Mining double-line spectroscopic candidates in the LAMOST medium-resolution spectroscopic survey using human-AI hybrid method

Shan-shan Li [ID],[1,2] Chun-qian Li [ID],[3,2] Chang-hua Li,[1] Dong-wei Fan [ID],[1] Yun-fei Xu [ID],[1] Lin-ying Mi,[1,2]
Chen-zhou Cui [ID],[1,2] and Jian-rong Shi [ID][3,2]

[1] *National Astronomical Data Center of China, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China*

[2] *School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China*

[3] *CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China*

ABSTRACT

We utilize a hybrid approach that integrates the traditional cross-correlation function (CCF) and machine learning to detect spectroscopic multi-systems, specifically focusing on double-line spectroscopic binary (SB2). Based on the ninth data release (DR9) of the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST), which includes a medium-resolution survey (MRS) containing 29,920,588 spectra, we identify 27,164 double-line and 3124 triple-line spectra, corresponding to 7096 SB2 candidates and 1903 triple-line spectroscopic binary (SB3) candidates, respectively, representing about 1% of the selection dataset from LAMOST-MRS DR9. Notably, 70.1% of the SB2 candidates and 89.6% of the SB3 candidates are newly identified. Compared to using only the traditional CCF technique, our method significantly improves the efficiency of detecting SB2, saves time on visual inspections by a factor of four.

Keywords: Catalogs – Spectroscopic binary stars – Radial velocity – Machine learning

1. INTRODUCTION

Multiple star systems, also called multi-systems, including binary star systems, are common and important celestial bodies in astronomical studies. Approximately half of the stars in the Milky Way reside in multiple star systems (Duchêne & Kraus 2013; Raghavan et al. 2010). Studying multiple star systems provides insights into the processes of stellar formation and evolution, wherein the details of material exchange and interactions clearly govern the evolutionary paths and final outcomes (Sana et al. 2011; Han et al. 2020). Additionally, binary star systems serve as the only known direct method to determine accurate stellar masses (Southworth 2020). As for high-energy astrophysics and frontier research fields, such as the multi-band and time-domain astronomy studies, a more complete and uniform sample set of multi-systems would help to constrain theoretical models.

Due to the Doppler shift resulting from the differing radial velocities (RVs) of each component in a binary star system, we can measure the splitting of spectral lines. Furthermore, if the binary stars have similar spectral types but significant differences in their RVs, they are more likely to be observed as double-lined spectroscopic binaries. Spectroscopic methods are used to search for such spectroscopic binaries (SBs), and the systems discovered are called double-line, triple-line, quadruple-line spectroscopic binaries and so on (SBns; $n > 1$) depending on the split number of spectral lines. There are also single-line spectroscopic binaries (SB1s) in which only the spectrum of one star can be clearly observed, often distinguished by variations in RV.

First published in 2004, the Ninth Catalog of Spectroscopic Binary Orbits (S_{B^9}) is one of the most popular and influential catalogs of SBs, including over 4000 SBs, about one-third of them are SB2s (Pourbaix et al. 2004, and the latest online version of the SB9 catalogue). Re-

ccz@nao.cas.cn

sjr@nao.cas.cn

cent large spectroscopic sky surveys make it possible to systematically find SBs in large quantities. Many works searching for SBs have utilized the survey data from the Apache Point Observatory Galaxy Evolution Experiment (APOGEE, Allende Prieto et al. 2008), which is part of the Sloan Digital Sky Survey (SDSS) (York et al. 2000), including Fernandez et al. (2017), El-Badry et al. (2018) and Price-Whelan et al. (2018, 2020). Moreover, a pipeline has been developed to autonomously find SB2s from the APOGEE data, and over 7000 SB2s have been identified from DR16 (Kounkel et al. 2021). From the Gaia-ESO survey (Gilmore et al. 2022; Randich et al. 2022), 641 SB1s, 342 SB2s, and 11 SB3s have been identified (Merle et al. 2017, 2020). Additionally, a recent study employing a new technique reports the identification of 322 SB2s, 10 SB3s, and 2 SB4s from a sample of 37,565 objects (Van der Swaelmen et al. 2023). Several studies on SBs have also been conducted based on the Radial Velocity Experiment (RAVE, Steinmetz et al. 2006), such as Matijevič et al. (2010) and Birko et al. (2019). The Galactic Archaeology with HERMES survey (GALAH, De Silva et al. 2015) has also contributed to this area, as demonstrated by Traven et al. (2020).

The Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST, Cui et al. 2012), also known as the Guo Shou Jing Telescope, is a special reflecting Schmidt telescope that enables a large spectroscopic sky survey with as many as 4000 optical fibers. It provides an ideal data source for SBs mining. The first stage of the LAMOST spectroscopic survey was from October 2011 to July 2017. In 2019, more than 250,000 spectroscopic binary or variable star candidates were discovered using this data (Qian et al. 2019). Tian et al. (2020) utilized data from LAMOST Data Release 4 (DR4) to construct a catalog that includes approximately 60,000 binary star system candidates. The LAMOST Medium-Resolution Spectroscopic Survey (LAMOST-MRS), which began in September 2018 (Liu et al. 2020), has a resolving power of R ∼ 7500. Li et al. (2021) identified 3133 SB2 and 132 SB3 candidates using LAMOST-MRS DR7 data. Zhang et al. (2022) used LAMOST-MRS DR8 data to discover 2198 SB2 candidates, and Kovalev et al. (2022) detected 2460 SB2 candidates.

In this paper, we use an innovative human and artificial intelligence (human-AI) hybrid method to measure RVs from LAMOST-MRS data and search for SB candidates. We use the data from the ninth data release (DR9) of LAMOST-MRS [1]. As shown in Figure 1, the LAMOST-MRS data includes a sufficient number

[1] https://www.lamost.org/dr9/v1.0

of high signal-to-noise ratio (S/N) spectra. Each exposure of LAMOST-MRS spectrum includes a blue arm (4950 ∼ 5350 Å) and a red arm (6300 ∼ 6800 Å). We use the blue arm of each spectrum to calculate the cross-correlation function (CCF) and RVs because it contains more absorption lines. Based on the number of RVs, the spectra can be identified and classified as SBs or not, but the precision is relatively low. Therefore, visual inspection is needed, which is inefficient. We use machine learning (ML) methods to develop automatic classifiers that can replace human inspection, greatly improving the efficiency of searching for SBs using LAMOST-MRS data.

In Section 2, we focus on the data selection. Section 3 describes the steps and details of our method, including conventional CCF technique and the construction of ML classifiers. In Section 4, we present the results of the data processing and all SB2, SB3 candidates identified. The discussion and conclusion are presented in Sections 5 and 6, respectively.

2. DATA SELECTION

The LAMOST-MRS DR9, released in March 2022, contains 29,920,588 spectra. The dataset includes all observational data newly processed by the latest pipeline, called LAMOST Stellar Parameter Pipeline (LASP; Luo et al. 2015). LAMOST-MRS DR9 includes both time-domain and non-time-domain surveys. In the time-domain survey, each target is observed multiple times over several nights, capturing both single-exposure spectra and coadded spectra (combined within one night). We focus exclusively on single-exposure data, indicated by a "0" in the *coadd* field of the catalog. Additionally, each exposure in LAMOST-MRS yields two spectra: one from the blue arm and another one from the red arm. We select the blue arm spectra from the MRS data due to their abundance of absorption lines, indicated by a *band* field value of "B" in the corresponding catalog files. Figure 1 displays the distribution of S/N versus *Gaia* G magnitude for LAMOST-MRS DR9 within the magnitude range of 10 to 15. Selecting spectra based on S/N is crucial for the search results. We choose spectra with S/N greater than 5 to fully exploit the potential of our method and maximize the discovery of SB2s and SB3s. Furthermore, we need to exclude fibers with issues caused by device malfunctions and problematic blue arm spectra. These issues are indicated in the catalog by the parameters *fibermask* and *band_b* with "0" indicating no issues and "1" indicating problems. Following these criteria, as shown in Table 1, we have obtained a final sample of 6,565,721 spectra from 930,783 stars.

For more information, please refer to the Data Release document of LAMOST-MRS DR9[2].

Table 1. The criteria we used for screening LAMOST-MRS DR9 data.

Field	Criteria	Notes
band	B	Only blue arm spectrum
coadd	0	No combined spectrum
fibermask	0	Delete spectra from bad fiber
bad_b	0	Delete bad blue arm spectra
S/N	≥ 5	

3. THE METHOD

Astronomical research has entered the era of big data, characterized by a rapid increase in data volume, with observational data nearly doubling every 16 months (Zhang & Zhao 2015). This massive increase in astronomical data presents challenges for research, potentially causing important discoveries to be missed or delayed (Smith & Geach 2023). Meanwhile, AI and ML methods have shown a certain degree of effectiveness in source detection and classification (Škoda et al. 2020), particularly for searching SB spectra in large spectroscopic sky survey data (Traven et al. 2020; Li et al. 2021; Zhang et al. 2022).

Notably, in large spectroscopic sky surveys like LAMOST-MRS, spectra from binary star systems account for only a small portion of the total spectra (1.2% as shown by Li et al. 2021). This imbalance makes it very challenging to construct a training dataset with a similar distribution of binary star systems and non-binary star systems, which can significantly affect the results of classification (He & Garcia 2009). Therefore, we use the traditional CCF technique for initial screening to detect SB2s and SB3s, which helps to increase the balance of the samples and improve the classification performance of the machine learning model. Additionally, the CCF technique is needed to measure the RV values of the SB candidates.

To maximize the advantages of conventional and ML methods, our study is mainly divided into two steps. First, we employ the CCF technique to calculate RVs using the observed and template spectra. This process results in a categorized list of spectra and the corresponding CCF data for each spectrum. Subsequently, multiple ML classifiers and an ensemble learning strategy are applied to the CCF data from the first step to identify SB2 and SB3 systems.

3.1. Cross-Correlation Function

The CCF technique is a conventional approach for detecting RV components and calculating RV values from spectra.

