COMPARATIVE STUDY OF UNSUPERVISED MACHINE LEARNING METHODS FOR MEMBERSHIP ASSOCIATION IN CLUSTERS USING GAIA DR3 DATA

Anish Kalsi[1*], Dhruv Tyagi[2], Harshit Choudhary[3], Ajeet Kumar[4]

Abstract

In observational astronomy, one of the primary challenges is to accurately identify stars that belong to a particular cluster or those that are part of field stars based solely on photometric data. This identification process is crucial as it serves as the initial step to isolate the target objects and precisely analyze the properties of a star cluster. Recently, machine learning algorithms and astrometric data have been utilized to tackle this problem. This study aims to compare the effectiveness of various unsupervised machine learning methods in the membership association of stars in open and globular clusters. A dataset consisting of three open and three globular star clusters from GAIA DR3 have been used to test the viability of these methods on different types of clusters. The study analyzed the sensitivity, precision, and false-discovery rate of methods such as GMM, DBSCAN, and pyUPMASK. The results showed that pyUPMASK had the highest precision, albeit with slightly higher computational time, while GMM and DBSCAN performed similarly. This study highlights the importance of selecting the appropriate machine-learning method to estimate the membership probability of a star belonging to a cluster based on cluster type, size, and composition.

Index Terms— Unsupervised Machine Learning, GAIA, Membership Classification, Open Clusters, Globular Clusters, DBSCAN, GMM, pyUPMASK, Photometry.

[1*,2,3,4] Department of Applied Physics, Delhi Technological University, Delhi-110042, India

***Corresponding Author:** Anish Kalsi

* Department of Applied Physics, Delhi Technological University, Delhi-110042, India.
Email: anishkalsi2k19ep014@dtu.ac.in, anish81.kalsi@gmail.com

DOI: 10.48047/ecb/2023.12.si10.00457

I. INTRODUCTION

Gigantic clouds of molecular gas and dust collapse under gravity at a few nucleation points, forming numerous stars. These newly born stars get gravitationally bound and form a group called Star Clusters. These clusters are generally categorized as open clusters (OC) or globular clusters (GC) based on the degree to which these clusters are gravitationally bound. Cluster membership classification is one of the most critical characteristics in studying star clusters.

We observe star-dense regions of the sky, looking for stars with comparable distances and velocities to identify cluster members. Such samples might be tainted because of the field stars with similar features. Therefore cataloging star clusters to study their specific properties become essential. Some prominent and reliable star cluster catalogs which have also been used to verify the results of our study include the Gaia DR2 OC catalog provided by [1] and the Gaia DR2 GC catalog [2].

OC population surveys [3] aim to obtain the list of members of a specific cluster and cluster parameters. The study carried out in [3] reported 60 newly discovered objects. It reclassified 2 OCs as GCs by better analysing the cluster parameters and members' positions. Studies like [4] show that between 10% and 43% of the stars do not have their proper motions measured in Gaia DR2. Eight GCs were studied in [4], and the contamination by field stars was found to be between 1% to 7%. Such contamination rates are detrimental to the accurate analysis of such clusters. Thus, making membership association a very vital step when studying star clusters' parameters.

Recently, Machine Learning (ML) Methods have been extensively researched and applied for membership classification in such clusters. The Unsupervised Photometric Membership Assignment in Stellar Clusters (UPMASK) package [5] has come to be one of the widely used clustering methods in recent years, being used in catalogs like [6]. Supervised ML Methods like Random Forest have shown to be an improvisation over the previous methods. When implemented to newer data releases like GAIA DR2 [7], they show a precision of up to 90%. [8].

Although different methods are continually being used, tried, and implemented in different catalogs, it is imperative to look into how these methods would vary in producing outputs when subjected to different clusters and varying identification parameters. The clustering algorithm's approach determines the members it will identify to a vast extent.

