IRSTI 41.27.19 https://doi.org/10.26577/RCPh.2024v90i3-01

A.A. Bissekenov[1,2], M.T. Kalambay[2,3,4*], Y.S. Abylkairov[2], B.T. Shukirgaliyev[3,2,4]
[1]Xi'an Jiaotong-Liverpool University, Suzhou, China
[2]Energetic Cosmos Laboratory, Nazarbayev University, Astana, Kazakhstan
[3]Heriot-Watt International Faculty, K.Zhubanov Aktobe Regional University, Aktobe, Kazakhstan
[4]Fesenkov Astrophysical Institute, Almaty, Kazakhstan
*e-mail: M.Kalambay@hw.ac.uk

EXPLORING OPEN STAR CLUSTER MEMBERSHIPS WITH N-BODY SIMULATIONS AND MACHINE LEARNING

This work explores the application of supervised machine learning algorithms on N-body simulations to analyze the membership of open star clusters. The simulations used in this study are based on the Plummer model, clusters formed with constant star-formation efficiency (SFE) per free-fall time. We use simulations with different SFE and initial random realization. The random forest model was trained using simulations based on a 15% SFE over a time period of 20-100 million years. Subsequently, the model was tested on other N-body simulations with SFEs ranging from 17% to 25%, demonstrating consistently high classification accuracy throughout the dynamic evolution of the tested simulations. The model successfully identified cluster members with minimal deviations despite variations in SFE. Additionally, the algorithm maintained robustness against noise and initial conditions. Most of the errors observed in the model were false positives (FP), often located within a 2 Jacobi radius, suggesting gravitational binding to the cluster's center. This framework and learning strategy exhibit effectiveness and hold promise for further application in analyzing mock observations obtained from N-body simulations. Future work will focus on extending this method to more realistic observational scenarios.

Key words: Star clusters, N-body simulation, Machine Learning, Supervised Learning.

А.А. Бисекенов[1,2], М.Т. Қаламбай[2,3,4*], Е.С. Абылкаиров[2], Б.Т. Шукиргалиев[3,2,4]
[1]Сиань Цзяотун-Ливерпуль университеті, Сучжоу қ., Қытай
[2]Энергетикалық ғарыш зертханасы, Назарбаев университеті, Астана қ., Қазақстан
[3]Хериот-Уатт Халықаралық факультеті, Қ. Жұбанов атындағы Ақтөбе өңірлік университеті, Ақтөбе қ., Қазақстан
[4]В.Г. Фесенков атындағы Астрофизикалық институт, Алматы қ., Қазақстан
*e-mail: M.Kalambay@hw.ac.uk

N-денені модельдеу және машиналық оқыту арқылы шашыраңқы жұлдыздық шоғырларға мүшелікті зерттеу

Бұл жұмыс шашыраңқы жұлдыздық шоғырға мүшелікті талдау үшін N-денелі модельдеулерге Машиналық оқыту алгоритмдерін қолдануды зерттейді. Осы зерттеуде пайдаланылған симуляциялар шоғырдың бір еркін құлау уақытында жұлдыз түзілу тиімділігі (ЖТТ) тұрақты болатын Пламмер моделіне негізделген. Біз әртүрлі ЖТТ мен бастапқы кездейсоқ таралулармен ерекшеленетін симуляцияларды қолданамыз. Рандом Форест (Кездейсоқ орман) моделі 20-100 млн жыл аралығындағы ЖТТ 15%-ға негізделген симуляцияға оқытылды. Кейіннен ол ЖТТ 17%-дан 25%-ға дейінгі басқа N-дене симуляцияларына сыналды, және бұл модель сыналған симуляциялардың динамикалық эволюциясы барысында дәйекті жоғары жіктеу дәлдігін көрсетті. Модель жұлдыздардың пайда болу тиімділігінің өзгеруіне қарамастан кластер мүшелерін минималды ауытқулармен сәтті анықтай алды. Сонымен қатар, алгоритм шуыл мен бастапқы шарттардың өзгерісіне қатысты тұрақтылықғын сақтап қалды. Модельде байқалған қателіктердің көпшілігі жалған позитивтер (FP) болды, олар көбінесе 2 Якоби радиусында орналасады, бұл шоғырдың центрімен гравитациялық байланысуының себебінен деп болжанады. Осы негіздемелік бағдарлама мен оқыту стратегиясы тиімді екен және N-дене симуляцияларын жорамал бақылауларын талдауда одан әрі қолданысқа ие болады. Болашақта осы жұмыстағы әдісті неғұрлым нақты бақылау сценарийлеріне дейін кеңейтуге бағытталады.

Түйін сөздер: жұлдыздық шоғырлар, N-денені модельдеу,машиналық оқыту, бақыланатын оқыту.

