEMP-r/s stars (T. C. Beers & N. Christlieb 2005). Although great effort has been devoted to understanding the abundance patterns of observed CEMP stars, there are still quite many unsolved questions regarding their origins. For CEMP-s stars, it is commonly believed that they originate from binary systems with an asymptotic giant branch companion. However, some of them were identified as isolated stars through radial velocity monitoring observations. The formation mechanisms of the abundance pattern of neutron-capture elements in these stars remain an enigma (T. T. Hansen et al. 2016; W. Shuang et al. 2022). For CEMP-r/s stars, researchers have successively proposed a number of scenarios concerning their origins, including binary systems (S. Lucatello et al. 2005; T. T. Hansen et al. 2016), dual core-flash neutron superburst theories (M. Lugaro et al. 2009), and enrichment through the so-called *i*-process (C. Abate et al. 2016; P. P. Goswami & A. Goswami 2020). However, none of these models can perfectly explain the abundance patterns observed in these stars. For CEMP-r stars, the lack of observational data significantly limits exploration of their origins (M. Safarzadeh et al. 2019; W. Shuang et al. 2022). CEMP-no stars tend to dominate the lowest metallicities, and thus their properties are very closely related to the nature of first stars and early chemical enrichment. The predominant hypotheses explaining the observed patterns in CEMP-no stars include rotating massive stars, faint supernovae, and inhomogeneous metal mixing. However, a definitive consensus remains elusive (H. Li et al. 2022).

Fully understanding the nature and origin of these most ancient CEMP stars would help us explore the first stars and the chemical enrichment of the early Universe. However, due to their rarity and some challenges in identifying them using low-resolution spectroscopy, there have been very few systematic searches for CEMP stars. There have been a few efforts concerning solving this problem. The first identification of CEMP stars was done by T. C. Beers et al. (1992), from HK

³ The standard notation of $[X/\text{H}] = \log(X/\text{H})_{\star} - \log(X/\text{H})_{\odot}$ is adopted.



low-resolution spectra, with quantitative carbon estimates provided by S. Rossi et al. (2005). Then V. M. Placco et al. (2010) selected CEMP candidates from Hamburg/ESO survey (HES) data through quantitatively estimating the strength of the G band, and later on, H. Li et al. (2018) adopted similar methods to identify CEMP candidates from Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) Data Release 3 (DR3). Both of them have resulted in over 600 candidate CEMP stars. Furthermore, M. Lucey et al. (2023) compiled a CEMP star catalog based on slitless spectra from Gaia. Additionally, advancements in photometric techniques, particularly with the use of narrowband filters in J-PLUS and S-PLUS, have enhanced our ability to identify CEMP stars and estimate $[C/Fe]$. For instance, D. D. Whitten et al. (2021) utilized S-PLUS Data Release 2 to estimate effective temperatures and metallicities for over 700,000 stars, constraining the metallicity distribution function of K dwarfs and identifying 364 candidate CEMP stars. Following this, L. Yang et al. (2022) leveraged J-PLUS data to provide effective temperatures, surface gravities, and elemental abundances for approximately two million Galactic stars, showcasing the potential of photometric data in studying chemical evolution. More recently, Y. Huang et al. (2024) provided effective temperature, surface gravity, age, and metallicity parameters for about five million stars, contributing valuable data for understanding the assembly and chemical evolution of the Milky Way. However, these studies did not quantitatively estimate the C abundance for these stars. To this end, this work is to find some CEMP star candidates by estimating C abundances based on LAMOST observations.

LAMOST (also called the Guo Shou Jing Telescope) is a reflecting Schmidt telescope with an effective aperture of 3.6–4.9 m and a field of view of 5° (X.-Q. Cui et al. 2012; L.-C. Deng et al. 2012; G. Zhao et al. 2012; X.-W. Liu et al. 2015). The focal surface of LAMOST can accommodate 4000 optical fibers and observe up to 4000 objects simultaneously. After a 2 yr commissioning phase, 1 yr pilot survey, and 11 yr of regular surveys, LAMOST has collected 11,581,542 star spectra with low resolution.⁴ In addition, the wavelength coverage of 3700–9100 Å and resolution of $R \sim 1800$ enable quite robust estimation of stellar parameters for FGK stars (A.-L. Luo et al. 2015; H.-B. Yuan et al. 2015), even down to very low metallicities (H. Li et al. 2018; X. Hou et al. 2024). This huge spectral database thus provides an excellent opportunity for searching for CEMP stars.

The organization of the remaining parts of this paper is as follows. Section 2 briefly reviews related works and discusses the motivation of this work. Section 3 introduces the overall process of the CEMP star search in this paper, the reference data used, and the data preprocessing procedures. Section 4 introduces the proposed method for estimating stellar spectral parameters. These parameter estimation results are used for screening and judging CEMP star candidates. Section 5 presents the CEMP star candidate search results. Section 6 compares the consistency of the parameter estimation results of this work with GALactic Archaeology with Hermes (GALAH). Section 7 provides a brief summary for the proposed CEMP star search scheme and its results.

2. Related Works and Motivation

The searching technologies for CEMP stars can be divided into the following three categories: line index methods, template-matching methods, and machine learning methods. The aforementioned classification criteria are established based on their style of using spectral information. Line index methods are characterized by a scalar description of one or more wavelength bands selected for spectral information, for example, the FWHM of a spectral line, or some measures based on the spectral flux integral. The template-matching method estimates parameters or recognize objects by comparing the corresponding flux of an observed spectrum on one or several specified wavelength bands with that of some reference spectra. The machine learning method is to treat a spectrum as a vector or matrix and establish a mapping from the spectrum to the parameter or type labels. These differences between these methods result in their limited applicability and characteristics. The line index method is the most interpretable due to the mandatory requirements for spectral lines and wavelength band selection, but it also requires the highest spectral resolution; the template-matching method is sensitive to spectral quality due to the direct dependence on individual flux. In essence, the key feature of machine learning methods is the automatic extraction of spectral features through algorithms; the automatic feature extraction can exploit the correlation between various wavelength bands and suppress some negative effects from noise. In theory, therefore, the machine learning method can deal with spectra with relatively low quality and resolution, but it needs a large amount of reference data as a carrier of empirical knowledge.

Therefore, the proposal, development, and selection of these schemes depend not only on the physical interpretability of the models, the signal-to-noise ratio (SNR) and resolution of the data, and the accumulation of CEMP star observation data, but also on the available computing resources. The following part of this section reviews studies based on these three types of methods. Although this work deals with low-resolution spectra, this review is dedicated to a comprehensive summary of CEMP searching techniques. Therefore, it is not limited to research works based on low-resolution spectroscopy.

2.1. Line Index Methods

CEMP stars are typically characterized by their low metallicity $[Fe/H]$ and high carbon abundance $[C/Fe]$. The line index method is based on the detection and description of spectral lines sensitive to metallicity and carbon abundance. The vast majority of early identified CEMP stars were discovered from the metal-poor star candidates in the OBJECTIVE-PRISM SURVEY, for example, the HK survey (T. C. Beers et al. 1985, 1992) and HES (A. Frebel et al. 2006; N. Christlieb et al. 2008). Both surveys search objects based on the presence of weak Ca II lines. N. Christlieb et al. (2001) published a HES catalog of some stars with strong carbon molecular lines, and subsequent medium-resolution spectroscopic observations and inspection of these objects showed that at least 50% have the characteristics of CEMP stars (A. Goswami et al. 2006). However, these carbon-rich candidates are sifted based on the aggregated evaluation of a series of carbon molecular lines, e.g., CN, C₂, CH, etc. Therefore, this approach overemphasizes cooler stars. Unfortunately, CEMP stars with effective temperatures above 5500 K usually exhibit only the unusual strength of a

⁴ <https://www.lamost.org/lmusers/>

single carbon signature, the CH G band near 4300 Å. Therefore, this sifting strategy likely results in the omission of such CEMP stars.