3.1.1. CCF calculation

We calculated the CCF based on the observed and template spectra with the classic normalized CCF calculation shown in Equation 1 (Gubner 2006; Zverko et al. 2007),

$$CCF(v) = \sum_{i=1}^{n} \left(\frac{O_i - \overline{O}}{\sigma_{\mathrm{O}}}\right)\left(\frac{T_{i,v} - \overline{T}}{\sigma_{\mathrm{T}}}\right) \tag{1}$$

here, O stands for observation, while T stands for template. O_i is the normalized flux of the spectrum at the wavelength-sampling point i, $\overline{O}$ and σ_{O} are the mean flux and scatter of flux values, respectively. $T_{i,v}$ represents the normalized flux of the template spectrum at the wavelength sampling point i under a Doppler shift corresponding to the velocity v. σ_{T} represents the standard deviation calculated from the total flux values of the template spectrum, and $\overline{T}$ is the arithmetic mean of the flux values of the template spectrum. The normalized CCF ranges from -1 to $+1$, with a fully correlated CCF at $+1$ and a fully anti-correlated CCF at -1. The normalized CCF allows for better detection of RV components and facilitates subsequent machine learning model training. Additionally, the CCF is calculated within the range of RV from -500 to +500 km/s with a step of 1 km/s.

We generate three spectral template using the stellar spectral synthesis program SPECTRUM (Gray & Corbally 1994) and ATLAS stellar atmospheric models (Castelli & Kurucz 2003) under the assumptions of 1D local thermodynamic equilibrium (LTE). These template spectra include three stellar parameters: hot dwarfs ($T_{\mathrm{eff}} = 8000\,\mathrm{K}$, $\log g = 4.0\,\mathrm{dex}$, [Fe/H] $= 0.0\,\mathrm{dex}$), cool dwarfs ($T_{\mathrm{eff}} = 5000\,\mathrm{K}$, $\log g = 4.0\,\mathrm{dex}$, [Fe/H] $= 0.0\,\mathrm{dex}$) and cool giants ($T_{\mathrm{eff}} = 5000\,\mathrm{K}$, $\log g = 2.0\,\mathrm{dex}$, [Fe/H] $= 0.0\,\mathrm{dex}$). The wavelength range and resolution of the template spectra match those of the LAMOST-MRS. After calculating the CCFs, the highest of the three CCFs is chosen for the next step in the multi-line spectral detection process.

3.1.2. Detection of RV components

We adopted the method proposed by Merle et al. (2017) to identify peaks in the CCF, corresponding

---

[2] https://www.lamost.org/dr9/v1.0/doc/mr-data-production-description
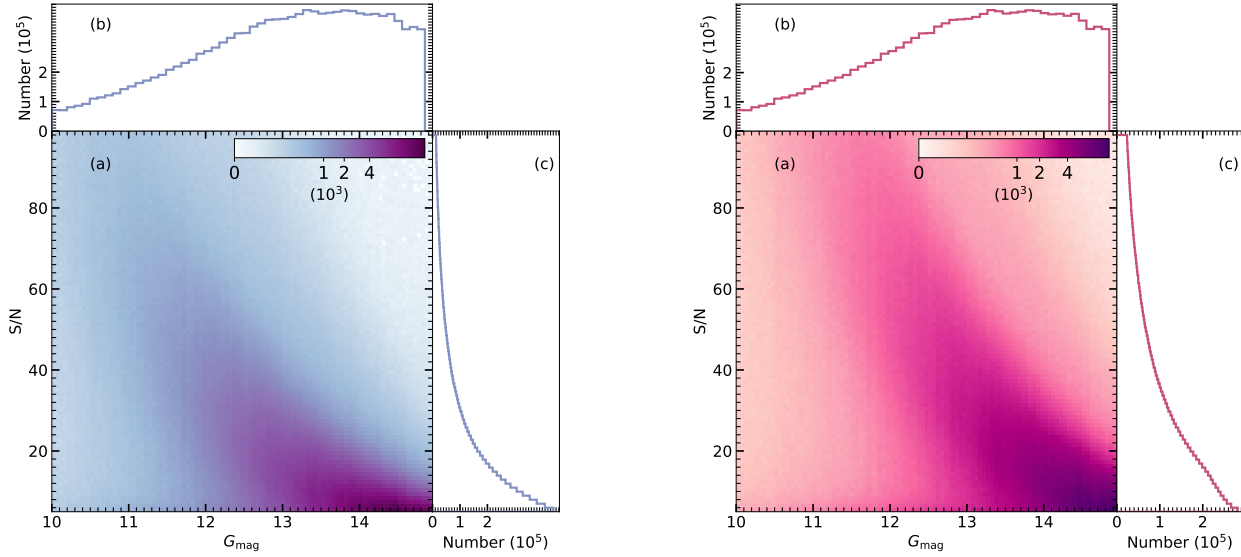
Figure 1. The left panel displays the spectra from the blue arm of the LAMOST-MRS DR9 data, while the right panel presents the spectra from the red arm. In each panel, the parts labeled as a, b, and c represent the distribution of S/N versus G magnitude, the distribution of G magnitude, and the distribution of S/N, respectively.

to RV components in the spectra of SB. This semi-automatic process computes the first three derivatives of the CCF for a given spectrum and locates the peak positions, including blended peaks. The RV values of these components can be measured by the positions where the third derivative of CCF crosses zero during the ascending phase.

To address the inherent discreteness of the CCF, we smooth its derivatives by convolving them with a Gaussian kernel. Following the method of Merle et al. (2017), we perform a Gaussian smooth and range selection for the CCF and its derivatives. We derive the successive derivatives of CCF, and use Python function *scipy.ndimage.gaussian_filter1d* (Virtanen et al. 2020) to smooth the derivatives of the CCFs. The σ of the Gaussian kernels is initially set as 13 km/s. This is based on our experience calculating CCFs using LAMOST-MRS spectra. A smaller initial value of sigma may lead to more false detections of CCF peaks, while a larger initial value may miss real peaks. We increase the σ by 1 km/s until the number of detected RV components matches the number of valleys in the second derivative or σ reaches 100 km/s. As shown in the left panel of Figure 2, an SB candidate is identified using this method with a final σ of 27 km/s. In the right panel, significant peak blending is demonstrated in the CCF, where the first derivative is insufficient to clearly distinguish the peaks. This shows the importance of calculating higher-order derivatives, as the peaks can be accurately identified using the third derivative.

We select the RV ranges with CCF values higher than 60% and the second derivative lower than 40% for RV component detection. The parameters of Gaussian smooth and range selection are based on the experience gained during the detection process for the LAMOST-MRS data. While we acknowledge that lowering these thresholds could potentially increase the detection rate, we prioritized the precision of detected targets over maximizing the number of candidates. This approach is guided by the findings of Merle et al. (2017), they presented that a too low threshold can lead to the detection of unrealistic velocity ranges. As shown in Figure 2, selected range of smoothed CCFs and derivatives are drawn with black solid lines and the original CCFs are illustrated in gray dashed lines. The thresholds are drawn in red horizontal lines and the RVs are indicated in black vertical lines. Additionally, the Figure 3 show the success of this method in identifying known SB2 candidates, and its applicability for searching SB3 candidates.

We adopt the Monte Carlo (MC) simulation to estimate the RV uncertainties for each spectrum of double- and triple-line candidates. In this process, we generate 100 simulated spectra by adding random noise to the flux at each wavelength point of observed spectrum. The noise values are randomly generated from a Gaussian distribution, where the mean and variance are defined by the flux and its error in the observed spectrum, respectively (Li et al. 2021). This approach allows us to conduct 100 simulations for each spectrum, thereby
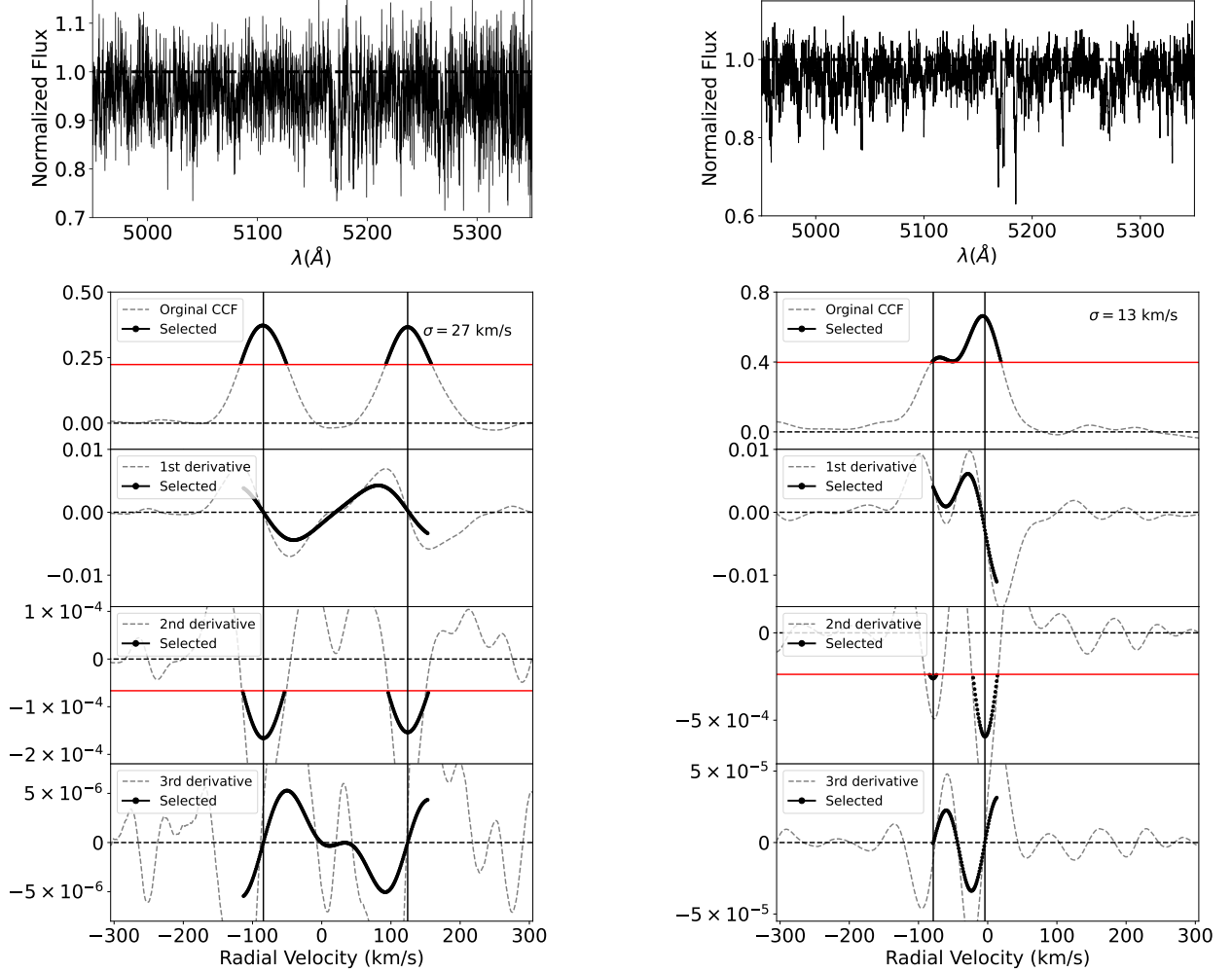
Figure 2. The normalized spectra, CCFs, and derivatives of two SB2 candidates. In the left panel, the final σ of the system is 27 km/s. In the right panel, the CCF exhibits significant peak blending; the first derivative cannot distinguish the peaks well, but they can be identified using the third derivative. Black solid lines are used to draw the selected range of smoothed CCFs and derivatives, while gray dashed lines illustrate the original CCFs. Red horizontal lines indicate the thresholds (above 60% for CCF values and below 40% for the second derivative of the CCF), and black vertical lines mark the RVs.

obtaining the errors and evaluating the reliability and stability of the RV values. The distribution of the uncertainties of RV against S/N is shown in Figure 4.