А.А. Бисекенов[1,2], М.Т. Каламбай[3,4,2*], Е.С. Абылкаиров[2], Б.Т. Шукиргалиев[3,2,4]
[1]Сиань Цзяотун-Ливерпульский университет, г. Сучжоу, Китай
[2]Энергетическая космическая лаборатория, Назарбаев университет, г. Астана, Казахстан
[3]Хериот-Уатт Международный факультет,
Актюбинский региональный университет им. К.Жубанова, г.Актобе, Казахстан
[4]Астрофизический Институт им. В. Г. Фесенкова, г. Алматы, Казахстан
*e-mail: M.Kalambay@hw.ac.uk

Исследование членства в рассеянных звездных скоплениях с помощью моделирования N-тел и машинного обучения

В данной работе приведены результаты исследования по применению алгоритмов машинного обучения с учителем на симуляциях N-тел для анализа членства в рассеянных звездных скоплениях. Моделирование, использованное в данном исследовании, основано на модели Пламмера, а кластеры сформированы с постоянной эффективностью звездообразования (ПЭЗ) на время свободного падения. Каждая симуляция отличается ПЭЗ и начальной случайной реализацией. Модель случайного леса была обучена с использованием симуляций на основе эффективности образования звезд на уровне 15% за период времени от 20 до 100 миллионов лет. Затем модель была протестирована на других симуляциях N-тел с эффективностями образования звезд от 17% до 25%, и показана постоянная высокая точность классификации на протяжении динамической эволюции протестированных симуляций. Модель успешно определила звезды, относящиеся к кластеру с минимальными отклонениями, несмотря на вариации в эффективности звездообразования. Кроме того, алгоритм сохранял устойчивость к шуму и начальным условиям. Большинство ошибок, обнаруженные в модели, были ложными срабатываниями (FP) часто находящимися в пределах двух радиусов Якоби, что указывает на гравитационную привязанность к центру скопления. Показаны эффективность данной структуры и стратегии обучения и возможное дальнейшее применение в анализе мнимых наблюдений, полученных из симуляций N-тел. Продолжение исследования будет сосредоточено на расширении этого метода на более реалистичные наблюдательные сценарии.

Ключевые слова: звездные скопления, моделирование N-тел, машинное обучение, обучение с учителем.

Introduction

In recent decades, observational data has been growing intensively with the development of technology. Analyzing these data gives us new concepts about the evolution of stellar systems, Galaxies, and the Universe [1]. The study of star formation reveals more about the evolution of galaxies and the universe as a whole [1, 2]. Stars form from interstellar gas and dust. Open clusters (OCs), characterized by their relatively young ages (<100 Myr) and consisting of up to few thousand of stars, serve as invaluable laboratories for studying stellar evolution and dynamics [3].

Currently, more than four thousand OCs have been found in our galaxy. *Gaia* observations have played a crucial role in this discovery, providing a comprehensive and accurate three-dimensional map of stars, including their motions, luminosity, temperature, and composition [3, 4]. Identifying the membership of stars within these clusters is essential for understanding star formation and stellar evolution processes. Various methods, such as machine learning (ML), have been applied to datasets such as *Gaia* observations [4, 5]. Commonly used machine learning methods are random forest, k-nearest neighbors (KNN), and unsupervised learning methods such as StarGo, UPMASK, and Gaussian mixture model (GMM) [6-10].

While these methods produce significant results and contribute to various catalogs, comparisons among such studies often reveal discrepancies. For instance, three different research (CG18 [11], KC19 [12], M21 [13]) investigated the membership of the NGC 2516 cluster. Each group reported a different count of member stars. However, only 25% of the stars identified by KC19, 41% by M21, and 68% by CG18 are consistent with one another [14]. These inconsistencies highlight the potential for errors in star classification, motivating our exploration of alternative approaches for membership analysis.

In this work, we explore the possibilities of using ML models on N-body simulations for the membership analysis of OCs. Previous discussions have indicated significant discrepancies in star memberships within catalogs, possibly due to observational limitations

and methodological differences. To address these challenges, we propose leveraging N-body simulations where stars are pre-labeled as members or non-members throughout their simulated lifespan. This approach provides a robust framework for applying and testing ML approaches, potentially leading to the development of more accurate and effective methods for membership analysis.

Methods

The study utilized diverse N-body simulations of star clusters featuring variations in positions, masses, types of stars, and other parameters. These simulations exhibit discrepancies stemming from factors such as the number of stars, random configurations, and the star formation efficiency, illustrating the percentage of the gas mass designated for generating stars. Each simulation category encompasses a range of numerical data concerning Open Clusters stars at each time interval during the cluster's evolution in N-body time. Notably, the simulations in this investigation are rooted in the Plummer model, where the density peaked in the center and the star formation efficiency is constant during one free fall time throughout the gas. And that time passes faster in the center than at the edge of the gas cloud, and more stars form in the center. [15, 16]. Random realizations represent probabilistic values allocated to the initial positions and masses at the start of the simulation. Simulations featuring distinct random realizations exhibit variations in either mass or position. Specifically, there are three different random realizations of both position and mass, as indicated in Table 1. For each Star Formation Efficiency, nine simulations exist encompassing all possible combinations of random realizations for position and mass, as referenced in Table 2. Four SFEs in total result in a cumulative count of 36 simulations.

Table 1 - All random realizations of position and mass

Positions (P)	Mass (M)
1	1
2	2
3	3

Table 2 - All combinations of random realizations, where the first number indicates the position and the second number indicates the mass

11	12	13
21	22	23
31	32	33

In this investigation, SFEs ranged from 15% to 25% and featured a population of 10454 stars before instantaneous gas expulsion. This choice was motivated by the understanding that star clusters with lower SFEs and fewer stars are prone to reduced stability and quicker dissolution compared to those with higher SFEs and larger star populations. Within the 10454-star ensemble, approximately 20% of faint stars were omitted from the training dataset. According to data from the Gaia databases, stars with apparent magnitudes exceeding 21 magnitude in G-band [17] would be indiscernible to observers; thus they were also omitted from both training and testing datasets, along with other faint stars. Consequently, approximately 6000 stars remained for further analysis. Nonetheless, further reduction may be feasible, as certain stars may drift too far from the cluster over time. Details regarding the methodology for such refinement will be provided in subsequent sections.