As a molecular absorption feature of C2, the Swan bands are spectral signatures of carbon overabundance, and can be used to identify CEMP stars. Therefore, D. F. Čotar et al. (2018) proposed a supervised method and an unsupervised detection method for the Swan bands. The supervised detection method enables the detection of Swan-band features near 4737 Å from a high-resolution spectrum, and a description of their shape and intensity. Using the intensity of the perceivable Swan-band features, CH stars can be detected. D. F. Čotar et al. (2018) then sifted the CEMP candidates based on visual detection and unsupervised machine learning methods (more details about the unsupervised detection method are explored in Section 2.3).

Therefore, the searching technologies based on the CH G band for CEMP stars is an important exploration direction. T. C. Beers et al. (1999) defined a line index, GP, for the wavelength range of 4297.5–4312.5 Å for the CH G band. However, some studies show that the line index GP does not fully capture all of the characteristics of the carbon absorption characteristics of the CH G band. To this end, N. Christlieb et al. (2008) defined a line index, GPES, using the wavelength range of 4281.0–4307.0 Å. V. M. Placco et al. (2010) proposed an extended line index, GPES Extended (GPE; as defined in Equation (1) in V. M. Placco et al. 2010) based on the wavelength band 4200.0–4400.0 Å, and accordingly computed a CEMP candidate catalog consisting of 669 previously unconfirmed CEMP candidates. The line index GPE describes the contrast between the CH G band and the corresponding fitted continuum. V. M. Placco et al. (2011) further improved the GPE and proposed the line index EGP (an extension with a red sideband for GPE, as defined in Equation (1) in V. M. Placco et al. 2011). Based on the line indexes G1 and EGP, H. Li et al. (2018) further proposed the criteria $G1 > 4.0$ Å and $EGP > -0.7$ mag, for searching for CEMP star candidates from stellar spectra with $4000 \text{ K} < T_{\text{eff}} < 7000 \text{ K}$ and found 636 CEMP star candidates from the LAMOST DR3 library. In the verification stage, the CH G band feature near 4300 Å is selected.

2.2. Template-matching Methods

The template-matching method, also known as χ^2 minimization, is widely used in astronomical observation signal identification and parameter extraction. Its principle is to calculate the cumulative sum of the difference between the corresponding fluxes of the observed data and some templates/empirical data as a measure of the dissimilarity between them, and then predict the observed data according to the type or parameter information of the template which can minimize the cumulative sum. For example, this method can be used to estimate both [Fe/H] and [C/Fe], and also to identify CEMP stars. The idea of the template-matching method is to make predictions based on the most consistent template or empirical sample with the observation data. It is simple and intuitive. Therefore, the template-matching method is widely adopted, and has become a benchmark method.

The keys to the template-matching method are the construction of the templates/empirical database and the selection of the wavelength bands. The former determines which empirical samples are used to predict the observed data, while the latter determines what characteristics are used to judge the similarity

between the observational and empirical data. For example, a rest wavelength range of 4000–4650 Å is used to estimate [C/Fe] from Sloan Digital Sky Survey (SDSS) data in Y. S. Lee et al. (2013), as this wavelength band contains the CH G band feature. To estimate [Fe/H], search for metal-poor star candidates, and CEMP star candidates from LAMOST Data Release 1 (DR1) data, H.-N. Li et al. (2015) proposed to use the 4500–5500 Å band. D. S. Aguado et al. (2017) proposed a two-step method for sifting CEMP star candidates from BOSS, SEGUE, and LAMOST spectra. This scheme was first used to find some CEMP candidates by estimating the [Fe/H] using the Ca II resonance line, where some outliers (false candidates) and obvious erroneous results were removed using a template-matching method based on the local bands around the Balmer line. To search for extremely metal-poor (EMP) stars and CEMP stars from LAMOST DR3, H. Li et al. (2018) used the wavelength range of 4360–5500 Å.

The abovementioned works indicate two classes of schemes for searching for CEMP star candidates: (1) first screen some carbon-enhanced stars, then identify some CEMP star candidates based on their metallicity; and (2) first detect some metal-poor star candidates from a stellar spectrum library, and then identify some CEMP star candidates based on their [C/Fe] characteristics. Therefore, the finding of CEMP stars in some of the abovementioned works is the by-product of searching for metal-poor stars, for example, in H.-N. Li et al. (2015) and D. S. Aguado et al. (2017). The wavelength range selection depends on the selected search idea and the parameter configurations of our interested CEMP stars. The selection of line index schemes share similar theories.

2.3. Machine Learning Investigations for Carbon-enhanced Metal-poor Candidate Searching

The fundamental idea of machine learning methods is to automatically discover and extract the discriminative features of CEMP stars from some empirical data through some algorithms. Although research of machine learning-based CEMP candidate searching has attracted more and more attention with the rapid increase of spectral data volume, it is still in its infancy. Related studies can be divided into unsupervised methods (D. F. Carbon et al. 2017; D. F. Čotar et al. 2018), regression methods (J. Xie et al. 2021), and supervised classification methods (M. Lucey et al. 2023).

The principle of the unsupervised CEMP star search method is to automatically find the aggregated relationship in some observed spectra through an algorithm. The spectra of a cluster are numerically very similar with each other, and there is a high probability that the corresponding objects come from a certain type of non-CEMP star, or a subtype of CEMP stars. Therefore, a spectrum is likely, with high probability, to be of a CEMP star observation if the star is in a cluster with a confirmed CEMP. For example, D. F. Čotar et al. (2018) searched for CEMP stars from the GALAH observation spectral library based on the wavelength band 4720–4890 Å and the t-SNE method. D. F. Carbon et al. (2017) searched for EMP, CEMP, and cataclysmic variable stars from SDSS data based on a linked scatterplot method. This approach helps to find the spectra of some novel subtypes of CEMP candidates.

The idea of the regression method is to formulate the CEMP search problem as a parameter extraction problem. J. Xie et al. (2021) made a very good pilot exploration in this area. That work represented a spectrum with a 64×64 matrix through the

s-shaped folding technique, and estimated the parameters $[\text{Fe}/\text{H}]$ and $[\text{C}/\text{Fe}]$ from a stellar spectrum by inputting the matrix/image into a convolutional neural network. Furthermore, VMP and CEMP star candidates were recognized from LAMOST Data Release 7 data based on the criteria $[\text{Fe}/\text{H}] < -2$ and $[\text{C}/\text{Fe}] > 1$. The model successfully found 260 of 414 known CEMP stars with a recall of 62.80%. Although there is a lot of room for improvement in the recall rate, the feasibility of such schemes was preliminarily explored. Compared to the CEMP star search works (H.-N. Li et al. 2015; H. Li et al. 2018) using LAMOST observations, the advantage of this scheme is that it can give $[\text{C}/\text{Fe}]$ estimates from low-resolution spectra. These estimated parameters help to select appropriate sources for high-resolution follow-up observations.

The supervised classification method characterizes the search problem for CEMP stars as a discrimination problem between a CEMP star and a non-CEMP star. The basic idea of this method is to project the observed spectrum into a certain space through an algorithm; the samples of CEMP stars and non-CEMP stars are in different regions of this space. Moreover, the mapping relationship from an observed spectrum to this space can be learned from some known CEMP star spectra and non-CEMP star spectra. Therefore, an observed spectrum can be first projected into this learned space and discriminated based on the region where they are located. For example, M. Lucey et al. (2023) explored the CEMP candidate search problem based on the XGBoost method and Gaia spectra, and obtained a recall rate of approximately 73%. The disadvantage of this work is the failure to give estimates of $[\text{C}/\text{Fe}]$ and other parameters for the CEMP star candidates from low-resolution spectra. However, the typical characteristic of this study is that it verified the feasibility of the machine learning CEMP search method for spectra with a resolution as low as $R = 50$. Therefore, the results of this work provide a feasible basis for CEMP searching from LAMOST low-resolution spectra with a resolution $R = 1800$ using a machine learning method.

