Exploring the Universe with Artificial Intelligence (UniversAI)
Proceedings IAU Symposium No. 397, 2026
I. Liodakis & V. Efthymiou, eds.
doi:10.1017/S1743921325104833

Comparison of Photometric and Spectroscopic Labels in Classifying Dusty Stellar Sources Using Machine Learning in the Magellanic Clouds

Sepideh Ghaziasgar[1], Mahdi Abdollahi[1], Atefeh Javadi[1], Jacco Th. van Loon[2], Iain McDonald[3], Joana Oliveira[2] and Habib G. Khosroshahi[1,4]

[1]School of Astronomy, Institute for Research in Fundamental Sciences (IPM), P.O. Box 195683-6613, Tehran, Iran
email: sepideh.ghaziasgar@ipm.ir

[2]Lennard-Jones Laboratories, Keele University, Keele, ST5 5BG, UK

[3]Jodrell Bank Centre for Astrophysics, Alan Turing Building, University of Manchester, Manchester, M13 9PL, UK

[4]Iranian National Observatory, Institute for Research in Fundamental Sciences (IPM), Tehran, Iran

Abstract. Dusty stellar sources, including young stellar objects (YSOs) and evolved stars such as oxygen- and carbon-rich AGBs (OAGBs, CAGBs), red supergiants (RSGs), and post-AGB stars (PAGBs), play a key role in the chemical enrichment of galaxies. Photometric surveys in the Magellanic Clouds have cataloged many such objects, but their classifications are often uncertain due to overlaps between populations. We trained machine learning models on spectroscopically labeled data from the SAGE project and applied them to photometric catalogs. The spectroscopic model achieves about 89% accuracy. Applied to photometric labels, nearly all OAGBs are correctly identified, and YSOs have a 95% confirmation rate. In contrast, 16% of CAGBs are reclassified as OAGBs, only 8% of RSGs retain their labels, and fewer than half of PAGBs are confirmed. Photometry is thus reliable for abundant populations with distinct signatures, but spectroscopic confirmation remains essential for rare or overlapping stellar classes.

Keywords. stars: classification, stars: AGB, RSG, and post-AGB, stars: YSOs, galaxies: spectral catalog, galaxies: photometric catalog, galaxies: Local Group, methods: machine learning

1. Introduction

Dusty stellar sources play a fundamental role in the life cycle of galaxies by returning heavy elements to the interstellar medium. These objects include young stellar objects (YSOs), which trace ongoing star formation, as well as evolved stars such as asymptotic giant branch (AGB) stars, red supergiants (RSGs), and post-AGB stars (PAGBs) (Abdollahi et al. 2023; Boyer et al. 2011, 2025; Mahani et al. 2025; Höfner & Olofsson 2018; Javadi et al. 2011, 2013, 2015).

Because of the high luminosities and red colors of dusty stellar sources, these populations are particularly well traced in the Magellanic Clouds (MCs), two nearby dwarf

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Astronomical Union.

https://doi.org/10.1017/S1743921325104833 Published online by Cambridge University Press

galaxies with low metallicities (Russell & Dopita 1992; Cioni 2009; Pietrzyński et al. 2013; Subramanian & Subramaniam 2011), making them ideal laboratories for studying dusty stellar evolution.

Traditionally, the identification of such objects has relied on photometric surveys using color-magnitude diagrams (CMDs) and infrared excesses. While photometry enables the construction of large catalogs at relatively low observational cost, it suffers from ambiguities: distinct stellar classes can occupy overlapping regions in photometric space. In contrast, spectroscopy provides a more secure diagnostic by revealing the chemical and physical properties of stellar atmospheres and circumstellar dust, but spectroscopic surveys are much more limited in size.

With the growing availability of both photometric and spectroscopic datasets from surveys such as the *Spitzer* SAGE and SAGE-Spec programs (Kemper et al. 2010; Woods et al. 2011; Ruffle et al. 2015; Jones et al. 2017), it has become possible to directly assess the reliability of photometric classifications by comparing them with spectroscopic benchmarks.

Unsupervised machine learning has also been applied in this context; for example, Pennock et al. (2022) used t-distributed Stochastic Neighbour Embedding (t-SNE) to identify sources dominated by AGN-dust emission in regions of the sky that include the Magellanic Clouds. In contrast, our study relies exclusively on supervised learning, training models on spectroscopically confirmed sources and then applying them to large photometric catalogs.

