International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Intelligent System for Prediction of Potentially Hazardous Nearest Earth Objects Using Machine Learning

[1]Dr. Pooja Bagane, [2]Dr. Srinivasa Rao Kandula, [3]Aditi Saxena, [4]Dr. Sanjit Das, [5]A. Deepak, [6]Dr. Sampathirao Govinda Rao

Submitted: 27/06/2023 **Revised**: 07/08/2023 **Accepted**: 26/08/2023

Abstract— Large potentially hazardous NEOs can cause a worldwide disaster in the occasion of a planetary colliding. These collisions with Earth are not uncommon. Every year, hundreds of asteroids hit the surface of our planet, the majorityof which are relatively small to cause any worry. Butoccasionally, big rocks can collide and harm anything. In orderto categorize the population of NEAs as potentially harmful or non-hazardous, this study proposes a method that uses differentalgorithms to learn complicated representations that are present in the distribution of accessible asteroid orbital data.

Keywords*—Asteroids, Prediction, NEOs, Machine Learning, Hazardous, Classification, Logistic Regression.*

I. Introduction

In space, there are an endless number of objects. A subgroup of asteroids known as Near-Earth Asteroids (NEAs) travel very close to the Earth. They are smaller, irregular in shape, and have shorter dynamical lives than main-belt asteroids (MBAs). They are frequently described as being of smaller size and having a perihelion distance thatis less or equal to 1.3 au. We do not acknowledge the dangerously close proximity of some of them. Even though we might assume that a distance of 70,000 kilometers cannot possibly have any effect on humans, this is a fairly minor distance on the scale of the universe and can affect a wide range of natural phenomena.

Thus, these objects or asteroids may turn out to be dangerous. One of the most crucial issues that needs to be solved to prevent humanity from suffering is classifying the population of Near-Earth asteroids concerning the identification well before an impact, of Potentially Hazardous Asteroids. To categorize the population of NEAs as potentially harmful or non-hazardous, this study proposes a method that uses different algorithms to learn complicated representations that can be found in the dispersion of asteroidorbital data. Therefore, we will be using a dataset that contains a list of asteroids designated by NASA as the closestobjects to the planet.

II. Literature Review

In the past decades, quite a few studies on Nearest earth objects, spatial classification and Asteroid dynamics have been conducted using various algorithms and tools. Some ofthe most prominent and relevant papers have been covered below.

John D. Hefele, Francesco Bortolussi and Simon Portegies Zwart *[1]* used a classifier that may affect the Earth in the next 20,000 years. An artificial neural network has been usedto detect asteroids that could collide with Earth. The Hazardous Object Identifier (HOI) was created by trying to launch objects from the Planet's surface and incorporating them back in time. HOI correctly identified 95.25% of known knockers simulated in the test set as possible traffic tools. Additionally, without direct

1Symbiosis Institute of Technology, Constituent of Symbiosis International University, Pune
poojabagane@gmail.com
2Professor, Department of ECE, Dhanekula Institute of Engineering and Technology,
Vijyawada, Andhra Pradesh, India
3Department of Electronics and Communication Engineering, GLA University Mathura
4Associate Professor, School of Advanced Sciences, VIT Chennai
5Department of Electronics and Communication Engineering, Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamilnadu. India
**deepakarun@saveetha.com*
6Professor, Department of Computer Science and Engineering,
Gokaraju Rangaraju Institute of Engineering and Technology (Griet)

training, 90.99% of total numberof the potentially hazardous objects identified by NASA wasdetected by the HOI.

Vadym Pasko *[2]* predicted the orbital parameter combinations of yet undiscovered Potentially Hazardous Asteroids (PHAs) using algorithms that come under machinelearning. The method that has been brought forward here aims to identify subsets of the main groups of Near Earth Asteroids (NEAs) that consist of high concentrations of PHAs. This method is intended to produce meaningful results and simple PHA sub group boundaries in 2 and 3 dimensional trajectory parameter subspaces. Svm algorithms with RBF kernels were primarily used to detect the boundary lines of these PHA subgroups. To process enough training and test data to simulate an undiscovered asteroid, an additional purely theoretical asteroid dataset was created.