2.4. Motivation

With the development of the large-scale spectroscopic surveys of SDSS, LAMOST, and Gaia, low-resolution spectroscopic CEMP star candidate searches have attracted much attention. Due to the limitations of the technology used, however, these survey data have not been fully exploited.

With a decrease of spectral resolution and SNR, the identifiability, detection, and description reliability of spectral lines are greatly reduced. Furthermore, in a large-scale low-resolution spectral survey, the SNR range of the spectrum is relatively wide. For example, there are over 5.4 million spectra with $\text{SNR}_g \leq 20$ in LAMOST Data Release 9 (DR9). Therefore, LASP, the official stellar spectral processing pipeline of LAMOST, uses the template-matching method rather than the line index method. However, the template-matching method is implemented by calculating the cumulative difference between the corresponding flux of the template data and an observation as a spectral dissimilarity measure. This metric fails to exploit the local morphological and global morphological features in the spectrum. The lack of morphological feature extraction means that the noise in individual pixels can seriously distort the final dissimilarity measure even if the observed spectra are very consistent with the template spectra on the whole. Therefore, application of the template-matching method is severely limited for low-SNR spectra. For example, although

LAMOST DR9 released 11.22 million low-resolution spectra, the LASP estimated stellar parameters only for 6.18 million spectra. Moreover, LASP only provides effective temperature, surface gravity, and $[\text{Fe}/\text{H}]$, but does not estimate $[\text{C}/\text{Fe}]$ or $[\text{C}/\text{H}]$ (Y. Wu et al. 2011; A.-L. Luo et al. 2015).

Fortunately, H.-N. Li et al. (2015), H. Li et al. (2018) made a very good pilot exploration and preliminarily showed the potential of searching for CEMP candidates from LAMOST low-resolution spectra. However, the aforementioned explorations were conducted based on the line index method and the template-matching method by limiting the analyzed spectral SNR to $\text{SNR}_g > 10$ and $\text{SNR}_r > 15$, and they did not estimate $[\text{C}/\text{Fe}]$ or $[\text{C}/\text{H}]$ from the low-resolution spectra. In the low-resolution spectral library of LAMOST DR9 v1.1, up to 3.58 million stellar spectra do not meet the abovementioned SNR requirements.

Therefore, it is necessary to explore for CEMP star search techniques that are more suitable for low-resolution, low-SNR spectroscopy to fully exploit the scientific value of large-scale low-resolution spectral libraries. Fortunately, this issue has been explored based on machine learning schemes, such as unsupervised clustering (D. F. Carbon et al. 2017; D. F. Čotar et al. 2018), supervised classification (M. Lucey et al. 2023), and supervised regression estimation (J. Xie et al. 2021). In terms of the resolution and wavelength coverage of the analyzed spectra, D. F. Carbon et al. (2017) and J. Xie et al. (2021) provide excellent feasibility support for exploring CEMP candidate search techniques from LAMOST low-resolution spectra based on machine learning schemes.

However, the aforementioned CEMP candidate search studies based on unsupervised clustering and supervised classification do not give parameter estimates for the candidates. The lack of spectral parameters has a negative effect on selecting objects with high scientific value for any future high-resolution observations. At the same time, it has been shown that the available search methods based on unsupervised clustering tend to miss rare CEMP candidates (D. F. Čotar et al. 2018); the CEMP candidate search scheme based on the supervised machine learning method in M. Lucey et al. (2023) has a recall ratio of only 73%, which means a high missing percentage. Although J. Xie et al. (2021) gave some preliminary results for CEMP candidate searching and parameter estimation from LAMOST spectroscopy using machine learning methods, their recall is $260/414 \approx 62.80\%$. These results indicate that there exists a lot of room for improving parameter estimation accuracy and CEMP candidate identification performance.

To this end, this work is to explore for a suitable machine learning method for CEMP candidate searches using low-resolution spectra, especially LAMOST observations.

3. Methodology, Data, and Preprocessing

3.1. Overall Process

This work is to search for CEMP candidates from the LAMOST DR8 low-resolution spectroscopic database. For a given LAMOST stellar spectrum, we first (1) estimate its effective temperature T_{eff} , surface gravity $\log g$, metallicity $[\text{Fe}/\text{H}]$, and carbon abundance $[\text{C}/\text{H}]$; (2) select the metal-poor star candidates based on the criterion $[\text{Fe}/\text{H}] < -1$; and lastly (3) identify some CEMP star candidates from the metal-poor star candidates using

the proposed criteria in W. Aoki et al. (2007):

$$\begin{cases} \log(L/L_{\odot}) \leq 2.3 \wedge [C/Fe] \geq +0.7, \\ \log(L/L_{\odot}) > 2.3 \wedge [C/Fe] \geq +3.0 - \log(L/L_{\odot}), \end{cases} \quad (1)$$

where

$$\begin{aligned} L/L_{\odot} &\propto (R/R_{\odot})^2 (T_{\text{eff}}/T_{\text{eff}\odot})^4 \\ &\propto (M/M_{\odot})(g/g_{\odot})^{-1} (T_{\text{eff}}/T_{\text{eff}\odot})^4, \end{aligned}$$

where M_{\odot} , g_{\odot} , and $T_{\text{eff}\odot}$ represent the mass, surface gravity, and effective temperature of the Sun, respectively.

3.2. Reference Data Set for the Carbon-enhanced Metal-poor Star Search

The parameter estimation procedure of the proposed scheme is implemented using a machine learning method. This method needs a reference set for estimating the model parameters. According to the definition of CEMP stars in W. Aoki et al. (2007; Section 3.1), the distinction between CEMP stars and carbon-normal metal-poor stars is described based on effective temperature, metallicity, and carbon abundance; there should be some significant differences on the spectral features between giants and dwarfs. This definition is designed based on the influence of stellar evolution on carbon abundance. Therefore, this reference set should not only consist of a series of stellar spectra, but also provide reference information for parameters such as T_{eff} , $\log g$, $[Fe/H]$, and $[C/H]$ for each spectrum.

For this purpose, we establish a reference set by cross-matching the LAMOST DR8 low-resolution stellar spectroscopic database with the APOGEE Data Release 17 (DR17) star catalog (Abdurro'uf et al. 2022), the LAMOST-Subaru star catalog (W. Aoki et al. 2022; H. Li et al. 2022), and the SAGA database (T. Suda et al. 2008, 2011; S. Yamada et al. 2013; T. Suda et al. 2017) with the Tool for Operations on Catalogs and Tables (TOPCAT; M. B. Taylor 2005). In TOPCAT, we limit the maximum angular separation (`max error`) between two stellar position coordinates (R.A., decl.) to 3", and set the `match selection` so that each LAMOST spectrum has the best-matching parameters in the APOGEE DR17 star catalog. The LAMOST-Subaru catalog and the SAGA catalog complement APOGEE DR17 in terms of low metal abundance sources and CEMP stars. Each reference sample consists of one observed spectrum, and its parameter information such as T_{eff} , $\log g$, $[Fe/H]$, and $[C/H]$; the spectral data are from LAMOST DR8, and the parameter information is from the APOGEE DR17 catalog, the LAMOST-Subaru catalog, or the SAGA catalog.

