Search for The Densest Galaxies In The Universe: Traditional Astronomy vs. Supervised Machine Learning

Teymoor Saifollahi[1]*, Reynier Peletier[1], Edwin Valentijn[1], Johan Knapen[2], FDS collaboration, SUNDIAL network **

[1]Kapteyn Astronomical Institute, University of Groningen (RUG), the Netherlands

[2]Instituto de Astrofísica de Canarias (IAC), Tenerife, Spain

*saifollahi@astro.rug.nl

** We acknowledge financial support from the European Union's Horizon 2020 research and innovation programme under Marie Sklodowska-Curie grant agreement No 721463 to the SUNDIAL ITN network.

In the late 90s and based on spectroscopic surveys of the Fornax galaxy cluster, Drinkwater et al. 1999 and Hilker et al. 1999 reported the detection of very compact galaxies at the redshift of the cluster. Since then, many studies have been carried out to understand the origin of the so called Ultra-Compact dwarf galaxies (UCD), their formation and evolution. As these studies shows that UCDs are larger, brighter and more massive than the biggest Milky Way star clusters and much more compact than typical dwarf galaxies of comparable magnitude (Misgeld and Hilker, 2011). Furthermore, they harbour old stellar populations (Frank et al. 2011, Francis et al. 2012) and have large mass-to-light ratios (Mieske et al 2008). As the result of all the research in the last two decades, two formation scenarios are favoured:

In the first scenario, UCDs are the remnant nuclei of tidally disrupted galaxies (Bekki et al. 2003). In this scenario, because of the tidal interactions, galaxies lost most of their stars, except of in the center of the galaxy where the gravitational potential is strong enough to keep the stars. In this case, the central black hole of the galaxy remains the same which leads to higher mass-to-light ratios (Seth et al. 2014, Afanasiev et al. 2018, Voggel et al. 2018). In the second scenario, UCDs are the result of mergers of star clusters. (Fellhauer and Kroupa et al. 2002). This scenario does not explain the higher mass-to-light ratios.

The uncertainties in our knowledge about UCDs arise from the limited (and biased) sample of confirmed UCDs since they are generally unresolved from the ground and therefore hard to find. Most of the known UCDs were found through spectroscopic surveys around massive galaxies or cores of galaxy clusters/groups. In the Fornax cluster, all the spectroscopically confirmed UCDs are located in the inner 300 kpc (half the virial radius) of the cluster, mostly around NGC1399, the bright cD galaxy in the center of the cluster.

Because of the small sizes of UCDs, photometric identification of these objects at the distance of the Fornax cluster is challenging. In ground-based images, UCDs appear as point-sources which makes them indistinguishable from foreground stars (Milkyway stars) or distant background galaxies. However, they can be detected in color-color diagrams, since for a given optical color, they are redder in optical-infrared colors. This work combines deep multi-wavelength observations in u,g,r,i from Fornax Deep Survey which we recently published (Venhola et al. 2018 and 2019) with J and Ks archival data (VISTA/VIRCAM ESO Saifollahi et al. in prep.) to identify the population of Ultra-Compact Dwarf Galaxies (UCDs). We followed two different methods: A tradiotnal method using a Color-Color Daigram (Muñoz et al. 2014, Powalka et al. 2017, Liu et al. 2018) and using Supervised Machine Learning. As the result, we identified ~900 UCD candidates within the virial radius of the Fornax cluster.

