# An Artificial-Intelligence-Based Spectral Visualization and Analysis Platform*

Lei Guohong, Pu Zhouhao, Xu Yang, Zhou Yunfan, Chen Xunzhou, Tian Haijun

2026-08-18

ChinaRxiv

## Abstract

Spectra carry vast amounts of astrophysical and chemical information and are an important means of revealing the nature of celestial objects and exploring the laws of the universe. With the continued advancement of major spectroscopic survey projects worldwide, the volume of spectral data has reached the scale of tens of millions, posing new challenges for data processing and analysis. To address this issue, the team designed and developed an artificial intelligence-based platform for spectral visualization and analysis, aiming to lower the barrier to the use of spectral data and improve the efficiency of scientific research and teaching. Integrating information technology with machine learning algorithms, the platform implements functions including user management, data management, visualization and analysis, spectral classification, parameter measurement, spectral annotation, and multi-band and multimodal data fusion, while supporting flexible user roles and data management strategies. The platform significantly reduces the difficulty of analyzing massive spectral data, improves the efficiency of scientific output, and, as an important tool of the National Astronomical Data Center, serves data processing and scientific research for large-scale scientific engineering projects such as LAMOST (Large Sky Area Multi-Object Fiber Spectroscopy Telescope) and MUST (MUltiplexed Survey Telescope). Meanwhile, the platform has potential for integrating science and education, providing a new approach to the cultivation of innovative talent.


## Full Text

doi: 10.15940/j.cnki.0001-5245.2026.04.007

# An Artificial-Intelligence-Based Spectral Visualization and Analysis Platform*


Lei Guohong[1] Pu Zhouhao[2,3] Xu Yang[4] Zhou Yunfan[2,3] Chen Xunzhou[2,3] Tian Haijun[2,3†]

(1 School of Ethnology, China Three Gorges University, Yichang 443002)
(2 School of Science, Hangzhou Dianzi University, Hangzhou 310018)
(3 Zhijiang Branch, National Astronomical Data Center, Hangzhou 310018)
(4 National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012)



**Abstract** Spectra carry abundant astrophysical and chemical information and are an important means for revealing the nature of celestial bodies and even exploring the laws of the universe. With the continuous advancement of major global spectroscopic survey projects, the volume of spectral data has reached the scale of tens of millions, posing new challenges for data processing and analysis. To address this problem, our team designed and developed an artificial-intelligence-based spectral visualization and analysis platform, aiming to lower the threshold for using spectral data and improve the efficiency of scientific research and teaching. The platform integrates information technology with machine-learning algorithms, and implements functions including user management, data management, visualization analysis, spectral classification, parameter measurement, spectral annotation, and fusion of multi-band and multi-modal data. It supports flexible user roles and data-management strategies. The platform significantly reduces the difficulty of analyzing massive spectral data, improves research output efficiency, and, as an important tool of the National Astronomical Data Center, serves data processing and scientific research for major science projects such as LAMOST (Large Sky Area Multi-Object Fiber Spectroscopy Telescope) and MUST (MUltiplexed Survey Telescope). At the same time, the platform has potential for integrating science and education, providing a new pathway for the cultivation of innovative talent.

**Keywords** astronomical database: astronomical spectral data, astrometry: big-data measurement, astronomical instruments: spectral visualization platform

**Chinese Library Classification No.:** P112; **Document identification code:** A

## 1 Introduction

The spectra of celestial bodies carry key information related to their physical and chemical properties. Through systematic analysis of spectral data, basic parameters such as chemical composition, effective temperature, surface gravity, age, and kinematic properties of celestial bodies can be obtained. These parameters not only help reveal the formation and evolutionary history of individual celestial bodies, but also provide important observational constraints for studying the structure and evolution of the Milky Way, and even for exploring the origin and evolutionary laws of the universe. Spectroscopic observations provide an important window for obtaining information from the deep universe and play a fundamental role in modern astronomical research.

The Sloan Digital Sky Survey (SDSS)[1] in the United States is one of the representative international survey projects. After more than twenty years of survey observations, the project has released more than 6 million spectra. To date, more than 25,000 high-quality academic papers have been published, and the articles have been cited more than 770,000 times (statistical data from the SAO (Smithsonian Astrophysical Observatory)/NASA (National Aeronautics and Space Administration) Astrophysics Data System (ADS)[2]). The major science project independently developed and built by China—the Guo Shoujing Telescope (Large Sky Area Multi-Object Fiber Spec-

Original manuscript received 2025-10-28; revised manuscript received 2026-03-16

* Supported by the National Natural Science Foundation of China (12373033,

12447204), and the first batch of scientific research projects of the China Space Station Telescope (CSST) survey (CMS-CSST-2021-A09/A08)
† hjtian@hdu.edu.cn

troscopy Telescope, LAMOST[^3]). Since its official acceptance in 2009, the amount of spectra it has obtained is three times the sum of spectra released by all other survey telescopes internationally, making it the world' s first sky-survey project to publish more than 20 million spectra. Over the next decade, next-generation telescope projects around the world are being accelerated, indicating that spectroscopic observing capability will achieve a new leap, and the total number of spectra is expected to increase by another order of magnitude. Among them, the Multiplexed Survey Telescope (MUST)[^4], led by Tsinghua University, is particularly eye-catching. Owing to its unique optical and mechanical design, with a 6.5 m aperture and a 7 deg$^2$ field of view, it can observe at one time a sky area equivalent to 30 full moons, while carrying out medium- and high-resolution spectroscopic observations of at least 20,000 celestial objects within the wavelength range of 0.36–1.0 $\mu$m. It is expected that within ten years after first light in 2029, MUST will be able to collect more than 300 million spectra. Such massive spectroscopic data provide abundant data resources for scientific research, teaching, and science popularization.