In conventional studies estimating stellar atmospheric physical parameters, researchers typically focus on the spectra in some parametric regions frequently being sampled. However, these spectra are usually not from metal-poor stars or CEMP stars. Reference samples of metal-poor stars are relatively rare, and their number is far less than that of non-metal-poor star samples. Consequently, the initial reference data exhibit a significant imbalance between metal-poor stars and non-metal-poor stars. This imbalance can lead to difficulties for the search model in perceiving the spectral features of metal-poor stars, including CEMP stars. To address this issue, we performed a downsampling of the spectra of non-metal-poor stars obtained from the matching. The final reference set consists of 9755 spectra along with their corresponding parameter information. In this reference set,

there are 4723 samples of non-metal-poor stars, and 5032 samples of metal-poor stars. In the metal-poor star samples, there are 167 observations of CEMP stars. In the 4723 non-metal-poor star samples, 4713 samples were randomly selected from common observations between the LAMOST DR8 and APOGEE DR17 catalogs. Additionally, nine samples are in common between LAMOST DR8 and the SAGA catalogs, while one sample is in common between LAMOST DR8 and the LAMOST-Subaru catalogs. In the 5061 metal-poor star samples, 4433 are in common between the LAMOST DR8 and APOGEE DR17 catalogs. Additionally, 486 samples are in common between the LAMOST DR8 and the LAMOST-Subaru catalogs, while 142 samples are in common between the LAMOST DR8 and the SAGA catalogs. The parameter ranges of this reference set are $[3615.00, 6772.54]$ K for T_{eff} , $[-0.15, 5.06]$ for $\log g$, $[-4.38, 0.59]$ for $[Fe/H]$, and $[-4.50, 0.90]$ for $[C/H]$.

In the obtained reference data, CEMP stars are significantly less common than carbon-normal metal-poor stars and non-metal-poor stars. To enhance the sensitivity of the parameter estimation model to the spectral features of CEMP stars, particularly those with extremely low metallicity, this study has assigned a label CEMP, non-metal-poor, or carbon-normal metal-poor to each reference spectrum (more about this can be found in Section 4.1). These labels are represented by the numerical values -1 , 0 , and 1 , respectively. The categorization labels for carbon-normal metal-poor and CEMP are calculated from the spectroscopic parameter information using the aforementioned criteria. The reference set consists of 167 CEMP samples, 4865 carbon-normal metal-poor samples, and 4723 non-metal-poor star samples.

The aforementioned reference set is randomly partitioned into a training set and a testing set at an approximate ratio of 8:2. These two sets are respectively denoted as $S_{\text{tr}}^{\text{stars}}$ and $S_{\text{te}}^{\text{stars}}$. The sample sizes of these two data sets are 7804 and 1951, respectively.

3.3. Data Preprocessing

The observational process is susceptible to various influences such as scattering, reflection, temperature variations, instrument anomalies, and other factors. These factors can introduce deviations between observed spectra and theoretical spectra. These deviations can reduce the accuracy of parameter estimation and the search for metal-poor stars and CEMP stars. Therefore, a series of preprocessing steps need to be conducted to minimize the adverse effects from these factors as much as possible.

The preprocessing steps used in this study are as follows.

1. The observed wavelengths of the spectra are transformed into the rest wavelength coordinate system based on the radial velocities estimated by the LAMOST pipeline.
2. According to the common wavelength range in the rest frame, each observed spectrum is truncated. The truncated spectra are then divided into a blue range of $[3800, 5700]$ Å and a red range of $[5900, 8800]$ Å to eliminate stitching areas. Each spectrum is resampled using a linear interpolation method with a step size of 1 Å for the blue end and 1.5 Å for the red end. The flux sequences of the spectra in the blue and red ends are denoted as f_b and f_r , respectively.
3. To accurately fit the pseudocontinua of the spectra, we first process the spectra using a median filter with a

Table 1
Parameter Estimation Model: Multiscale Feature Learning and Cross-band Information Mining Network

A Spectrum	
(1)	<i>Multiscale feature extraction.</i> Consecutively perform four layers of convolution followed by average pooling with a scale of two. Each convolutional layer is computed based on $4N_i$ convolution kernels, where the number of kernels with scales 1, 3, 5, and 7 is N_i , and i represents the index of the convolutional layer, ranging from 1 to 4. Here, $N_1 = 20$, $N_2 = 40$, $N_3 = 80$, and $N_4 = 160$.
(2)	<i>Multiscale feature fusion.</i> Based on 100 convolution kernels with a scale of 12, feature fusion is performed with a convolution operation using a stride of 12.
(3)	<i>Information exploitation across wavelength.</i> Segment the spectral information into 20 subbands with equal intervals according to wavelength, and utilize two long short-term memory learning modules to explore interband information. These two LSTM modules respectively extract low-band spectral features to complement high-band spectral features and high-band spectral features to complement low-band spectral features.
(4)	Fuse the spectral features using a two-layer fully connected network respectively consisting of 1025 and 512 neurons. Each neuron adopts the rectified linear unit (ReLU) activation function, and the batch normalization is applied after each fully connected layer computation.
(5)	<div style="display: flex; justify-content: space-between;"> <div style="width: 48%;"> A two-layer fully connected network respectively consist of 256 neurons and four neurons. The first layer incorporates batch normalization, the ReLU activation function, and dropout regularization, while the second layer consists of a linear prediction output unit. </div> <div style="width: 48%;"> A two-layer fully connected network respectively consist of 256 neurons and three neurons. The first layer incorporates batch normalization, the ReLU activation function, and dropout regularization, while the second layer is a softmax regression estimation layer. </div> </div>
(6)	<div style="display: flex; justify-content: space-between;"> <div style="width: 48%;">Spectral parameters are T_{eff}, $\log g$, $[\text{Fe}/\text{H}]$, and $[\text{C}/\text{H}]$.</div> <div style="width: 48%;">Categories are carbon-normal metal-poor stars, CEMP, and non-metal-poor.</div> </div>

Note. This model incorporates two branches, namely parameter estimation (described on the left side of the table) and classification (described on the right side of the table), in the last two rows (steps (5) and (6)).

window size of three to eliminate some noise. Subsequently, we estimate the pseudocontinua respectively for the blue end and red end using the asymmetric least-squares method (S. He et al. 2014). These fitted pseudocontinua are denoted as f_b^c and f_r^c , respectively.

4. Finally, we obtain the normalized spectra, $\tilde{f}_b(i) = f_b(i)/f_b^c(i)$ and $\tilde{f}_r(i) = f_r(i)/f_r^c(i)$, by dividing the flux of the spectra by the corresponding flux of the fitted pseudocontinuum.

4. Stellar Spectroscopic Parameter Estimation

In the CEMP star search workflow of this work (Section 3.1), the key procedure is to estimate the stellar spectroscopic parameters T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, and $[\text{C}/\text{H}]$. Therefore, this section first introduces the parameter estimation scheme (Table 1).

4.1. Multiscale Feature Learning and Information Exploitation across Wavelength Space

The key characteristics of this approach are its ability to learn multiscale features and exploit cross-wavelength information. Therefore, it is referred to as the Multi-Scale Feature Learning and Information Exploitation across Wavelength Network (MSFL-IECW Net). In the output end of this model, there are two branches: one parameter estimation branch and one classification branch. According to the proposed workflow for CEMP star search (Section 3.1), theoretically, MSFL-IECW Net does not need a classification branch. However, in the reference data, CEMP stars are significantly less prevalent than carbon-normal metal-poor stars and non-metal-poor stars. Experimental studies indicate that the addition of a classification branch can enhance the sensitivity of the model to the spectral features of the samples with particularly low metallicity during computation steps (1)–(4) (Table 1), and improve the reliability of the model’s parameter estimation for such samples.