GAIA data has previously been used in conjunction with Gaussian Mixture Model (GMM) for cluster identification to unravel the incompleteness of cluster population [9]. This makes GMM a very likely candidate when comparing different algorithms for cluster member identification, especially working with GAIA data releases. However, in some cases, slow computation speed and low sensitivity of GMM to star cluster data have also been seen [9]. Six OCs were studied using Ultra Violet Imaging Telescope (UVIT), implementing Gaussian Mixture Modelling from GAIA EDR3 astrometry to find 2000–2500 additional members in the six clusters, making the UVIT cluster catalog more expansive. [10] This portrays the versatility of the GAIA astrometry dataset and the GMM on different kinds of data ranging from UV–Optical colour bands. Certain instances include the combination of two machine learning methods being used in conjunction to provide better results. [11] A combination of the Gaussian Mixture Model and Random Forest applied on members of NGC 6405 resulted in the detection of 518 high-probability members. [11]

Methods like Density-Based Spatial Clustering of Applications with Noise (DBSCAN) rely solely on the density of clusters, i.e., the distance between nearest points. In contrast, an algorithm like Agglomerative Clustering relies on hierarchical clustering,i.e., building clusters from the bottom up. Such approach variance induces differences in the cluster members identified by each method. The first instance of DBSCAN being used for star clustering in GAIA data was used to study the spatial structure and sub-structure of regions rich in Hipparcos stars [12]. In this study, 35 agglomerates of early-type Hipparcos stars were found. This helped to discover some stellar overdensities in Orion Molecular Cloud Complex, along with accurate calculations of parallactic distances to Pleiades, NGC 2451 A, and IC 2391, along with several field stars. DBSCAN has previously shown better sensitivity when compared to algorithms like GMM, 50-62% as compared to 33% for GMM [13]. Along with high sensitivity, this method also offers excellent specificity and precision. DBSCAN can be used for pattern match procedures. This is effective when distinguishing young star clusters from contaminated regions containing old clusters and photometrically unphysical clusters [14]. This favors the versatility of such algorithms in membership identification and specific identification procedures, such as finding Young Stellar Objects (YSO) from Star Forming Regions (SFR). A spatial Structure study of OC NGC 2112 using the DBSCAN algorithm reveals 1193 likely cluster members. [15] DBSCAN divided the cluster into 865 core stars and 328 border stars. Based on the distance and kinematical data of stars identified by the DBSCAN method were used to estimate the galactic orbit of the cluster.

The upgradation of the existing UPMASK package to pyUPMASK [16] and GAIA's new data release GAIA DR3 [17] provides a whole new array of high-resolution data and improved techniques to work upon. This provides a better look at previously cataloged member stars in clusters and an excellent opportunity to find more cluster members. This also means that various previously available Machine Learning methods being used in membership classification must be compared and analyzed against each other against various parameters. This includes the already included packages in pyUPMASK and some more clustering methods. The initial use of UPMASK for membership analysis on actual data can be found in [18], where UPMASK was used to determine the properties of the Local (Orion) spiral arm, NGC 2302. Despite being a sparsely populated cluster, UPMASK was deployed at a higher k value of k = 31, where "k" is the number of stars present per k-means cluster. Despite having low members, UPMASK performed reasonably well here in a recent study conducted to identify blue straggler stars (BSS), 50 OCs used pyUPMASK to detect 138 new BSS. [19] This increased the number of BSS in OCs by nearly 10%.

This denotes the viability of these three ML methods for star membership identification and also how this process is a crucial part of facilitating further studies of the star clusters and their evolution. In this paper, we have discussed the source of the data we have used in Section II. The working of GMM, DBSCAN, and pyUPMASK when used for astronomical data has been discussed in brief in Section III. Further, results obtained from all these methods have been compared, visualised, and discussed in detail in Section IV. Moreover, the conclusion of our study is presented in Section V.

II. DATA SAMPLE

We have selected six lesser-studied star clusters (three OCs and three GCs) for our tests from the GAIA Data Release 3 [17] archive made public on 13th June 2022, which catalogs astrometric and photometric data of up to 1.59 billion sources. [17] The preliminary selection of data to obtain all potential members in the vicinity of the clusters was made by performing ADQL queries for all sources in a circular region of a suitable radius beyond the angular size of the specific clusters. For our sample data sets, we chose 3 OCs: NGC188, NGC6031, NGC6756, and likewise 3 GCs: NGC6139, NGC6362, and NGC3201. Only sources with parallax error $\leq$ 0.5 have been included in the initial samples. For our study, we assume that a star cluster is a densely populated collection of stars roughly similar in age and chemical composition. We implement membership assignment to our sample clusters based on a few characteristic parameters, i.e., components of the proper motion of sources, Right Ascension (RA), Declination (DEC), parallax, and their corresponding errors.