For learning and testing, characteristics should be accessible and observable through observations, necessitating the selection of specific features. These features include the 2D galactocentric coordinates, velocities in corresponding directions, color index, and apparent magnitude. The primary rationale behind this choice is that these features are readily available in various datasets, making them suitable for study. This approach enables the exclusion of background stars from the testing and training processes, as background stars have minimal impact on the data when viewed from 150 parsecs above the galactic plane. We opted for 2D coordinates to simulate the scenario of observing the cluster from a vantage point situated 150 parsecs above the galactic plane along the Z-axis. Additionally, the presence of other stars in the Galaxy, known as field stars, necessitates their exclusion from consideration when observed from this elevated perspective. The color index ($B–V$) and apparent magnitude (m) serve as indirect measurements for the distance (Z-axis). The apparent magnitude is not directly incorporated into the simulations but can be computed using the formula:

$$m = M + 5(lgd - 5), \qquad (1)$$

where $d = 150 - Z$.

Learning and testing

The selection and overall approach to training the model on simulated star clusters posed the most critical and challenging aspect of the research. Given the tendency of open clusters to begin dissolving in their early stages of evolution and considering the

general similarity across small time intervals and specific stages of cluster evolution, determining which snapshot model to train proved to be particularly difficult. Moreover, imbalance became a concern as the cluster dissolved, resulting in fewer members and more non-members.

The imbalance issue was addressed by removing stars located more than 3 Jacobi radii from the cluster's center, ensuring they would not influence the learning process and maintaining a more balanced dataset. This approach facilitated effective training and testing, as stars distant from the cluster are easily identified as non-members due to their significant distance from the cluster's actual domain. This approach enabled the development of new learning strategies incorporating additional timeframes for training. While this resulted in a reduction in the number of stars in the datasets, the extent of this reduction varies across different timeframes and is not fixed.

The strategy for learning is to train on the snapshot after violent relaxation. This is the stage of the cluster equilibrium and will save the dynamic that would continue throughout the whole life-cycle of the cluster. In our simulations, the period is from 20 Myr to 100 Myr [17, 18]. The learning datasets are randomly chosen from 20 snapshot datasets. Learning on more snapshots is possible but will lead to overfitting, which would decrease performance. This approach was used for simulation with 15% SFE and 1st random realization for both position and mass, as shown in Table 3 with gray color. Further simulations with randomizations are given as simple numbers after the SFE number and "-" like "17-13", meaning 17% SFE with the 1st random realization of position and the 3rd random realization of mass.

Table 3 – Types of simulations used for testing. Gray is the simulation that was not tested but used for training. Pink is a simulation similar to training datasets, either by position or mass. Violet are the simulations that are completely different from learning data sets by position and mass.

11	12	13
21	22	23
31	32	33

For the testing, the trained model is used to predict membership of OC simulations with different SFEs of 17%, 20%, and 25% with position and mass random realizations as indicated in Table 3 with pink and violet colors. The time period used was from 20 Myr (end of violent relaxation) until the time-step when the cluster would have a Jacobi mass of less than 100 solar mass since it will be considered a dissolved cluster. Despite the absence of a mass feature in learning and testing, the state of the cluster with 100 solar mass can be calculated by the corresponding Jacobi radius formula [19]:

$$R_j = \frac{GM_j}{(4-\beta^2)\Omega^2}, \qquad (2)$$

where G is the gravitational constant, M_j is Jacobi mass, β is normalized epicyclic frequency, and Ω is the angular speed of the star cluster [16]. Another stopping condition is the number of member stars, and an OC with less than 50 stars would not be considered OC.

Machine learning algorithm

We tested multiple supervised algorithms for the classification but only presented one most successful model, Random Forest (RF). The RF algorithm is the decision tree algorithm for various high-dimensional data. It works by bootstrapping (randomly creating new datasets from input datasets) and building a predictive tree. After that, each sample goes through these trees, and the algorithm selects the average output of those trees [20]. Despite being an ML model for datasets with multiple features, it proved to be very good at predicting the membership of the OCs [21], [22], [23]. Thus, testing and application of this algorithm were a high priority of the paper.

RF classifiers from the scikit-learn package have multiple hyperparameters, and major hyperparameters are the number of trees, max tree depth, max number of features, etc. The number of trees or estimators depends on the dataset, and a more significant number of trees can lead to a more accurate classification but also may cause longer runtime and overfitting. Max depth accounts for how many depth of leaves the tree should have when running the algorithm. Max depth affects the computational complexity of the learning and prediction.

The maximum number of features is the number used for the node split of the tree, and it can be either square root or log_2 type. As for the criterion, the Gini index is used to measure the impurity in the values of datasets. Also, it is possible to use random states. In this study, the random forest has 100 trees, a max depth of 10, a square root type of max features (6 features ~ 2.49), and a random state of 42. It was possible to use a higher number of trees that would have also been deeper, but it was performing well as it is, and there was no real need to increase the complexity of both learning and testing. Other hyperparameters are left as default.

Evaluation metrics

The primary evaluation metric is the accuracy on other N-body simulation snapshots. It shows a percentage of how well the prediction was done broadly.