To date, there still remains internationally a lack of an integrated platform for spectroscopic data visualization, analysis, and management that can simultaneously meet the needs of scientific research, teaching applications, and science communication. In the current field of astronomical spectral analysis, many specialized software tools emerged in the early period, such as the Virtual Observatory Spectrum Tool (VOSpec)[^5], SpecView[^6], Spectral Analysis Tool (SPLAT)[^7], Centre d' Analyse Scientifique de Spectres Infrarouges et Submillimetriques (CASSIS)[^8], and others. Most are based on the Java framework, are complex to deploy, depend on the local environment, and have relatively single functions; they have not formed a unified collaborative platform integrating scientific research and teaching, resulting in their failure to become widely popular in practical applications. In contrast, the artificial-intelligence-based spectrum visualization and analysis platform built by our team has undergone systematic upgrading and optimization on the basis of the original version[^9]. Centered on a modern Web architecture with front-end/back-end separation, users do not need to install any plug-ins or configure a complex environment; full-featured access can be achieved simply through a browser, greatly lowering the threshold for use. Although the old platform implemented basic spectrum visualization and template matching, it had limitations such as an outdated architecture, insufficient artificial intelligence (AI) capability, and weak collaboration functions, restricting its scope of use. The new platform deeply integrates multiple artificial intelligence algorithms, possesses intelligent analysis capabilities such as automatic spectral classification, parameter measurement, redshift estimation, and anomaly detection, and integrates multimodal data fusion, interactive visualization, and group annotation functions. At the same time, the system supports flexible user role management and a team collaboration mechanism, enabling integration of task distribution, annotation review, and result statistics. It shows significant advantages in efficient screening and multidimensional verification of special objects, thereby taking the lead comprehensively over existing international software in functionality, ease of use, and collaboration.

Compared with photometric image data that possess intuitive visual characteristics, the degree of visualization of spectroscopic data is relatively low; their form of presentation does not contain rich color or direct spatial distribution information, and thus they have limitations in intuitiveness and universality. In addition, the storage formats of spectroscopic data are usually rather complex and involve specialized knowledge of physics and data processing; for users lacking the relevant background, understanding and using them present considerable difficulties. This characteristic means that in science-education integration activities, especially in teaching practice, the learning and application of spectroscopic data often face major challenges.

With the rapid development of information technology, AI is playing an increasingly important role in scientific research. By virtue of its powerful capabilities in data processing, pattern recognition, predictive analysis, and automated execution, AI is driving a profound transformation of the scientific research paradigm. At the same time, through interactive science education tools and platforms, AI technology can present complex scientific concepts in a more intuitive form, not only promoting scientific discovery by researchers but also helping to enhance the public's scientific understanding and interest.

Against this background, this study proposes and designs a set of artificial-intelligence-based spectral identification and analysis platform (hereinafter referred to as the "spectral identification platform"). Relying on information technology, the platform integrates the visualization, analysis, and management of spectroscopic data, aiming to build a team-collaborative science-education integration system oriented toward the dual needs of scientific research and education. With spectroscopic data as its core, the platform combines multimodal data such as multi-band images, time-domain information, and literature, realizing comprehensive analysis and correlated exploration of data. Astronomical researchers can use the platform to conduct scientific research work such as searches for special celestial objects and sample construction; teachers can implement inquiry-based teaching activities with astronomical characteristics based on platform resources. In addition, with the assistance of artificial intelligence algorithms, the platform can also complete annotation tasks for large-scale spectroscopic data through student group collaboration. The construction and promotion of the spectral identification platform are expected to become an important data-processing and science-education support platform for China's large spectroscopic survey projects (such as LAMOST and MUST).

Fig. 1 The overall architecture of an AI based spectral recognition platform system

Figure 1: Fig. 1 The overall architecture of an AI based spectral recognition platform system

## 2 Artificial-Intelligence-Based Spectral Identification Platform

By introducing multiple machine-learning algorithms, adopting a front-end/back-end separated architecture, and integrating collaborative analysis of multimodal data, the new spectral identification platform has realized a leap from a "tool" to an "integrated science-education platform." To facilitate a clear explanation of the platform' s overall design ideas, the following briefly introduces its architectural design and main functions.

### 2.1 Overall System Architecture Design

In its architectural design, the new spectral-recognition platform contains the following four core components: the artificial intelligence layer (AI Layer), the data node layer (Data Node Layer), the spectral visualization and labeling layer (Visualization&Label Layer), and the user roles layer (User Roles Layer), as shown in Figure 1:

Figure 1 Overall architecture of the AI-based spectral-recognition platform system

Fig. 1 The overall architecture of an AI based spectral recognition platform system

**Artificial intelligence layer:** This layer integrates multiple machine-learning algorithms, covering core functions such as spectral classification, stellar spectral feature extraction, inversion of stellar atmospheric parameters, and redshift measurement for galaxies and quasars (e.g., Support Vector Machine, SVM; Artificial Neural Network, ANN; Principal Component Analysis, PCA; Deep Learning, DL). Given the relatively large amount of computation, the algorithm modules in this layer are independently deployed on dedicated servers to centrally process computation requests and return the results to the spectral visualization and labeling layer for further analysis and verification.

**Data node layer:** This layer is responsible for managing the tabular and spectral data uploaded by users. It uses Java Database Connectivity (JDBC) connections and supports three data-visibility modes: PublicDB (accessible to all registered users), GroupDB (accessible only to group members), and MyDB (private to the user). The upload process adopts a large-file chunked-transfer and breakpoint-resume mechanism to improve the efficiency and stability of data transmission. FITS (Flexible Image Transport System) files are an important format for the storage and transmission of astronomical data; they contain header records and data records, with related information described in textual form in the header. CSV (Comma-Separated Values) files are a commonly used format in astronomy; they store tabular data (numbers and text) in plain text, facilitating data exchange and sharing among different software and programming languages.

**Spectral visualization and labeling layer:** This layer implements online visualization and analysis of spectral data. Visualization functions include tabular and spectral-data display, interactive visualization of spectral images, and integrated display of multiband data. Data-analysis functions cover AI-assisted spectral classification, spectral-template matching, spectral-line labeling, real-time calculation of redshift and line-of-sight velocity, and submission and management of spectral-recognition results. The visualization module is implemented based on the Echarts[^10] chart component, and integrates wavelet-transform denoising algorithms and a $3\sigma$ sky-filtering algorithm, thereby improving the accuracy and reliability of data processing.

**User roles layer:** This layer is responsible for management of user roles and work groups, assignment of spectral-recognition tasks, and review and statistics of labeling results. This module adopts the OAuth[^11] single sign-on authorization mechanism, providing convenient registration-free access for users of the National Astronomical Data Center (NADC)[1].