T.Yildirim, H.K.Cigizoglu *[3]* classified hazardous and non-hazardous asteroids using various machine learning techniques i.e.; Logistic Regression, Decision Trees, SVM, K-Nearest Neighbors, Random Forest, Naive Bay, Adaboost,and Xgboost methods. Random Forest and Xgboost methods provide the highest accuracy while the Naive Bayes method provides the lowest accuracy of 80.70%.

Anish Si *[4]* compared MLP (Multilayer Perceptron) and GRNN (Generalized Regression Neural Network) techniques using some hydrological data. It is found that the GRNN estimates significantly outperformed the MLP estimates during the testing period. GRNN produces no negative assumptions and better fits the observations. This finding is significant because accurate assessment of hydrological data, such as suspended particles concentration, is critical for manywater resource projects. On the other hand, the researchers have the impression that GRNN is regarded as a reliable alternative to traditional MLP in future -ANN evaluation andprediction.

Richard Linares and Roberto *[5]* implemented data driven classification for spatial object(SO) classifying using light curve data based on the convolutional neural network (CNN) schema. The classification method uses SO light curve observational data to determine shape classes. These are the missile body, loading, and wreckage classes. CNN is a featurelearning and classification architecture. Convolutional layers are built on a series of filters that extract local features. This work uses three layers of convolution with kernels of 32, 12,and 6 unit sizes respectively. The advantage of this data- driven CNN approach is that it is a simple implementation that does not require 272 datum modeling. For the shape classification problem, the -CNN approach achieved an overall accuracy of 99.6% correct classification. This work demonstrated that CNNs provide an accurate and computationally efficient method for classifying SOs.

Victor Basu *[6]* considered an asteroid's diameter to be one of the most crucial physical parameters of asteroids, and thiswas used to calculate many other dynamic features, such as computation of its rotation period, and to determine whether an asteroid is truly dangerous or not if it was discovered to be a NEO. It discusses the theory of using an neural network thatis artificial to guesstimate the an asteroid’s diameter by employing the MLP algorithm as the foundational supervised learning technique to predict the diameter. Some of the metrics used to evaluate and compare different model performances were the Mean-Absolute Error, Mean-Squared Error, Median-Absolute Error, Variance score as well as the R2 Score.

Analysis of different algorithms regressors such as that of Gradient Boosting, XG Boost , Random Forest and AdaBoost were done for prediction of diameter of an asteroid followed by comparison with the base algorithm (MLP). Multilayer Perceptron algorithm proved to out pass all of theother regressors in this study.

V. Carruba along with S.Aljbaae, R.C.Domingos, M.Huaman,and W.Barletta *[7]* used different ML methods in asteroid dynamics to define representatives of space debris families, spatial images in astronomical fields, and lines of argument images of asteroids in 3 resonances. This paper conducted athorough review of the available literature, namely 21 papers,with the most recent effort focusing on the application of ANN, for classification tasks of asteroid resonance frequency arguments, detection of Solar System small bodies, or classification of asteroids into spectral classes based on theirtaxonomy and others that explored the suitability of numerous ML and DL algorithms for a variety of applications concerned with the field of asteroid dynamics.

O. Naoya , C. Kanta , C. Takuya , P. Nishanth , T. Naoya , H.Ryuki *[8]* proposed an approach of a new trajectory design for asteroid flyby cyclers using deep neural network and its model (surrogate). The aforementioned cycler orbits allow multiple flybys of asteroids with very little Δv (change in velocity between pre-collision and post-collision trajectories of an asteroid). To improve prediction accuracy, the suggested architecture constructs the Earth-asteroid Earth blocks by trying to integrate the surrogacy arrangement model with astrodynamics knowledge.