Table 4 – Confusion matrix of binary classification of stars

		Prediction	
		Non-Member	Member
Actual value	Non-Member	TN	FP
	Member	FN	TP

Despite having the general metric, we need to analyze true negatives (TN) and true positives (TP) for binary classification to understand performance better. The performance of the binary classification can be summarized with the confusion matrix shown in Table 4. The diagonal represents the number of correctly classified samples, and other values represent the number of samples classified incorrectly either as False Negative (FN) or False Positive (FP) [24]. Considering that a significant proportion of the testing samples would be either TP or TN, the true performance will be seen in the rates of FP and FN: False Positive Rate (FPR) and False Negative Rate (FNR). It is simply the rate of how many real positives or negatives are classified wrongly on the test. Formulas can be seen below:

$$FPR = \frac{FP}{TN + FP}, \quad (3)$$

$$FNR = \frac{FN}{TP + FN}. \quad (4)$$

Results and discussions

In this chapter, we would like to present the results of this work and show the performance of models according to evaluation metrics and further analysis of results. All dependent plots are given with the logarithmic scale because for each Myr on the x-axis until 175 Myr has six datasets. This is the reason why the plot is thicker until 200 Myr.

In plots, it is noticeable that there are several trends in all the tests for all SFEs. Firstly, it is evident that performance is very high, exceeding 90%. However, this trend is only universal in earlier time frames of the classification. Mostly, it is because OC dissolves, and as time goes by, there will be fewer and fewer member star samples.

Classification performance on simulations with 17% SFE can be seen on the top panel of Figures 1 and 2. The accuracy stays high, but the actual drop starts at different late snapshots on different simulations. This is because OCs with lesser SFE tend to be less stable and dissolve faster. Early dissolving ones are the SFE 17-13 and 17-33. These random realizations have similar masses and start dissolving at the similar age of 400-500 Myr. Other simulations are more stable, and the actual drop in performance is less than 90% after 600 Myrs. However, even until the end, they maintained a high performance of closer to 80%.

Further, we've done an analysis of FPs and FNs throughout the dynamic evolution. In the case of simulations with 17% SFE in the early timeframes, the rate of FN stays highest and, over time, decreases, eventually reaching 0. This means that the model makes a few mistakes with star members of the cluster, and we mostly do not lose member stars during the classification. On the other hand, the amount of FP stays low at 20 Myr, but on further snapshots, it only increases and almost reaches 50% at the end. This means the model mistakes up to 50% of the non-members as members. Considering that the cluster is cut to a 3 Jacobi radius, this is the expected result.

Cluster classification examples at different ages are shown in Figure 3. The colors represent the type of classification from Table 3. In Figure 3 (top right panel), you can see how the model identified membership of stars at 20 Myr. Classification accuracy is around 92.7%, with 157 FPs and 83 FNs. In simulations, violent relaxation ends at 20 Myr, which is the starting testing period. The picture shows that the cluster's shape is generally captured, and it correctly classified both members and non-members for the most part. This is evident from the circular shape of the identified cluster and the location of the TPs and TNs. However, you may also see that all the FP and FN are located at the 2D plane or on the borders of the cluster. This means that mistakes mostly happen because of the absence of the Z plane because most of the FPs are located upper or lower to the X and Y planes on the Z axis. FN is located on borders, meaning that they are marginal statistical errors of the ML model. Despite these errors, these are promising results for young OC with 17% SFE and 20 Myrs.

Similar results can be seen with the same cluster after 100 Myr later in Figure 3 (top left panel), but the accuracy is higher and reached 95%. The numbers of FP and FN are 39 and 33, respectively. Considering the lower number of non-members, this

can be regarded as a good performance. Placements of the FPs and FNs are similar, but mistakes are considerably lower now than before.

The classification at 500 Myr can be seen in Figure 3 (bottom panel). The accuracy at that age is 93%, with no FNs and 42 FPs. FPs are both on borders and on the Z plane, higher or lower than the cluster.

Overall, the number of stars is decreasing, and this kind of performance will continue for the whole dynamic evolution.

From the analysis of the layout of the stars, we identified that most of the mistakes happen to stars on cluster borders and on stars that are higher or lower than the Z plane. Most mistakes are the FPs, and FNs are mostly statistical mistakes that are not present in the further testing age. This raises the question of whether this number of FPs is acceptable for the membership analysis. To explore this, we need to know how far these stars are from the cluster's center. The cluster border is found by the Jacobi radius, and FPs are not within the Jacobi radius.

However, they might be within a two or 3 Jacobi radius, and at a 2 Jacobi radius distance, stars might still be gravitationally bound to the cluster's center.

The number of FPs within 2 and 3 Jacobi radii as a function of time can be seen in Figure 4.

From these results, we can see that most of the FPs are within a 2 Jacobi radius in all the classifications with different SFEs, and they may still be gravitationally bound to the center of the cluster. Thus, even FPs' mistakes can be considered member candidates in the future.