The new-generation spectral platform adopts a system architecture with front-end and back-end separation. The front-end part is built on the Vue.js framework[^12]. This framework is a progressive JavaScript development environment that can achieve efficient user-interface rendering and component-based management. The back-end part adopts the Spring-Boot framework[^13]. On the basis of the traditional Spring system[^14], this framework is highly integrated and optimized, simplifying the configuration process and dependency management and providing a flexible and efficient runtime environment. In addition, the back-end data persistence layer uses the MyBatis framework[^15] to implement object-relational mapping (Object-Relational Mapping, ORM), thereby supporting efficient mapping and interaction between Java objects and database records. The system database uses MySQL to ensure the stability and scalability of data management.

Supported by the above technical system, the new platform has higher performance stability and functional extensibility. At present, the system has been officially launched and is operating with the support of the National Astronomical Data Center (NADC).

[1] https://nadc.china-vo.org/speczoo-system/

Flux template visualization. Sub category: O. TemplateName: O9V_mk. Plot of flux versus wavelength with marked H, H$\gamma$, and H$\beta$ lines.

Figure 2: Flux template visualization. Sub category: O. TemplateName: O9V_mk. Plot of flux versus wavelength with marked H, H$\gamma$, and H$\beta$ lines.

## 2.2 Basic Functional Overview

### 1. User Roles and Workgroup Management

The platform is designed with four categories of user roles: system administrator, workgroup leader, workgroup administrator, and spectral expert (ordinary user). Each user may participate in one or more workgroups according to their responsibilities. Workgroup leaders are usually senior researchers with experience in relevant fields; they are responsible for establishing workgroups and assigning spectral-identification tasks. Workgroup administrators assist the leaders in task management and progress supervision; ordinary users mainly undertake specific work such as spectral annotation and data verification. This multi-level role and permission design improves the organizational flexibility and collaborative efficiency of the system.

### 2. Spectral-Type Management and Support for Characteristic Spectral Lines

The system presets spectral templates for common celestial-body types such as stars, galaxies, and quasars, and integrates a characteristic spectral-line wavelength library for common chemical elements. Figure 2 shows an example of a spectral template: it is a spectral template for an O-type star, for which the platform presets three characteristic spectral lines. This module provides users with standardized classification references and spectral-line retrieval support, helps improve the scientific rigor and consistency of spectral identification, and lowers the threshold for beginners.

### 3. Catalog Data Upload and Spectral Visualization

Figure 3 shows the platform visualization interface. Users can obtain catalog data from data centers such as LAMOST and SDSS, and upload them to the system according to the CSV specification defined by the platform. After the upload is completed, the system automatically performs data integrity checks and import operations. Users can browse data, analyze spectra, or perform overplot processing of multiple spectra on the platform. The platform supports functions such as spectral visualization, local feature magnification, spectral annotation, and wavelet denoising, without the need to install additional plug-ins. In addition, reading data from CSV files and assembling them into URLs to obtain files from major survey cloud servers entails network access latency; uploading local FITS files can further improve access and operation efficiency.

Figure 2 A stellar spectral template with spectral type O9V; this template is from Morgan-Keenan Classification (MKCLASS)[^12].

Fig. 2 A stellar spectral template with spectral type O9V, sourced from Morgan-Keenan Classification (MKCLASS)[^12].

#### 4. Task Deployment and Submission of Annotation Results

As shown in Fig. 4, on the platform, a workgroup administrator can assign scientific research tasks or research-based learning tasks to members or students within the group and collect their completion status. After a task is completed, the annotation results are mainly reflected in the annotation of the physical parameters of target celestial bodies. Different types of celestial bodies correspond to different annotation contents. For example, quasar-like objects require annotation of redshift, emission-line width, and continuous-spectrum flux; galaxies require annotation of redshift, age, metallicity, and mass; and stars require annotation of radial velocity, surface gravity, and metallicity. To improve system flexibility, the platform incorporates the physical parameters to be annotated into the scope of dynamic management, so as to support diversified task requirements and continuously updated scientific research objectives.

**Fig. 3** The platform' s visualization interface is demonstrated using a quasar spectrum as an example (screen capture; the same applies hereafter). The interface comprises three tabs: "Data Browser," used for paginating through all data rows of the target table; "Visualization," used for spectral analysis; and "Overlay Spectra," used for superimposing and comparing different spectral data. In the "Visualization" tab, a dark curve represents the spectrum of a quasar at a redshift of 1.89646; the light-colored curve above it is a quasar template spectrum, and the inserted subfigure is the quasar' s SDSS image.

Fig. 3 The platform' s visualization interface is demonstrated using a quasar spectrum as an example (screen capture, the same applies hereafter). The interface comprises three tabs: "Data Browser" for paginating through all data rows of the target table; "Visualization" for spectral analysis; and "Overlay Spectra" for superimposing and comparing different spectral data. In the "Visualization" tab, a dark curve represents the spectrum of a quasar at a redshift of 1.89646. A light-colored template spectrum of a quasar is overlaid for comparison, and an inserted subplot displays the quasar' s SDSS image.

### 2.3 Advanced Functions Based on Artificial Intelligence

The spectral-recognition platform integrates multiple machine-learning algorithms to improve the efficiency and precision of astronomical spectral analysis. In terms of spectral classification, Wu et al.[16] introduced partial convolution and grouped convolution techniques, solving the problem that traditional convolutional neural networks are prone to overfitting in astronomical spectral classification and significantly improving model performance. In terms of measuring spectral parameters, Zhang et al.[17] proposed a support-vector-regression-based

Figure 4 interface screenshot: "Add new marking task," with fields including GroupName, TaskName, Description, SheetName, BeginDate, EndDate, RepeatLimit, MarkQuantity, and OK/Cancel buttons.

Figure 3: Figure 4 interface screenshot: "Add new marking task," with fields including GroupName, TaskName, Description, SheetName, BeginDate, EndDate, RepeatLimit, MarkQuantity, and OK/Cancel buttons.