Since ML based trajectory design necessitates a computationally massive database, this study devised a good strategy to generate a database that, by introduction of a pseudo asteroid, can enhance the size of the ideal path by its magnitude order. These Earth asteroid Earth blocks based onthe surrogate allow us to efficiently look for optimal sequences using beam search. The recommended method demonstrates effective asteroid flyby cycler movements and is realistically applicable to space mission design.

Proposed System

In this paper, we bring forward a system that makes use of a combination of three machine learning methods; Logistic Regression, Random Forest and Support Vector Machines for the prediction of potentially hazardous Nearest Earth Objectsto get better accuracy for prediction. The data used in this process has been gathered from the National Aeronautics Space Administration-(NASA) and officially maintained by its Jet Propulsion Laboratory-(JPL).

Python is used for implementing the above algorithms. Exploratory Data analysis will be carried out to inspect the asteroid density/weight, distance from the Earth and other factors that play an important role for the problem considered and classify them from the dataset and display them in a visual manner.

Further research will be done to examine the correlation between the core features of the data.

III. Algorithm Used

A. *Random Forest*

Random forest, a supervised learning clustering algorithm, combines numerous decision trees to form a forest, as well asthe bagging concept, which introduces randomization into the model building process. The individual tree is split using a random selection of features, while the training data subset for each decision tree is created using a random selection of instances.

The variable from the random number of features is examinedfor the optimal split at each decision node in the tree. Random forests will choose the most frequent as its forecast if the target attribute is categorical. If it's a numerical question, theaverage of all forecasts will be chosen.

Running each test data point through each decision tree in the forest generates a prediction for it. This prediction is then built based on the majority vote of the models, creating a more potent and flexible single learner.

The trees then vote on an outcome. The prediction average will resemble the true value or ground truth (classification), allowing random forests to conquer the prediction variabilitythat each decision tree has (regression).

B. *Support Vector Machines (SVM)*

This is a popular Supervised Learning technique that maybe used to solve both classification and regression issues. However, it is more extensively used for Classification.

The algorithm's objective is to determine the optimal decision boundary or line for categorising n-dimensional space into classes so that subsequent information points maybe quickly assigned to the appropriate category. Support vectors are the extreme points or vectors that SVM selects tobuild the hyperplane.

C. *Logistic Regression*

Logistic regression is a well-known Machine Learning algorithm from the Supervised Learning technique. It forecasts the categorical dependent variable based on a setof independent variables.

A categorical dependent variable's output is predicted using this algorithm. Therefore, its output must be discrete or categorical. It can be True or False, Yes or No, 0 or 1, and so on, but rather than providing the precise values of 0 and 1, it provides the values of probability that lie between 0 and 1.This method can be used to classify observations using various types of data and can quickly determine which variables are most effective for classification.

IV. Dataset

Source:	Kaggle
Topic:	NASA - Nearest Earth Objects
Size of the dataset:	9.48 MB

	name	est_diameter_min	est_diameter_max	relative_velocity	miss_distance	absolute_magnitude	hazar
0	162635 (2000 SS164)	1.198271	2.679415	13569.249224	5.483974e+07	16.73	
1	277475 (2005 WK4)	0.265800	0.594347	73588.726663	6.143813e+07	20.00	
2	512244 (2015 YE18)	0.722030	1.614507	114258.692129	4.979872e+07	17.83	
3	(2012 BV13)	0.096506	0.215794	24764.303138	2.543497e+07	22.20	
4	(2014 GE35)	0.255009	0.570217	42737.733765	4.627557e+07	20.09	

Fig1: Excerpt of NASA dataset

Parameters available:

➢ ***id :*** identifier (the same object can have severalrows in the dataset, as it has been observed multiple times)

➢ ***name :*** name given by NASA (including the yearthe asteroid was discovered)

➢ ***est_diameter_min*** : minimum estimated diameterin kilometers

➢ ***est_diameter_max :*** maximum estimated diameterin kilometers

➢ ***relative_velocity*** : velocity relative to earth

➢ ***miss_distance*** : distance in kilometers it missesEarth

➢ ***orbiting_body*** : planet that the asteroid orbits

➢ ***sentry_object*** : whether it is included in sentry - anautomated collision monitoring system

➢ ***absolute_magnitude*** : intrinsic luminosity

➢ ***hazardous*** : whether the asteroid is potentiallyharmful or not

Libraries used:

● **Pandas -** For working with structured data sets common to statistics, finance, the social sciences, and many other subjects, this library offers a variety of integrated, intuitive routines for performing common data manipulations and analysis on such data sets.

