Highlights of Spanish Astrophysics XII, Proceedings of the XVI Scientific Meeting of the Spanish Astronomical Society held on July 15 - 19, 2024, in Granada, Spain. M. Manteiga, F. González Galindo, A. Labiano Ortega, M. Martínez González, N. Rea, M. Romero Gómez, A. Ulla Miguel, G. Yepes, C. Rodríguez López, A. Gómez García and C. Dafonte (eds.), 2025

White Dwarf Spectral Analysis: Applying Unsupervised Machine Learning to the Gaia XP coefficients

Pérez-Couto, X.[1,2], Pallas-Quintela, L.[1,2], Manteiga, M.[3,2], Villaver, E.[4,5], ... and Dafonte, C.[1,2]

[1] Universidade da Coruña (UDC), Department of Computer Science and Information Technologies, Campus de Elviña s/n, 15071, A Coruña, Galiza, Spain
[2] CIGUS CITIC, Centre for Information and Communications Technologies Research, Universidade da Coruña, Campus de Elviña s/n, 15071 A Coruña, Galiza, Spain
[3] Universidade da Coruña (UDC), Department of Nautical Sciences and Marine Engineering, Paseo de Ronda 51, 15011, A Coruña, Galiza, Spain
[4] Instituto de Astrofísica de Canarias, 38200 La Laguna, Tenerife, Spain
[5] Universidad de La Laguna (ULL), Astrophysics Department, 38206 La Laguna, Tenerife, Spain

Abstract

Identifying new white dwarfs (WDs) heavy elements is crucial, as they serve as valuable tools for deducing the chemical characteristics of potential planetary systems accreting material onto their surfaces. To detect metallic WDs, we propose a methodology based on an unsupervised learning technique known as Self-Organizing Maps (SOM). This approach projects a high-dimensional dataset onto a two-dimensional grid, where similar elements are grouped into the same neuron.

Using this method, we uncovered 143 bona fide WD candidates in the Gaia space mission with several metallic lines in their spectra, including Ca, Mg, Na, Li, and K. The precision metrics achieved with our method are comparable to those of recent supervised techniques.

1 Introduction

White dwarfs (WD) are the degenerate stellar remnants of low-to-intermediate-mass stars ($\leq$ 8 solar masses) (Iben et al., 1997). Due to their high density ($\sim 10^3 kg/m^3$), WDs have fully stratified interiors, containing a degenerate core composed of C and O. This core is encased in a thin He mantle, which constitutes at most about 1% of the white dwarf's mass. Surrounding this He layer is an even thinner H envelope, which makes up no more than

approximately 0.01% of the mass.

More exciting are those WDs that show heavy metal lines in their atmospheres (mainly, Ca and Mg). In cool WDs (below approximately 25,000 K), those heavy elements tend to diffuse downward in the atmospheres due to gravitational settling in the presence of strong gravitational fields (Koester, 2009). Since the diffusion timescales due to gravitational settling are much shorter than the evolutionary time scales of WDs, those metals cannot be primordial; they must have been accreted, with the accretion of rocky material from planetesimals being the most widely accepted explanation (Farihi et al., 2010). For this reason, the detection of those polluted WDs is nowadays an effervescent field and a valuable tool to infer the presence and physical properties of exoplanets.

The Gaia space mission has provided us an unprecedented amount of positions, distances, and proper motions for 2000 million stars. Moreover, Gaia low resolution mean spectra (hereafter, XP spectra) is also provided for $\sim$ 220 million sources by using spectrophotometry (Carrasco et al., 2021). Since classifying such a large number of spectra by human visual inspection is not feasible, we aim in this work to use unsupervised machine learning to classify the Gaia catalog of 100,000 white dwarfs [5]. Instead of being in flux units per wavelength units, this spectra is expressed as a linear combination of 110 Hermite polynomials, so that each spectrum is defined as a vector of 110 coefficients.

2 Methods

In this work, we use a neural network-based dimensionality reduction algorithm called Self-Organizing Maps (SOM) (Kohonen, 1982) where, given a nonlinear high-dimensional dataset, the input data is projected on a 2D grid map where similar elements fall into the same neuron. Here, the similarity is defined by a metric (e.g. Euclidean distance) so the unsupervised learning process aims to maximize the similarity between objects belonging to the same neuron at the same time it minimizes the similarity between objects within different neurons. Therefore, topology is naturally preserved: similar neurons are also grouped next to each other. Consequently, SOMs combine the two major utilities of unsupervised learning: dimensionality reduction and clustering, unlike other algorithms that either perform only clustering (e.g. K-means) or only dimensionality reduction (e.g. t-SNE, UMAP).