Stellar LAbel Machine (SLAM) model, enabling precise measurement of stellar atmospheric physical parameters. In terms of large-language-model-assisted analysis, the platform has integrated the Aligning developed by the Zhijiang Laboratory.

and Translating Spectroscopic Measurements for Stars (SpecCLIP) model[^18], which supports remote invocation and returns stellar parameters; Zhong et al.' s[^19] Galaxy Spectra neural Networks (GaSNet) algorithm has added new computing power to the platform, further enriching the platform algorithms. The integration of these algorithms provides strong support for automatic spectral classification and parameter measurement, significantly enhancing the functional richness and usability of the platform. Figure 5 shows application examples of three algorithms. Through "one-click" operations, users can complete automatic identification of celestial-object types and automatic measurement of basic stellar atmospheric parameters, effectively assisting spectral analysis and visualization in scientific research and teaching tasks. The integration of the above algorithms lays a foundation for automated spectral classification and parameter measurement on the platform, enabling users to efficiently complete celestial-object classification and measurement of stellar atmospheric parameters on the platform, and assisting spectral analysis and visualization work.

Figure 4 Interface for assigning annotation tasks. During task assignment, the group administrator can decide which members of the group will participate in task allocation, how many spectral annotation tasks each member will complete, how many random repetitions each spectrum will have, and the deadline for task completion.

Fig. 4 The interface for assigning annotation tasks. During the task assignment process, the group administrator can decide which members of the group will participate in task allocation, how many spectral annotation tasks each member will complete, how many random repetitions each spectral will have, and the deadline for task completion.

(1) Intelligent classification of celestial objects

Multiscale Partial Convolution (MSPC-Net)[^16] is a spectral classification method based on convolutional neural networks (CNNs), aimed at improving the accuracy, efficiency, and generalization capability of astronomical spectral classification. By introducing techniques such as partial convolution, large-kernel convolution, and grouped convolution, this method achieves multi-scale feature extraction, thereby alleviating the overfitting and feature-redundancy problems that traditional methods may encounter when processing complex spectral data. MSPC-Net uses convolution kernels of different sizes to perform convolution operations on spectral data, simultaneously capturing local and global features and comprehensively characterizing the complex structure of spectra; partial convolution performs convolution only on some channels, which can reduce the number of model parameters and lower the risk of overfitting; grouped convolution divides the input channels into multiple groups for separate convolution processing, further reducing computation and the number of parameters while enhancing each channel' s ability to learn different features. Compared with traditional CNN models, MSPC-Net maintains high performance while significantly reducing parameter scale, thereby lowering storage and computational costs and improving the model' s generalization performance on different datasets, making it suitable for diverse astronomical spectral classification tasks. Through the innovative combination of multi-scale feature extraction and convolution strategies, this method provides an efficient and reliable tool for automated astronomical spectral classification.

(2) Prediction of stellar atmospheric parameters

Stellar LAbel Machine (SLAM)[^17] is a data-driven method based on support vector regression (SVR), dedicated to extracting accurate stellar atmospheric parameters from large-scale spectroscopic surveys. As a nonlinear regression method, SVR maximally compresses prediction errors by constructing an optimal regression hyperplane in a high-dimensional feature space, and has good generalization capability. SLAM can adapt to multiple spectral types and dynamically adjust model complexity according to data characteristics, thereby achieving high precision and high stability in parameter extraction. Test results on the LAMOST Data Release 5 (DR5) dataset show that this model can achieve a relatively high level of prediction accuracy under high signal-to-noise-ratio conditions. In addition, SLAM can handle high-dimensional problems; for example, when combining APOGEE (Apache Point Observatory Galactic Evolution Experiment) and LAMOST stellar data, it provides accurate stellar-label predictions for about 1 million K-type giants in LAMOST DR5. Compared with traditional data-driven methods,

SLAM has outstanding performance in terms of model flexibility, adaptability, and extensibility, and can be extended to prediction tasks with more stellar labels.

(3) Large-language-model linkage

The platform has integrated the SpecCLIP large language model[^18]. This model was jointly developed by the Zhijiang Laboratory and the National Astronomical Observatories, Chinese Academy of Sciences, and is one of the achievements of the "Astronomical Large Model Training Program." The training program aims to cultivate interdisciplinary "AI + astronomy" talent by integrating astronomical knowledge and large-model technologies through cross-disciplinary

Fig. 5

Figure 4: Fig. 5

collaboration. During development, the team relied on the Tianji artificial-intelligence open-source platform of the Zhijiang Laboratory, using distributed computing engines and efficient training frameworks to accelerate model iteration and optimization. SpecCLIP is mainly used for stellar census work; it can systematically analyze large volumes of stellar spectral data, identify stellar types and physical parameters (such as temperature, chemical composition, and motion state) and evolutionary characteristics, extract key features such as spectral-line intensity and multi-spectral-band displacement, and complete stellar classification (such as main-sequence stars, red giants, and white dwarfs) as well as estimation of parameters including mass, age, and metal abundance through deep-learning algorithms. Compared with the SLAM method corresponding to SpecCLIP, SLAM has greater flexibility in resource-constrained scenarios and small-data tasks, whereas SpecCLIP relies on large-scale training data and is suitable for end-to-end complex pattern mining and automated analysis, such as anomaly detection in low-signal-to-noise spectra. Combining the two can achieve complementary advantages: it not only optimizes the allocation of computing resources, but also provides a complete solution from data screening to in-depth analysis for research on stellar formation and chemical evolution.

Fig. 5 Using a metal-poor star (LAMOST obsid: 51216196) as an example to demonstrate three algorithms. MSPC-Net shows it as an F7 star, while SLAM and SpecCLIP measure its atmospheric parameters. The atmospheric physical parameters of the star measured by high-resolution spectroscopy are $T_{\text{eff}} = (5560.69 \pm 171.69)$ K, $\log g = 3.56 \pm 0.47$, $[\text{Fe/H}] = -1.98 \pm 0.2$ ($T_{\text{eff}}$ is the effective temperature, and $\log g$ is the surface gravity $\log g$, where $g$ is in units of $\text{cm} \cdot \text{s}^{-2}$).