● **Seaborn-** Seaborn is a matplotlib-based Python data visualization library. It gives you a high-level interface for creating visually appealing and informative statistical graphics.

● **Matplotlib** - Matplotlib is a state-based matplotlib interface. It has a plotting interface that is similar toMATLAB. Pyplot is primarily designed for engaging plots and simple programmatic plots.

● **Plotly**- More than 40 distinct chart types are supported by this dynamic python library, an open-source plotting framework with support for 3-dimensional, statistical, financial, geographic, and scientific use-cases.

● **Scikit Learn**- It is one of the most important machine learning libraries used for statistical modeling including classification, regression, clustering and dimensionality reduction consisting of many components.

V. Experimental Analysis

A. PRE-PROCESSING THE DATA

The dataset is cleaned and processed beforehand with thefollowing steps:

- Importing required libraries and dataset
- Cleared up missing values (none found)
- Conversion of category value into numerical data
- Finding the correlation between the attributes
- Dropping insignificant features
- Dividing the database into a training and test set
- Testing of different models

B. DATA VISUALIZATION

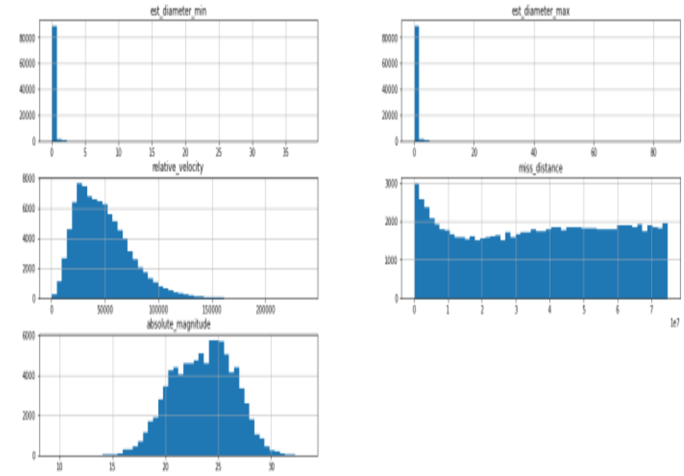


Fig1: Histogram distribution of all relevant columns

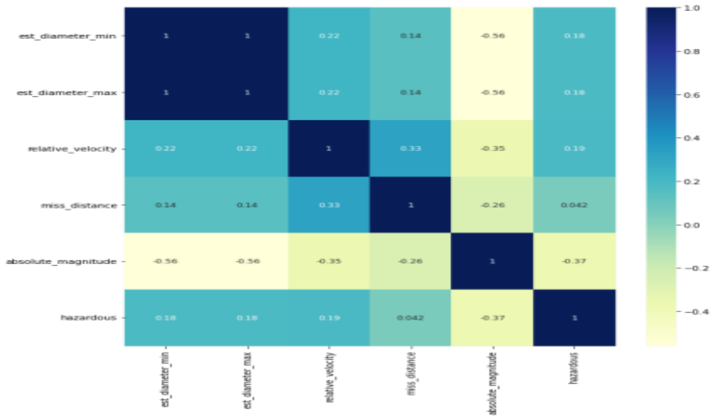


Fig 2: Heatmap of the NEO dataset

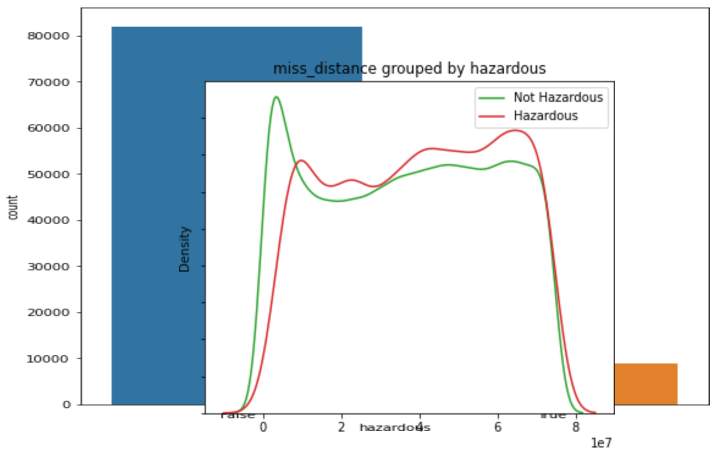


Fig 3: Plotting the data in the "hazardous" feature We see that the number of hazardous objects is extremely small compared to our non-hazardous objects.

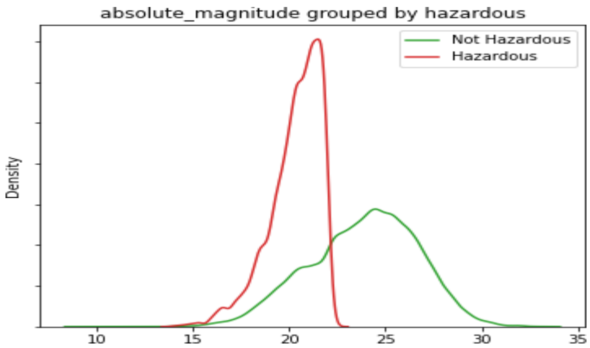


Fig4: For a certain threshold, that is, for an absolute magnitudeof 10-15 and 23 onwards, no NEOs are proven to be hazardous.

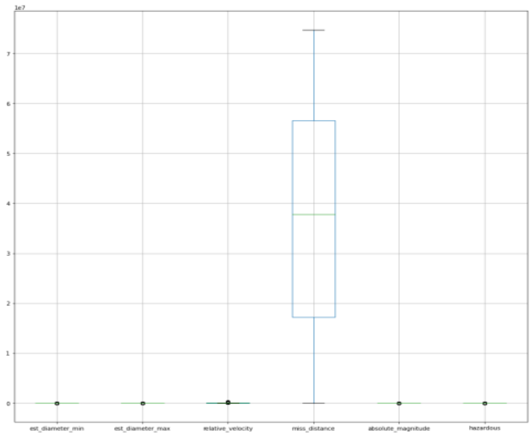


Fig 5: Box Plot to check for outliers in chosen columns

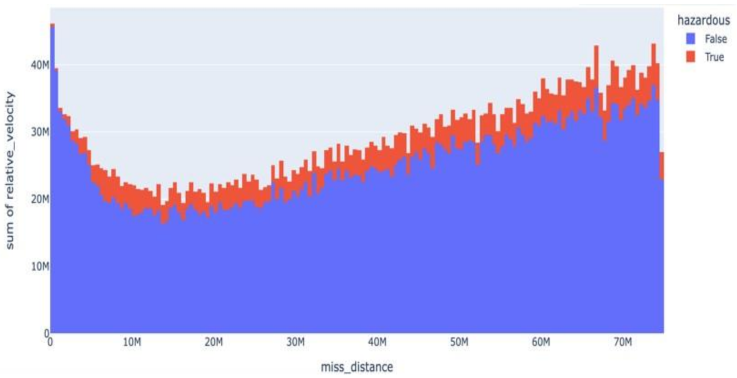


Fig 6 : The hazardous NEOs have a higher velocity andmove faster than non- hazardous NEOs

