*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p-ISSN: 2278-8727, Volume 26, Issue 6, Ser. 1 (Nov. – Dec. 2024), PP 37-44*
*www.iosrjournals.org*

# AI Based Hazardous Asteroid Detection

Ananya A
*Computer Science And Engineering Rv College Of Engineering Bengaluru, Karnataka*

Srividya M S
*Computer Science And Engineering Rv College Of Engineering Bengaluru, Karnataka*

***Abstract—***
*As the need for early detection and mitigation of potential threats from near-Earth objects grows, this study presents a comprehensive approach to predicting hazardous asteroids using machine learning techniques. Given the increasing importance of protecting our planet from potential impact events, accurately classifying and predicting hazardous asteroids is cru- cial. The study proposes a comprehensive system for predicting and mitigating asteroid hazards using advanced data processing, machine learning, and simulation techniques. The proposed study collect real-time and historical asteroid data, process and clean this data, extract relevant features, and train machine learning models to predict potentially hazardous asteroids. Additionally, Monte Carlo simulations are employed to assess the collision risks of these asteroids with Earth. A forecasting module predicts future asteroid behavior based on historical and simulated data. The system also includes an alert mechanism that notifies relevant stakeholders when a potentially hazardous asteroid is detected.*
***Index Terms—****Asteroid Impact, Prediction, Potentially Haz- ardous Asteroids, Machine Learning, Random Forest Classifica- tion, – minor planets, Planetary defense.*

---------------------------------------------------------------------------------------------------------------------------------------
Date of Submission: 15-11-2024 Date of Acceptance: 25-11-2024
---------------------------------------------------------------------------------------------------------------------------------------

## I. Introduction

Asteroids, often referred to as the building blocks of planets, are rocky remnants from the early solar system's formation, dating back 4.6 billion years. These objects vary widely in size, from small pebbles to massive bodies spanning several kilometers in diameter. While most asteroids are located in the asteroid belt between Mars and Jupiter, some, known as near- Earth asteroids (NEAs), have orbits that bring them close to our planet, creating a potential collision risk. Understanding their behavior such as composition, orbital paths, and im- pact potential is essential for developing effective mitigation strategies. Traditional asteroid monitoring methods have been effective to some extent, but advancements in technology have opened new avenues for more accurate and timely detection. The proposed project introduces a novel approach that utilizes machine learning and deep learning models, such as Random Forest and Long Short-Term Memory (LSTM) networks, to enhance the accuracy of asteroid hazard prediction. By analyzing real-time and historical asteroid data, performing Monte Carlo simulations to estimate collision probabilities, and generating synthetic data to balance the dataset using Generative Adversarial Networks (GANs), this system aims to improve the accuracy and timeliness of asteroid hazard predictions. Additionally, the project integrates forecasting techniques to predict future asteroid trajectories and behaviors. An alert notification system further enhances this approach by providing real-time warnings to relevant stakeholders when a potentially hazardous asteroid is identified. This approach not only enhances planetary defense but also contributes to our understanding of asteroid behavior and the evolution of our solar system.

## II. Literature Review

Surekha Nemmaluri Purna, et al [1] proposed a study to enhance planetary defense by predicting hazardous asteroids using machine learning. The research employs a diverse dataset of asteroid characteristics, combining feature engineering and advanced algorithms to improve prediction accuracy. The study explores models like Logistic Regression, Random Forest, and Light Gradient Boosting, finding that a well-tuned Random Forest model offers the best accuracy in predicting hazardous asteroids.

Vedant Bahel, et al [2] explored the application of AI in detecting Potentially Hazardous Asteroids (PHAs) using machine learning. Their study evaluated four classification algorithms—Logistic Regression, K-Nearest Neighbor, Decision Tree, and Random Forest—on data from NASA's "Small-Body Database." The Random Forest classifier emerged as the most effective. The research underscores the challenges of imbalanced data and suggests that tree- based algorithms are particularly well-suited for such tasks. Future work will focus on incorporating additional asteroid characteristics to further enhance prediction accuracy.

Sandeep Avala, et al [3] investigated the use of machine learning to predict the orbits of potentially hazardous asteroids. They employed various algorithms, including XGBoost, K-Nearest Neighbors (KNN), and Multi-Layer Perceptron (MLP), to classify asteroids based on their orbit classes. XGBoost demonstrated the highest accuracy. The study emphasizes the potential of integrating these models into planetary defense systems for enhanced risk assessment. Future research may focus on incorporating real-time data to improve prediction accuracy.

Bhavsar, et al [4] introduced a Quantum Machine Learning (QML) approach to predict asteroid hazard potential, addressing the limitations of classical methods in accuracy and efficiency. Using a Kaggle dataset with 958,524 records (only 2,066 labeled as hazardous), they balanced it to 4,132 records through random sampling. Preprocessing involved cleaning, normalization, and dimensionality reduction, with classical algorithms used for feature selection. The QML techniques, including Variational Quantum Circuits (VQC) and Pegasos Quantum Support Vector Classifier (PQSVC), achieved an impressive accuracy, outperforming traditional methods.