Ideally, several astrometric and photometric parameters such as radial velocities, galactic coordinates, and g-bp-rp band mean magnitudes also contribute to determining the membership association criteria for clusters. However, due to the unavailability of these variables for all the sources, we did not use such an ample parameter space.

III. RESEARCH METHODOLOGY

In this section, the methodology adopted to classify and distinguish between members and field stars of distinct clusters is discussed. A brief description of the used unsupervised algorithms implemented to identify members of clusters has been provided. The distinct approaches employed by each algorithm in the identification of clusters lead to different performances of algorithms.

A. DBSCAN

Density-based spatial clustering is a clustering algorithm that assigns points that are packed tightly in close proximity to one another to the same cluster. DBSCAN assigns clusters to data points based on two user-defined parameters, namely Epsilon (ϵ) and Minimum points (N). Epsilon (ϵ) corresponds to the maximum distance between 2 points to label that one is in the neighborhood of the other, and minimum points (N) is the number of points in a neighborhood of a point to label it as a core point. Based on the selected values of Epsilon and Minimum points, the DBSCAN algorithm operates as follows: The nearest neighboring points in the Epsilon range around every point are located, and consequently, the points with neighbors more significant than (N) are identified as ”core points”. Points that possess less than $'N'$ neighbors are identified as ”boundary points”. Meanwhile, points that possess 0 neighbors are labeled as noise. We used Python’s scikit-learn library to analyse our cluster data using DBSCAN.

B. GMM

Gaussian Mixture Models (GMM) is a type of unsupervised clustering algorithm that assumes that clusters are distributed in a Gaussian or normal distribution and, based on this assumption, estimates the probability density of data points belonging to a particular cluster. Gaussian Mixture Model-based clustering assigns data points to clusters with some ’soft’ probability and hence allows for the determination of the membership probability of stars in clusters.

The Gaussian mixture model forms different mixture components based on the different features of the data set being used. The number of clusters formed is denoted by the set K. It then assigns a set of cluster weights for k-th distribution (denoted by π_k) to each Gaussian component. A distinct cluster is represented by each mixture component. This distinct cluster is specified by the following 3 cluster parameters:

$\{\pi_k,\mu_k,\sigma_k\}$

Where π_k are the mixture weights, π_k is the mean vector, and σ_k is the covariance matrix for the k-th distribution.

Thus the probability that the i^{th} point in a dataset ’X’ belongs to the kth cluster by GMM is given as:

$p(z_i = k) = \pi_k$

Where z_i is the cluster assignment for observation ’X_i’ of the dataset. We used Python’s scikit-learn library to analyze our data using GMM.

C. pyUPMASK

UPMASK stands for Unsupervised Photometric Membership Assignment In Stellar Clusters. As the name suggests, it is an unsupervised machine learning algorithm to determine the membership probabilities of stars in the given field of view. Principal Component Analysis (PCA), a clustering technique, and Kernel Density Estimations (KDE) are the foundations of the methodology used in this study for membership evaluation. Arbitrary error models can be taken into consideration using the approach. We will be making use of pyUPMASK, which is an algorithm built upon the UPMASK package and written entirely in Python.