4.2. Model Evaluation

The learning process of the parameter estimation model proceeds as follows: First, the reference set $S_{\text{tr}}^{\text{stars}}$ (Section 3.2) undergoes preprocessing (Section 3.3) to obtain the preprocessed training set $S_{\text{tr,pre}}^{\text{stars}}$. Subsequently, the preprocessed training set is fed into MSFL-IECW Net to learn model parameters. The evaluation of the learning results on the test set $S_{\text{te}}^{\text{stars}}$ (Section 3.2) is presented in Figure 1.

For T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, and $[\text{C}/\text{H}]$, only 1.89%, 1.58%, 1.74%, and 1.58% of the predictions deviate from their theoretical consistency over 3σ , respectively. This phenomenon indicates that this work is extremely consistent with the reference set. Due to the poor quality of a subset of the LAMOST spectra, it is unavoidable that some of the parameter estimations have relatively evident deviations. This is the main reason for the outliers. Therefore, an additional high-confidence catalog has been compiled in this work in Section 5.1 in order to facilitate more accurate follow-up studies of CEMP stars. More related evaluations are conducted in Section 6.

5. Carbon-enhanced Metal-poor Star Search

5.1. Search Results for Carbon-enhanced Metal-poor Star Candidates

Following the CEMP star search process in Section 3.1, we estimated stellar parameters T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, and $[\text{C}/\text{H}]$ (Section 4). Subsequently, based on the $[\text{Fe}/\text{H}] \leq -1$ criterion, we obtained some metal-poor star candidates. Then, according to the criteria in Equation (1), we classified the metal-poor star candidates into CEMP star candidates and carbon-normal metal-poor star candidates.

By processing 8,651,552 low-resolution, $\text{SNR}_g > 5$ stellar spectra in LAMOST DR8, 12,766 CEMP star candidates were discovered. In these computed CEMP star candidates, 9461 are VMP star candidates ($[\text{Fe}/\text{H}] < -2$), and 164 are EMP star

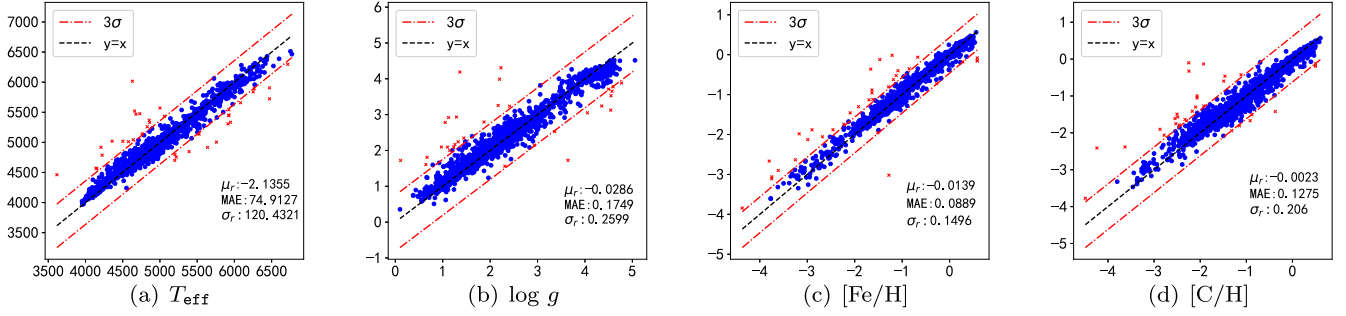


Figure 1. The performance of parameter estimation. The horizontal axis represents the reference values and the vertical axis represents the estimated values. These results are computed from the test set. The black dashed lines represent the reference line for theoretical consistency, while the red dashed lines represent the 3σ reference lines. The parameters μ_r and σ_r are, respectively, the mean and standard deviation of the difference between the parameter estimation and its corresponding reference on test spectra. The mean absolute error/difference between the parameter estimation and its corresponding reference on test spectra is calculated.

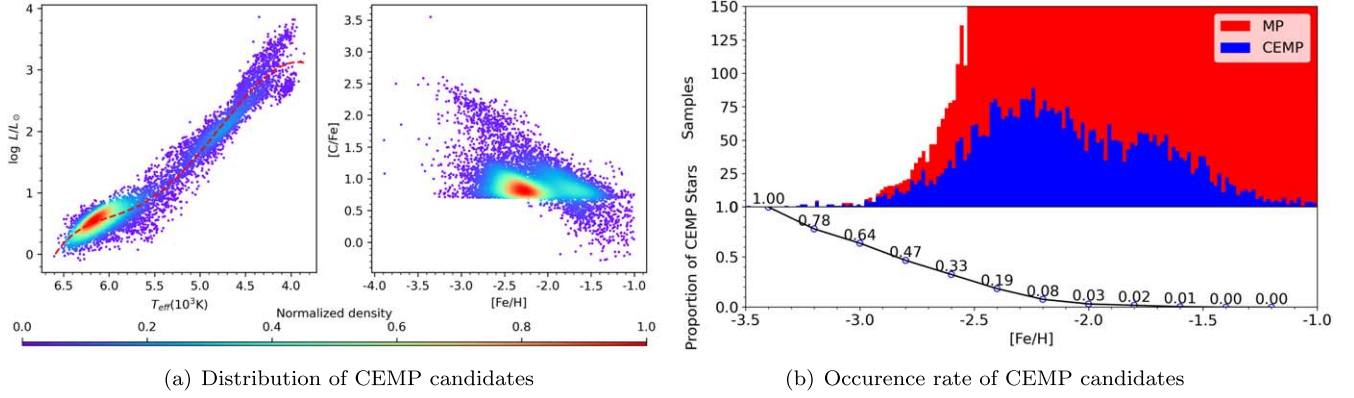


Figure 2. The distribution of 12,766 CEMP candidates and the occurrence rate of CEMPs. In (b), proportions exceeding 150 are not displayed. $[\text{C}/\text{Fe}] = [\text{C}/\text{H}] - [\text{Fe}/\text{H}]$.

candidates ($[\text{Fe}/\text{H}] < -3$). The distribution of these CEMP star candidates in parameter space is presented in Figure 2(a).

According to the official documentation from LAMOST,⁵ some spectra are deemed unsuitable for deriving high-precision atmospheric physical parameters due to their quality issues. These spectral quality issues could potentially reduce the accuracy of the CEMP search. Therefore, the CEMP star candidates corresponding to such spectra are annotated with low confidence, while the other 4396 CEMP candidates are annotated with high confidence.

The computed CEMP candidates are presented in an H-R diagram (the left panel of Figure 2(a)). This subfigure illustrates the evolutionary sequence of the CEMP star candidates, transitioning from main sequence to subgiant, to red giant phases. While Equation (1) does not directly constrain the relationship between $[\text{C}/\text{Fe}]$ and $[\text{Fe}/\text{H}]$, the supplementary application of the CEMP classification formula to stars with $[\text{C}/\text{Fe}] < 0.7$ is effective only for stars with $[\text{Fe}/\text{H}] > -2.5$ as depicted in the right panel of Figure 2(a). This phenomenon is consistent with the observed distribution of the reference data set. The variation in carbon abundance is a result of the different evolutionary stages of the stars. During evolution on the upper red giant branch, carbon in the stellar atmosphere can be converted into nitrogen through the carbon–nitrogen cycle, leading to its mixing into the stellar surface. This process results in carbon depletion, which consequently causes a decrease in the observable carbon abundance (V. M. Placco et al. 2014). As a result, some stars with lower carbon

abundance in the later evolutionary stages have been reclassified as CEMP stars.

Figure 2(b) illustrates the relationship between the proportion of CEMP stars and $[\text{Fe}/\text{H}]$. As the metallicity decreases, the proportion of CEMP stars increases significantly. This trend is consistent with previous studies, such as V. M. Placco et al. (2014).