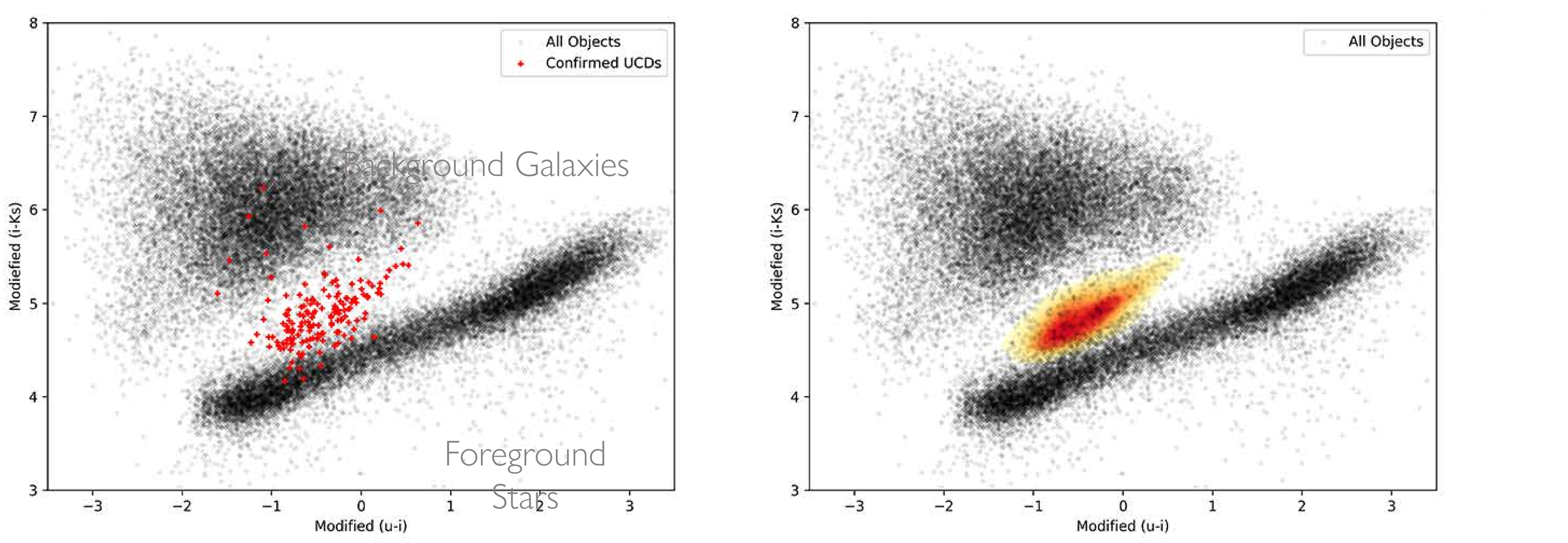

Figure-1: Modified color-color diagram of all the detected stellar-like objects in our dataset (Saifollahi et al. in prep). From this figure, it is seen that by combining "u - i" and "i - Ks", we can separate foreground stars and background galaxies (left). In this diagram, the red points in the left figure indicate confirmed UCDs from the literature and define their own sequence (right).

TRADITIONAL ASTRONOMY

In the absence of spectroscopic data, the **optical/near-infrared** color-color diagram is a powerful tool to identify UCDs and GCs (Munoz et al. 2014, Powalka et al. 2017, Liu et al. 2018). In figure-1, the **modified** u-i/i-Ks color-color diagram of all the small objects is shown. Instead of normal colors, modified colors are used to intensify the separation between compact objects (foreground stars and UCDs) and extended sources (background galaxies). The definition of modified colors is as follows:

$$\text{modiefied} (u - i) = (u - i) + \exp (1 - \text{compactness})$$
$$\text{modiefied} (i - Ks) = (i - Ks) + \exp (1 - \text{compactness})$$

The compactness is defined as the difference of aperture magnitudes measured by 2" and 4" apertures. As is seen from figure-1 (left), the optical/near-infrared color combination separates foreground stars, UCDs (red points) and background galaxies. Using the modiefied color-color diagram, we defined the **sequence of UCDs** (figure-1 right) and selected all the objects on the sequence as UCD candidates.

SUPERVISED MACHINE LEARNING

To improve the selection, we applied **KNN** (K-Nearest Neighbour) to the 21-dimensional parameter space that includes 15 colors, angular sizes and compactness (in g, r and i). For each unknown object in the parameter space, KNN looks at the K nearest known data points or **neighbours** and based on their class (foreground-star, UCD/GC or background galaxy, it classifies the unknown object.