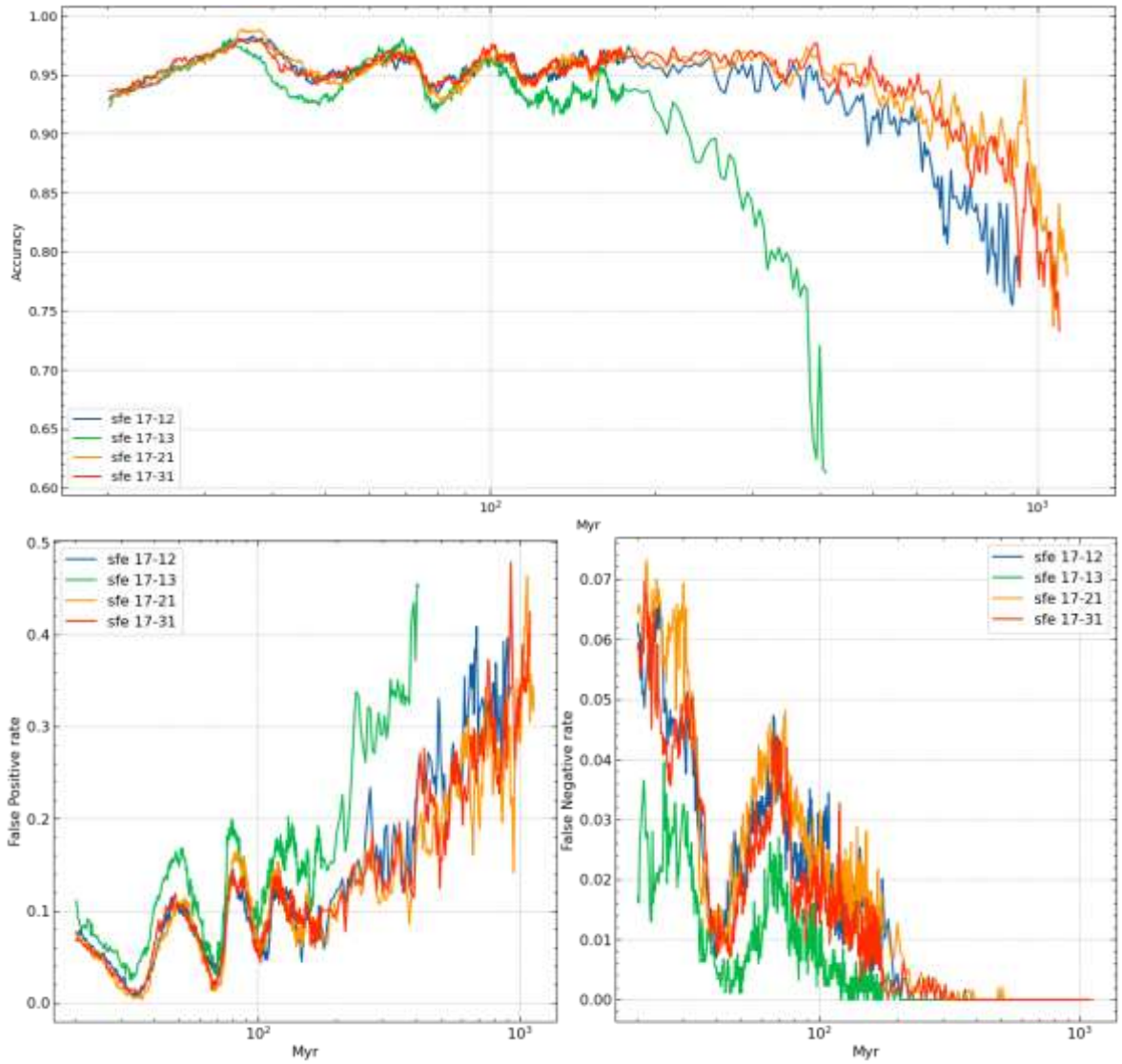

Figure 1 – Performance evaluation of RF model on simulations with 17% SFE on pink models from Table 2 in terms of Accuracy (top panel), False Positive Rate (bottom left panel), and False Negative Rate (bottom right panel) throughout the dynamic evolution.

