Using Machine-Learning to obtain Be star candidates

Yael Aidelman[1,2,*] & Carlos Escudero[2,]**

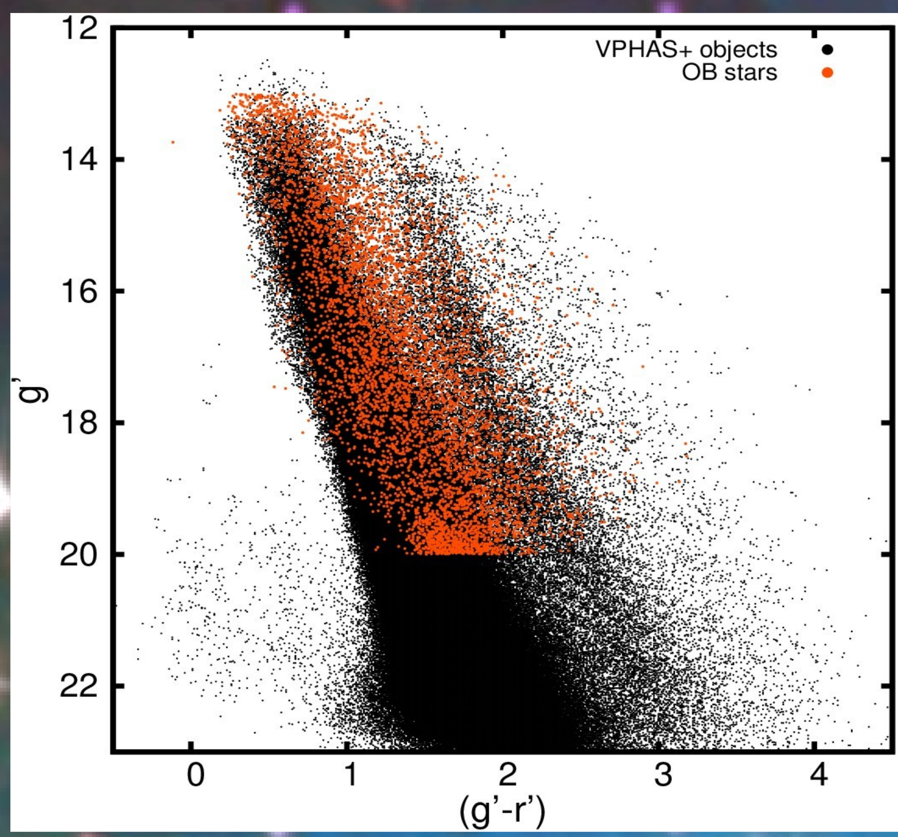
1- Facultad de Ciencias Astronómicas y Geofísicas de La Plata, (FCAGLP-UNLP)
2- Instituto de Astrofísica de La Plata (IALP-CONICET)
*- Grupo Modelos de Estrellas Peculiares (MEP)
**- Grupo Sistemas Estelares Extragalácticos y su Contexto Cosmológico (SEECC)