In the traditional approach, the next step is to manually inspect all targets identified as SB candidates by the CCF calculations for verification. However, using our method and data screening criteria, the number of SB2s identified through the CCF technique could exceed one hundred thousand, making direct visual identification impractical. To enhance efficiency, we introduced the ML method to reclassify the results and expedite the identification process.

3.2. *Machine learning optimization*

We use the CCF data corresponding to each spectrum obtained from the first step for ML classification. ML algorithms tend to perform better in identifying data with clear and distinctive features. Compared to spectral data, the CCFs are considerably smoother and possess more conspicuous features. We utilized deep neural networks (DNNs), a type of multi-layer supervised learning model. DNNs can map input CCFs to output types by leveraging the interconnectedness and weights among multiple layers of neurons to learn complex nonlinear features. Previous research suggests that increasing the number and diversity of classifiers and adopting multiple methods can enhance classification efficiency and accuracy (Dietterich 2000; Brown et al. 2005; Musehane et al. 2008), Therefore, we construct multiple classifiers and adopt an ensemble learning strategy to improve the performance of ML.
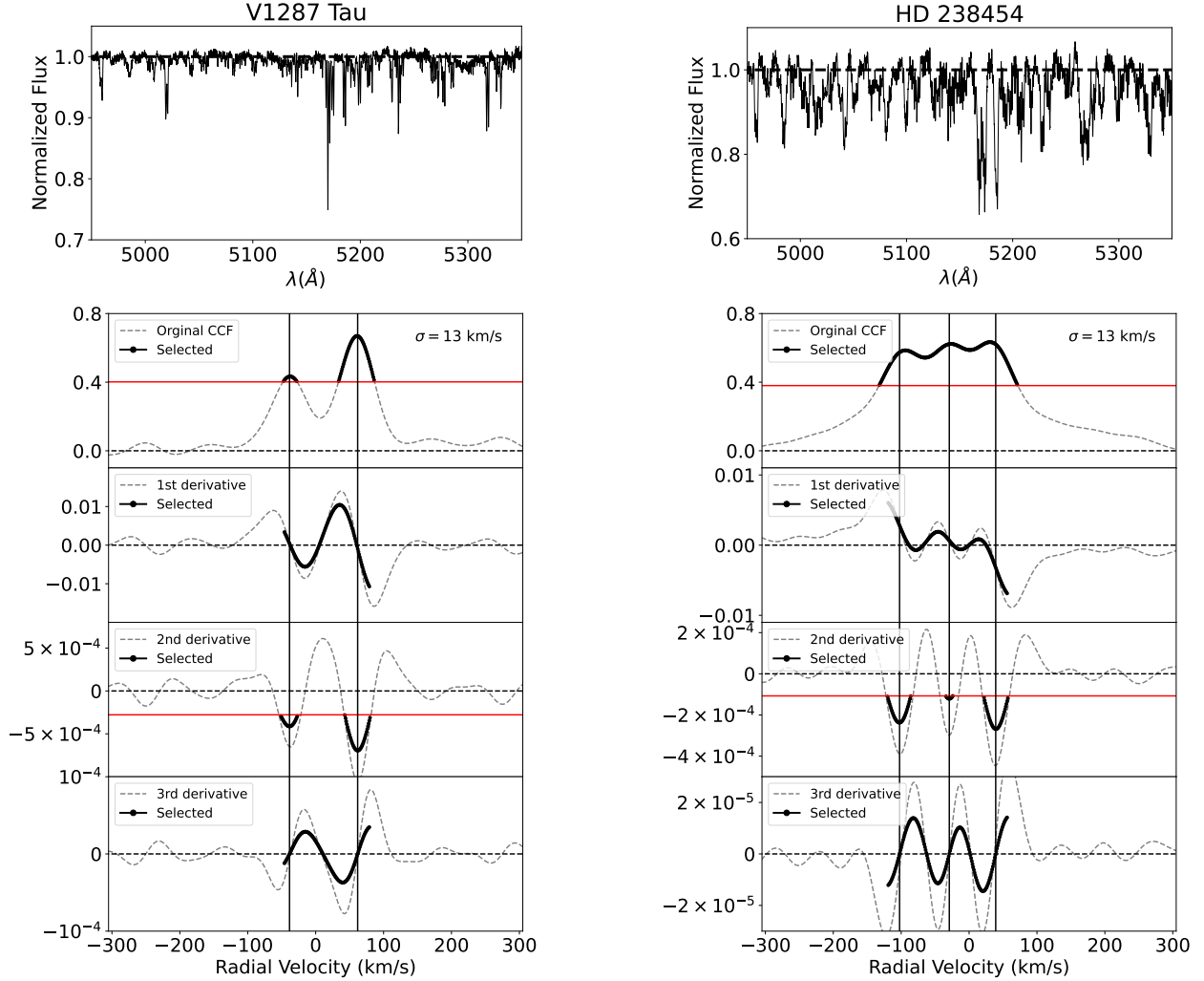
Figure 3. The normalized spectra, CCFs and derivatives of SB2 (V1287 Tau) and SB3 candidates (HD 238454) selected from LAMOST-MRS. The Black solid lines, gray dashed lines, red horizontal lines, and black vertical lines are used as in Figure 2.

3.2.1. *Data preparation*

We use different classification methods based on the characteristics of the data to build four different Classifiers. The final judgment is based on the comprehensive classification results of these Classifiers. The training dataset is composed of four distinct categories, labeled as L0, L1, L2, and L3. L0 represents CCF data with no significant peaks, while L1, L2, and L3 correspond to CCFs of single, double, and triple-line spectra of SBs, respectively. The L0 category indicates either no detectable signal or very weak signals that do not stand out against the noise. It helps the model identify noise data, improving classification performance and reducing false positives. However, the verified number of SB2s and SB3s from LAMOST-MRS observational data is not sufficient for L2 or L3 categories, especially for L3. To address these issues, we use a synthetic spectral library

to generate spectra of simulated SBs, and calculate the CCF of these spectra as the training samples.

By applying the ATLAS stellar atmospheric models with the new opacity distribution functions (Castelli & Kurucz 2003), we calculate the spectra with absolute flux. The wavelength range matches the LAMOST-MRS spectra, spanning $4900 \sim 5400\,\text{Å}$ with a step of $0.1\,\text{Å}$. The grids are in the parameter space with T_{eff} between $3500\,\text{K}$ and $8000\,\text{K}$ (step $100\,\text{K}$ for $3500 \sim 7500\,\text{K}$, while $250\,\text{K}$ for $7500 \sim 8000\,\text{K}$), $\log g$ from $0.0\,\text{dex}$ to $5.0\,\text{dex}$ (step $0.25\,\text{dex}$) and [Fe/H] from -4.0 to $0.5\,\text{dex}$ (step $0.2\,\text{dex}$ for $-4.0 \sim -1.0\,\text{dex}$, while $0.1\,\text{dex}$ for $-0.1 \sim 0.5\,\text{dex}$).

We first generate the spectra of single stars by randomly selecting stellar parameters from the LAMOST-LRS DR9 stellar parameter catalog and interpolate the spectral library in the stellar parameter space to generate the spectra (Xiang et al. 2015). Then, we select two
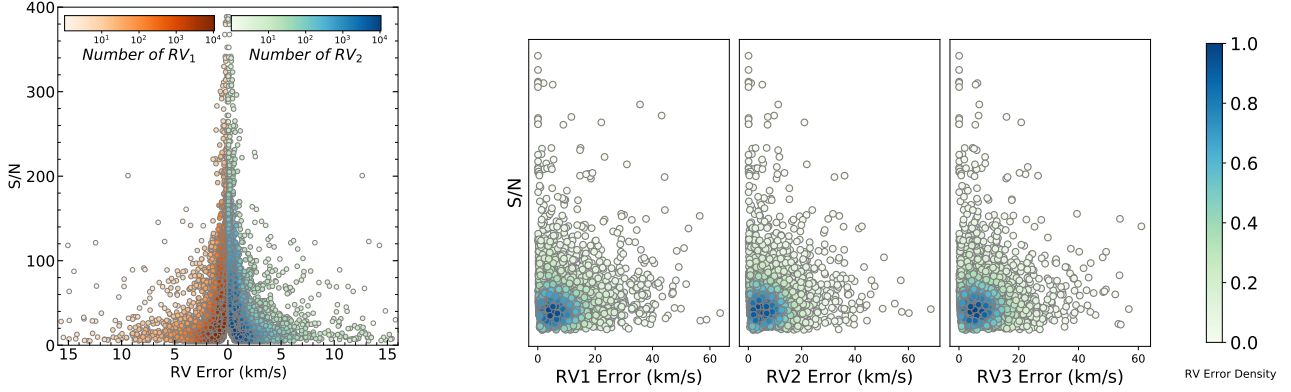
Figure 4. The distribution of RV error versus S/N is illustrated, with the left plot corresponding to the error of RV_1 and RV_2 of SB2 classification and the right plot corresponding to the RV_1, RV_2 and RV_1 in SB3 classification. The color gradient shows the number density of RV error.

and three spectra to generate the simulated double- and triple-line spectra, respectively.