There are multiple features introduced in pyUPMASK to improve its performance and accuracy, but the basic workflow of the algorithm remains the same. Firstly, PCA is used to identify two features that contribute most to the clustering from the given parameter space, apart from the positional parameters. Then the chosen clustering algorithm is used to cluster the data into a user-defined range of clusters. pyUPMASK offers a dozen of clustering methods, and in our study, we chose GMM. Simultaneously, a large dataset of random field stars is initialized and compared with the members of the clusters obtained. A density estimation function ($\Phi(x,y)$) is sampled for the clusters obtained using GMM, which outputs a set (Φ). From this, a distance ($D(\Phi)$) between its maximum value and its mean value normalized with standard deviations (σ_Φ) is computed:

$$D(\Phi) = \frac{max(\Phi) - mean(\Phi)}{\sigma_\Phi} \quad (1)$$

Similarly, a KDE ($\Psi(x,y)$) is computed and sampled for the random field stars generated. Later, a collection of parameters ($D(\Psi)$) obtained from each independent random realization is utilized to create a set D_Ψ. Lastly, the data is deemed unsuitable for a uniform field’s random realization if

$$D(\Phi) \geq mean(D(\Psi)) + T \times \sigma_{D\Psi} \quad (2)$$

Where $\sigma_{D\Psi}$ is the standard deviation of random realization, and T is the threshold level above σ. All this procedure occurs in the inner loop called the UPMASK kernel. This gives us a binary classification. In order to incorporate errors associated with the observations and reduce the bias caused by the initial conditions of the chosen clustering algorithms, an outer loop is introduced with some new features in pyUPMASK to iterate over this process and get refined results. Its ability to successfully segregate cluster and field populations is demonstrated by running the algorithm on simulated data clusters [5]. Under a wide range of circumstances, it is possible to reconstruct the general spatial structure and distribution of cluster member stars in the color-magnitude diagram. We used Python’s pyUPMASK library to analyze our data.

IV. RESULTS

We have used DBSCAN, GMM, and pyUPMASK to obtain a binary classification between members and field stars for

TABLE I PERFORMANCE OF DBSCAN

Cluster	Radius (arcmin)	Predicted Members Members	Catalog	Sensitivity	Precision	False-discovery Rate
NGC6031	7	10195	78	0.0074	0.4310	0.5689
NGC6756	7.4	7262	266	0.0363	0.5280	0.4720
NGC188	36.6	3479	864	0.2432	0.9814	0.0186
NGC6139	11.1	8607	2072	0.1775	0.1792	0.8208
NGC6362	14.7	16105	16149	0.6435	0.7016	0.2984
NGC3201	22.2	38828	29668	0.5650	0.5837	0.4163

TABLE II PERFORMANCE OF GMM

Cluster	Radius (arcmin)	Predicted Members Members	Catalog	Sensitivity	Precision	False-discovery Rate
NGC6031	7	9134	78	0.0082	0.4310	0.5689
NGC6756	7.4	6052	266	0.0436	0.5280	0.4720
NGC188	36.6	8651	864	0.0989	0.9794	0.0206
NGC6139	11.1	8405	2072	0.1767	0.1783	0.8217
NGC6362	14.7	9562	16149	0.9264	0.9919	0.0081
NGC3201	22.2	21538	29668	0.9585	0.9859	0.0141

TABLE III PERFORMANCE OF PYUPMASK

Cluster	Radius (arcmin)	Predicted Members Members	Catalog	Sensitivity	Precision	False-discovery Rate
NGC6031	7	4064	78	0.0182	0.4539	0.5460
NGC6756	7.4	957	266	0.2633	0.6478	0.3522
NGC188	36.6	1642	864	0.5201	0.9827	0.0173
NGC6139	11.1	1960	2072	0.7026	0.7102	0.2898
NGC6362	14.7	10514	16149	0.9096	0.9765	0.0235
NGC3201	22.2	21640	29668	0.9319	0.9590	0.0409

OCs NGC6031, NGC6756, and NGC188 and GCs NGC6139, NGC6362, and NGC3201. Since GMM and pyUPMASK provided the associated membership probabilities along with the classification for sources, we only considered cluster members with membership probabilities > 0.8. An overview of the number of predicted members by DBSCAN, GMM, and pyUPMASK is presented in Table I, Table II, and Table III, respectively. Tables I, II, and III also illustrate the difference in the number of predicted members for the same clusters based on the clustering algorithm(s) used.