(4) GaSNet-III spectral analysis

The spectrum-identification platform has integrated the GaSNet-III (Galaxy Spectral Network) function[^19]. GaSNet-III is a generative deep-learning framework that integrates spectral reconstruction, redshift estimation, and anomaly detection into a unified analysis pipeline, and is suitable for large-scale spectroscopic surveys. This method adopts two types of generative neural-network architectures: one is inspired by autoencoders and is used to extract PCA-like interpretable eigenspectra and coefficients; the other is an end-to-end denoising-optimized U-Net model, used to exploit translational symmetry in logarithmic wavelength space and spectral similarity among observational data to reconstruct rest-frame spectra. GaSNet-III, through rapid scanning

across the entire $\chi^2$ space, across multiple classes of celestial objects such as stars, galaxies, and quasars, identifying the globally optimal solution for object class and redshift range, thereby achieving object classification, high-precision redshift determination, and robust marking of anomalous spectra based on the $\chi^2$ curve. On stellar samples, its classification accuracy can exceed 0.999. The computational efficiency of this method is several orders of magnitude higher than that of classical template-fitting methods, while maintaining comparable or even higher precision. GaSNet-III improves computational efficiency by several orders of magnitude, without compromising prediction accuracy, providing an extensible analysis solution for next-generation spectroscopic surveys such as the Dark Energy Spectroscopic Instrument (DESI) [20] and the 4-metre Multi-Object Spectroscopic Telescope (4MOST) [21].

## 3 Large-Sample Search and Construction Practices Based on Spectroscopic Data

Relying on the knowledge platform, researchers can independently or collaboratively conduct diversified scientific studies, showing particular advantages in the search for special celestial objects and the construction of large samples. Taking the search practices for three typical classes of celestial objects as examples, the following demonstrates the platform' s capabilities in multi-band data integration and analysis.

### 3.1 Identification of Strong Gravitational-Lens Candidates

The strong gravitational-lensing effect is an important phenomenon predicted by general relativity. When a massive celestial body (such as a galaxy or black hole) lies between the observer and a more distant luminous object, its gravity bends the light from the distant object, forming a magnified and sometimes multiply distorted image. Gravitational lensing changes the path by which light from a background galaxy reaches Earth, but does not change its spectral features; therefore, it has broad applications in astronomy. For example, gravitational lensing can be used to detect quantum dark matter in the Milky Way, especially massive compact halo objects (MACHOs); by observing light variations caused by microlensing effects, the total mass of compact objects in the Milky Way can be estimated [22]. Strong gravitational lensing can also be used to search for supermassive binary black holes at the centers of galaxies, thereby studying the dynamics of galactic nuclei and the formation mechanisms of black holes [23]. In addition, gravitational lensing can assist in measuring cosmological parameters, such as the density and expansion rate of the Universe [24]. Most known galaxy-scale gravitational-lens systems have been discovered through imaging observations, while only a small number of systems have been identified through spectroscopic observations.

The spectra of strong gravitational-lensing objects are formed by the superposition of the spectra of the lensing object and the source object, exhibiting multiple-redshift characteristics. For identifying strong gravitational-lens candidates using spectroscopic data, this platform integrates automated parameter measurement and spectral annotation functions for both the lensing object and the source object, thereby systematically improving the discovery efficiency of strong-lensing candidates. The platform can refer to the method proposed by Zhong et al. [25], and, based on spectroscopic observational data, achieve efficient screening through the following steps: 1. perform preliminary screening of all spectra, calculate the probability that each spectrum is a lens candidate, and select high-probability candidates according to a preset probability threshold; 2. separately predict the background-source redshift and foreground-lens redshift of the candidates, requiring that the background-source redshift be at least 0.1 larger than the foreground-lens redshift; 3. include the high-quality candidates satisfying the criteria in manual visual inspection, to further confirm their authenticity and completeness.

In the process of carrying out systematic screening of strong gravitational-lens candidates in survey data such as LAMOST, the knowledge platform provides key technical support. Its advantages are mainly reflected in multimodal data collaboration and visualization analysis capabilities: researchers can use the platform' s spectral-interaction functions to identify typical emission lines in spectra that may originate from background galaxies (such as [O III] $\lambda$5007 Å), thereby determining whether multiple redshift features are present.

At present, the team is manually reviewing strong gravitational-lens candidates in survey data such as LAMOST and, with the aid of the knowledge platform, constructing a systematic screening workflow. The platform' s core advantages lie in its multimodal collaborative data-analysis capabilities: first, researchers can clearly identify, through the interactive spectral-visualization function, the characteristic emission lines of background galaxies in the candidate spectra (such as [O III] $\lambda$5007 Å and Ly$\alpha$), confirming the multiple-redshift features of the light source; second, the platform integrates cross-survey data linkage modules, which can directly retrieve and compare deep-field images from high-resolution space telescopes such as the Hubble Space Telescope (HST). When a candidate simultaneously satisfies the following conditions, preliminary authentication can be completed: 1. the foreground emission lines in the spectrum can be clearly identified, and their redshift is markedly different from that of the background emission lines; 2. HST or other survey images display typical strong-lensing morphological features (such as an Einstein ring, multiple images, or prominent arc-like structures).

By further combining other observational methods (such as Very Long Baseline Interferometry, VLBI), strong gravitational-lens candidates can be analyzed in depth to obtain more information supporting confirmation. Within the "model screening–manual review" framework, the knowledge platform provides efficient visualization, data linkage, and collaborative management support, thereby significantly reducing the cost of manual inspection, improving the accuracy of candidate screening, and playing a key role in multidimensionally verifying strong gravitational-lensing phenomena.

Figure 6 shows an example of a typical strong gravitational-lens system on the

Fig. 6 spectrum and HST image of a strong gravitational lens source

Figure 5: Fig. 6 spectrum and HST image of a strong gravitational lens source

knowledge platform, with the corresponding HST image shown in the subfigure. From the spectral features

···indicates that the main signal comes from the foreground galaxy. This galaxy is an old bright red galaxy; because star-formation activity has almost ceased, its spectrum contains almost no emission lines and mainly shows absorption-line features, with a redshift of $z = 0.1196$. The emission lines in the spectrum mainly come from the background galaxy. As shown in Fig. 7, its redshift is $z = 0.1965$, further confirming that the redshift of the background source is greater than that of the foreground galaxy. This workflow fully demonstrates how the spectrum-recognition platform can be used to identify and verify strong gravitational-lens candidates.