To generate spectra for SB2 system, we set the RV difference between the two components of the binary is randomly set between 60 and 250 km/s. For SB3, the relation of the RV of the three components is $RV_1 < RV_2 < RV_3$. The RV differences between RV_1 and RV_2, as well as between RV_2 and RV_3, are randomly set within the range of 60 to 250 km/s. The lower limit is set due to the resolution of the LAMOST-MRS, which make it difficult to detect stars with RV differences smaller than 60 km/s as shown in Figure 5. Previous studies have shown that the RV difference between the two components in SBs is rarely greater than 250 km/s (Li et al. 2021). Therefore, this upper limit is empirically selected to improve computational efficiency.

The limitation of the CCF technique also needs to be considered, as it determines the range of spectral flux ratios between the two stars in the binary system that the method can resolve. "Successful detection" is defined as using the CCF technique to effectively separate two or three peaks, which represent the simulated components of the SB2 or SB3 systems, respectively. Figure 6 shows the relationship between successful detection and the spectral flux ratios between the two stars in the system. We generate this figure by simulating SB2s, with each component having the same $\log g = 4.5\,$dex, [Fe/H] $= 0.0\,$dex and RV differences are 200km/s. The T_{eff} ranges from 3500 K to 7500 K, with a step of 100 K. Most spectra with a flux ratio between 1/3 and 3 can be detected, and for a higher probability of successful detection, the flux ratio between stars in a binary system should fall within this range. We observe that the flux ratio limit drops below 3 around 5000 K, which can be explained by the strong and broad MgH molecular lines in K5-type stars, which cause the spectral flux to change more rapidly with T_{eff} around 4500 K. The increased flux

ratio causes greater signal blending, making it more difficult to distinguish the individual spectral features of the stars.

The L0 dataset is selected from CCFs with a maximum CCF value lower than three times the noise level, defined as the standard deviation of the CCF values, to dynamically account for the S/N and spectral quality of different spectra, thereby ensuring a more precise selection. The L1 dataset is formed with single-line spectra. The CCF technique is used on SBs spectra to detect double-line and triple-line features. Only spectra with successful detection are selected to form the L2 and L3 training datasets. Finally, the training dataset includes 1500 samples from the L0 dataset, along with 1500 samples each from the L1, L2, and L3 datasets, for a total of 6000 samples.

3.2.2. *Training and evaluation*

For training DNNs, labeled data are used to adjust the weights and biases of the neural network to minimize the discrepancy between predicted outputs and true labels, thereby improving the prediction accuracy. We adopt Keras with TensorFlow (Abadi et al. 2016) to achieve DNN training.

The neural network architecture includes three fully connected layers implemented on the training set. The first layer consists of 600 neurons and uses Rectified Linear Unit (ReLU, Nair & Hinton 2010) as the activation function. This is followed by a second layer of 300 neurons, also using ReLU. The final layer comprises n neurons, where n corresponds to the number of output classes, and employs the normalization exponential function softmax as the activation function. When dealing with n-class classification problems, the output results of the softmax function, which are the probability values of each class, are obtained by Equation 2,
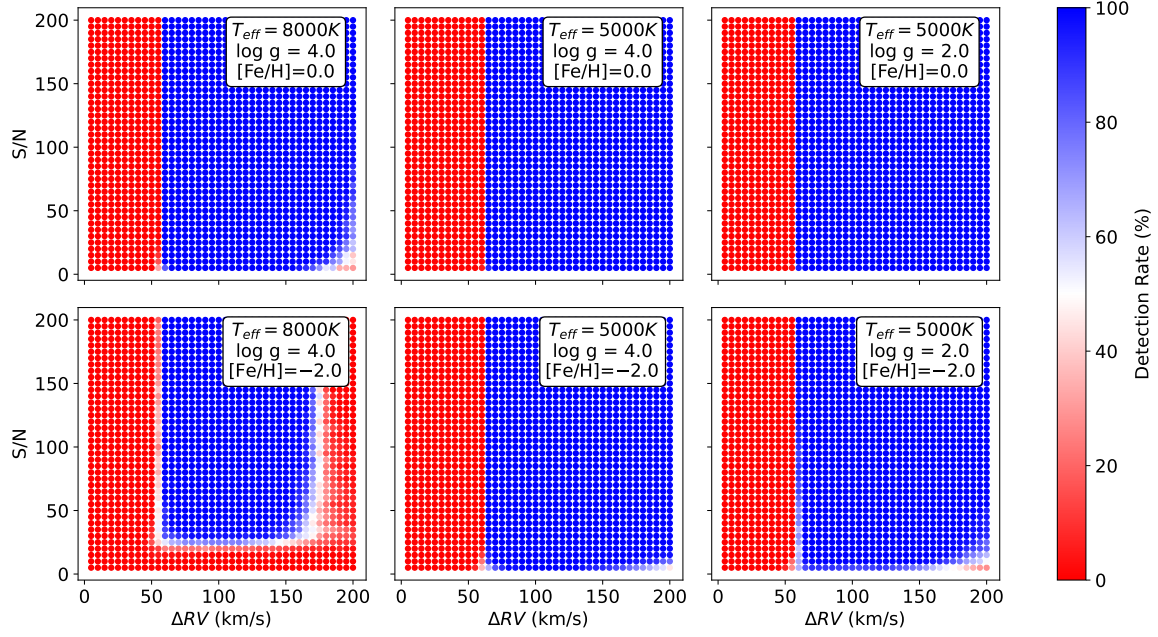
Figure 5. Detection rates of twin stars with different atmospheric parameters, S/Ns and RVs. The minimum S/N used is 5, with a step size of 5.
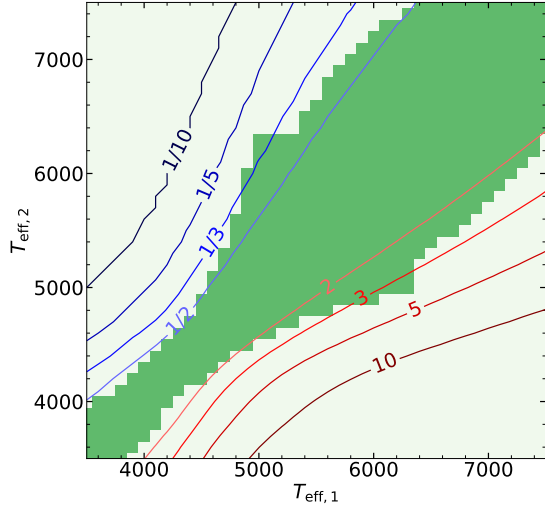
Figure 6. The relationship between successful detection and the spectral flux ratios of the two stars in the binary system. The green blocks indicate successful detection of double-line spectra, while the contour lines depict the spectral flux ratio between the components of the binary stars.

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}, \quad i = 1, 2, \cdots, n \quad (2)$$

here, x is an n-dimensional vector composed logits output by the model for different categories, where n corresponds to the number of output classes, taking values of 2, 3, or 4 depending on the classification model used by each classifier. After applying softmax function, these logits are converted into normalized probabilities

(scores), where each value is between 0 and 1, and the total sums to 1, representing the model's confidence in each category.

The model training adopts the Adaptive Moment Estimation (Adam, Kingma & Ba 2014) as optimizer and uses the Sparse Categorical Cross Entropy function shown in Equation 3 as the loss function,

$$LOSS = -\frac{1}{N} \sum_{i=1}^{N} \log(P_{y_j}) \quad (3)$$

where N is the number of samples in the batch. y_i is the true class label (an integer index), P_{y_i} is the predicted score for the true class y_i from the softmax function output of the model. A test set is used during the training process to evaluate the performance of the model.

Based on different classification methods of training data, we constructed four classifiers, named C1 to C4. All four types of data samples (labeled L1 to L3) are directly used to train C1, so the output of this classifier consists of four categories. C2 is a two-step classifier. Initially, it combines samples labeled as L2 and L3 into one category called "LN", which, along with L0 and L1 samples, undergoes three-class classification training. Then, it uses L2 and L3 data for binary classification. When using the C2 classifier, the data are first divided into three categories, then those classified as LN are reclassified into either L2 or L3. Thus, the final output of C2 is also four categories. C3 uses only L0, L1, and L2 data for three-class classification training, yielding three classification results. Finally, C4 uses only L0,

Table 2. The classification results of all calculated CCFs. The last column represents the number of spectra classified as SB2 or SB3 by all machine learning classifiers. For SB2, normalized probabilities (scores) must be greater than 95%. For SB3, the selection criterion is set at 99%.

RV calculation classification	Classifier	L0	L1	L2	L3	Selected results
SB2 (118,274)	C1	52,661	181	37,913	27,519	
	C2	44,219	1157	38,290	34,608	27,233 ($P > 95\%$)
	C3	57,436	208	60,630	-	
SB3 (49,847)	C1	22,445	6	549	20,519	
	C2	20,758	2	676	22,083	11,904 ($P > 99\%$)
	C4	23,365	44	-	20,110	

L1, and L3 data for three-class classification training. To maintain data balance across different training categories, each category has 1500 samples. We can use classifiers C1, C2, and C3 to identify SB2 samples from the CCFs obtained with the traditional method, corresponding to CCFs with double peaks. Similarly, we can use C1, C2, and C4 to identify SB3 samples, corresponding to CCFs with three peaks.

We conduct a 10-fold cross-validation process for each classifier, meaning the entire dataset is evenly divided into 10 subsets. In each iteration, one subset is selected as the test dataset, while the remaining nine are combined to form the training dataset. This ensures that each subset serves as the test set once, providing a comprehensive evaluation of our model's performance. This approach evaluates the flexibility and reliability of our model by testing it with different data sets.

After each training iteration, we calculate the values of accuracy, precision, recall, and F1 score for the overall ML method, as well as the precision, recall, and F1 score for L2 and L3 with test dataset for evaluation. We use the *sklearn.metrics* package (Pedregosa et al. 2011) to calculate these metrics for each classifier. A detailed description and explanation of the evaluation metrics are provided in Appendix A. The evaluation results are very promising, as all values are greater than 99%. This indicates that our classifiers perform very well on the simulated data.

3.3. *Classification criteria*

The final output layer of our DNN models uses the softmax function, which outputs normalized probabilities for each category, with values ranging between 0 and 1. For instance, classifier C1 provides the normalized probabilities that the input CCF data belongs to categories L0, L1, L2, or L3. Typically, the data is assigned to the category with the highest probability. However, to avoid misclassification and improve classification precision, we focus only on data where the nor-

malized probability of L2 or L3 is significantly higher than that of the other categories. We empirically set this threshold at 95% for L2 selection and 99% for L3 selection. Additionally, we use the majority voting method from ensemble learning approaches (Dietterich 2000) to integrate the outputs of all four classifiers to obtain the final result.