3 Results

A SOM of 8x8 neurons was employed to classify the spectra of 66,337 WDs. To give labels to each neuron, the primary classes of WDs candidates that are also present in the Montreal White Dwarf Database (MWDD) (Dufour et al., 2016) are painted and used as trackers. The SOM is shown in Figure 1.

Once we have assigned a label to each neuron (and therefore to each WD falling in that neuron), we can compare the predicted class with the true class given by the MWDD by using a confusion matrix C such the one shown in Figure 2. The number in each cell $C_{i,j}$ shows the number of WDs with a true label i and a predicted class j. Immediately below is

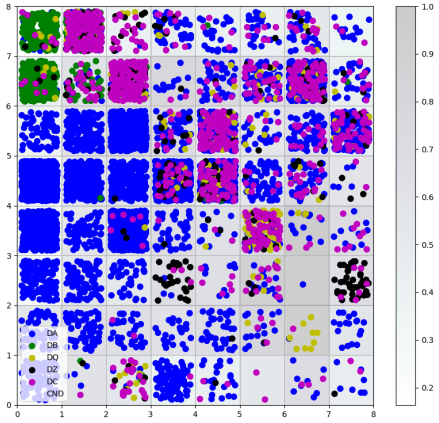

Figure 1: SOM map with our sample of 66, 337 clean sources. WDs with known spectral type in MWDD catalog appear in different colors, and candidates are invisible to enhance visualization.

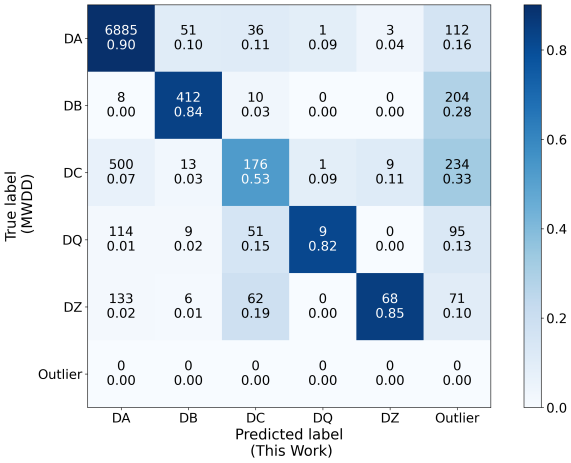

Figure 2: Confusion matrix of the SOM classification for WD primary spectral types with respect to MWDD classification.

shown the same quantity, but normalized over the predicted labels (columns) so the precision of the classification for each class appears in the diagonal. Thus, the confusion matrix would be diagonal for an ideal classification. Our confusion matrix shows excellent precision for DA and DZ classes ($\geq 85\%$), very good precision for DB and DQ classes ($\geq 80\%$), and poor metrics for DC class that is mainly confused with DAs and Outliers. In addition, Outlier neurons are mainly populated with a mixture of all classes but mainly with DCs and DBs. Recall is excellent for DAs but poor for the rest of the classes. This metrics are equal or better than those shown by recent supervised machine learning techniques such as Random Forest (García-Zamora et al., 2023) and Neural Networks (Vincent et al., 2024).

Regarding the polluted WDs, that are the main goal of this work, if we look at the two neurons in black, we can confirm, by drawing their median combined spectrum, the expected Ca II H&K absorption line at 393-396 nm. This is true either for the 68 confirmed WDs and for the 399 new polluted WD candidates we report here. Furthermore, more metals such as Mg, Na, Li, and K are also observed, as it is shown in Figure 2, where we plotted the normalized median spectra of the neuron (3, 2) (hereafter, DZA neuron, in the left side) and neuron (7,2) (hereafter, DZ neuron, in the right side). While the DZA neuron shows spectra with H Balmer lines, thus indicating a hydrogen-rich atmosphere, in the second one those H lines are hidden, arguably due to low temperatures. In addition to this, more metallic species are identified. Indeed, the mean temperature in these neurons are 9200 K (DAZ neuron) and 7200 K (DZ neuron). As a consequence, those metals had to be accreted.