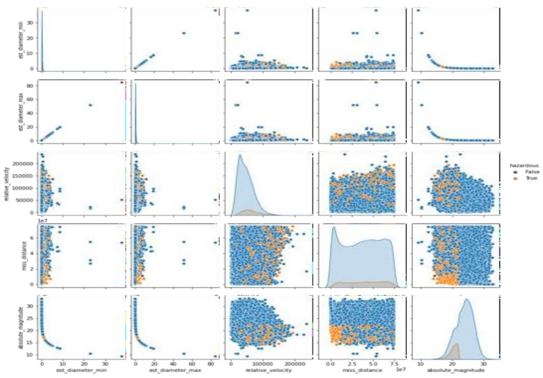


Fig 7 : Plotting of multiple pairwise bivariate distributionswith the hazardous feature.

C. RESULT OF PREDICTION OF ML METHODS USED :

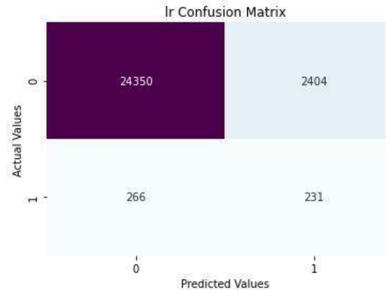
- ***Logistic Regression*** :

❖ Accuracy Score :
Training Set : 91.53%Test Set : 90.20%

❖ Precision:
For`0: 0.91
For 1: 0.46

❖ Recall:
For 0: 0.99
For 1: 0.09

❖ F-1 Score:
For 0: 0.95
For 1: 0.15



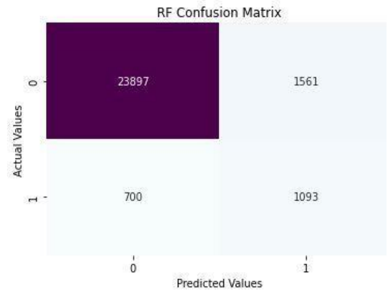
- ***Random Forest Classifier:***

❖ Accuracy Score :
Training Set : 92%Test Set : 91.68%

❖ Precision:
For`0: 0.94
For 1: 0.61

❖ Recall:
For 0: 0.97
For 1: 0.49

❖ F-1 Score:
For 0: 0.95
For 1: 0.49



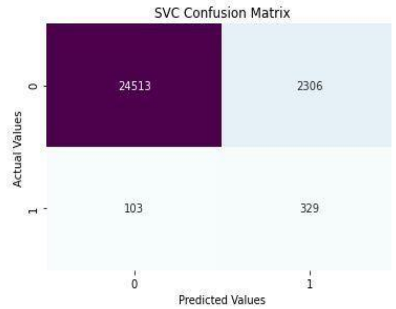
- ***Support Vector Machines:***

❖ Accuracy Score :
Training Set : 91.04%Test Set : 91.15%

❖ Precision:
For`0: 0.91
For 1: 0.76

❖ Recall:
For 0: 1.00
For 1: 0.12

❖ F-1 Score:
For 0: 0.95
For 1: 0.21



D. INSIGHTS

From our model, we have observed the following:

- We should be focused more towards larger asteroidsas they seem to be more hazardous than smaller asteroids.
- We see that most objects have a relative velocity between 0 and 150000 km/h and the absolute magnitude is majorly in the 15 to 30 region. Hazardous objects seem to be having a higher relative velocity on average.
- We see that most objects miss Earth by a distance between 0.1x10^7 and 7.8x10^7 units. Also, the PHAs seem to be missing earth by a higher distance on average
- By the current repository of data, hazardous objects detected seem to be much less in number compared to non-hazardous objects.