Therefore, to classify a spectrum as SB2, its CCF must be categorized as L2 in classifiers C1, C2, and C3, with normalized probabilities exceeding 95% in each. For SB3, we set the threshold at 99% for L3 classification in C1, C2, and C4, with reasons discussed in Section 5. Ultimately, as recorded in Table 2, a total of 27,233 CCFs labeled as L2 and 11,904 CCFs labeled as L3 were selected.

4. RESULT

Based on the set of 6,565,721 blue arm spectra selected from the LAMOST-MRS DR9 database, we search for SBs. We obtain the RV component number using the conventional CCF technique, leading to the selection of 118,274 double-line and 43,519 triple-line spectra. The CCF data are also calculated to enable subsequent classification with ML classifiers.

After the ML approach, we identify 27,233 SB2 spectra and 11,904 SB3 spectra. The results are then examined visually to verify the precision of this method, resulting in 27,164 double-line spectra and 3124 triple-line spectra being selected. The screening criterion is that the double-line or triple-line signal in the peak area must be significantly stronger than that in the wing part. CCF data that are rejected typically exhibit broader peaks with greater uncertainties. In the case of double lines, clearly asymmetric CCF peaks are also removed, as they may not be caused by binary stars. As shown in Figure 7, the left is a double-line case while the right a triple-line case, both identified by CCF technique and ensemble learning but rejected through visual inspection.
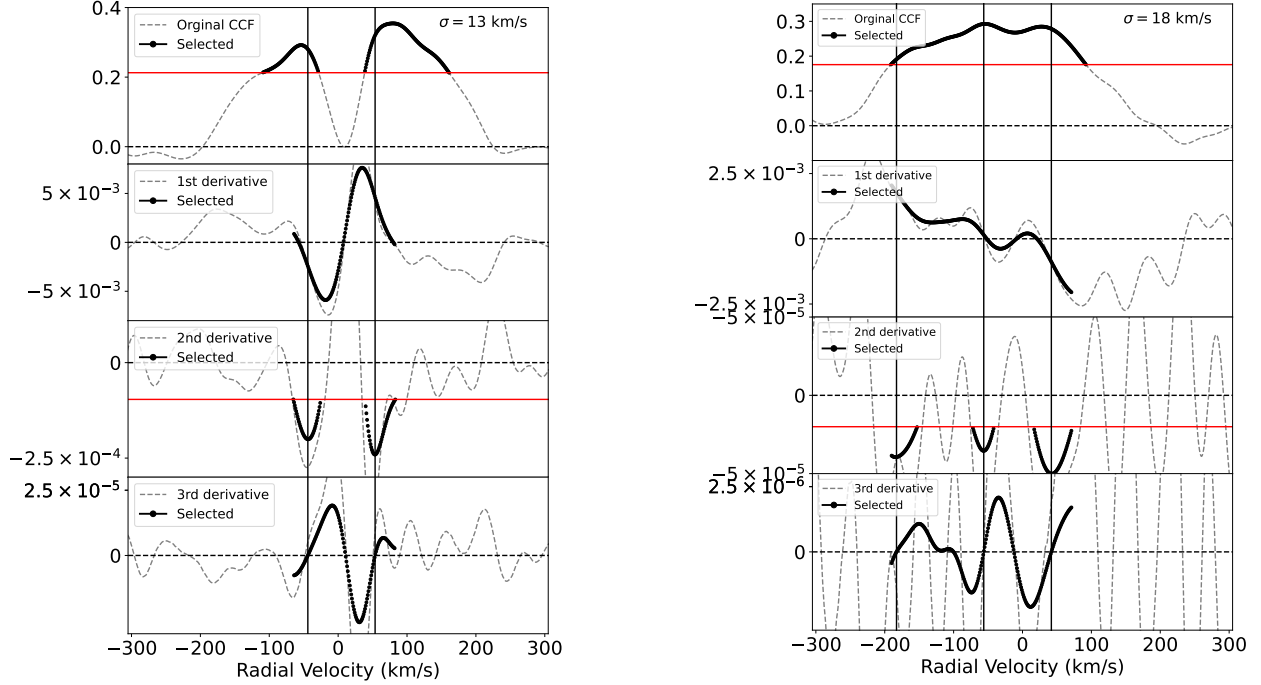
Figure 7. Two examples of cases rejected by visual inspection, including the CCF and its derivatives. The left shows a double-line case identified by the CCF technique and ensemble learning but rejected based on the visual inspection criterion, while the right shows a similar triple-line case.

Given that a target may be observed multiple times with multiple spectra, it is possible for it to exhibit different CCF patterns and consequently be classified into different categories. After careful checking, we find 76 candidates have spectra from different observations classified into both double-line and triple-line. These candidates are ultimately classified as an SB3 candidate and removed from the SB2 category.

After detection and classification using the CCF and ML methods, we identify 7096 SB2 and 1903 SB3 candidates, accounting for 0.8% and 0.2% of our selection dataset of LAMOST-MRS DR9 , respectively, using the criteria shown in Table 1. We present the results, including RV values and uncertainties of each spectrum, in Table 3. The table includes plan name

(planID), spectrograph ID (spID), fiber ID of the target (fiberID), local modified Julian Minute at the start of exposure (LMJM), median S/N of all pixels (S/N) from the LAMOST-MRS data release. The Local Modified Julian Minute (LMJM) is obtained by multiplying the Local Modified Julian Day (LMJD) of the observation start time by 1440. The LMJD refers to the Modified Julian Day at Xinglong Observatory, which operates in the +8 time zone. *Gaia* source ID (Gaia source id) and G magnitude (G) are cross-matched from *Gaia* DR3 (Gaia Collaboration et al. 2023) using coordinates with a radius of 3.3". The radius is determined from the the diameters of the fiber of LAMOST (Cui et al. 2012). The RVs in the table are ordered by increasing values.

Table 3. The information of the spectra identified as double- and triple-line, along with their RVs.

RA(J2000)	DEC(J2000)	Gaia source id	G(mag)	Multiplicity	planID	spID	fiberID	LMJM	S/N	RV$_1$ (km/s)	RV$_2$ (km/s)	RV$_3$ (km/s)
0.011929	55.573441	4205125900347739840	12.6	SB2	HIP1178420I	8	117	83604867	24.47	−87.13 ± 11.07	30.77 ± 8.49	-
0.012862	58.019821	4226322723527299984	12.1	SB3	NT235302N592517N01	7	118	84169112	40.75	−118.4 ± 7.03	−21.92 ± 2.46	64.26 ± 5.25
0.012862	58.019821	4226322723527299984	12.1	SB3	NT235302N592517N01	7	118	84169131	39.8	−94.07 ± 10.9	2.5 ± 2.66	94.65 ± 2.88
0.087823	55.456971	4205115936042626240	9.1	SB2	HIP472H382601	3	185	85224695	172.86	−46.5 ± 3.32	38.41 ± 4.72	-
0.090428	38.651075	2880989157927180800	12.9	SB2	NT001046N372826M01	16	6	84706218	17.14	−112.81 ± 0.47	−29.7 ± 0.78	-
0.090428	38.651075	2880989157927180800	12.9	SB2	NT001046N372826M01	16	6	84706241	18.8	−114.05 ± 0.6	−30.64 ± 0.6	-
0.090428	38.651075	2880989157927180800	12.9	SB2	NT001046N372826M01	16	6	84706265	19.4	−112.91 ± 0.38	−30.87 ± 0.74	-
0.10414	34.114326	2875121652781859200	13.0	SB2	NT235421N333536M01	9	180	85231820	38.18	−65.59 ± 0.44	22.11 ± 0.33	-
0.10414	34.114326	2875121652781859200	13.0	SB2	NT235421N333536M01	9	180	85231844	35.85	−66.32 ± 0.41	23.44 ± 0.35	-
0.146393	55.574784	4205131354852523456	12.6	SB2	NGC778901	8	116	83647886	18.47	−49.62 ± 10.99	46.19 ± 8.33	-
0.240085	40.398735	2881983837993175552	12.3	SB2	HIP1177690I	6	11	83604832	15.6	−53.43 ± 8.0	31.16 ± 7.99	-
0.291179	58.695535	4227686461509736996	12.2	SB2	NGC778901	12	240	83647859	15.02	−14.56 ± 1.58	65.22 ± 1.36	-
0.291179	58.695535	4227686461509736996	12.2	SB2	NGC778901	12	240	83647873	13.84	−15.11 ± 2.2	67.53 ± 2.1	-
0.291179	58.695535	4227686461509736996	12.2	SB2	NGC778901	12	240	83647886	16.41	−16.21 ± 1.64	65.19 ± 1.43	-
0.347601	41.11409	2882225867991453056	12.2	SB2	HIP1177690I	6	39	83609157	21.39	−104.36 ± 0.66	26.61 ± 0.42	-
0.348095	62.072807	4299109749134782272	10.1	SB3	NGC778801	8	78	83650805	117.37	−157.33 ± 36.57	−31.58 ± 20.73	65.03 ± 13.86
0.387174	63.517781	4315933951963335488	12.8	SB2	NGC778801	12	23	83650762	15.39	−83.4 ± 0.65	−6.5 ± 1.01	-
0.40717	37.245455	2880290044627019264	12.4	SB3	NT001046N372826M01	10	237	84706241	47.87	−188.89 ± 12.42	−65.57 ± 10.11	63.72 ± 22.25
0.40717	37.245455	2880290044627019264	12.4	SB3	NT001046N372826M01	10	237	84706265	44.31	−187.82 ± 17.33	−34.73 ± 30.52	73.13 ± 33.9
0.459621	61.623334	4295224006344115840	10.9	SB2	NGC778802	6	245	83650843	54.99	−63.65 ± 0.66	23.94 ± 0.46	-
0.502711	62.14576	4299122118640024832	11.5	SB2	NGC778802	8	99	83650826	40.49	−59.58 ± 3.97	26.38 ± 3.84	-
0.527144	55.92513	4208906158682176000	10.8	SB2	NGC778901	8	129	83647886	75.73	−67.09 ± 0.39	−0.12 ± 0.44	-
0.564065	60.307202	4293261378069834240	12.7	SB2	NGC778801	7	202	83650792	16.14	−121.21 ± 6.22	20.87 ± 18.73	-
0.573322	36.428159	2880099653021400832	12.4	SB3	TD000246N354855T01	15	185	84704935	51.91	−79.75 ± 7.35	11.44 ± 2.68	90.9 ± 2.88
0.57379	62.551303	4300411252947777088	12.2	SB3	NGC778802	13	243	83650826	33.35	−100.19 ± 12.77	2.1 ± 15.27	80.54 ± 19.52
0.590771	36.817228	2880214040885803264	14.8	SB2	TD000246N354855T01	15	24	85227526	7.48	−53.97 ± 2.72	21.99 ± 2.23	-
0.60622	61.149567	4294690743192305920	12.8	SB3	NGC778802	6	23	83650857	15.88	−98.87 ± 5.67	−21.61 ± 7.25	66.09 ± 16.79
0.615598	35.827237	2877066104735658752	13.7	SB2	NT001046N372826M01	2	85	84706218	15.8	−135.35 ± 0.79	−35.23 ± 1.07	-
0.615598	35.827237	2877066104735658752	13.7	SB2	NT001046N372826M01	2	85	84706241	15.12	−134.49 ± 0.8	−34.49 ± 0.98	-
0.615598	35.827237	2877066104735658752	13.7	SB2	NT001046N372826M01	2	85	84706265	13.64	−132.74 ± 0.9	−36.23 ± 1.12	-
0.664305	33.941192	2875153676057907456	11.1	SB3	TD000246N354855T01	1	155	85227572	36.08	−89.28 ± 14.73	44.5 ± 42.53	128.19 ± 38.21
0.673242	54.984642	4204046631205888	12.2	SB2	NGC778901	7	59	83647873	28.19	−161.98 ± 1.01	56.83 ± 4.2	-
0.673242	54.984642	4204046631205888	12.2	SB2	NGC778901	7	59	83647886	27.12	−161.8 ± 1.16	56.2 ± 3.14	-
...	...	...	...	...	...	...	...	...	...	...	...	...