As a result, we report the identification of 143 metal polluted WD candidates that have not been previously classified in the literature. Follow-up high resolution spectroscopy from

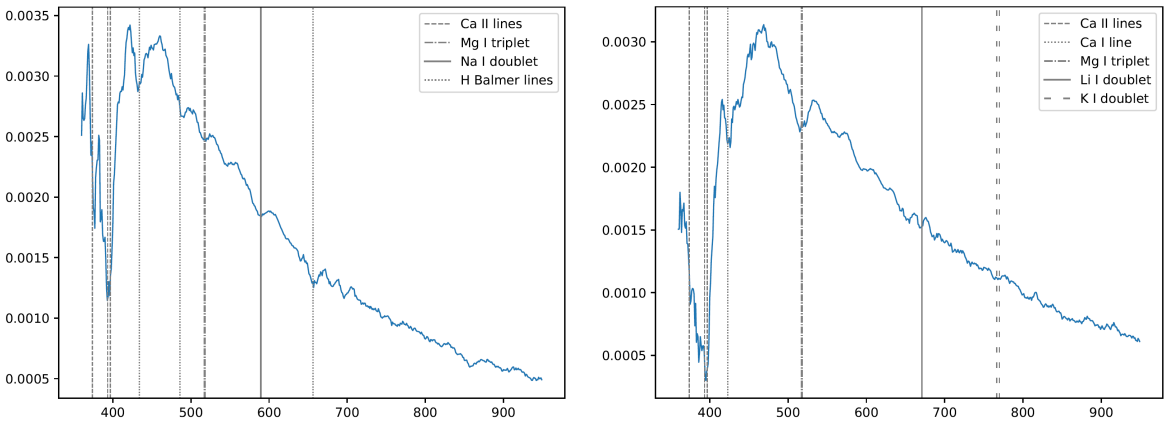

Figure 3: The normalized median spectra of the two neurons populated with polluted WDs: DZA neuron (left) and DZ neuron (right) are shown with the several metallic features identified with vertical lines.

Earth will be performed to study them.

4 Conclusions

We have demonstrated in this work the power of unsupervised learning in white dwarf spectral classification. Self-Organizing Maps have shown a similar performance that in recent supervised learning works, with high precision for DA, DB, DQ, and DZ white dwarfs. This strongly justifies the use of Self-Organizing Maps as they provide a natural and useful way to group similar spectra and, at the same time, to label new data, by performing statistics within each neuron.

This method allowed us to identify, with high confidence, 143 new polluted WD candidates that show spectral features of several metals (namely, Ca, Mg, Na, Li, and K), and even to distinguish between DZ and DZA subtypes. In order to confirm those candidates and to delve into other interesting neurons (such as DQ and DXZ subtypes) follow-up spectroscopy of the best candidates will be performed in the near future.

Acknowledgements

This work has made use of data from the European Space Agency (ESA) Gaia mission and processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement. This research was funded by the Horizon Europe [HORIZON-CL4-2023-SPACE-01-71] SPACIOUS project, Grant Agreement no. 101135205, the Spanish Ministry of Science MCIN / AEI / 10.13039 / 501100011033, and the European Union FEDER through the coordinated grant PID2021-122842OB-C22. We also acknowl-

edge support from the Xunta de Galicia and the European Union (FEDER Galicia 2021-2027 Program) Ref. ED431B 2024/21, ED431B 2024/02, and CITIC ED431G 2023/01. X.P. acknowledges financial support from the Spanish National Programme for the Promotion of Talent and its Employability grant PRE2022-104959 cofunded by the European Social Fund. Funding from Spanish Ministry project PID2021-127289NB-100 is also acknowledged.

References

Carrasco, J. M., Weiler, M., Jordi, C., et al. 2021, A&A, 652, A86

Dufour, P., Blouin, S., et al. 2016, arXiv:1610.00986 [astro-ph.SR]

Farihi, J., Barstow, M. A., Redfield, S., et al. 2010, MNRAS, 404, 2123

Garcia-Zamora E. M., Torres S., Rebassa-Mansergas A., 2023, A&A, 679, A127

Gentile Fusillo, N. P., Tremblay, P.-E., Cukanovaite, E., et al. 2021, MNRAS, 508, 3877. doi:10.1093/mnras/stab2672

Iben I. J., Ritossa C., Garcia-Berro E., 1997, ApJ, 489, 772

Koester, D. 2009, A&A, 498(2), 517-525.

Kohonen, T. 1982, Biol. Cybern. 43, 59-69

Vincent, O., Barstow, M. A., Jordan, S., et al. 2024, A&A, 682, A5