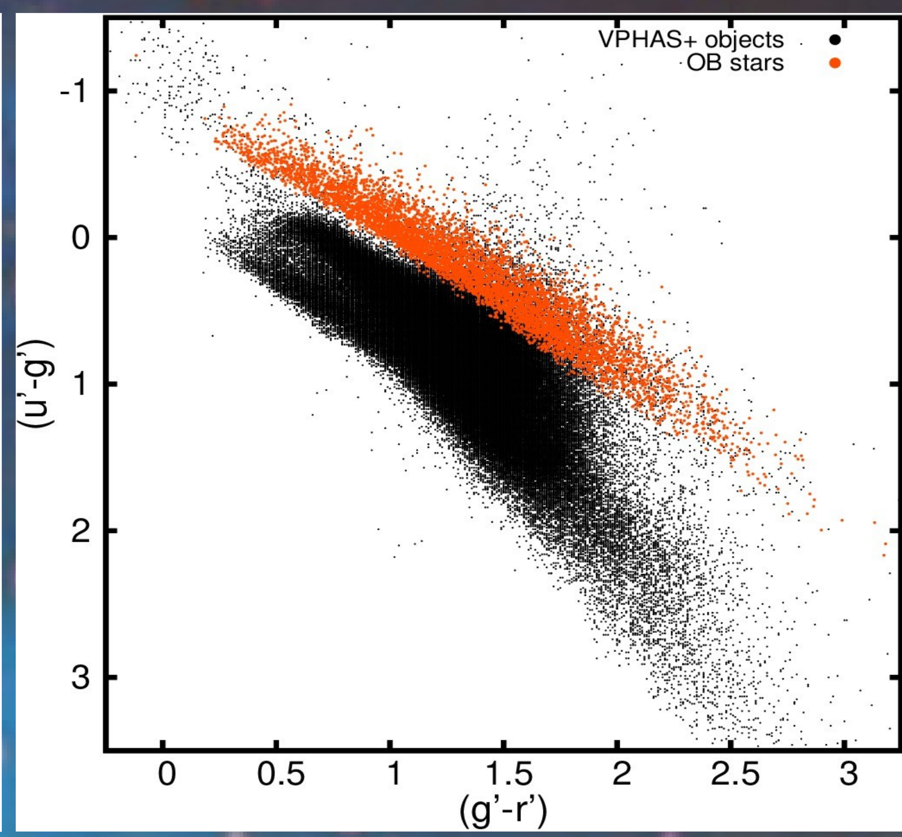
INTRODUCTION

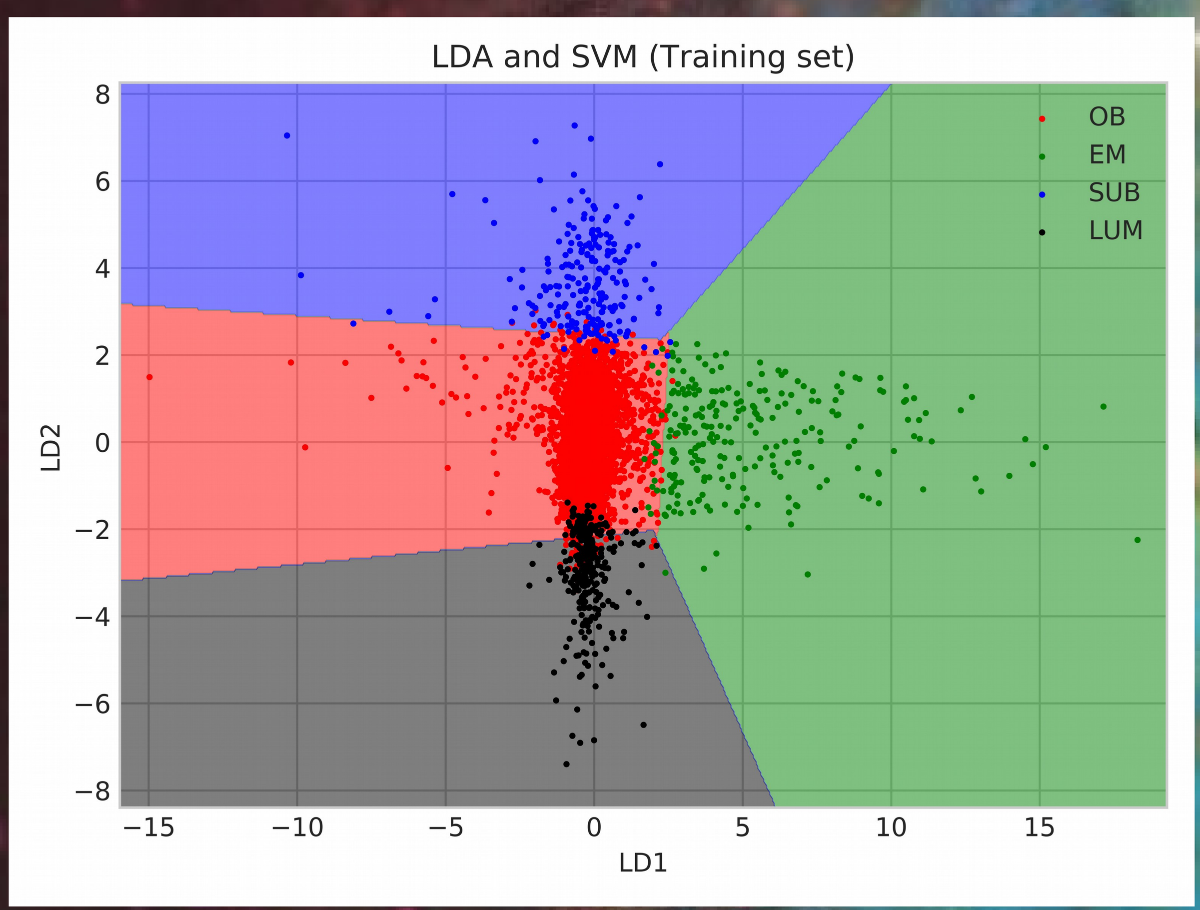
Classical B-type emission line (Be) stars are non-supergiant stars whose spectra exhibit or have ever exhibited Hα or more Balmer lines in emission (Jaschek et al. 1981). The properties observed in the Hα line profile of Be stars are consistent with the presence of an optically-thin gaseous circumstellar disk (flattened in the plane perpendicular to the axis of rotation) in which the gas is ionized by ultraviolet radiation from the central object. Although the definition of Be star is based on its spectroscopic properties, it is possible to detect them photometrically as long as the emission in the Hα line is very intense. One technique commonly used to identify classical Be stars candidates is the use of photometric 2-colour diagrams (2-CDs) that utilize a narrow-band filter centered on Hα and an associated filter that samples the nearby continuum region (e.g., Raddi et al. 2015), for example (r – Hα) versus (r – i).