NOTE—The table presents the classification outcomes, including RV values and associated errors. The columns are: RA(J2000); DEC(J2000); Gaia source ID from Gaia DR3; G(mag) is the G magnitude from Gaia; Multiplicity, which is the result of the classification; planID is the plan name from LAMOST-MRS; spID is the spectrograph ID from LAMOST-MRS; fiberID is the fiber ID of the target from LAMOST-MRS; LMJM, showing the local modified Julian Minute at the start of exposure; S/N representing the median S/N of all pixels; RV$_1$ (km/s), RV$_2$ (km/s), and RV$_3$ (km/s), which are the RVs calculated from the first step using the conventional CCF technique and arranged in order of their values. The full version of this table with 30,288 rows.

## 4.1. SB2 and SB3 Candidates

We cross-match our SB2 and SB3 candidates with other binary catalogs including the Kepler Eclipsing Binary Stars (KEBC, Kirk et al. 2016), the TESS Eclipsing Binary stars (TESS-EBs, Prša et al. 2022), SB in the APOGEE DR16 and DR17 Data (Kounkel et al. 2021), Gaia DR3 Non-single stars (Gaia Collaboration 2022), SB candidates from Gaia-ESO Survey (Merle et al. 2017), FGK binary stars from the GALAH survey (Traven et al. 2020), the S_{B^9} catalog(Pourbaix et al. 2004), Wide binaries from Gaia eDR3 (El-Badry 2021) and the Washington Visual Double Star Catalog (WDS, Mason et al. 2001). As the diameter of the LAMOST fiber is about 3.3", we cross-match catalogs and obtain the common targets by coordinates within a 3.3" radius. A total of 690 common SB2 and 151 common SB3 candidates have been obtained.

We also cross match the candidates obtained by Li et al. (2021) using LAMOST-MRS DR7 data and find 1637 SB2 and 58 SB3 candidates identified in both works. Taking into account of all the cross match results, 2121 SB2 and 197 candidates identified in this work have been included in other catalogs or studies to our knowledge (Table 4). Therefore, 4975 SB2 and 1706 SB3 candidates are newly found. The proportion of new discoveries in all identified SB2 and SB3 candidates is 70.1% and 89.6%, respectively.

4.2. Statistical analysis

We perform a statistical analysis on all the 7096 SB2 and 1903 SB3 candidates identified. For each candidate, we determine the number of spectra (blue arm) in the LAMOST-MRS DR9 database that can yield RV values using our methods. The distribution of SBs as a function of the number of exposures is illustrated in Figure 8. Among these candidates, 108 SB2 and 7 SB3 candidates have single exposures, with the highest number of exposure times being 121 for SB2 and 116 for SB3. There are 3650 SB2 and 1312 SB3 candidates have exposures over 6 and the orbital parameters for this sample of SB2 candidates could be obtained.

In Figure 9, the distribution of RV differences for each spectrum of the SB2 candidates are illustrated, and the RV differences ΔRV are basically larger than 60 km/s, which is related to the resolving power of LAMOST-MRS. Also, due to the resolution limitations, detecting all SBs with $\Delta RV < 80$ km/s is challenging. The distribution of $\Delta RV > 80$ km/s follows an exponential pattern. Figure 10 shows the distribution of S/N versus G magnitude of SB2 and SB3, and its distribution has the same trend as Figure 1.
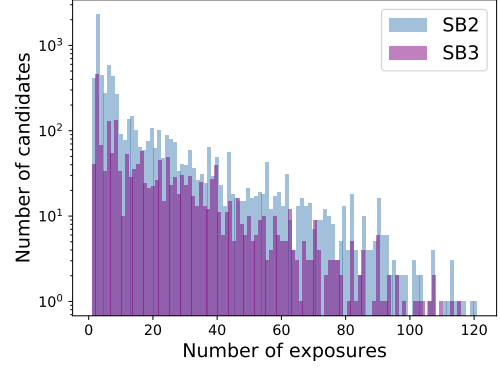
Figure 8. Number of detected SB candidates versus number of exposures (Blue arm). The number of candidates is in a logarithmic scale.
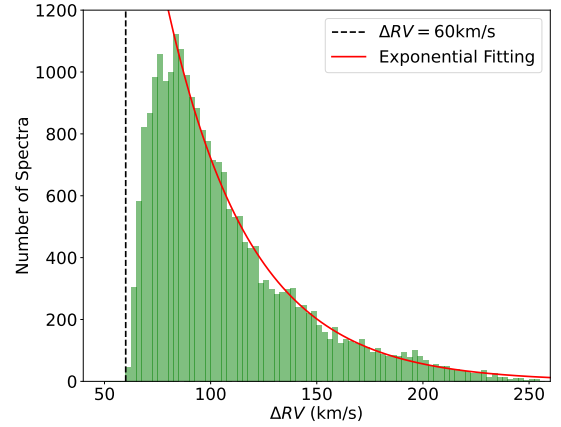
Figure 9. Distribution of RV differences (ΔRV) of all observed SB2 candidates in LAMOST-MRS DR9. The vertical dashed black line indicates the detection limit of ΔRV for SB2 in LAMOST-MRS spectra, which is about 60 km/s. The red line represents the exponential fitting curve. We applied an exponential function to fit the distribution of $\Delta RV > 80$ km/s.

5. DISCUSSION

Combining the traditional CCF technique with the ML methods, our human-AI hybrid process significantly improved the precision and efficiency of searching and classification SB candidates from the LAMOST-MRS spectra. In the conventional CCF technique step, we identify 118,274 spectra as double-line candidates. This number significantly exceeds the results from Li et al. (2021) and manual examination of the result at this scale is impractical. After ML process, the number of spectral classified as double-line candidates have been reduced from 118,274 to 27,233, and 27,164 are confirmed visually. Assuming the spectra confirmed by human validation as ground truth, the ML method improves the

Table 4. The list of common SB2 and SB3 candidates with other binary catalogs.

RA(J2000)	DEC(J2000)	Gaia Source ID	G(mag)	Multiplicity	N_{Exp}	Comment
0.291179	58.695535	422768646150973696	12.2	SB2	3	LC_2021
0.347601	41.114090	2882225867991453056	12.2	SB2	1	Gaia WB,LC_2021
0.387174	63.517781	431593395196335488	12.8	SB2	1	LC_2021
0.407170	37.245455	2880290044627019264	12.4	SB3	2	WDS
0.459621	61.623334	429522400634115840	10.9	SB2	1	Gaia WB
0.527144	55.925130	420890615868217600	10.8	SB2	1	WDS
0.673242	54.984642	420460466311205888	12.2	SB2	2	LC_2021
0.863206	29.936055	2861247637104271360	11.7	SB2	1	Gaia WB
1.354247	58.314799	422698316070242816	11.3	SB2	2	LC_2021
1.417710	56.414781	420974865952459904	11.6	SB2	3	Gaia WB,LC_2021
1.513075	-0.538712	2545718964216072320	13.7	SB2	3	LC_2021
1.650851	57.137139	422510368300513792	9.6	SB2	3	LC_2021
2.144856	58.668384	422906772301985408	12.5	SB3	1	Gaia NSS
2.184212	37.497105	2877637747703176704	11.4	SB2	1	WDS
2.275254	61.646542	429848096591774208	12.4	SB2	1	LC_2021
2.482804	-0.206435	2545836543240450816	13.4	SB2	1	LC_2021
2.742505	54.741034	420218882982724224	13.9	SB2	4	WDS
2.796898	57.199411	422425740266770048	11.3	SB2	3	LC_2021
2.870514	56.897016	422327918091786752	11.8	SB2	3	LC_2021
2.882273	57.712906	422469097952109568	13.5	SB2	5	LC_2021
2.905883	59.259216	423035316384989824	13.3	SB2	5	LC_2021
2.942407	58.377895	422884717652219776	12.9	SB2	3	Gaia WB,WDS
3.332230	2.609186	2548317385070710656	13.9	SB2	3	LC_2021
3.397440	2.621560	2548316904034412416	14.2	SB2	1	LC_2021
3.495918	-1.915404	2541576382359366784	13.8	SB2	3	LC_2021
3.648931	53.828180	420110443650955904	11.1	SB2	4	WDS
3.931877	58.901100	422964977698391168	10.1	SB2	6	LC_2021
3.943690	57.117233	422389284583657472	14.0	SB2	1	LC_2021
4.059039	58.118580	422821942408901376	10.8	SB2	1	Gaia WB
4.070429	59.209457	428978108012689792	11.1	SB3	1	Gaia NSS
4.275290	39.547891	379956794398809472	11.9	SB2	3	Gaia WB
4.651834	2.108786	2548075939189064320	13.9	SB2	3	LC_2021
5.081090	58.854170	428204189266719616	13.6	SB2	6	LC_2021
5.198236	60.060925	428383890690176512	10.0	SB2	4	WDS
5.316476	60.702454	428799368656549120	13.3	SB2	2	LC_2021
5.519838	57.614777	421996449694222720	13.6	SB2	1	Gaia WB
5.868530	24.272019	2801215161221470720	13.9	SB3	1	Gaia NSS
6.140866	59.600082	428307612078727936	13.4	SB2	4	LC_2021
6.256972	23.492209	2800927638930707712	12.2	SB2	3	Gaia WB
6.355069	59.116074	428267308105062528	11.1	SB2	6	Gaia NSS,LC_2021,WDS
6.508332	32.691329	2862935284373463168	13.7	SB2	3	Gaia WB
...	...	...	...	...	...	...

