

Strategies for the creation of the VVV near-IR Galaxy catalogue in a northern part of the galactic disc

I. V. Daza-Perilla^{1,2}, M. A. Sgró^{1,3}, L. D. Baravalle^{1,3}, M. V. Alonso^{1,3}, C. Villalon¹, M. Lares^{1,3}, M. Soto⁶, J. L. Nilo Castellón^{4,5}, C. Valotto^{1,3}, P. Marchant Cortés^{4,5}, D. Minniti^{7,8,9} and M. Hempel^{7,10}

¹Instituto de Astronomía Teórica y Experimental, CONICET-UNC, Córdoba, X5000BGR, Argentina.

²Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba (UNC), Córdoba, CP:X5000HUA, Argentina.

³Observatorio Astronómico de Córdoba, Universidad Nacional de Córdoba, Laprida 854, Córdoba, X5000BGR, Argentina.

⁴Instituto Multidisciplinario en Investigación y Postgrado (IMIP), Universidad de La Serena. Av. Raúl Bitrán Nachary No 1305, La Serena, 1720236, Chile.

⁵Departamento de Astronomía, Universidad de La Serena. Av. Juan Cisternas 1200, La Serena, 1720236, Chile.

⁶Instituto de Astronomía y Ciencias Planetarias, Universidad de Atacama, Copayapu 485, Copiapó, 1532297, Chile.

⁷Instituto de Astrofísica, Facultad de Ciencias Exactas, Universidad Andrés Bello, Av. Fernández Concha 700, Las Condes, Santiago, 7550000, Chile.

⁸Vatican Observatory, V00120 Vatican City State, Italy.

⁹Departamento de Física, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, 88040 970, Brasil.

¹⁰Max Planck Institute for Astronomy, Königstuhl 17, 69117 Heidelberg, Germany.

Keywords: *galaxies: statistics, infrared: galaxies*

We presented the methodology for the automatic classification of galaxies and non-galaxies in low Galactic latitude regions using both images and photometric/morphological near-IR data from the VVVX survey. Using the VVV NIRGC (Baravalle et al., 2021), we analyzed the most relevant features for galaxy identification through classical and un-supervised statistical methods. This catalogue was also used to generate datasets of extragalactic candidates from a pipeline (Baravalle et al., 2018) and to train a Convolutional Neural Network (CNN) with image data and an Extreme Gradient Boosting (XGBoost) architecture with photometric and morphological data. This allows us to derive probability catalogues used to analyze completeness and purity as a function of configuration parameters and to explore the best combinations of the models. The resulting classifier achieves an F1-score of 0.67, a purity of 65 percent, and a completeness of 69 percent. As a test case, we applied this methodology to the VVVX survey in part of the Northern disc region, generating a dataset of 172,396 extragalactic candidates with probabilities of being galaxies. We studied the impact of implementing supervised models trained with data from different regions of the Galactic disc. The methodology chosen to create a catalogue of galaxies automatically is subject to a decision threshold of 0.6 and the combination of using imaging and photometric/morphological information. We compared

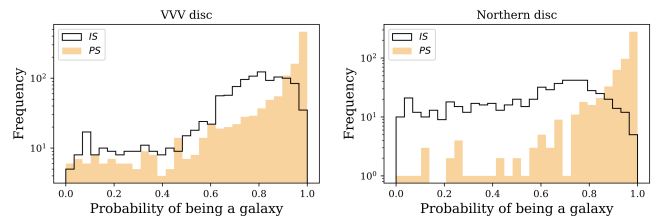


Figure 1. Probability distribution of galaxies. The probabilities of being a galaxy for the CNN model trained with IS are shown in black, and in orange the distribution of probabilities of being a galaxy for the XGBoost model trained with PS. The y-axis is shown in logarithmic scale. Upper panel: Galaxies belonging to VVV. Lower panel: Galaxies belonging to the Northern disc (Figure 13 in Daza-Perilla et al. (2023)).

the precision metrics of galaxies obtained in the VVV disc and Northern disc. We took two sets of 1,682 candidates to be galaxies that had probabilities greater than or equal to 0.6 for either model (CNN and XGBoost). The probability distribution of galaxies for each model in the VVV disc and Northern disc regions are shown in Figure 1. We found that the models tend to reproduce the identification characteristics in regions with visual inspection. These results confirm the estimated purity in the galaxy catalogues generated from the methodology detailed in Daza-Perilla et al. (2023). Currently, we are

performing transfer learning to improve identification in low Galactic latitude regions and to generate a new catalogue of galaxies across the entire Galactic disc.

■ References

- Baravalle, L. D., Alonso, M. V., Nilo Castellón, J. L., Beamín, J. C., & Minniti, D. 2018, *AJ*, 155, 46, doi: [10.3847/1538-3881/aa9ee2](https://doi.org/10.3847/1538-3881/aa9ee2)
- Baravalle, L. D., Alonso, M. V., Minniti, D., et al. 2021, *MNRAS*, 502, 601, doi: [10.1093/mnras/staa4020](https://doi.org/10.1093/mnras/staa4020)
- Daza-Perilla, I. V., Sgró, M. A., Baravalle, L. D., et al. 2023, *MNRAS*, 524, 678, doi: [10.1093/mnras/stad1767](https://doi.org/10.1093/mnras/stad1767)