5.2. Evaluation of the Carbon-enhanced Metal-poor Star Searching Scheme

Based on the CEMP star search process in Section 3.1, along with the definition criteria for CEMP stars (Equation (1)), the effectiveness of CEMP star searches depends on the parameter estimation accuracy for T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, and $[\text{C}/\text{H}]$. Relevant evaluation results are presented in Figure 1.

In this study, CEMP star spectra are regarded as positive samples, while carbon-normal metal-poor star spectra are considered as negative samples. We evaluated the effectiveness of the CEMP star search scheme using such metrics as accuracy, precision, recall, and F1 score. Accuracy is used to measure the proportion of correctly identified CEMP and carbon-normal metal-poor stars. Recall is used to measure the proportion of correctly identified CEMP stars. Precision is used to measure the proportion of correctly identified samples as CEMP among all samples classified as CEMP. In the context of this paper, the number of carbon-normal metal-poor stars is much higher than that of CEMP stars. In such cases, even if the accuracy is high, the precision or recall is possible very low. A low recall means that many CEMP stars fail to be detected. A low precision means that many carbon-normal metal-poor stars

⁵ <https://www.lamost.org/dr8/v2.0/doc/lr-data-production-description>

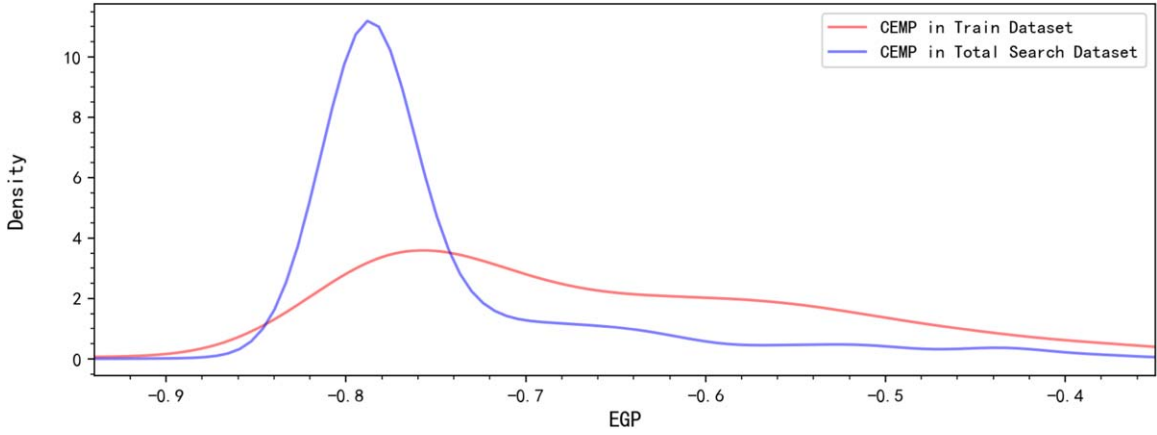


Figure 3. Distribution of EGP indices for the CEMP stars, with blue for all CEMP stars searched and red for the CEMP stars in the training set.

are mistakenly recognized as CEMP star candidates. Therefore, we need a measure to comprehensively evaluate the recall and precision. The F1 score provides a comprehensive measure of precision and recall; it is the harmonic mean of these two metrics. Further discussion of the aforementioned metrics can be found in Q. Zeng et al. (2020).

This work achieved an accuracy of 0.9554, a recall of 0.6970, a precision of 0.8519, and an F1 score of 0.7667 for CEMP searching using the test set. It can be observed that even though CEMP stars constitute a very small proportion of the total samples, this work can still identify approximately 69.7% of the CEMP stars from a large number of carbon-normal metal-poor stars and non-metal-poor stars. In particular, among the detected candidates, approximately 85.19% of the sources are correctly identified as CEMP stars, which is highly valuable for subsequent high-resolution observation source selection. In the future, as the reference data for confirmed CEMP stars increases, there is potential to further improve the identification effectiveness by expanding the training data and retraining the model presented in this paper.

5.3. Carbon-enhanced Metal-poor Searching Performance Evaluation Based on the CH G band

This work employs a data-driven approach to search for CEMP stars; we expect the model's predictions to be consistent with astrophysical research. H. Li et al. (2018) demonstrate that the G1 band line index is highly sensitive to carbon enhancement. Therefore, it helps evaluate the accuracy of CEMP star candidate search to analyze the flux in the G1 band. V. M. Placco et al. (2011) proposed the EGP index to measure the CH G band. This measure helps identify CEMP star candidates from low-resolution spectra while avoiding contamination from the strong H γ line. The calculation method for the EGP is provided in Equation (2):

$$\text{EGP} = -2.5 \log \left(\frac{\int_{4200}^{4400} I_{\lambda} d\lambda}{\int_{4425}^{4520} I_{\lambda} d\lambda} \right). \quad (2)$$

The experimental results in Figure 3 show the distribution of the discovered CEMP star candidates in the EGP index space, where the blue curve represents the distribution of the CEMP stars in the training set and the red curve represents the distribution of the CEMP star candidates discovered in this work. The CEMP stars in the training set were carefully

selected from several works, and verified by medium-resolution or high-resolution spectroscopic observations. Both CEMP star data sets have a peak of sample occurrence probability of $-0.85 \leq \text{EGP} \leq -0.7$, and their occurrence probability of CEMP stars gradually decreases as the distance increases from the peak in the cases of $\text{EGP} > -0.7$ and $\text{EGP} < -0.85$. Therefore, their EGP occurrence ranges and distribution characteristics are consistent to some degree. This consistency indicates the reliability of the CEMP star candidates searched in this work. At the same time, there is a very strong selection effect in the medium-resolution or high-resolution observation verification stages. Therefore, there exist some differences in their distributions.

Furthermore, the dependencies of the EGP index on $[\text{Fe}/\text{H}]$ and $[\text{C}/\text{H}]$ are investigated respectively in Figure 4. It is shown that the EGP index linearly depends on $[\text{C}/\text{H}]$ and $[\text{Fe}/\text{H}]$ excellently in cases of $4300 \text{ K} \leq T_{\text{eff}} \leq 5000 \text{ K}$ and $\log g \leq 2.5$, and $T_{\text{eff}} > 5000 \text{ K}$ and $\log g \leq 4.5$.

6. Comparing the Parameter Estimation Results with GALAH

GALAH DR3 (S. Buder et al. 2021) contains high-quality spectroscopic data for nearly 600,000 stars, covering a wide range of stellar types from dwarfs to giants, as well as a broad range of metal abundances from metal-poor to metal-rich stars. These spectra have a resolution of up to 28,000 and allow for precise measurements of the abundances of up to 30 elements. GALAH utilizes the spectral grid from AMBRE for global fitting to obtain initial parameters, and refines parameter estimates by linearly combining parameters from the 10 closest model spectra.

According to the discriminant criterion in Equation (1), the effectiveness of CEMP star searches depends on the estimation of four parameters: T_{eff} , $\log g$, $[\text{Fe}/\text{H}]$, and $[\text{C}/\text{H}]$. Consistency with parameter estimates from high-resolution spectroscopic catalogs can be a good indication of the credibility of this work. In Section 4.2, our work demonstrates a high level of consistency with reference catalogs, such as APOGEE DR17 (Abdurro'uf et al. 2022), SAGA (T. Suda et al. 2017), and LAMOST-Subaru (X. Li et al. 2022). Furthermore, in Section 5.2, we compare the parameter estimation results with GALAH (S. Buder et al. 2021) to indirectly evaluate the reliability of our search results. The comparison between GALAH and our work is presented in Figure 5.