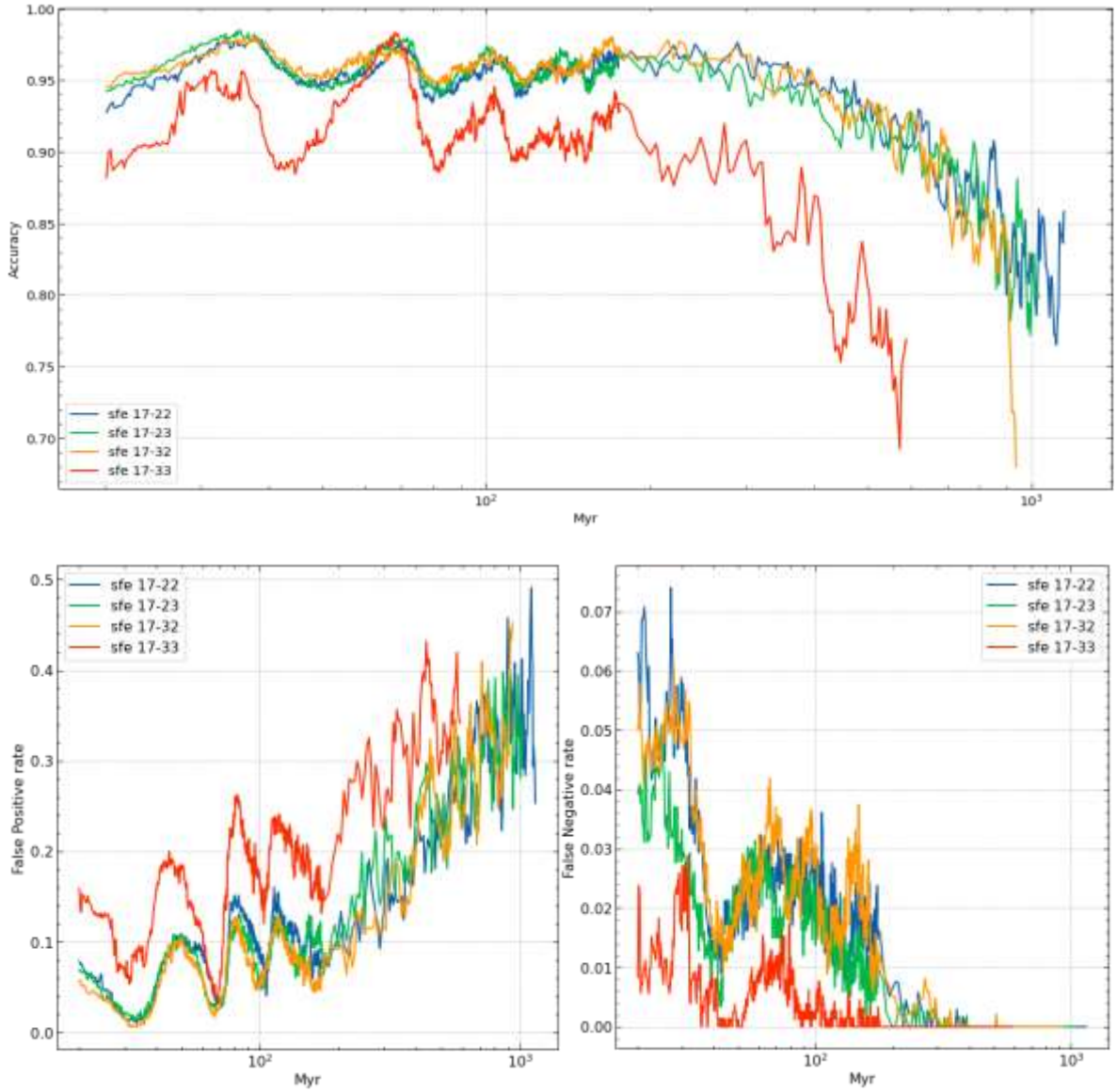

Figure 2 – Performance evaluation of RF model on simulations with 17% SFE on violet models from Table 2 in terms of Accuracy (top panel), False Positive Rate (bottom left panel), and False Negative Rate (bottom right panel) throughout the dynamic evolution.

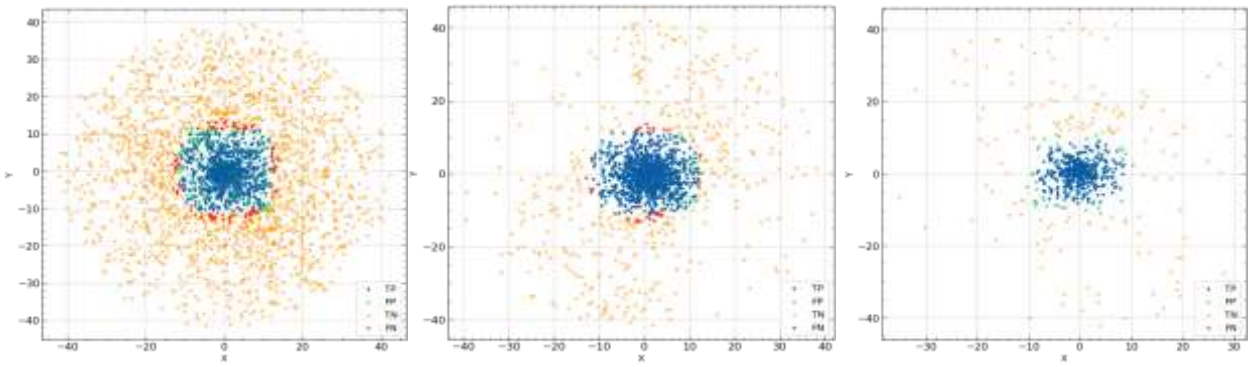

Figure 3 – Classification results on 17% SFE cluster at 20 Myr (1st panel), 100 Myr (2nd), and 500 Myr (3rd panel). Crosses are the negatives that are either true or false. Triangles are the positives, both true and false.

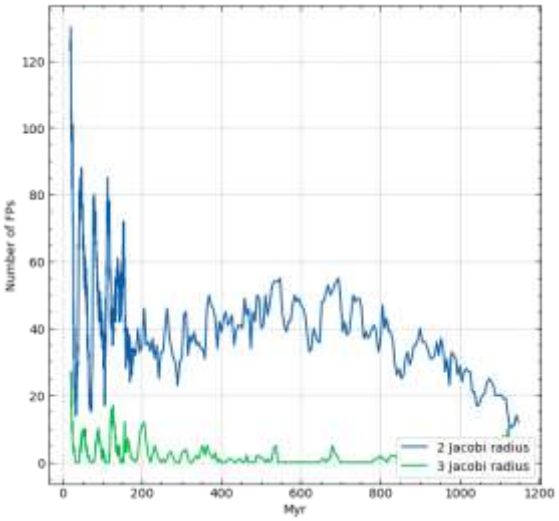

Figure 4 – The number of FPs throughout the test was within 2 Jacobi radius (blue) and between 2 and 3 Jacobi radius (green) on simulation with 17% SFE and 22 random realizations. There were no FPs beyond these distances

Conclusion

This work explores the possibilities of using the supervised learning approach combined with N-body simulations for membership determination. We trained the RF model on simulation with 15% SFE on the 20 random snapshots of its dynamic evolution's 20-100 Myr timeframe. This specific age of the cluster is chosen because it is the end of violent relaxation. Validation tests were done on N-body simulations with 17% SFE and different randomizations.

The model can predict the membership of all snapshots from 20 Myr until its dissolution with varying accuracy. The majority of snapshots, until getting close to dissolution, exceed 90%. Despite having such performance, it is identified that the model tends to mistake non-member stars as members (FP), and the number of FPs only increased due to a lesser number of stars in testing and closer to the complete dissolution. However, most of the FPs are located inside the 2 Jacobi radius, which might indicate that those stars are still gravitationally bound to the cluster's center and may still be considered member stars. This framework shows that it is possible to train a supervised ML model on N-body simulation to predict the membership of similar and utterly different N-body simulations with six easily obtainable features, such as 2D coordinates, velocities, color index, and apparent magnitude. Nonetheless, this is just a proof of idea and cannot be used for observation with described interpretation.