Machine learning provides a powerful means of conducting such comparisons (Baron 2019). In this study, we used supervised models trained on spectroscopic labels of dusty stellar populations in the MCs (Ghaziasgar et al. 2024, 2025a,b). The tested algorithms, as described in Ghaziasgar et al. (2025b), included Probabilistic Random Forest (PRF), Random Forest (RF), K-Nearest Neighbors (KNN), C-Support Vector Classification (SVC; polynomial and RBF kernels), and Gaussian Naive Bayes (GNB). Among these, PRF (Reis et al. 2019; Kinson et al. 2021, 2022; Pennock et al. 2025) provided the best performance, as it naturally accounts for uncertainties in both features and labels. By comparing the labels predicted by the model trained on spectroscopic labels and photometric labels, we aimed to quantify both the strengths and the limitations of photometric classification, especially for less numerous or overlapping stellar classes.

2. Data and Method

The spectroscopic dataset originates from the SAGE-Spec program and archival *Spitzer*/IRS observations, encompassing more than 600 dusty stellar sources in the LMC and SMC (Kemper et al. 2010; Woods et al. 2011; Ruffle et al. 2015; Jones et al. 2017). These sources were spectroscopically labeled into five classes: YSO, CAGB, OAGB, RSG, and PAGB.

For the photometric sample (Table 1), we compiled catalogs from major surveys of the LMC and SMC (Whitney et al. 2008; Srinivasan et al. 2009; Gruendl & Chu 2009; Riebel et al. 2010; Boyer et al. 2011; Kamath et al. 2014, 2015; Yang et al. 2018, 2019, 2020, 2021). Photometric classifications are generally based on positions in CMDs and infrared excess, but they may suffer from ambiguities where stellar classes overlap. We cross-matched the photometric catalogs with the spectroscopic set, removed duplicates, and applied the same preprocessing steps (feature selection, imputation, extinction, and distance corrections) as described in Ghaziasgar et al. (2025b).

To assess the reliability of the photometric classifications reported in those mentioned above, we apply models trained on spectroscopic labels presented by Ghaziasgar et al. (2024, 2025b). As discussed earlier, several models were tested and validated, and the probabilistic random forest (PRF) model was ultimately identified as the best performer.

https://doi.org/10.1017/S1743921325104833 Published online by Cambridge University Press

Table 1. Photometric classification of dusty stars in the Magellanic Clouds after cross-matching all catalogs and removing duplicates.

Stellar Type	Number
CAGB	8,537
OAGB	37,950
PAGB	45
RSG	6,132
YSO	1,642
Total	**54,306**

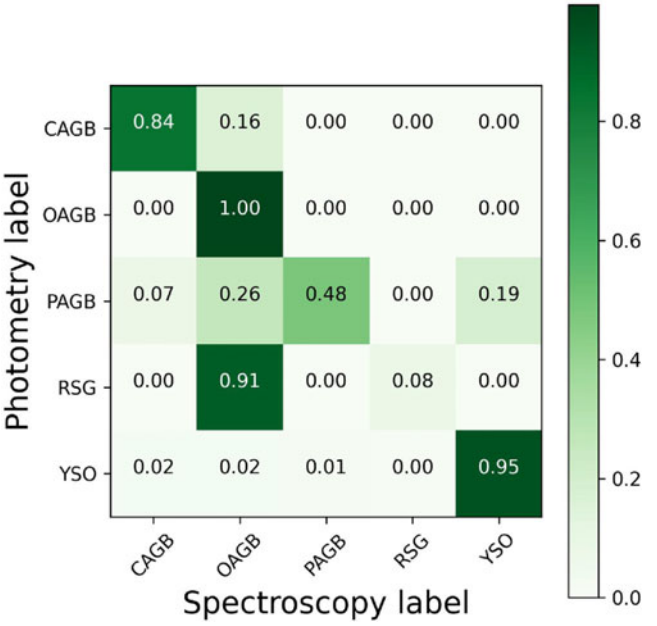


Figure 1. Confusion matrix comparing photometric and spectroscopic labels.

3. Comparison Results

The comparison between the predicted data and the photometric labels reveals that oxygen-rich AGB stars are almost perfectly recovered, and young stellar objects show a high confirmation rate of about 95%. In contrast, roughly 16% of carbon-rich AGB stars are reassigned as OAGBs, reflecting overlaps in CMD space. Red supergiants are particularly problematic: only about 8% of photometric RSG labels are confirmed, with most misclassified as OAGBs. Post-AGB stars show the highest variability, with fewer than half confirmed. A more detailed view of these misclassifications, based on confusion matrices for each class, is presented in Figure 1.