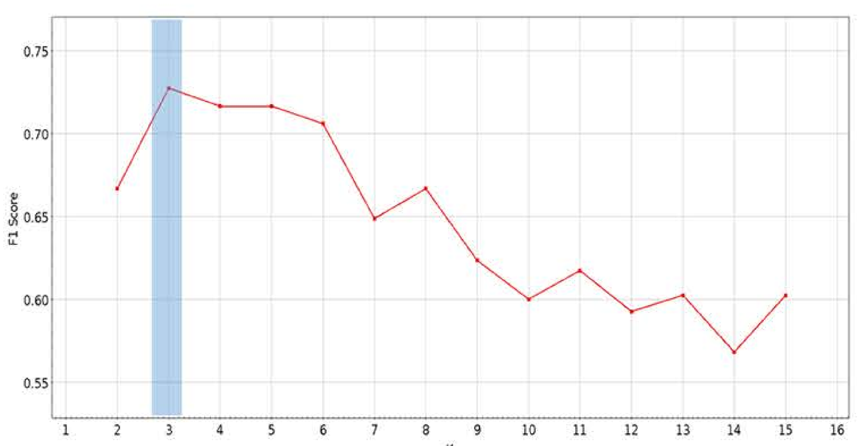

Figure-2: KNN was applied for different values of K from 2 to 15 and F1 score of each run was measured to find the best value for K.

For the training/validation/test set, we used the catalogues of kinematically confirmed UCDs/GCs in the Fornax cluster (Wittmann et al. 2016) and spectroscopic surveys of the Fornax cluster (Maddox et al. 2019). In total we selected about 4000 foreground stars, 160 UCDs and 2000 background galaxies. We used 6/8 of the whole sample as the training (reference) set, 1/8 as the validation set and the other 1/8 as the test set. In the following and before classification using KNN, all the parameters were **normalized.** Futhurmore, to take into acount the imbalance in the training set, KNN were applied 100 times with **randomly** selected training set and the final decision was made based on the results of all the runs. The training was done for different values of K from 2 to 15 and F1 scores were measured. This experiment led to the K=3 as the best value (figure-2).

After fidning the best value of K, we classified our objects 100 times with a randomly selected training/validation set. We measured the number of runs that an object is classified as a UCD (**UCD_SCORE**), foreground star (STAR_SCORE) or background galaxy (GAL_SCORE) and selected objects with UCD_SCORE higher than their STAR_SCORE and GAL_SCORE as UCD candidates. As the result, we identified **~900 UCD candidates** with recall of > 0.9 and precision of > 0.6. Itmeans that this mothed was able to recover 90% of UCDs (10% false negatives) and 60% true-positive (40% false positives) among all the identified UCD candidates.