As for future plans, we will apply this strategy to the simulations converted to mock observed simulations. Current simulations are in cluster-centered coordinates, and mock observation versions of the simulations would have features in a format that can be seen in actual observational data. Also, the clusters will be placed on the Galactic disk viewed from the position of the Sun. This should make classification more complex. Under these conditions, background stars will play a considerable role, and distinguishing field stars from cluster members will be hard for the supervised ML model. Additionally, we plan to use N-body simulations as a testing ground for other ML models, such as Star Galactic Origins, various density-based scanners, and more.

Acknowledgments

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant AP13067834 and AP19677351).

References

1 Madau, P., and Mark, D., Cosmic star-formation history // Annual Review of Astronomy and Astrophysicsıo - 2014. - Vol.52. - P.415-486.

2 Robert, C.K., and Neal, J.E., Star formation in the Milky Way and nearby galaxies. // Annual Review of Astronomy and Astrophysics. - 2012. - Vol.50. - P.531–608.

3 Krumholz, M,R., Christopher, F.M., and Joss, B.,. Star clusters across cosmic time. // Annual Review of Astronomy and Astrophysics. - 2019. - Vol.57. - P.227-303.

4 Vallenari, A., et al. Gaia data release 3 summary of the content and survey properties. // Astronomy & Astrophysics. - 2023. - Vol.674. - Art.No.A1.

5 Brown, A.G.A., et al. Gaia data release 2-summary of the contents and survey properties. // Astronomy & Astrophysics. - 2018. - Vol.616. - Art.No.A1.

6 Agarwal, M., et al. ML-MOC: machine learning (kNN and GMM) based membership determination for open clusters. // Monthly Notices of the Royal Astronomical Society. - 2021. - Vol.502.2. - P.2582-2599.

7 Hunt, E.L., and Sabine, R., Improving the open cluster census-I. Comparison of clustering algorithms applied to Gaia DR2 data. // Astronomy & Astrophysics - 2021. - Vol.646. - Art.No.A104.

8 Gao, X., Memberships of the open cluster NGC 6405 based on a combined method: Gaussian mixture model and random forest. // The Astronomical Journal. - 2018. - Vol.156.3. - P.121.

9 Krone-Martins, A., and Andre, M., UPMASK: unsupervised photometric membership assignment in stellar clusters. // Astronomy & Astrophysics. - 2014. - Vol.561. - Art.No.A57.

10 Yuan, Zh., et al. StarGO: A New Method to Identify the Galactic Origins of Halo Stars. // The Astrophysical Journal. - 2018. - Vol.863.1. - P.26.

11 Cantat-Gaudin, T., et al. A Gaia DR2 view of the open cluster population in the Milky Way. // Astronomy & Astrophysics. - 2018. - Vol.618. - Art.No.A93.

12 Marina, K., and Kevin, C., Untangling the galaxy. i. local structure and star formation history of the Milky Way. // The Astronomical Journal. - 2019. - Vol.158.3. - P.122.

13 Stefan, M., João, A., and Alena, R., Extended stellar systems in the solar neighborhood-v. Discovery of coronae of nearby star clusters. // Astronomy & Astrophysics. - 2021. - Vol.645. - Art.No.A84.

14 Bouma, L.G., et al. Rotation and lithium confirmation of a 500 pc halo for the open cluster NGC 2516 // The Astronomical Journal. - 2021. - Vol.162.5. - P.197.

15 Shukirgaliyev, B., Parmentier, G., Berczik, P., & Just, A. The star cluster survivability after gas expulsion is independent of the impact of the Galactic tidal field // Monthly Notices of the Royal Astronomical Society. - 2019. – 486. - 1045.

16 Shukirgaliyev, B., Parmentier, G., Just, A., & Berczik, P. The Long-term Evolution of Star Clusters Formed with a Centrally Peaked Star Formation Efficiency Profile // The Astrophysical Journal. - 2018. – 863. - 171.

17 Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., Bailer-Jones, C. A. L., Biermann, M., Evans, D. W., Eyer, L., et al. Gaia Data Release 2. Summary of the contents and survey properties // Astronomy and Astrophysics. - 2018. – 616. - A1.

18 Shukirgaliyev, B., Parmentier, G., Berczik, P., & Just, A. Impact of a star formation efficiency profile on the evolution of open clusters // Astronomy and Astrophysics. - 2017. – 605. - A119.

19 Just, A., Berczik, P., Petrov, M. I., & Ernst, A. Quantitative analysis of clumps in the tidal tails of star clusters // Monthly Notices of the Royal Astronomical Society. - 2009. – 392. - P.969.

20 Biau G., Scornet E. A random forest guided tour //Test. – 2016. – Vol.25. – P.197-227.

21 Gao X. A Machine-learning-based Investigation of the Open Cluster M67 //The Astrophysical Journal. – 2018. – Vol.869. – No 1. – P.9.

22 Gao X. H. Memberships, distance and proper-motion of the open cluster NGC 188 based on a machine learning method //Astrophysics and Space Science. – 2018. – Vol.363. – P.1-8.

23 Gao X. An Investigation of the Pleiades Cluster Using Machine Learning //Publications of the Astronomical Society of the Pacific. – 2019. – Vol.131. – No 998. – P.044101.