The results obtained after applying DBSCAN, GMM, and pyUPMASK clustering have been plotted for an OC NGC6756 and are shown in Figure 1. This is a positional plot between Right Ascension (α) v/s Declination (δ). The Color-Magnitude Diagram (CMD) for NGC6756 has also been plotted for predicted cluster members obtained using the studied algorithms in Figure 2. The predicted cluster members are assigned the color orange, while the blue data points represent field stars. A CMD is a graphical plot between the apparent magnitude of the G band and the difference of magnitude in BP and RP band filters of a star cluster. The G pass band filter covers wavelengths ranging from UV to near Infrared

(330-1050 nm). Blue Photometer (BP) and Red Photometer (RP) are photometric instruments aboard the GAIA observatory used to provide the associated BP (330–680 nm) and RP (630–1050 nm) spectra of wavelengths for a star cluster. CMDs are used to study the evolution of stars in a cluster. Therefore, it becomes crucial to identify and distinguish cluster members from the field stars to study the evolution of stars in the cluster accurately. Similarly, positional and CMD were plotted for GC NGC6362 in Figure 3 and 4.

Tables I, II, and III present an overview of the performance of DBSCAN, GMM, and pyUPMASK, respectively, on the different OC and GC data sets.

To validate the obtained results, we compare the results for the OCs with the OC Member Catalog provided by [1]. We obtained the OC cluster catalog data from [20]. The clustering results obtained for GCs are compared with the GC Member Catalog provided by [2]. We obtained the GC cluster catalog data from [21].

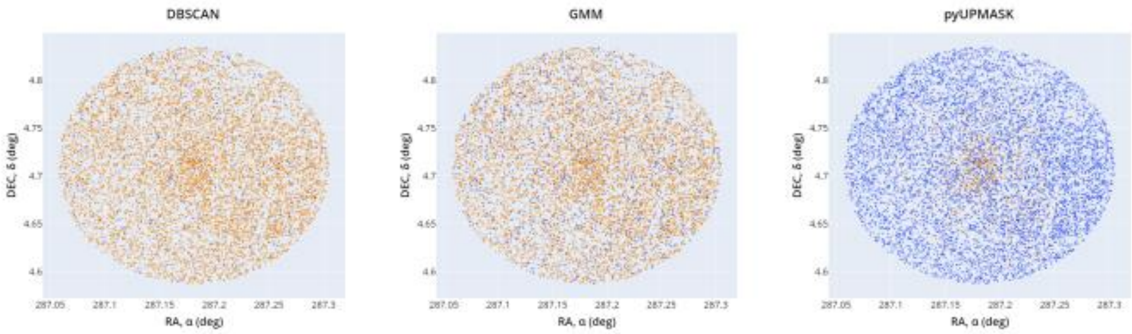

Fig. 1. RA vs Dec distribution plots of field stars and predicted cluster members obtained using DBSCAN, GMM and pyUPMASK for Open Cluster NGC6756 region.

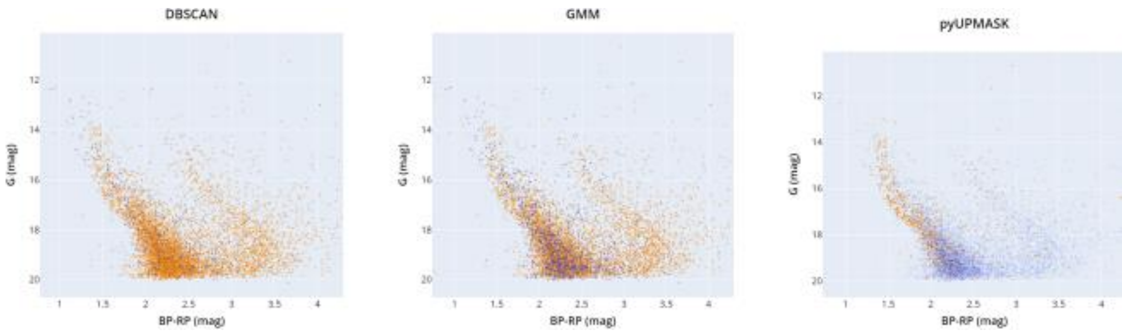

Fig. 2. Color-Magnitude Diagram plots of field stars and predicted cluster members obtained using DBSCAN, GMM and pyUPMASK for Open Cluster NGC6756 region.