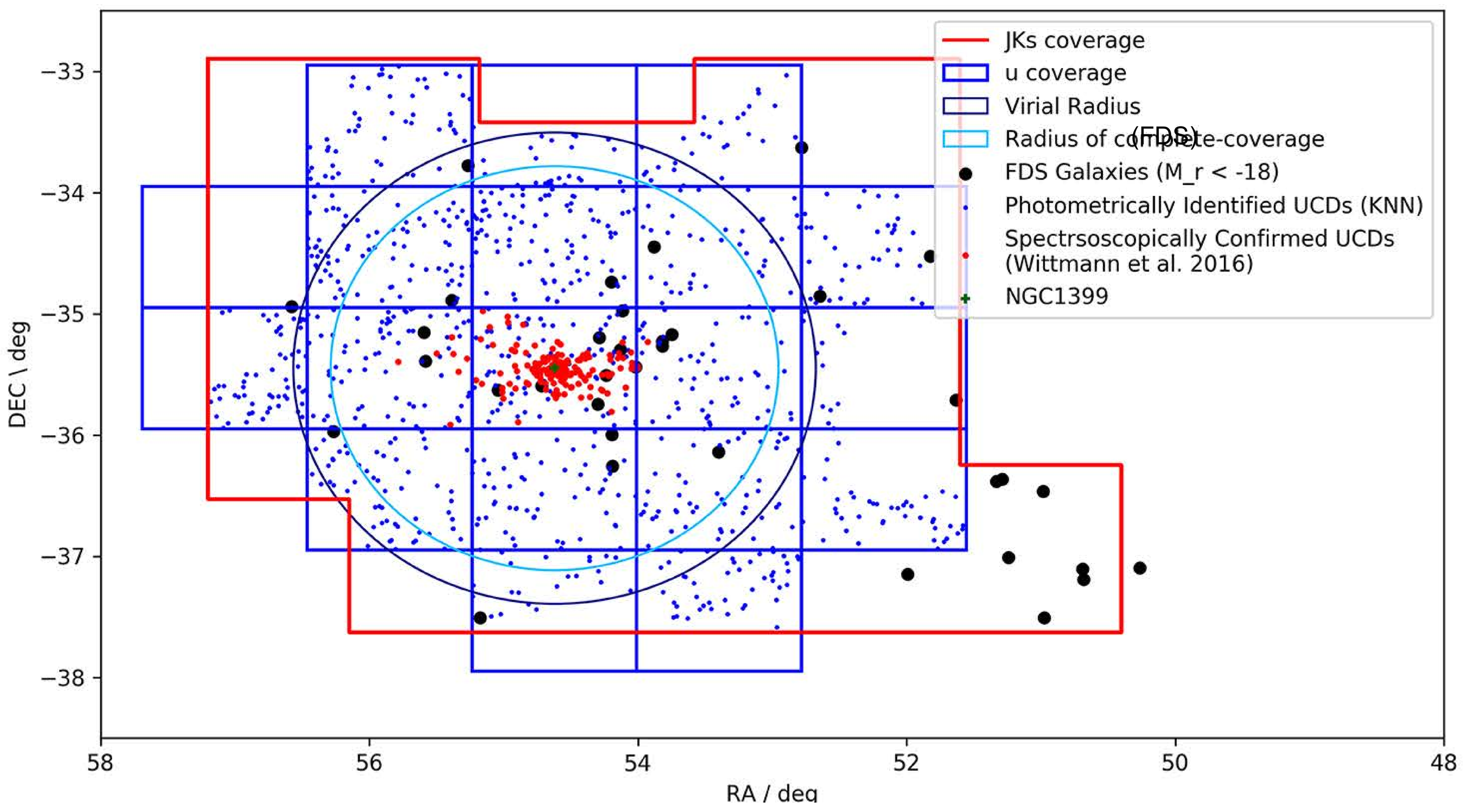

Figure-3: The spatial distribution of the identified UCD candidates in the Fornax Cluster (This work, Saifollahi et al. in prep)

RESULTS

In this work we used the deep multi-wavelength observations to identify Ultra Ccompact Dwarf Galaxies (UCDs) in the Fornax galaxy cluster. We started with a traditional astronomical methods (color-color diagram) and continued with a KNN (supervised machine learning) classifier. This led to a catalogue of about 900 new UCD candidates in the Fornax cluster within approximately its virial radius (figure-3) This catalogue will be used for **follow-up spectroscopy** in future.

Our analysis on the test set shows that objects with **UCD_SCORE=100** are UCD (probability of **100%**) and objects with **UCD_SCORE=99** are very likely to be UCD (probability of **90%**). Interestingly, all the best UCD candidates which overlap with ACS observations (Jordan et al. 2015) were categorized as UCDs/GCs (figure-4) Therefore, we selected the best UCD candidates (52 candidates with UCD_SCORE=100 and 116 Candidates with UCD_SOCRE=99) and studied their spatial distribution in detail. Many of the best UCD candidates have been found **outside of the Fornax cluster core** which implies that UCDs are not necessarily concentrated around big and massive galaxies. Furthurmore, we have detected 15 of the best cnaidates in **GALEX-NUV** observations.

