Revista Mexicana de Astronomía y Astrofísica Serie de Conferencias, 60, 118-119, (2026)

OPEN ACCESS https://doi.org/10.22201/ia.14052059p.2026.60.106

"XVII LARIM", Montevideo, Uruguay, November 27 - December 1, 2023.Volume Editors: Gonzalo Tancredi, Cecilia Mateu & Juan José Downes

Identifying T Tauri stars in the Milky Way using machine learning and LAMOST spectra

C. D. Millan-Valderrama[1], **B. Sabogal**[1] **and J. Hernández**[2]

[1]*Universidad de los Andes, Bogotá, Colombia.*
[2]*Universidad Nacional Autónoma de México, Instituto de Astronomía.*

Keywords: *stars: pre-main-sequence, stars: variable: T Tauri, Herbig Ae/Be*

Corresponding author: cd.millan@uniandes.edu.co

OB associations are low-density groups of young stars. Because their OBA-type members are more massive and therefore brighter, the OB associations are primarily characterized by these stars (Wright, 2020). In contrast, their low-mass members are less identified, despite being more numerous. Improving our understanding about stellar formation requires the identification of low-mass members, such as T Tauri stars (TTS).

TTS are low-mass stars, $< 2M_{\odot}$, in the pre-main sequence. They are considered precursors of planetary systems (Briceño et al., 2019). Their optical spectra show an absorption line of LiI, which is an indicator of youth. TTS are classified into two categories: classical TTS (CTTS) and weak-line TTS (WTTS). The CTTS exhibit H_α line emission caused by a circumstellar disk, while in WTTS, the H_α line is relatively weak and is attributed to chromospheric activity. In general, this emission depends on their evolutionary state. Other significant optical spectral features of TTS include TiO and VO bands, which give an account of the stellar effective temperature.

The Chinese Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST)[1], generates 4000 spectra in a single exposure providing a large amount of spectra in the wavelength range of 3700 to 9000 Å. In the public data release DR8, there are approximately 10 million low-resolution spectra (R≈1800).

The large amount of data generated in modern telescopes, like LAMOST, demands modern approaches to data treatment. Machine learning (ML) is an essential tool that offers several solutions to this problem (Zhang & Zhao, 2015).

We present results on the detection and spectral classification of TTS in the Orion Star Forming Complex (OSFC) using ML techniques and LAMOST spectra.

To detect TTS, we conducted automated equivalent width (EW) measurements on H_α and LiI lines, as well as TiO and VO bands for 342 TTS from the Ori OB1a substellar association, as reported by Hernández et al. (2023). We incorporated a set of non-TTS (field stars) in the training sample. Subsequently, we implemented a Logistic Regression model (LR) for binary classification. This supervised ML algorithm assigns a value probability of being TTS or not through a sigmoid function:

$$p(m_i, \theta_i) = \frac{1}{1 + e^{-m_i \theta_i}}. \quad (1)$$

Where m_i represents the EW measures of low-correlated variables, such as LiI, H_α, TiO5, TiO6, and VO1; and θ_i denotes a set of parameters to be determined. The model assessment was conducted using the F1 score, resulting in a value of 0.92 and an accuracy of 0.94 for the positive class, indicating good performance.

In the cross-match between Gaia DR3 and LAMOST DR8, 5800 spectra were obtained from the OSFC, and measurements were successfully obtained for 5046 of them. By considering measurement errors, a probability greater than 0.90 was determined for 1331 spectra, indicating their likelihood of being T Tauri stars (TTS).

Finally, for the spectral classification of TTS, we employed EW measures on synthetic spectra ranging from spectral type K0 to K7 and M0 to M9, obtained from the stellar atmosphere code PHOENIX. Subsequently, by applying a Principal Components Analysis transformation to the measures, we implemented various multi-class models, among which a Gradient Boosting Machine model exhibited superior performance metrics in preliminary assessments. The aforementioned work is still in the development stage.

[1] http://www.lamost.org

References

Briceño, C., Calvet, N., Hernández, J., et al. 2019, AJ, 157, 85, doi: 10.3847/1538-3881/aaf79b

Hernández, J., Zamudio, L. F., Briceño, C., et al. 2023, AJ, 165, 205, doi: 10.3847/1538-3881/acc467

Wright, N. J. 2020, NewAR, 90, 101549, doi: 10.1016/j.newar.2020.101549

Zhang, Y., & Zhao, Y. 2015, Data Science Journal, 14, 11, doi: 10.5334/dsj-2015-011

“XVII LARIM”, Montevideo, Uruguay, November 27 - December 1, 2023.Volume Editors: Gonzalo Tancredi, Cecilia Mateu & Juan José Downes