VI. Conclusion And Future Enhancements

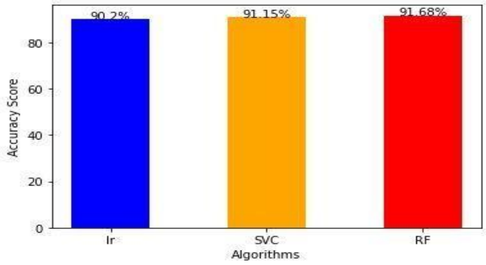


Fig8: The hazardous NEOs are expected to have a smallermiss distance but here it is observed that it is not the case

Using the three algorithms Random Forest, SVM, and Logistic Regression we predicted the Nearest Earth objects that prove to be hazardous and analyzed thepredictive accuracy of all three out of which the random forest algorithm had a maximum test accuracy score of almost 91.68% in prediction. To improve translation and prediction performance we will further use real time nearest earth objects that are currently listed as hazardouson NASA's Asteroid Watch.

Acknowledgements

We would like to express our appreciation to Symbiosis International University for providing us with an immense amount of valuable and constructive material to assist us during the planning and development of this research analysis.

References

[1] Hefele, John D., Francesco Bortolussi, and Simon Portegies Zwart. "Identifying Earth-impacting asteroids using an artificial neural network." Astronomy & Astrophysics 634 (2020): A45.

[2] Pasko, Vadym. "Prediction of Orbital Parameters for Undiscovered Potentially Hazardous Asteroids Using Machine Learning." Stardust Final Conference. Springer, Cham, 2018.

[3] Anish Si(2020). Hazardous Asteroid Classification through Various Machine Learning Techniques, International Research Journal of Engineering and Technology (IRJET),Volume 7.

[4] Yidirim, T., and H. K. Cigizoglu. "Comparison of generalized regression neural network and MLP performances on hydrologic data forecasting." Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02.. Vol. 5. IEEE, 2002.

[5] Linares, Richard, and Roberto Furfaro. "Space object classification using deep convolutional neural networks." 2016 19th International Conference on Information Fusion (FUSION). IEEE, 2016.

[6] V. Basu, "Prediction of asteroid diameter with the help of multi-layer perceptron regressor" (2019). Int. J. Adv. Electron. Comput. Sci., ISSN(p): 2394-2835

[7] Carruba, Valerio & Aljbaae, Safwan & Domingos, Rita De Cássia & Barletta, W.. (2022). "Machine learning applied to asteroid dynamics." Celestial Mechanics and Dynamical Astronomy. 134. 10.1007/s10569-022- 10088-2.

[8] Naoya Ozaki, Kanta Yanagida, Takuya Chikazawa, Nishanth Pushparaj, Naoya Takeishi, and Ryuki Hyodo, "Asteroid Flyby Cycler Trajectory Design Using Deep Neural Networks," Journal of Guidance Control and Dynamics, Vol.45, No.8, August 2022, pp. 1496-1511.

[9] R. N. Ranaweera and T. Fernando, "Prediction of Potentially Hazardous Asteroids using Deep Learning," 2022 2nd International Conference on Advanced Research in Computing (ICARC), 2022, pp. 31-36, doi: 10.1109/ICARC54489.2022.9753945.

[10] Bancelin, David & Colas, Francois & Thuillot, William & Hestroffer, D. & Assafin, M.. (2012). Asteroid (99942) Apophis: New predictions of Earth encounters for this

potentially hazardous asteroid. Astronomy and Astrophysics. 544. 15-. 10.1051/0004-6361/201117981.

[11] Nesvold, Erika & Greenberg, Adam & Erasmus, Nicolas & Heerden, Elmarie & Galache, J. & Dahlstrom, Eric & Marchis, Franck. (2018). The Deflector Selector: A Machine Learning Framework for Prioritizing Hazardous Object Deflection Technology Development. Acta Astronautica. 146.

[12] 10.1016/j.actaastro.2018.01.049.

[13] Prasad, L. & Anjali Shivani Reddy, Thadisina & Kashi, Bhuvaneshwari. (2020). Asteroid Detection using Machine Learning Algorithm. Communications of the Byurakan Astrophysical Observatory. 329-334. 10.52526/25792776-2020.67.2-329.

[14] Khalil, Mohamed & Said, M. & Ibrahim, A. & Gamal, M. & Mobarak, M.. (2021). Tracking Comets and Asteroids Using Machine Learning and Deep Learning: A Review. International Journal of Advanced Astronomy. 9. 26. 10.14419/ijaa.v9i1.31466.