NOTE—The list of common SB2 and SB3 candidates with other binary catalogs. The columns are: RA (J2000); DEC (J2000); *Gaia* source ID from *Gaia* DR3; G(mag) is the G magnitude from *Gaia*; Multiplicity, which is the result of the classification; N_{Exp} is the number of exposures of this candidate in LAMOST-MRS, Comment listing the catalogs in which this candidate is included. In the table, abbreviations represent various star catalogs: KEBC (Kepler Eclipsing Binary Stars), TESS (TESS Eclipsing Binary Stars), APOGEE (SB in APOGEE DR16 and DR17 Data), Gaia NSS (Gaia DR3 Non-single stars), Gaia-ESO (SB candidates from Gaia-ESO Survey), GALAH (FGK binary stars from GALAH survey), SB9 (SB9 catalog), Gaia WB (Wide binaries from Gaia eDR3), WDS (Washington Visual Double Star Catalog), and LC_2021 (Li et al. 2021). The full version of this table with 2318 rows.

precision from 23.0% to 99.7%. As for triple-line spectra, after ML process, the number has been reduced from 43,519 to 11,904, and 3124 spectra are identified visually. The ML method improves the precision from 7.2% to 26.3%.

From our work, the identified 7096 SB2 and 1903 SB3 candidates account for 1.0% of the LAMOST-MRS DR9 data we selected for SB searching. This fraction is slightly lower than the fraction reported by Li et al. (2021) using LAMOST-MRS DR7 data. This is mainly because we set the spectral screening criteria of S/N to greater than 5 and processed a significantly larger number of spectra initially.

The reliability and interpretability of ML methods applied in astronomical research remains a subject of debate. This skepticism is largely due to the perceived
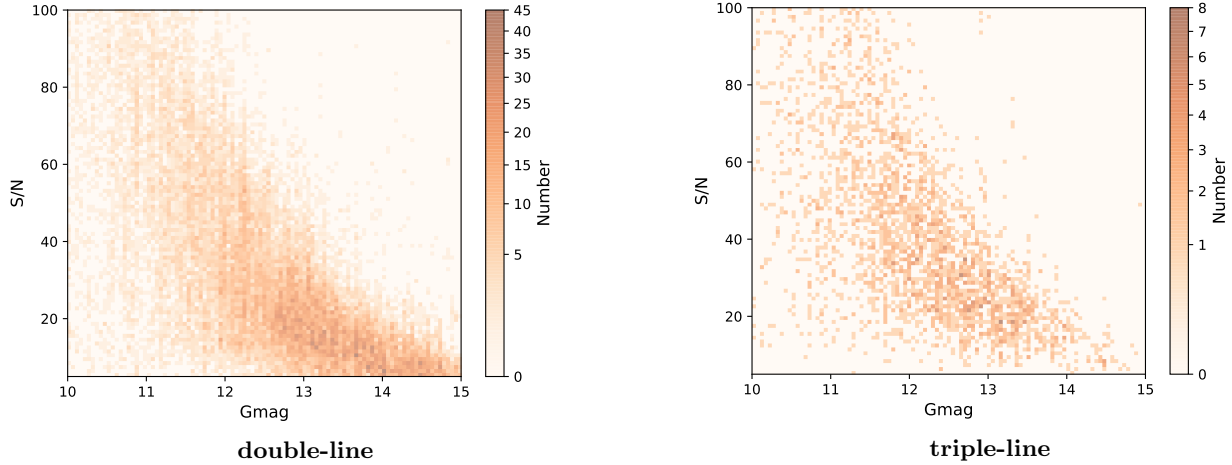
Figure 10. Left panel: The S/N versus G magnitude distribution of 27,164 manually confirmed double-line spectra in LAMOST-MRS DR9. Right panel: The S/N versus G magnitude distribution of 3124 manually confirmed triple-line spectra in LAMOST-MRS DR9.

"black box" nature of these methods (Smith & Geach 2023). Nonetheless, human-AI collaboration, which integrates ML approaches with traditional methods based on rigorous mathematical rules, has the potential to enhance the practicality and accuracy of ML methods. This topic is gaining increasing attention (Djorgovski et al. 2022).

We can evaluate the precision of the ensemble learning method and each DNN classifier using classification result in Table 2 and the spectra identified visually. The precision of L2 is 99.7% for ensemble learning and are 71.6%, 70.1% and 44.8% for classifier C1, C2 and C3, respectively. The precision of L3 is 19.4% for ensemble learning result and 11.3%, 10.5% and 11.5%. As shown in the Figure 11, the two CCFs are respectively identified as double-line and triple-line by the CCF technique, but they are filtered out by the ensemble learning classifier. The ML classifiers likely filtered out the CCFs with the peak signals that are not significantly greater than the noise. This indicates that the model is effectively prioritizing precision and successfully eliminating data that could otherwise introduce false positives.

For our SBs search efforts, we focus more on the precision of ML in identifying L2 and L3 classes. For L2, we can see that our ensemble learning method achieves very high overall precision, nearly matching the results obtained during model training with theoretical spectra. This is because our ML approach is specifically designed to prioritize precision, aiming to reduce the need for manual inspection and enhance overall efficiency. We apply high selection thresholds across multiple classifiers and use their intersection for final selection. As a result, the precision in detecting SB2 candidates approaches 100%. However, true SB2 candidates may still be in-

cluded in the spectra that are filtered out, highlighting the need for further improvement in accuracy.

For the L3 category, both precision and accuracy are still require further improvements. This may be due to the fact that our training data does not fully reflect the distribution of real L3 samples, as well as the random selection of RV differences, which does not fully account for the hierarchical relationship of three components in SB3 system. Moreover, the impact of training samples on classification results cannot be ignored. In our study, we performed multiple training sessions using training sets formed from different data types: pure observational data, a combination of observational and theoretical data, pure theoretical data, and theoretical data with physical constraints. When applying these trained models to the spectra to be classified, some spectra, especially those of SB3, could only be identified by a single model but were missed by others. The influence of different training data on the results is particularly significant for SB3 spectra.

6. CONCLUSION AND PROSPECTS

We employ a hybrid human-AI approach, consisting of three steps: traditional CCF analysis, ML methods, and human-eye inspection, to detect SB2 and SB3 candidates using the blue-arm spectra from the LAMOST-MRS DR9 data with S/N higher than 5. Initially, number of components and RV values are obtained through CCF calculations producing a preliminary classification. Subsequently, ML methods are introduced to further screen the data and improve the precision of classification. We employ the DNN to build four classifiers and adopt ensemble learning approach. With the human inspection processes, we finally identify a total of 27,164
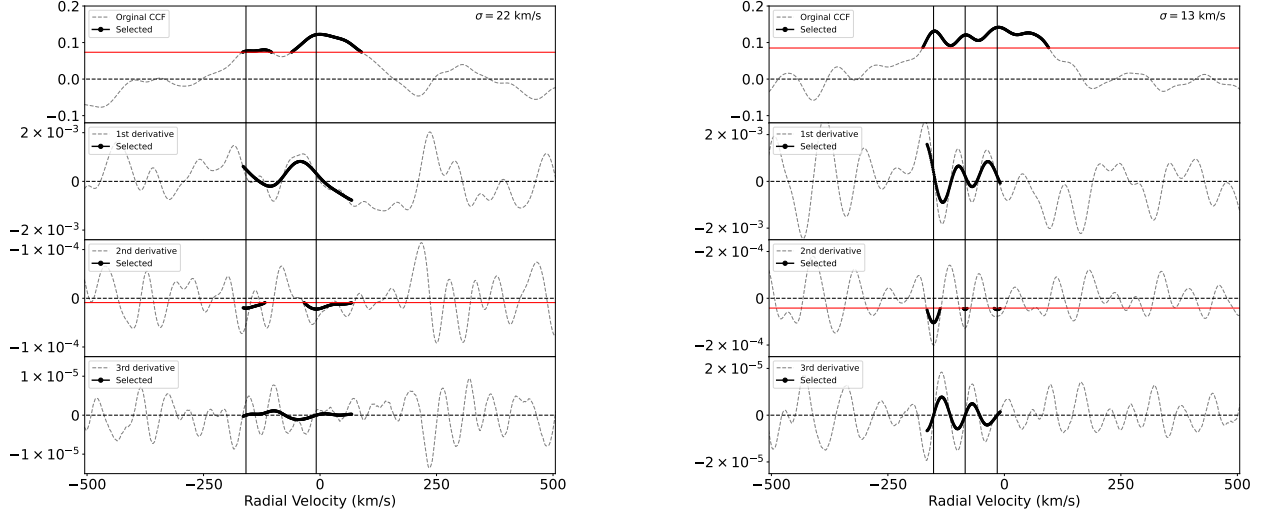
Figure 11. Two CCF data identified as double-line (left) and triple-line (right) by the CCF technique, but filtered out by the ensemble learning classifier.

double-line spectra and 3124 triple-line spectra from the LAMOST-MRS DR9 data, corresponding to 7096 SB2 and 1903 SB3 candidates, respectively. Notably, 70.1% of the SB2 candidates and 89.6% of the SB3 candidates are newly identified.