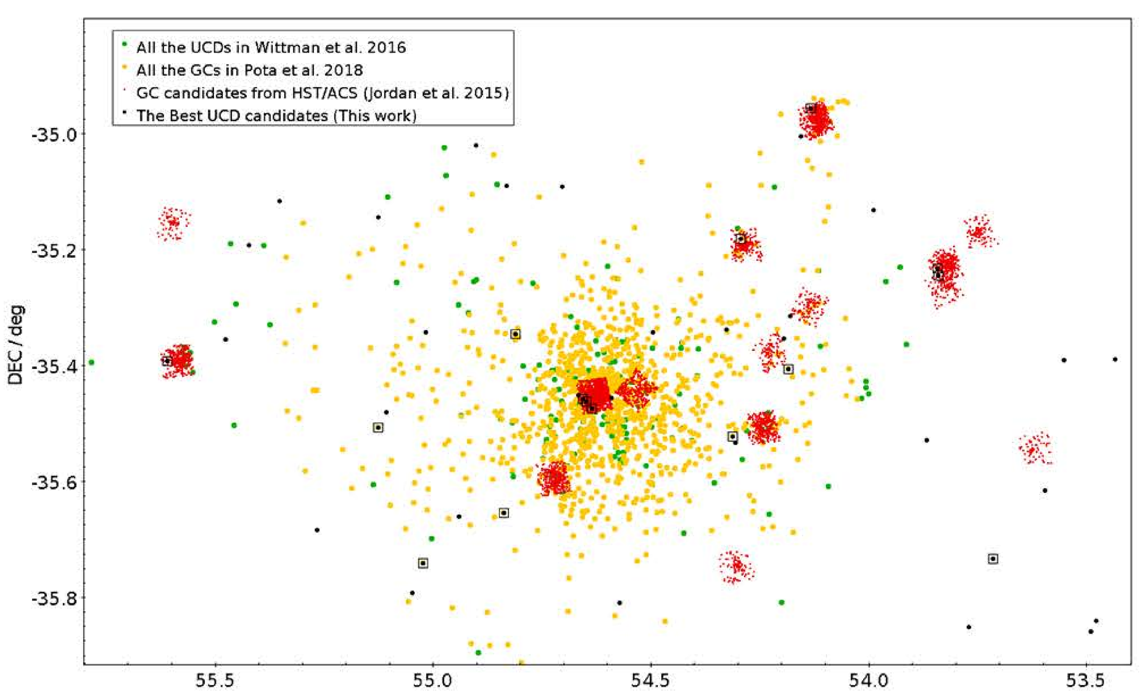

Figure-4: Distribution of the known UCDs/GCs in the cluster from previous works (red, yellow and green dots) and the best UCD candidates from this work (black dots). When compare the UCD candidates with literature, in the core of the cluster where most of the observations were done, we found that most of the UCD candidates are categorized as UCDs using spectroscopic surveys (6 objects) or HST/ACS observations (10 objects). All the candidates which overlap with HST/ACS Observations (Jordan et al. 2015) are categorized as UCDs/GCs. For objects with spectroscopic data, 5 objects out of 6 are UCDs/GCs at redshift of the cluster and 1 object is background galaxy.

REFRENCES

Afanasiev, A. V., Chilingarian, I. V., Mieske, S., et al. 2018, MNRAS, 477, 4856
Bekki, K., Couch, W. J., Drinkwater, M. J., & Shioya, Y. 2003, MNRAS, 344, 399
Drinkwater, M. J., Jones, J. B., Gregg, M. D., & Phillipps, S. 2000, PASA, 17, 227
Fellhauer, M. & Kroupa, P. 2002, MNRAS, 330, 642
Francis K. J., Drinkwater M. J., Chilingarian I. V., 2012, MNRAS, 425, 325
Frank M. J., Hilker M., Mieske S. et al., 2011, MNRAS, 414, L70
Hilker, M., Infante, L., Vieira, G., Kissler-Patig, M., 1999b, A&AS, 134, 75
Misgeld, I. & Hilker, M. 2011, MNRAS, 414, 3699
Pota V., et al., 2018, MNRAS, 481, 1744
Seth, A. C., van den Bosch, R., Mieske, S., et al. 2014, Nature, 513, 398
Venhola A., et al., 2018, A&A, 620, A165
Venhola A., et al., 2019, A&A, 625, A143
Voggel, K. T., Seth, A. C., Neumayer, N., et al. 2018, ApJ, 858, 20
Wittmann, C., Lisker, T., Pasquali, A., Hilker, M., et al. 2016, MNRAS, 459, 4450