[15] V. Bahel, P. Bhongade, J. Sharma, S. Shukla and M. Gaikwad, "Supervised Classification for Analysis and Detection of Potentially Hazardous Asteroid," 2021 International Conference on Computational Intelligence and Computing Applications (ICCICA), 2021, pp. 1-4, doi: 10.1109/ICCICA52458.2021.9697222.

[16] L. Cheng, Z. Wang, F. Jiang and J. Li, "Fast Generation of Optimal Asteroid Landing Trajectories Using Deep Neural Networks," in IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 4, pp. 2642-2655,

[17] Aug. 2020, doi: 10.1109/TAES.2019.295

[18] Ďurech, J., et al. "Detection of the YORP effect in asteroid (1620) Geographos." Astronomy & Astrophysics 489.2 (2008): L25-L28.

[19] Carruba, V., et al. "Machine learning applied to asteroid dynamics." Celestial Mechanics and Dynamical Astronomy 134.4 (2022): 1-27.

[20] Cazeneuve, Dorian, et al. "ODNet: A Convolutional Neural Network for Asteroid Occultation Detection." arXiv preprint arXiv:2210.16440 (2022).

[21] Meß, Jan-Gerd, Frank Dannemann, and Fabian Greif. "Techniques of artificial intelligence for space applications-a survey." European workshop on on-board data processing (OBDP2019). European Space Agency, 2019.

[22] Ravani, Khilan, S. Mathavaraj, and Radhakant Padhi. "Site detection for autonomous soft-landing on asteroids using deep learning." Transactions of the Indian National Academy of Engineering 6.2 (2021): 365-375.

[23] Carruba, V., et al. "Machine Learning applied to asteroid dynamics: an emerging research field." arXiv preprint arXiv:2110.06611 (2021).

[24] Zhou, Dong, Gunaghui Sun, and Xiaopeng Hong. "3D Visual Tracking Framework with Deep Learning for AsteroidExploration." arXiv preprint arXiv:2111.10737 (2021).

[25] Wang, Franklin, Jian Ge, and Kevin Willis. "Discovering faint and high apparent motion rate near-Earth asteroids using a deep learning program." Monthly Notices of the Royal Astronomical Society 516.4 (2022): 5785-5798.

[26] Song, Yu, and Shengping Gong. "Solar-sail trajectory design for multiple near-Earth asteroid exploration based on deep neural networks." Aerospace Science and Technology 91 (2019): 28-40.

[27] Carruba, V., S. Aljbaae, and A. Lucchini. "Machine- learning identification of asteroid groups." Monthly Notices of the Royal Astronomical Society 488.1 (2019): 1377-1386.

[28] Rabeendran, Amandin Chyba, and Larry Denneau. "A two-stage deep learning detection classifier for the atlas asteroid survey." Publications of the Astronomical Society of the Pacific 133.1021 (2021): 034501.

[29] Jedicke, Robert, Jeffrey Larsen, and Timothy Spahr. "Observational selection effects in asteroid surveys and estimates of asteroid population sizes." Asteroids III (2002): 71-87.

[30] Mohapatra, S. K. ., Patnaik, S. ., &

Kumar Mohapatra, S. . (2023). An Enhanced Automated Epileptic Seizure Detection Using ANFIS, FFA and EPSO Algorithms . International Journal on Recent and Innovation Trends in Computing and Communication, 11(4s), 57–67. https://doi.org/10.17762/ijritcc.v11i4s.6307

[31] Aisha Ahmed, Machine Learning in Agriculture: Crop Yield Prediction and Disease Detection , Machine Learning Applications Conference Proceedings, Vol 2 2022.

[32] Dhabliya, D., & Sharma, R. (2019). Cloud computing based mobile devices for distributed computing. International Journal of Control and Automation, 12(6 Special Issue), 1-4. doi:10.33832/ijca.2019.12.6.01