Overall, photometric classification is reliable for abundant populations with distinctive infrared signatures, such as OAGBs and YSOs; however, it is less effective for rare or overlapping stellar classes. These results underscore the importance of spectroscopic confirmation in such cases, as well as the value of spectroscopically trained machine learning models in refining large photometric catalogs.

References

Abdollahi, H., Javadi, A., Mirtorabi, M. T., et al. 2023, ApJ, 948, 63
Baron, D. 2019, arXiv:1904.07248
Boyer, M. L., Srinivasan, S., van Loon, J. T., et al. 2011, AJ, 142, 103

https://doi.org/10.1017/S1743921325104833 Published online by Cambridge University Press

Boyer, M. L., Sloan, G. C., Nanni, A., et al. 2025, arXiv:2507.16766
Cioni, M.-R. L. 2009, A&A, 496, 399
Ghaziasgar, S., Abdollahi, M., Javadi, A., et al. 2024, Communications of BAO, 71, 377
Ghaziasgar, S., Masoudnezhad, A., Javadi, A., et al. 2025, Proc. IAU Symp., 368, 67
Ghaziasgar, S., Abdollahi, M., Javadi, A., et al. 2025, ApJ, 986, 168
Gordon, K. D., Meixner, M., Meade, M. R., et al. 2011, AJ, 142, 102
Gruendl, R. A., & Chu, Y.-H. 2009, ApJS, 184, 172
Höfner, S., & Olofsson, H. 2018, A&ARv, 26, 1
Javadi, A., van Loon, J. Th., & Mirtorabi, M. T. 2011, MNRAS, 411, 263
Javadi, A., van Loon, J. Th., Khosroshahi, H., et al. 2013, MNRAS, 432, 2824
Javadi, A., Saberi, M., van Loon, J. Th., et al. 2015, MNRAS, 447, 3973
Jones, O. C., Woods, P. M., Kemper, F., et al. 2017, MNRAS, 470, 3250
Kamath, D., Wood, P. R., & Van Winckel, H. 2014, MNRAS, 439, 2211
Kamath, D., Wood, P. R., & Van Winckel, H. 2015, MNRAS, 454, 1468
Kemper, F., Woods, P. M., Antoniou, V., et al. 2010, PASP, 122, 683
Kinson, D. A., Oliveira, J. M., & van Loon, J. Th. 2021, MNRAS, 507, 5106
Kinson, D. A., Oliveira, J. M., & van Loon, J. Th. 2022, MNRAS, 517, 140
Mahani, H., Javadi, A., van Loon, J., et al. 2025, arXiv:2508.05596
Meixner, M., Gordon, K. D., Indebetouw, R., et al. 2006, AJ, 132, 2268
Pennock, C. M., van Loon, J. Th., Anih, J. O., et al. 2022, MNRAS, 515, 6046
Pennock, C. M., van Loon, J. Th., Cioni, M.-R. L., et al. 2025, MNRAS, 537, 1028
Pietrzyński, G., Graczyk, D., Gieren, W., et al. 2013, Nature, 495, 76
Reis, I., Baron, D., & Shahaf, S. 2019, AJ, 157, 16
Riebel, D., Meixner, M., Fraser, O., et al. 2010, ApJ, 723, 1195
Ruffle, P. M. E., Kemper, F., Jones, O. C., et al. 2015, MNRAS, 451, 3504
Russell, S. C., & Dopita, M. A. 1992, ApJ, 384, 508
Srinivasan, S., et al. 2009, AJ, 137, 4810
Subramanian, S., & Subramaniam, A. 2011, ASInC, 3, 144
Whitney, B. A., Sewilo, M., Indebetouw, R., et al. 2008, AJ, 136, 18
Woods, P. M., Oliveira, J. M., Kemper, F., et al. 2011, MNRAS, 411, 1597
Yang, M., Bonanos, A. Z., Jiang, B.-W., et al. 2018, A&A, 616, A175
Yang, M., Bonanos, A. Z., Jiang, B.-W., et al. 2019, A&A, 629, A91
Yang, M., Bonanos, A. Z., Jiang, B.-W., et al. 2020, A&A, 639, A116
Yang, M., Bonanos, A. Z., Jiang, B., et al. 2021, A&A, 646, A141

https://doi.org/10.1017/S1743921325104833 Published online by Cambridge University Press