By combining conventional CCF and ML methods, the precision and efficiency of identifying multi-system candidates are significantly improved. Specifically, the precision for SB2 candidates has improved from 23.0% to 99.7%. This indicates that the combined method can effectively automate the search for binary stars in LAMOST-MRS data, saving a significant amount of time on visual inspection. However, the results for SB3 candidates are not satisfactory. Our future plans involve further optimizing the process to enhance classification efficiency and precision, particularly for SB3 candidates.

To improve classification performance, we plan to add penalty parameters and apply multi-channel training while increasing the number of samples. Additionally, we are examining the spectra misclassified by the CCF technique (as shown in Figure 11) and those misclassified by the ensemble learning method but rejected through visual inspection (as shown in Figure 7), attempting to identify their patterns.

REFERENCES

Abadi, M., Agarwal, A., Barham, P., et al. 2016, arXiv e-prints, arXiv:1603.04467,
doi: 10.48550/arXiv.1603.04467

Allende Prieto, C., Majewski, S. R., Schiavon, R., et al. 2008, Astronomische Nachrichten, 329, 1018,
doi: 10.1002/asna.200811080

Birko, D., Zwitter, T., Grebel, E. K., et al. 2019, AJ, 158, 155, doi: 10.3847/1538-3881/ab3cc1

Brown, G., Wyatt, J., Harris, R., & Yao, X. 2005, Information Fusion, 6, 5,
doi: https://doi.org/10.1016/j.inffus.2004.04.004

Castelli, F., & Kurucz, R. L. 2003, in Modelling of Stellar Atmospheres, ed. N. Piskunov, W. W. Weiss, & D. F. Gray, Vol. 210, A20. https://arxiv.org/abs/astro-ph/0405087

Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, Research in Astronomy and Astrophysics, 12, 1197, doi: 10.1088/1674-4527/12/9/003

De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, MNRAS, 449, 2604, doi: 10.1093/mnras/stv327

Dieterich, T. G. 2000, in Multiple Classifier Systems (Berlin, Heidelberg: Springer Berlin Heidelberg), 1–15

Djorgovski, S. G., Mahabal, A. A., Graham, M. J., Polsterer, K., & Krone-Martins, A. 2022, arXiv e-prints, arXiv:2212.01493. https://arxiv.org/abs/2212.01493

Duchêne, G., & Kraus, A. 2013, ARA&A, 51, 269, doi: 10.1146/annurev-astro-081710-102602

El-Badry, K. 2021, Wide binaries from Gaia eDR3, Zenodo, doi: 10.5281/zenodo.4435257

El-Badry, K., Ting, Y.-S., Rix, H.-W., et al. 2018, MNRAS, 476, 528, doi: 10.1093/mnras/sty240

Fernandez, M. A., Covey, K. R., De Lee, N., et al. 2017, PASP, 129, 084201, doi: 10.1088/1538-3873/aa77e0

Gaia Collaboration. 2022, VizieR Online Data Catalog: Gaia DR3 Part 3. Non-single stars (Gaia Collaboration, 2022), VizieR On-line Data Catalog: I/357. Originally published in: Astron. Astrophys., in prep. (2022). https://ui.adsabs.harvard.edu/abs/2022yCat.1357....0G

Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1, doi: 10.1051/0004-6361/202243940

Gilmore, G., Randich, S., Worley, C. C., et al. 2022, A&A, 666, A120, doi: 10.1051/0004-6361/202243134

Gray, R. O., & Corbally, C. J. 1994, AJ, 107, 742, doi: 10.1086/116893

Gubner, J. A. 2006, Probability and random processes for electrical and computer engineers (Cambridge University Press)

Han, Z.-W., Ge, H.-W., Chen, X.-F., & Chen, H.-L. 2020, Research in Astronomy and Astrophysics, 20, 161, doi: 10.1088/1674-4527/20/10/161

He, H., & Garcia, E. A. 2009, IEEE Transactions on Knowledge and Data Engineering, 21, 1263, doi: 10.1109/TKDE.2008.239

Kingma, D. P., & Ba, J. 2014, arXiv e-prints, arXiv:1412.6980, doi: 10.48550/arXiv.1412.6980

Kirk, B., Conroy, K., Prša, A., et al. 2016, AJ, 151, 68, doi: 10.3847/0004-6256/151/3/68

Kounkel, M., Covey, K. R., Stassun, K. G., et al. 2021, AJ, 162, 184, doi: 10.3847/1538-3881/ac1798

Kovalev, M., Chen, X., & Han, Z. 2022, MNRAS, 517, 356, doi: 10.1093/mnras/stac2513

Li, C.-q., Shi, J.-r., Yan, H.-l., et al. 2021, ApJS, 256, 31, doi: 10.3847/1538-4365/ac22a8

Liu, C., Fu, J., Shi, J., et al. 2020, arXiv e-prints, arXiv:2005.07210. https://arxiv.org/abs/2005.07210

Luo, A. L., Zhao, Y.-H., Zhao, G., et al. 2015, Research in Astronomy and Astrophysics, 15, 1095, doi: 10.1088/1674-4527/15/8/002

Mason, B. D., Wycoff, G. L., Hartkopf, W. I., Douglass, G. G., & Worley, C. E. 2001, AJ, 122, 3466, doi: 10.1086/323920

Matijevič, G., Zwitter, T., Munari, U., et al. 2010, AJ, 140, 184, doi: 10.1088/0004-6256/140/1/184

Merle, T., Van Eck, S., Jorissen, A., et al. 2017, A&A, 608, A95, doi: 10.1051/0004-6361/201730442

Merle, T., Van der Swaelmen, M., Van Eck, S., et al. 2020, A&A, 635, A155, doi: 10.1051/0004-6361/201935819

Musehane, R., Netshiongolwe, F., Nelwamondo, F. V., Masisi, L., & Marwala, T. 2008, arXiv e-prints, arXiv:0810.3865, doi: 10.48550/arXiv.0810.3865

Nair, V., & Hinton, G. E. 2010, in International Conference on Machine Learning

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, J. Mach. Learn. Res., 12, 2825–2830

Pourbaix, D., Tokovinin, A. A., Batten, A. H., et al. 2004, A&A, 424, 727, doi: 10.1051/0004-6361:20041213

Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., et al. 2018, AJ, 156, 18, doi: 10.3847/1538-3881/aac387

—. 2020, ApJ, 895, 2, doi: 10.3847/1538-4357/ab8acc

Prša, A., Kochoska, A., Conroy, K. E., et al. 2022, ApJS, 258, 16, doi: 10.3847/1538-4365/ac324a

Qian, S.-B., Shi, X.-D., Zhu, L.-Y., et al. 2019, Research in Astronomy and Astrophysics, 19, 064, doi: 10.1088/1674-4527/19/5/64

Raghavan, D., McAlister, H. A., Henry, T. J., et al. 2010, ApJS, 190, 1, doi: 10.1088/0067-0049/190/1/1

Randich, S., Gilmore, G., Magrini, L., et al. 2022, A&A, 666, A121, doi: 10.1051/0004-6361/202243141

Sana, H., James, G., & Gosset, E. 2011, MNRAS, 416, 817, doi: 10.1111/j.1365-2966.2011.18698.x

Smith, M. J., & Geach, J. E. 2023, Royal Society Open Science, 10, 221454, doi: 10.1098/rsos.221454

Southworth, J. 2020, in Stars and their Variability Observed from Space, ed. C. Neiner, W. W. Weiss, D. Baade, R. E. Griffin, C. C. Lovekin, & A. F. J. Moffat, 329–334, doi: 10.48550/arXiv.1912.13400

Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, AJ, 132, 1645, doi: 10.1086/506564

Tian, Z., Liu, X., Yuan, H., et al. 2020, ApJS, 249, 22, doi: 10.3847/1538-4365/ab9904

Traven, G., Feltzing, S., Merle, T., et al. 2020, A&A, 638, A145, doi: 10.1051/0004-6361/202037484

Van der Swaelmen, M., Merle, T., Van Eck, S., et al. 2023, arXiv e-prints, arXiv:2312.04721, doi: 10.48550/arXiv.2312.04721

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2

Škoda, P., Podsztavek, O., & Tvrdík, P. 2020, A&A, 643, A122, doi: 10.1051/0004-6361/201936090

Xiang, M. S., Liu, X. W., Yuan, H. B., et al. 2015, MNRAS, 448, 822, doi: 10.1093/mnras/stu2692

York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579, doi: 10.1086/301513

Zhang, B., Jing, Y.-J., Yang, F., et al. 2022, ApJS, 258, 26, doi: 10.3847/1538-4365/ac42d1

Zhang, Y., & Zhao, Y. 2015, Data Science Journal, doi: 10.5334/dsj-2015-011

Zverko, J., Žižňovský, J., Mikulášek, Z., & Iliev, I. K. 2007, Contributions of the Astronomical Observatory Skalnate Pleso, 37, 49

APPENDIX

A. MODEL PERFORMANCE EVALUATION

We employed a standard method for model evaluation, utilizing performance metrics accuracy, precision, recall, and F1 score, as delineated in Equations A1 through A4. Where, true positive (TP) represents the positive sample been successfully identified as positive, true negative (TN) is the negative sample correctly identified as negative, false positive (FP) is the negative sample mistakenly classified as positive, and false negative (FN) is the positive sample categorized as negative by the model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{A1}$$

$$Precision = \frac{TP}{TP + FP} \tag{A2}$$

$$Recall = \frac{TP}{TP + FN} \tag{A3}$$

$$F1\ score = \frac{2 * (Precision * Recall)}{(Precision + Recall)} \tag{A4}$$

Using L2 as an example, when considering L2 as the positive sample, it indicates that our analysis focuses on samples labeled as L2:

- A true positive (TP) occurs when a sample labeled as L2 is correctly identified as L2 by the model.

- A true negative (TN) happens when a sample not labeled as L2 is correctly recognized as not L2.

- A false positive (FP) is when a sample not labeled as L2 is incorrectly classified as L2.

- A false negative (FN) occurs when a sample labeled as L2 is mistakenly identified as not L2 by the model.

In machine learning (ML), accuracy, precision, recall, and F1 score are key indicators for evaluating the performance of classification models. Each metric measures the performance of the model from a different perspective. A high accuracy indicates that the model has a high rate of correct predictions overall or for a specific category. A high precision means the model rarely misclassifies negative samples as positive. A high recall indicates the model captures most of the positive samples, with few misses. The F1 score is the harmonic mean of precision and recall, providing a balance between them. A high F1 score signifies good performance in both aspects, making it especially suitable for evaluating imbalanced datasets.