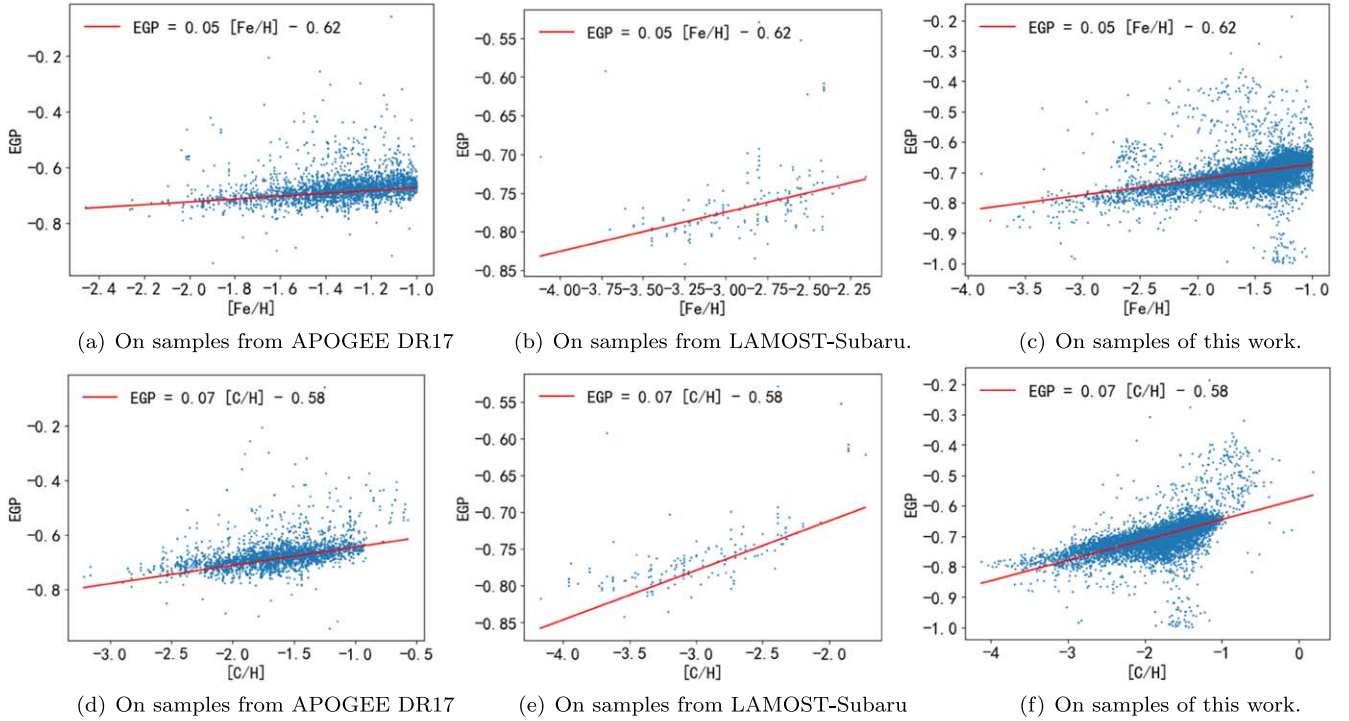


Figure 4. Dependences of the EGP index on $[\text{Fe}/\text{H}]$ and $[\text{C}/\text{H}]$ on the cases of $4300 \text{ K} \leq T_{\text{eff}} \leq 5000 \text{ K}$ and $\log g \leq 2.5$. The lines in (a), (b), and (c) are estimated using the samples in (c), similarly for the lines in (d), (e), and (f).

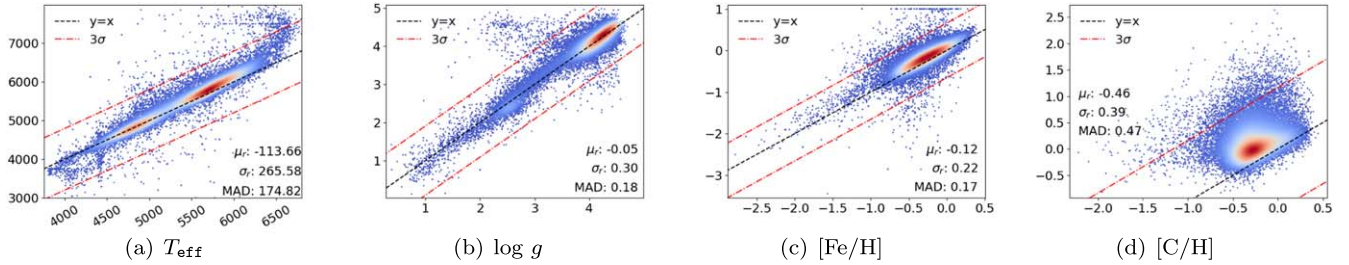


Figure 5. The comparisons between this work and GALAH. The x -axis represents the parameter estimates from this work, while the y -axis represents the GALAH estimates. The color intensity corresponds to the density of samples, with brighter colors indicating higher sample densities. The black dashed line represents the line of theoretical consistency, while the red dashed line represents the 3σ reference line. The parameters μ_r and σ_r are the mean and standard deviation of the difference between two catalogs, respectively. The mean absolute difference between the two catalogs is shown. The horizontal coordinates of the graphs indicate the EGP values, the vertical coordinates indicate the density of the distribution, and the area of each bar graph sums to 1.

For the estimation of effective temperature, there is a divergence branch on the GALAH parameter range of $[6773, 8000] \text{ K}$ (Figure 5(a)). This discrepancy arises from the fact that our training samples predominantly concentrate within the temperature range of $[3615, 6773] \text{ K}$; consequently, our model may exhibit a certain degree of underestimation for samples out of this range. Significant temperature discrepancies between the two studies occur for only about 0.4% of the samples. These discrepant samples primarily concentrate in the high-temperature region. In the high-temperature discrepancy region, no CEMP stars are found; thus, this phenomenon is consistent with expectations. Carbon signals notably become weak as temperature increases. This case is consistent with the findings of C. E. C. Witten et al. (2022), which indicate that $[\text{C}/\text{H}]$ precision can only reach 0.5 in the case of $T_{\text{eff}} < 6000 \text{ K}$. It is challenging to estimate carbon enhancement for stars with excessively high temperatures. Therefore, the discrepancy between this work and GALAH in the high-temperature region does not affect the result for the CEMP star searches.

For the estimation of $\log g$, this work exhibits a relatively high consistency with GALAH, although approximately 0.4% of the stars are underestimated near $\log g \approx 5.0$. No CEMP stars were found at the divergence area. This result is predictable. This predictability is due to the fact that the largest $\log g$ of CEMP stars in the reference set of this work is about four. Therefore, in the case of a CEMP sample with large $\log g$ values, this work may suffer from underestimation.

In terms of $[\text{Fe}/\text{H}]$ estimation, the two works are in good agreement overall and especially for $[\text{Fe}/\text{H}] < -2$. This excellent consistency indicates a high credibility of the selected VMP candidates. CEMP stars are predominantly found among VMP stars. Therefore, the high accuracy of VMP stars enables us to confidently identify CEMP stars in this work.

For the estimation of $[\text{C}/\text{H}]$, this work exhibits significant systematic biases compared to GALAH; similar biases are also observed between APOGEE DR17 and GALAH. This may stem from the inconsistencies in the standards used for $[\text{C}/\text{H}]$ between the two studies.

In summary, this work demonstrates a high level of consistency with GALAH in T_{eff} for [3615.00, 6772.54] K, $\log g$ for [−0.15, 5.06], [Fe/H] for [−4.38, 0.59], and [C/H] for [−4.50, 0.90]. In particular, the accuracy of parameter estimation for VMP stars ensures the credibility of the CEMP stars identified in our study.