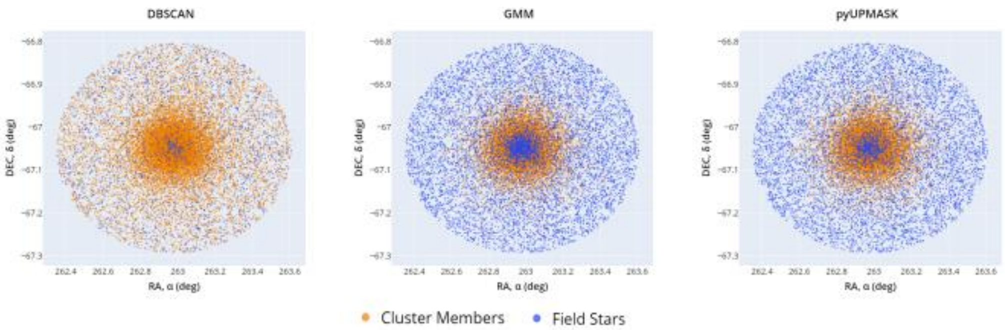

Fig. 3. RA vs Dec distribution plots of field stars and predicted cluster members obtained using DBSCAN, GMM and pyUPMASK for Globular Cluster NGC6362 region.

A variety of parameters are widely used to gauge the algorithms. The confusion matrices are created by comparing the performance of machine learning algorithms, such as sensitivity the predicted members with confirmed and verified cluster, precision, and false-discovery rates for all three algorithms members by the catalogs. These parameters were calculated, are computed by creating confusion matrices for these

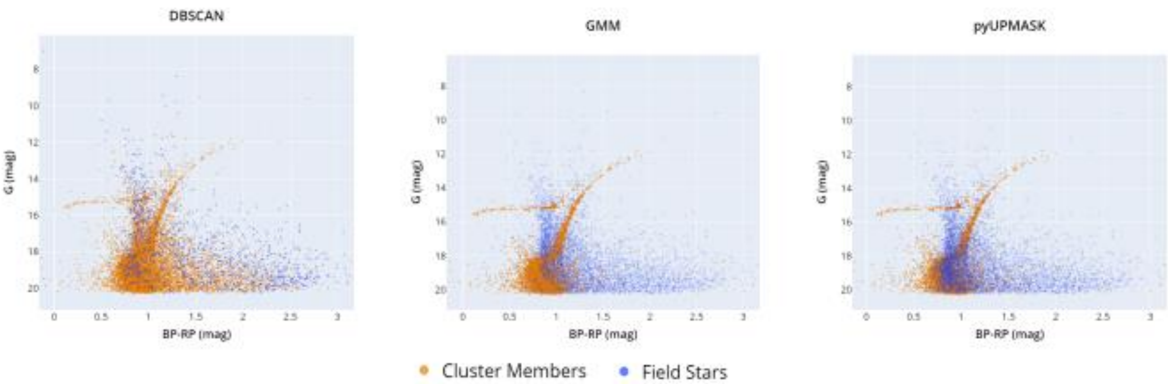

Fig. 4. Color-Magnitude Diagram plots of field stars and predicted cluster members obtained using DBSCAN, GMM and pyUPMASK for Globular Cluster NGC6362 region.

using the following formulas:

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (3)$$

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (4)$$

$$False\ discovery\ rate = \frac{False\ Positives}{False\ Positives + True\ Positives} \quad (5)$$

All these performance parameters varied with the type of cluster as illustrated by Tables I, II, and III. For OCs, we obtained poor sensitivity, low precision, and a high false-discovery rate using DBSCAN and GMM. With the exception of NGC188, for which we obtained a precision of 0.9814 and 0.9794, respectively, albeit with extremely poor sensitivities of 0.2432 and 0.0989, respectively, which leads to the obtained results being unreliable. On the other hand, pyUPMASK shows slightly better precision and false-discovery rate with much better sensitivity. For GCs, DBSCAN performed the worst with low sensitivity and precision with a high false-discovery rate. GMM performed exceptionally with NGC6362 and NGC3201, giving the best sensitivity and precision with the least false-discovery rate. Overall, pyUPMASK performed well, giving reliable and consistent results.