Fig. 6 The spectrum and HST image (subplot) of a strong gravitational lens source. The foreground is a bright red galaxy system, and in the spectrum the absorption lines K, H, Ge, Mg I, Na I, H$\alpha$, etc., clearly show that the redshift of the foreground galaxy is 0.1196.

Fig. 6 The spectrum and HST image (subplot) of a strong gravitational lens source. The foreground is a bright red star system, and in the spectrum, absorption lines such as K, H, Ge, Mg I, Na I, H$\alpha$ clearly show a redshift of 0.1196 for the foreground galaxy.

## 3.2 White-Dwarf–Main-Sequence Binary Search

White-dwarf–main-sequence binaries (White Dwarf-Main Sequence Binary, WDMS) are a common type of compact binary system in the Milky Way, consisting of a white dwarf and a low-mass main-sequence companion star, and usually originate from the evolution of main-sequence binaries. In most cases, because the initial orbital separation is relatively large, the components of the binary can evolve independently and no significant mass exchange occurs. However, for some main-sequence binaries with relatively small initial orbital separations, when the more massive primary leaves the main sequence and enters the red-giant or asymptotic-giant branch, its expanded envelope may engulf the companion star, causing the system to enter the common-envelope (Common Envelope, CE) evolutionary phase. During this phase, friction within the common envelope causes the orbital separation of the binary to gradually shrink; meanwhile, the orbital energy and angular momentum decrease accordingly, ultimately leading to the ejection of the common envelope and the formation of a post-common-envelope binary (Post-Common-Envelope Binary, PCEB). Because WDMS binaries are numerous and easy to observe, they have become an ideal class of astronomical systems for studying the mechanisms of common-envelope evolution [26].

Fig. 7

Figure 6: Fig. 7

To efficiently identify and study such systems, this platform integrates spectral analysis with secondary-template cross-validation functions, significantly improving the discovery and confirmation efficiency of WDMS binaries. As shown in Fig. 8, a pair of WDMS binaries identified through the spectrum-recognition platform is presented; its optical image comes from SDSS (subplot), and the dark curve is the corresponding LAMOST optical···

spectrum. Because the white dwarf and the main-sequence star have a small angular separation, LAMOST cannot separate the optical signals of the two stars in spectroscopic observations, causing the spectrum to superpose the features of both stars. The Spectra Platform provides a two-template matching function, which can fit different wavelength ranges of a spectrum and thereby identify the superposed components: the blue end of the spectrum matches a white-dwarf spectral template, while the red end matches an M-type main-sequence-star template. To confirm that these two stars constitute a physical binary (i.e., that they have a real mutual interaction), it is necessary to verify that their distances from Earth are essentially the same and that their kinematic states are consistent (neglecting the effect of mutual orbital motion). The Spectra Platform supports one-click access to the parallax and proper-motion data of the ESA Gaia satellite, with a parallax difference within 5%. If the above conditions are satisfied, the system can be preliminarily identified as a white-dwarf–main-sequence binary.

**Fig. 7** The emission lines of the background galaxy superimposed on the spectrum of the strong gravitational lensing source, where the subgraph is an enlarged image of the spectrum near wavelengths 8000 Å. These emission lines clearly show a redshift value of 0.1965 for the background galaxy.

By integrating complex spectral decomposition and multisource data-verification procedures into a coherent and automated analysis workflow, this platform has fundamentally changed the traditional research mode that relied on manual comparison, greatly improving the efficiency and reliability of large-sample screening and study of peculiar binary stars. Based on the massive accumulated spectroscopic data from surveys such as SDSS and LAMOST, the systematic search for WDMS binary samples is of great scientific value for in-depth understanding of the evolutionary processes of compact binaries[^27], especially common-envelope evolution mechanisms[^28]. In addition, for general types of binary systems, the Spectra Platform, by leveraging its advantage in multimodal data fusion, can also efficiently identify and capture targets with special spectral features, providing strong support for the rapid screening, classification, and certification of rare celestial objects.

Fig. 8

Figure 7: Fig. 8

### 3.3 Efficient Screening and Subtype Determination of Carbon Stars

Carbon stars are a class of special stars whose atmospheric carbon abundance is higher than oxygen. Their optical spectra show prominent molecular carbon absorption bands (such as $C_2$, CN, and CH). Such stars generally have temperatures and luminosities similar to those of late-type G-, K-, and M-type giant stars, but their chemical

···exhibit anomalous composition. According to spectral features, they can be further divided into subtypes such as C–H, C–R, and C–N. The formation of carbon stars is closely related to mass transfer in binary systems, and they are important tracers for studying the chemical evolution and stellar evolution of galaxies. However, the accurate identification of carbon stars from massive spectral data faces the challenges of feature discrimination and low efficiency in manual screening; moreover, in practical applications, the final confirmation of carbon stars is highly dependent on manual inspection of spectra.

To address this challenge systematically, this platform integrates spectral analysis and cross-validation with multi-band photometric data, which can greatly improve the efficiency and reliability of carbon-star candidate screening. Based on the functions of the platform, researchers can efficiently reproduce or optimize the multi-stage screening procedure constructed by Li et al.[29] based on 10,599,979 low-resolution spectra from LAMOST DR7. The procedure includes: 1. selecting spectra with signal-to-noise ratio $S/N(i) > 10$ ($i$ denotes the $i$-th pixel); 2. screening candidates using threshold values of carbon molecular-band indices; 3. combining 2MASS infrared photometric data to divide the candidates into warm and cool groups and measure them separately; 4. finally, the candidates enter the manual inspection stage.

Fig. 8 A spectrum of a WDMS binary from LAMOST, with its SDSS optical image shown in the inset. The thin line and the thick black curve are the spectral templates of the white dwarf and the M-type main-sequence star, respectively.

Fig. 8 The spectra of a WDMS binary from LAMOST, with its SDSS optical image shown in the subfigure. The thin line and the thick black curve are the spectral templates of the white dwarf and the M-type main-sequence star, respectively.