7. Conclusion

This study addresses the search for CEMP stars based on low-resolution spectra and proposes a machine learning solution based on deep learning/neural networks. Initially, the method estimates stellar parameters T_{eff} , $\log g$, [Fe/H], and [C/H] from observed spectra. The CEMP star candidates are identified based on the estimated parameters. Using this approach, we discovered 819,671 metal-poor star candidates and 12,766 CEMP star candidates from the LAMOST DR8 low-resolution stellar spectrum database. Among these CEMP star candidates, there are 9461 VMP star candidates and 164 EMP star candidates. For ease of reference and use, we provide estimates of such parameters as T_{eff} , $\log g$, [Fe/H], and [C/H] for the identified targets. The computed catalog is available via doi:10.12149/101500.

Acknowledgments

This work is supported by the National Key R&D Program of China (grant No. 2024YFA1611903) and the National Natural Science Foundation of China (grant Nos. 12373108 and 12222305). The Guo Shou Jing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope, LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences.

ORCID iDs

Xiangru Li  <https://orcid.org/0000-0003-3182-6959>
Haining Li  <https://orcid.org/0000-0002-0389-9264>

References

- Abate, C., Stancliffe, R. J., & Liu, Z.-W. 2016, *A&A*, **587**, A50
 Abdurro'uf, Accetta, K., Aerts, C., et al. 2022, *ApJS*, **259**, 35
 Aguado, D. S., González Hernández, J. I., Allende Prieto, C., & Rebolo, R. 2017, *A&A*, **605**, A40
 Aoki, W., Beers, T. C., Christlieb, N., et al. 2007, *ApJ*, **655**, 492
 Aoki, W., Beers, T. C., Lee, Y. S., et al. 2012, *AJ*, **145**, 13
 Aoki, W., Li, H., Matsuno, T., et al. 2022, *ApJ*, **931**, 146
 Beers, T. C., & Christlieb, N. 2005, *ARA&A*, **43**, 531
 Beers, T. C., Preston, G. W., & Shectman, S. A. 1992, *AJ*, **103**, 1987
 Beers, T. C., Preston, G. W., & Shectman, S. A. 1985, *AJ*, **90**, 2089
 Beers, T. C., Rossi, S., Norris, J. E., Ryan, S. G., & Shefler, T. 1999, *AJ*, **117**, 981
 Bonifacio, P., Sbordone, L., Caffau, E., et al. 2012, *A&A*, **542**, A87
 Buder, S., Sharma, S., Kos, J., et al. 2021, *MNRAS*, **506**, 150
 Carbon, D. F., Henze, C., & Nelson, B. C. 2017, *ApJS*, **228**, 19
 Čotar, D. F., Zwitter, T., & Kos, J. 2018, *MNRAS*, **483**, 3196
 Christlieb, N., Green, P. J., Wisotzki, L., & Reimers, D. 2001, *A&A*, **375**, 366
 Christlieb, N., Schörck, T., Frebel, A., et al. 2008, *A&A*, **484**, 721
 Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *RAA*, **12**, 1197
 Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, *RAA*, **12**, 735
 Frebel, A., Christlieb, N., Norris, J. E., et al. 2006, *ApJ*, **652**, 1585
 Frebel, A., & Norris, J. E. 2015, *ARA&A*, **53**, 631
 Goswami, A., Aoki, W., Beers, T. C., et al. 2006, *MNRAS*, **372**, 343
 Goswami, P. P., & Goswami, A. 2020, *JApA*, **41**, 47
 Hansen, T. T., Andersen, J., Nordström, B., et al. 2016, *A&A*, **586**, A160
 He, S., Zhang, W., Liu, L., et al. 2014, *Anal. Methods*, **6**, 4402
 Hou, X., Zhao, G., & Li, H. 2024, *MNRAS*, **532**, 1099
 Huang, Y., Beers, T. C., Xiao, K., et al. 2024, *ApJ*, **974**, 192
 Lee, Y. S., Beers, T. C., & Kim, Y. K. 2019, *ApJ*, **885**, 102
 Lee, Y. S., Beers, T. C., Kim, Y. K., et al. 2017, *ApJ*, **836**, 91
 Lee, Y. S., Beers, T. C., Masseron, T., et al. 2013, *AJ*, **146**, 132
 Li, H., Aoki, W., Matsuno, T., et al. 2022, *ApJ*, **931**, 147
 Li, H., Tan, K., & Zhao, G. 2018, *ApJS*, **238**, 16
 Li, H.-N., Zhao, G., Christlieb, N., et al. 2015, *ApJ*, **798**, 110
 Li, X., Wang, Z., Zeng, S., et al. 2022, *RAA*, **22**, 065018
 Liu, X.-W., Zhao, G., & Hou, J.-L. 2015, *RAA*, **15**, 1089
 Lucatello, S., Gratton, R. G., Beers, T. C., & Carretta, E. 2005, *ApJ*, **625**, 4049
 Lucey, M., Al Kharusi, N., Hawkins, K., et al. 2023, *MNRAS*, **523**, 6049
 Lugaro, M., Campbell, S. W., & de Mink, S. E. 2009, *PASA*, **26**, 322
 Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, *RAA*, **15**, 1095
 Nomoto, K., Kobayashi, C., & Tominaga, N. 2013, *ARA&A*, **51**, 457
 Norris, J. E., Bessell, M. S., Yong, D., et al. 2012, *ApJ*, **762**, 25
 Placco, V. M., Frebel, A., Beers, T. C., & Stancliffe, R. J. 2014, *ApJ*, **797**, 21
 Placco, V. M., Kennedy, C. R., Beers, T. C., et al. 2011, *AJ*, **142**, 188
 Placco, V. M., Kennedy, C. R., Rossi, S., et al. 2010, *AJ*, **139**, 1051
 Rossi, S., Beers, T. C., Sneden, C., et al. 2005, *AJ*, **130**, 2804
 Safarzadeh, M., Sarmiento, R., & Scannapieco, E. 2019, *ApJ*, **876**, 28
 Shuang, W., Guo-chao, Y., Lu, Z., et al. 2022, *ChA&A*, **46**, 49
 Suda, T., Hidaka, J., Aoki, W., et al. 2017, *PASJ*, **69**, 76
 Suda, T., Katsuta, Y., Yamada, S., et al. 2008, *PASJ*, **60**, 1159
 Suda, T., Yamada, S., Katsuta, Y., et al. 2011, *MNRAS*, **412**, 843
 Taylor, M. B. 2005, in ASP Conf. Ser. 347, Astronomical Data Analysis Software and Systems XIV, ed. P. L. Shopbell, M. C. Britton, & R. Ebert (San Francisco, CA: ASP), **29**
 Venn, K. A., Kieft, C. L., Sestito, F., et al. 2019, *MNRAS*, **492**, 3241
 Whitten, D. D., Placco, V. M., Beers, T. C., et al. 2021, *ApJ*, **912**, 147
 Witten, C. E. C., Aguado, D. S., Sanders, J. L., et al. 2022, *MNRAS*, **516**, 3254
 Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, *RAA*, **11**, 924
 Xie, J., Bu, Y., Liang, J., et al. 2021, *AJ*, **162**, 155
 Yamada, S., Suda, T., Komiya, Y., Aoki, W., & Fujimoto, M. Y. 2013, *MNRAS*, **436**, 1362
 Yang, L., Yuan, H., Xiang, M., et al. 2022, *A&A*, **659**, A181
 Yoon, J., Beers, T. C., Dietz, S., et al. 2018, *ApJ*, **861**, 146
 Yoon, J., Beers, T. C., Placco, V. M., et al. 2016, *ApJ*, **833**, 20
 Yuan, H.-B., Liu, X.-W., Huo, Z.-Y., et al. 2015, *MNRAS*, **448**, 855
 Zeng, Q., Li, X., & Lin, H. 2020, *MNRAS*, **494**, 3110
 Zhao, G., Chu, Y.-H. Z. Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, *RAA*, **12**, 723