V. CONCLUSION

The GMM algorithm provides the membership status of stars in terms of a membership probability matrix which determines the likelihood of any star within the vicinity of the search radius belonging to the cluster. The GMM method is a statistical approach that involves calculating the weighted sum of probability density functions derived from several Gaussian distributions. Meanwhile, DBSCAN is a nonparametric algorithm that groups point into clusters based on their proximity in the feature space from one another. Due to these reasons, GMM and DBSCAN on their own are not ideal techniques to definitively identify cluster members and only provide a means for a rough classification of the underlying cluster structure.

The primary reason behind this disparity in the performance of sensitivity among the three algorithms arises due to the fact that the pyUPMASK algorithm utilizes adaptive thresholding to distinguish the foreground stars from the background stars, which allows it to identify a higher percentage of true cluster members while minimizing the number of false positives. Also, unlike GMM, pyUPMASK can use a model-free algorithm that does not make any assumption regarding the distribution of the data set. Hence it is also more suited for member identification in clusters with non-uniform or irregular density distributions. However, as may be seen from Tables I, II, and III for certain globular clusters with large sizes, such as NGC6362 and NGC3201, GMM provides marginally higher sensitivity and precision, while also showing the lowest false-discovery rates. This is primarily due to the fact that globular clusters often exhibit a centrally concentrated distribution of member stars which GMM is capable of effectively modeling due to its assumption of Gaussian components.

In conclusion, despite exhibiting comparable levels of precision and false discovery rates, the sensitivity of the pyUPMASK method is notably superior as compared to both DBSCAN as well as GMM. Although due to the iterative approach of pyUPMASK, it takes slightly more computational time than other methods, in return, it provides precise results. This algorithm proves to be robust and precise and can be used to study any star cluster in depth. On the other hand, DBSCAN and GMM are faster and give decent approximate results. These algorithms can be used to study and estimate members for a large number of star clusters, especially while studying a catalog of GCs, where the number of members is enormous and needs to be classified quickly and accurately.

ACKNOWLEDGMENT

We acknowledge the GAIA scientific mission for providing the GAIA DR3 data to the public, enabling researchers to conduct significant scientific studies and further our understanding of the universe.