The application of the spectrum-identification platform significantly reduced the time required for manual inspection. On the basis of automated screening (line indices and infrared colors), the team carried out systematic manual inspection of the candidate spectra. The specific procedure was as follows:

1. Remove low signal-to-noise spectra ($S/N < 10$), remove quasars (broad emission lines), M dwarfs (TiO/VO molecular bands), recog-

Fig. 9 Spectrum of a carbon-star candidate from LAMOST

Figure 8: Fig. 9 Spectrum of a carbon-star candidate from LAMOST

nizer/instrumental noise (abnormal spikes/fractures), and interstellar extinction distortions; 2. visually inspect the presence of key carbon molecular absorption bands, including the depth and profile of the $C_2$ molecular bands (4737 Å, 5165 Å, 5635 Å), the intensity ratios of the CN molecular bands (7065 Å, 7820 Å), and continuum cutoff features (C-N stars have no continuum at $\lambda < 4400$ Å); 3. according to the classification criteria proposed by Barnbaum et al.[30] in 1996, determine the subtype through diagnostic spectral lines; 4. finally, all spectra are independently judged by at least two researchers, and conflicting cases are resolved by team discussion.

Fig. 9 shows the spectrum of a carbon-star candidate identified through the spectral platform. This target comes from LAMOST sky-survey data and, after automatic preliminary screening, entered the manual verification stage. By comparison with classical templates, it can be clearly recognized that its spectrum contains prominent carbon-molecular absorption features, especially the $C_2$ molecular bands at 4737 Å and 5165 Å. At the same time, strong absorption lines of Sr II (4077 Å) and Ba II (4554 Å, 6496 Å) are present, indicating enrichment in strong *s*-process (slow neutron-capture process) elements, confirming it as a Ba-type carbon star.

Figure 9 Spectrum of a carbon-star candidate from LAMOST. Strong $C_2$ molecular absorption is visible at 4737 Å and 5165 Å, together with strong absorption lines of Sr II (4077 Å) and Ba II (4554 Å, 6496 Å), indicating that it is a Ba-type carbon star. According to the results returned by SpecCLIP (subplot), the abundance of carbon is several times that of oxygen.

Fig. 9 The spectrum of a carbon star candidate from LAMOST reveals strong $C_2$ molecular absorption bands at 4737 Å and 5165 Å along with strong absorption lines of Sr II (4077 Å) and Ba II (4554 Å, 6496 Å), identifying it as a Ba-type carbon star. According to the results returned by SpecCLIP (subplot), the abundance of carbon is several times that of oxygen.

## 4 Summary and Outlook

Spectra carry key information about the physical and chemical properties of celestial objects and are an important means of revealing the nature of celestial bodies and exploring the laws of the universe. With the continued advancement of large spectroscopic survey projects such as SDSS and LAMOST, the volume of spectroscopic data has reached the scale of tens of millions. In the next decade, next-generation survey telescopes under development internationally (such as MUST) are expected to raise the total amount of spectroscopic data by another order of magnitude, exceeding one hundred million spectra.

This wealth of data provides unprecedented opportunities for scientific research and science education. Through spectroscopic-data visualization, on the one hand, researchers can be helped to understand spectral features intuitively and accelerate scientific discovery; on the other hand, abstract concepts can be made concrete, improving students' interest in learning and teaching effectiveness, and cultivating scientific literacy and innovative ability.

However, unlike photometric images, spectroscopic data are not immediately intuitive and easy to understand. For users lacking relevant foundational knowledge, there is a certain difficulty in understanding and using them. To lower the threshold for using spectroscopic data and improve the efficiency of scientific research and teaching, our team has designed and developed a set of artificial-intelligence-based spectral visualization and analysis platforms. This platform uses advanced information technology and machine-learning algorithms to implement integrated functions including user management, data management, visual analysis, spectral classification, parameter measurement, spectral annotation, and fusion of multi-band and multimodal data.

In terms of research applications, the spectral platform supports researchers in independently or collaboratively carrying out searches for special celestial objects and constructing samples. For example, in strong gravitational lensing research, the platform can automatically screen spectroscopic survey data, calculate the probability that each spectrum is a lens candidate, and predict the redshifts of the foreground lens and the background source, ensuring that the redshift of the background source is higher than that of the foreground galaxy. High-probability candidates will enter manual review, including comparison of spectral features with high-resolution images (such as HST) to confirm typical strong-lensing morphologies, such as Einstein rings, multiple images, or arc-shaped structures. The platform provides multimodal data linkage and interactive visualization functions, enabling researchers to identify the absorption lines of foreground galaxies and the emission lines of background galaxies, and to complete high-reliability candidate confirmation. In binary-system research, the platform can be used for efficient identification of white-dwarf–main-sequence binaries (WDMS). WDMS binaries consist of a white dwarf and a low-mass main-sequence companion; some systems evolve through a common envelope (CE), during which orbital energy and angular momentum decrease, forming post-common-envelope binaries (PCEB). Through dual-template matching, the platform decomposes composite spectra into white-dwarf and main-sequence components, and combines Gaia trigonometric parallax with proprietary data to verify the physical association of the binary, significantly improving the screening efficiency and reliability of large WDMS samples. In studies of chemically peculiar stars, carbon stars have a higher carbon abundance than oxygen in their atmospheres, and their spectra show obvious carbon molecular absorption bands (such as $C_2$, CN, and CH), which can be subdivided into subtypes such as C-H, C-R, and C-N. By combining spectral features, infrared photometry, and manual review, the spectral platform enables systematic screening and certification of large-scale carbon-star candidates, significantly reducing manual workload while ensuring classification accuracy.

The platform' s multimodal and collaborative capabilities also provide important support for research and teaching. Through the cloud-based architecture, learners can access massive spectral data with no threshold, conduct interactive analyses, and visualize complex processes such as spectral diagnosis and redshift measurement. In the course practice of "Research Training and Practical Innovation" at Hangzhou Dianzi University, with AI assistance students can independently complete special-object identification tasks and experience the full-chain training from literature reproduction and data mining to outcome production. This model significantly lowers the entry threshold for astronomical research, enabling undergraduates to master spectral diagnostic capabilities within a short period, while also supporting the integrated cultivation of talent pipelines from undergraduate to master' s and doctoral levels.