TRAINING SET

To disentangle OB-type stars in four groups: emission line stars (EM), sub- and over-luminous stars (SUB and LUM, respectively), and normal stars (OB), we use supervised machine learning algorithms. For this aim, initially we use the published catalogue of Mohr-Smith et al. (2017) on the Carina Arm region (282° ≤ l ≤ 293°). These authors used combined data from the VST Photometric Hα Survey (VPHAS+; Drew et al. 2014) in $u,g,r,H\alpha,i$ and 2MASS J,H,K magnitudes. Using the spectral energy distribution fittings in these photometric data, subsequently confirming with spectroscopy of hundreds of objects, Mohr-Smith et al. grouped the sample of objects in the four groups mentioned above. The figures show the colour-magnitude and colour-colour diagrams of the objects detected in a region of Carina's Arm (black points) along with the OB stars studied by Mohr-Smith et al (orange points). For more detail see Mohr-Smith et al. (2015,2017).





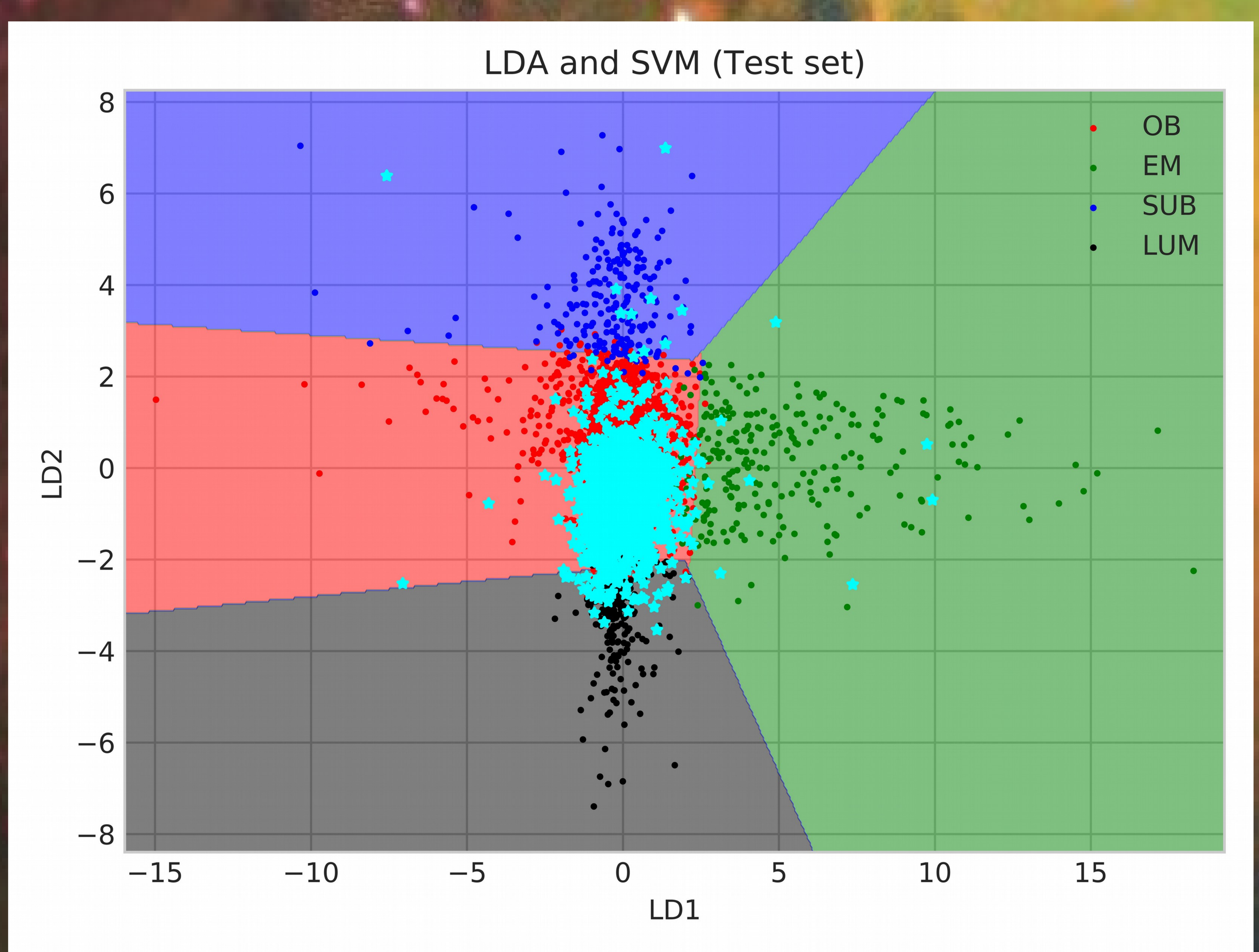


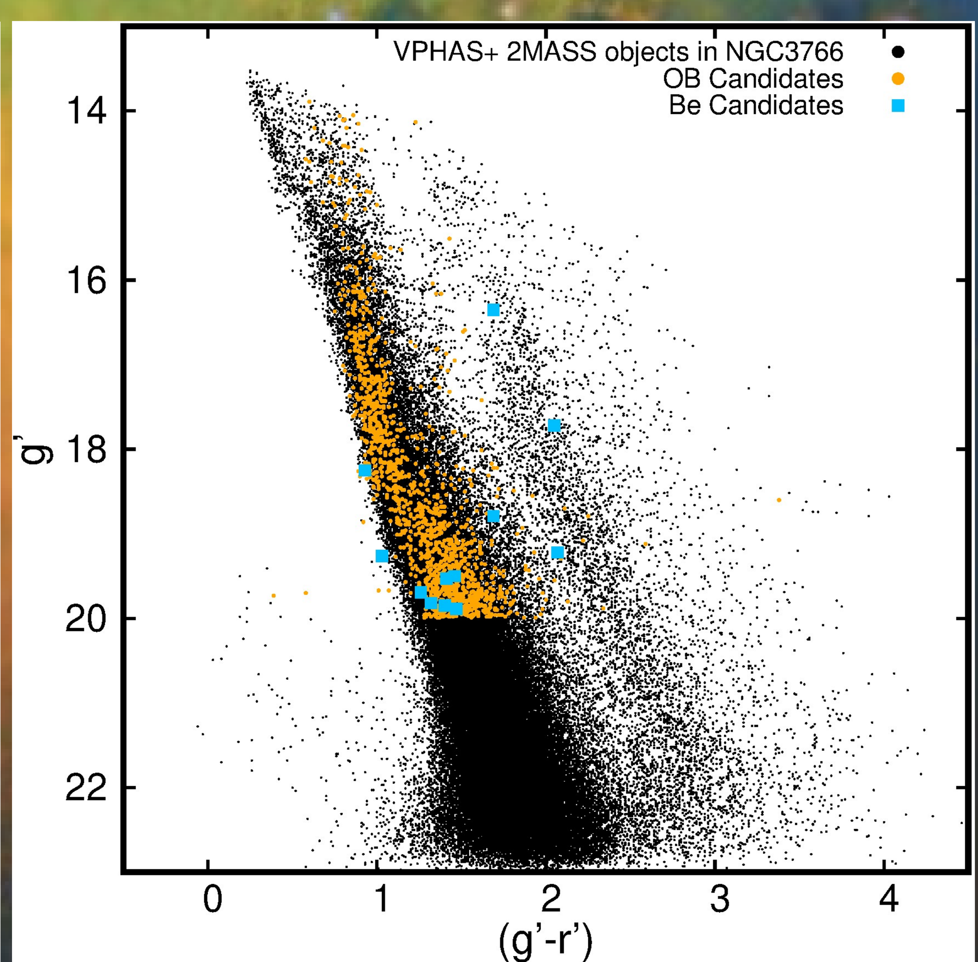
TRAINED ALGORITHM

In this work, we use the Linear Discriminant Analysis (LDA) technique together with Support Vector Machine (SVM) algorithm on the dataset of Mohr-Smith et al. (2017), in order to achieve a better separation between these different groups of stars. LDA is most commonly used as dimensionality reduction technique in the pre-processing step for pattern-classification and machine learning applications. The goal of this method is to project a feature space (a dataset n-dimensional samples) onto a smaller subspace k (where k≤n−1) while maintaining the class-discriminatory information. In general, dimensionality reduction does not only help to reduce computational costs for a given classification task, but it can also be helpful to avoid overfitting by minimizing the error in parameter estimation. On the other hand, SVM is a supervised Machine-Learning algorithm capable of performing classification, regression and even outlier detection. The goal of SVM is to find the best splitting boundary between data. Figure (on the left) shows the two LDA components obtained with the four delimited regions corresponding to each group of OB stars in the Mohr-Smith et al. catalogue.