REFERENCES

1. Cantat-Gaudin, T., Jordi, C., Vallenari, A., Bragaglia, A., BalaguerNu´nez, L., Soubiran, C., Bossini, D., Moitinho, A., Castro-Ginard, A.,˜ Krone-Martins, A. and Casamiquela, L., 2018. A Gaia DR2 view of the open cluster population in the Milky Way. Astronomy & Astrophysics, 618, p.A93.
2. Bustos Fierro, I.H. and Calderon, J.H., 2019. Extraction of globular´ clusters members with Gaia DR2 astrometry. Monthly Notices of the Royal Astronomical Society, 488(3), pp.3024-3034.
3. T. Cantat-Gaudin et al., “A Gaia DR2 view of the open cluster population in the Milky Way,” Astronomy & Astrophysics, vol. 618, p. A93, Oct. 2018, https://doi.org/10.1051/0004-6361/201833476.
4. I. H. B. Fierro and J. H. Calderon, “Extraction of globular clusters´ members with Gaia DR2 astrometry,” Monthly Notices of the Royal Astronomical Society, Jul. 2019, https://doi.org/10.1093/mnras/stz1879.
5. A. Krone-Martins and A. Moitinho, “UPMASK: unsupervised photometric membership assignment in stellar clusters,” Astronomy & Astrophysics, vol. 561, p. A57, Dec. 2013, https://doi.org/10.1051/00046361/201321143.
6. N.V. Kharchenko, A.E. Piskunov, E. Schilbach, S.R¨oser, R.D. Scholz, Astron.Astrophys. 558, A53 (2013). https://doi.org/10.1051/00046361/201322302.
7. Gaia Collaboration et al., ‘Gaia Data Release 2’, Astron. Astrophys., vol. 616, p. A1, Aug. 2018.
8. Mahmudunnobe, M., Hasan, P., Raja, M. et al. Membership of stars in open clusters using random forest with gaia data. Eur. Phys. J. Spec. Top. 230, 2177–2191 (2021). https://doi.org/10.1140/epjs/s11734-02100205-x
9. T. Cantat-Gaudin et al., “Gaia DR2 unravels incompleteness of nearby cluster population: new open clusters in the direction of Perseus,” Astronomy & Astrophysics, vol. 624, p. A126, Apr. 2019, https://doi.org/10.1051/0004-6361/201834453.
10. V. V. Jadhav, C. M. Pennock, A. Subramaniam, R. Sagar, and P. K. Nayak, “UOCS – III. UVIT catalogue of open clusters with machine learning-based membership using Gaia EDR3 astrometry,” Monthly Notices of the Royal Astronomical Society, vol. 503, no. 1, pp. 236–253, Jan. 2021, doi: https://doi.org/10.1093/mnras/stab213.
11. X. Gao, “Memberships of the Open Cluster NGC 6405 Based on a Combined Method: Gaussian Mixture Model and Random Forest,” The Astronomical Journal, vol. 156, no. 3, p. 121, Aug. 2018, doi: https://doi.org/10.3847/1538-3881/aad690.
12. J. A. Caballero and L. Dinis, “A revisit to agglomerates of earlytype Hipparcos stars,” Astronomische Nachrichten, vol. 329, no. 8, pp. 801–834, Oct. 2008, https://doi.org/10.1002/asna.200811024.
13. E. L. Hunt and S. Reffert, “Improving the open cluster census,” Astronomy & Astrophysics, vol. 646, p. A104, Feb. 2021, https://doi.org/10.1051/0004-6361/202039341.
14. L. Prisinzano et al., “Low-mass young stars in the Milky Way unveiled by DBSCAN and Gaia EDR3: Mapping the star forming regions within 1.5 kpc,” Astronomy & Astrophysics, vol. 664, p. A175, Aug. 2022, https://doi.org/10.1051/0004-6361/202243580.
15. X. Gao, S. Xu, and L. Xue, “The spatial structure and dynamical state of the open cluster NGC 2112,” Publications of the Astronomical Society of Japan, vol. 73, no. 3, pp. 652–659, Apr. 2021, doi: https://doi.org/10.1093/pasj/psab027.
16. M. S. Pera, G. I. Perren, A. Moitinho, H. D. Navone, and R. A. Vazquez, “pyUPMASK: an improved unsupervised clustering algorithm,” Astronomy & Astrophysics, vol. 650, p. A109, Jun. 2021, https://doi.org/10.1051/0004-6361/202040252.
17. Gaia Collaboration et al., ‘Gaia Data Release 3: Summary of the content and survey properties’, 2022.
18. E. Costa et al., “Insights into the properties of the Local (Orion) spiral arm. NGC 2302: First results and description of the program,” Astronomy & Astrophysics, vol. 580, p. A4, Jul. 2015, https://doi.org/10.1051/0004-6361/201525784.
19. C. Li, J. Zhong, S. Qin, and L. Chen, “The new detection of blue straggler stars in 50 open clusters using Gaia DR3,” Astronomy &

Astrophysics, Feb. 2023, doi: https://doi.org/10.1051/0004-6361/202244998.

20. Cantat-Gaudin, T., Jordi, C., Vallenari, A., Bragaglia, A., BalaguerNu´nez, L., Soubiran, C., Bossini, D., Moitinho, A., Castro-Ginard,˜ A., Krone-Martins, A. and Casamiquela, L., 2018. ”Gaia DR2 open clusters in the Milky Way : J/A+A/618/A93” Centre de donnees´ astronomiques de Strasbourg (CDS). https://cdsarc.cds.unistra.fr/vizbin/cat/J/A+A/618/A93 (accessed May. 09, 2023).
21. Bustos Fierro, I.H. and Calderon, J.H., 2019. ”Globular clusters members with Gaia DR2 : J/MNRAS/488/3024” Centre de donnees´ astronomiques de Strasbourg (CDS). https://cdsarc.cds.unistra.fr/vizbin/cat/J/MNRAS/488/3024 (accessed May. 09, 2023).