In the future, the spectral platform will continue to integrate advanced algorithms and multi-band analysis functions, enhance global data-sharing capabilities, and expand educational applications, providing sustained support for astronomy and interdisciplinary research. At the same time, by establishing an active user community, it will encourage the sharing of research and teaching experiences, thereby achieving the platform' s continuous optimization and knowledge exchange.

**Acknowledgments** The authors thank Wu Jingjing, Zhang Yanxia, Zhang Bo, Wang Yi, Chen Huawei, Zhong Fumin, Li Changhua, Fan Dongwei, and others for their professional guidance and valuable support in key technical aspects such as the development of intelligent celestial-object classification algorithms, construction of basic parameter measurement models, and deployment of machine-learning services; they also thank Wu Xuebing, Shen Shiyin, Shi Jianrong, Luo Ali, Cui Chenzhou, Huang Xiang, He Zhihong, Li Rui, Tian Xiaoman, Cui Wenyuan, Ren Juanjuan, and others for their patient guidance and strong support during the development and practical exploration of innovative cases integrating science and education.

## References

[1] Almeida A, Anderson S F, Argudo-Fernandez M, et al. ApJS, 2023, 267: 44

[2] Grant C, Accomazzi A, Eichhorn G, et al. Physi, 2002, 143: 111

[3] Cui X Q, Zhao Y H, Chu Y Q, et al. A&A, 2012, 12: 1197

[4] Cai Feng, Zhang Chao, Fan Fan. Experimental Technology and Management, 2021, 38: 10

[5] Osuna P, Barbarisi I, Salgado J, et al. Astronomical Data Analysis Software and Systems XIV. San Francisco: ASP, 2005, 347: 198

[6] Busko I. Astronomical Data Analysis Software and Systems IX. San Francisco: ASP, 2000, 216: 79

[7] Škoda P, Draper P W, Neves M C, et al. A&C, 2014, 7: 108

[8] Lebouteiller V, Barry D J, Spoon H W W, et al. ApJS, 2011, 196: 8

[9] Lei Guohong, Xu Yang, Niu Chenhui, et al. Astronomical Research & Technology, 2018, 15: 216

[10] Xie J, Yang Z, Liu J, et al. Big Data and Algorithms. Piscataway: IEEE, 2023: 1347

[11] Guo Danqing. Computer Systems & Applications, 2014, 11: 7

[12] Doernhoefer M. ACM SIGSOFT Software Engineering Notes, 2006, 31: 16

[13] Wang Yonghe, Zhang Jinsong, Deng Anming, et al. Information and Communications, 2016, 10: 4

[14] Hu Qimin, Xue Jinyun, Zhong Linhui. Computer Engineering and Applications, 2008, 44: 5

[15] Ouyang Hongji, Ge Meng, Cheng Haibo. Microcomputer Applications, 2023, 39: 73

[16] Wu J J, He Y C, Zhang H T, et al. AJ, 2024, 167: 260

[17] Zhang B, Liu C, Deng L C. ApJS, 2020, 246: 1

[18] Zhao X F, Huang Y, Zhang H W, et al. ApJ, 2025, 998: 189

[19] Zhong F, Napolitano N R, Li R, et al. MNRAS, 2025, 543: 691

[20] Dey A, Schlegel D J, Lang D, et al. AJ, 2019, 157: 168

[21] De Jong R S, Agertz O, Agudo Berbel A, et al. Physi, 2012, 8446: 389

[22] Fu Jian, Yang Xiaofeng, Yan Zhen. Annals of Shanghai Astronomical Observatory, Chinese Academy of Sciences, 2007, 28: 66

[23] Su Yi. *Chinese Journal of Nature*, 2010, 32: 5

[24] Luo Liaofu, Xu Chongming. *Chinese Science Bulletin*, 1984, 23: 1422

[25] Zhong F, Napolitano N R., Li R., et al. *RAA*, 2022, 22: 28

[26] Ren Shanshan, Luo Ali, Zhao Yongheng, et al. *Progress in Astronomy*, 2014, 32: 462

[27] Willems B, Kolb U. *A&A*, 2004, 419: 1057

[28] Wang Wenbao, Guo Gemu, Ma Chunyu, et al. *Optics and Spectroscopy*, 2018, 38: 4

[29] Li L L, Wang S, Luo A L, et al. *ApJS*, 2024, 271: 12

[30] Barnbaum C, Stone R P S, Keenan P C. *ApJS*, 1996, 105: 419

# Spectral Visualization and Analysis Platform Based on Artificial Intelligence


LEI Guo-hong[1] PU Yuan-hao[2,3] XU Yang[4] ZHOU Yun-fan[2,3] CHEN Xun-zhou[2,3] TIAN Hai-jun[2,3]

*(1 College for Nationalities, China Three Gorges University, Yichang 443002)*
*(2 School of Science, Hangzhou Dianzi University, Hangzhou 310018)*
*(3 Zhijiang Branch, National Astronomical Data Center, Hangzhou 310018)*
*(4 National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012)*



**Abstract** Spectra, rich in the physical and chemical signatures of celestial objects, play a pivotal role in revealing the nature of astronomical bodies and exploring the mysteries of the universe. With the advancement of major spectral survey projects worldwide, we have ushered in an era of spectral data at the tens-of-millions scale. Confronted by the challenges posed by massive spectral datasets, we have designed and developed an artificial intelligence-powered spectral visualization and analysis platform. This platform aims to lower barriers to spectral data utilization while enhancing efficiency in both research and education. Leveraging cutting-edge information technology and machine learning algorithms, the platform integrates multiple functions including user management, data governance, visual analytics, spectral classification, parameter measurement, spectral annotation, and multi-band/multi-modal data integration. Its flexible user role and data management architecture significantly reduce the analytical threshold for massive spectral datasets, accelerating scientific productivity. As a core tool of the National Astronomical Data Center, it supports data processing and scientific output for major projects such as LAMOST (Large Sky Area Multi-Object Fiber Spectroscopy Telescope) and MUST (MUltiplexed Survey Telescope). Simultaneously, its potential to bridge scientific research and education opens new pathways for cultivating innovative talent.



**Key words** astronomical data bases: astronomical spectral data, astrometry: big data measurement, astronomical instrumentation: spectral visualization platform


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