24 Kohl M. Performance measures in binary classification //International Journal of Statistics in Medical Research. – 2012. – Vol.1. – No. 1. – P.79.

References

1 P. Madau, & D., Mark, Annu. Rev. Astron. Astrophys., 52, 415-486 (2014).
2 C.K. Robert, & J.E., Neal, Annu. Rev. Astron. Astrophys., 50, 531–608 (2012).
3 M.R. Krumholz, et al, Annu. Rev. Astron. Astrophys., 57, 227-303 (2019).
4 A. Vallenari, et al, Astron. & Astrophys., 674, A1 (2023).
5 A.G.A. Brown, et al., Astron. & Astrophys., 616, A1 (2018).
6 M. Agarwal, et al., MNRAS, .502.2., 2582-2599 (2021).
7 E.L. Hunt, & R., Sabine, Astron. & Astrophys., 646, A104 (2021).
8 X. Gao, The Astronomical Journal, 156.3, 121 (2018).
9 A. Krone-Martins, & M., Andre, Astron. & Astrophys., 561, A57 (2014).
10 Zh. Yuan, et al., The Astrophysical Journal, 863.1, 26 (2018).
11 T. Cantat-Gaudin, et al., Astron. & Astrophys., 618, A93 (2018).
12 K. Marina, & C., Kevin, The Astronomical Journal, 158.3, 122 (2019).
13 M. Stefan, A., João, & R., Alena, Astron. & Astrophys., 645, A84 (2021).
14 L.G. Bouma, et al., The Astronomical Journal, 162.5, 197 (2021).
15 B. Shukirgaliyev, et al., MNRAS, 486, 1045 (2019).
16 B. Shukirgaliyev et al., The Astrophysical Journal, 863, 171 (2018).
17 Gaia Collaboration, Astron. & Astrophys., 616, A1 (2018).
18 B. Shukirgaliyev et al., Astron. & Astrophys., 605, - A119 (2017).
19 A. Just et al., MNRAS, 392, 969 (2009).
20 G. Biau et al., Test, 25, 197-227 (2016).
21 X.H. Gao et al., The Astrophysical Journal, 869, 1, 9 (2018).
22 X.H. Gao et al., Astrophys. & Space Sci., 363, 1-8 (2018).
23 X.H. Gao et al., Pub.Astron. Soc. of the Pacific., 131, 998, 044101 (2019).
24 M. Kohl et al. Inter. Journal of Stati. in Med. Resch., 1, 1, 79 (2012).

Article history:
Received 9 May 2024
Received in revised form 22 June 2024
Received in revised form 5 September 2024
Accepted 8 September 2024

Мақала тарихы:
Түсті – 09.05.2024
Түзетілген түрде түсті – 22.06.2024
Түзетілген түрде түсті – 05.09.2024
Қабылданды – 08.09.2024

Information about authors:

1. **A.A. Bissekenov** – PhD student, Xi'an Jiaotong-Liverpool University, Suzhou, China; Energetic Cosmos Laboratory, Nazarbayev University, Astana, Kazakhstan: e-mail: Abhinav.Varma21@student.xjtlu.edu.cn

2. **M.T. Kalambay** (corresponding author) – PhD, Energetic Cosmos Laboratory, Nazarbayev University, Astana; Heriot-Watt International Faculty, K.Zhubanov Aktobe Regional University, Aktobe; Fesenkov Astrophysical Institute, Almaty, Kazakhstan. e-mail: M.Kalambay@hw.ac.uk

3. **Y.S. Abylkairov** – PhD, Energetic Cosmos Laboratory, Nazarbayev University, Astana, Kazakhstan. e-mail: sultan.abylkairov@nu.edu.kz

4. **B.T. Shukirgaliyev** – PhD, Ass.Prof., Heriot-Watt International Faculty, K.Zhubanov Aktobe Regional University, Aktobe; Energetic Cosmos Laboratory, Nazarbayev University, Astana; Fesenkov Astrophysical Institute, Almaty, Kazakhstan. e-mail: b.shukirgaliyev@hw.ac.uk

Авторлар туралы мәлімет:

1. **А.А. Бисекенов** – PhD студенті, Сиань Цзяотун-Ливерпуль университеті, Сучжоу қ., Қытай; Энергетикалық ғарыш зертханасы, Назарбаев университеті, Астана қ., Қазақстан, e-mail: Abhinav.Varma21@student.xjtlu.edu.cn

2. **М.Т. Қаламбай** (автор корреспондент) – PhD, Энергетикалық ғарыш зертханасы, Назарбаев университеті, Астана қ.; В.Г. Фесенков ат. Астрофизикалық институт, Алматы қ.; Хериот-Уатт Халықаралық факультеті, Қ. Жұбанов ат. Ақтөбе өңірлік университеті, Ақтөбе қ., Қазақстан. E-mail: M.Kalambay@hw.ac.uk

3. **Е.С. Абылкаиров** – PhD, Энергетикалық ғарыш зертханасы, Назарбаев университеті, Астана қ., Қазақстан. e-mail: sultan.abylkairov@nu.edu.kz

4. **Б.Т. Шукиргалиев** – PhD, қауым.проф., Хериот-Уатт Халықаралық факультеті, Қ. Жұбанов атындағы Ақтөбе өңірлік университеті, Ақтөбе қ.; Энергетикалық ғарыш зертханасы, Назарбаев университеті, Астана қ.; В.Г. Фесенков атындағы Астрофизикалық институт, Алматы қ., Қазақстан. e-mail: b.shukirgaliyev@hw.ac.uk.