TEST SET

Subsequently, we use this trained algorithm on a new VPHAS+ and 2MASS dataset corresponding to the cluster NGC 3766 (α_{J2000}=11h36m14s; δ_{J2000}=-61°36'30'') close to the region studied by Mohr-Smith et al (2017). Previously, we selected the candidates for OB stars of this region following the guidelines of Mohr-Smith et al. (2015): we define a limit magnitude of g'=20 mag, and we also use the reddening curve in the colour plane (u' – g') vs (g' – r'). On this sample we apply the classification algorithm. The Figures show the location of these objects in the LDA component diagram (cyan stars) as well as their location in the colour-magnitude diagram (orange points), respectively. The cyan squares in the colour-magnitude diagram indicate those OB star candidates that could present emission in Hα, i.e. candidates for Be stars.





RESULTS

Using VPHAS+ $u,g,r,H\alpha,i$ photometry and Machine Learning algorithms, we catalogue 1370 stars around the NGC3766 cluster region (r < 17 arcmin) with magnitudes between 13.6 < g' < 20 mag. From this sample, 1285 objects were catalogued as normal OB stars, 9 as subluminous, 64 as superluminous, and 12 candidates as Be stars. The goal of this work was to use the potential of broadband filters together with the narrowband Hα filter offered by VPHAS+ in order to obtain Be star candidates that can then be observed and confirmed spectroscopically. In this context, the Southern Photometric Local Universe Survey (S-PLUS; Mendes de Oliveira et al. 2019) with its 12 optical bands will allow obtaining a greater number of combinations of filters and therefore a possible better separation between the groups studied here.

REFERENCES

Drew et al. 2014, MNRAS, 440, 2036
Jaschek et al. 1981, Be star terminology., Be Star Newsletter
Mendes de Oliveira et al. 2019, MNRAS, 489, 241
Mohr-Smith et al. 2015, MNRAS, 450, 3855
Mohr-Smith et al. 2017, MNRAS, 465, 1807
Raddi et al. 2015, MNRAS, 446, 274