Anish Si [5] proposed a novel model for identifying hazardous asteroids among near-Earth objects using various machine learning algorithms, including Logistic Regression, Support Vector Machine, Decision Tree, K-Nearest Neighbor, Random Forest, Na¨ıve Bayes, AdaBoost, and XGBoost. Among these, Random Forest and XGBoost achieved the highest accuracy, with Random Forest preferred for its faster training time. The study highlights that Random Forest with 15 trees offers an optimal balance between efficiency and accuracy, making it a valuable tool for enhancing the identification of hazardous near-Earth asteroids and contributing to planetary defense efforts.

Tushar Sharma, et al [6] developed a machine learning-based model to predict asteroid hazards by analyzing factors such as asteroid dimensions, speed, path, and atmospheric conditions. They employed algorithms like XGBoost and Random Forest to analyze past collision records and simulate impact scenarios. The study achieved an impressive accuracy , with XGBoost outperforming in precision and overall performance. This model's efficiency in prediction and comparison with traditional methods underscores its potential to significantly enhance asteroid risk assessment and preparedness efforts.

Chomette, et al [7] Mathias developed a machine learning (ML) approach to assess asteroid impact hazards, aiming to predict damage radii more efficiently than traditional methods like NASA's PAIR model. By training ML models, including neural networks and random forests, they significantly reduced computation time while maintaining accuracy. Their approach cuts computational time by a factor of 1000 and incorporates Shapley sensitivity analysis to prioritize asteroid observation missions based on impact parameters.

Malakouti, et al [8] study explored asteroid hazard classification using five machine learning algorithms: Extra Trees, Random Forest, LightGBM, Gradient Boosting, and AdaBoost. They focused on identifying hazardous asteroids using NASA's dataset of 90,836 asteroids, analyzing features like estimated diameter, relative velocity, and miss distance. The study found that Random Forest achieved the highest accuracy , making it the most reliable model for this task, while AdaBoost had the lowest accuracy . This research highlights the value of advanced machine learning techniques in enhancing asteroid hazard assessment.

Pasko's [9] study uses machine learning to predict orbital parameter combinations for undiscovered potentially hazardous asteroids (PHAs). By applying Support Vector Machines (SVM) with a Radial Basis Function (RBF) kernel, the research identifies high-concentration regions of PHAs within the near-Earth asteroid (NEA) population. The study utilized synthetic datasets of virtual asteroids, revealing an 'XX'-shaped region in the orbital parameter plane, containing about 50% of known PHAs with approximately 90% purity. Additionally, the study employed DBSCAN for clustering NEAs into distinct domains with high PHA concentrations, offering valuable insights for future PHA discovery surveys and asteroid-hunting missions.

Prasad, et al [10] developed a machine learning algorithm to enhance asteroid detection efficiency. By analyzing Pan-STARRS telescope data, their Python-based approach improves detection speed to minutes and achieves over 60% efficiency, compared to traditional methods. This new method, which processes FITS files and uses machine learning for photometry, not only accelerates asteroid identification but also increases accuracy, offering a significant advancement in tracking potentially hazardous asteroids.

## III. Methodology

The proposed system outlines a systematic approach to predicting hazardous asteroids using machine learning simulation and forecasting techniques. The Fig 1 depicts the methodology to detect and forecast the hazardous asteroids to earth.

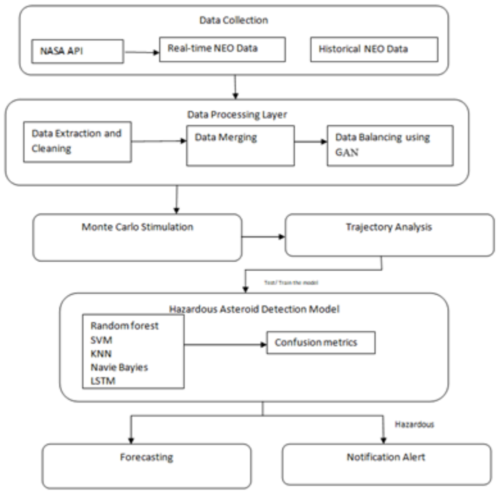


**Fig. 1. Block diagram to detect and forecast the hazardous asteroids to earth**

**Data Collection**

Real-time asteroid data is sourced from NASA's Near Earth Object (NEO) database, offering orbital, physical, and approach details. This data is collected using Python's requests library and stored in JSON format. Additionally, historical asteroid data from NASA's Jet Propulsion Laboratory, available on Kaggle, provides a comprehensive dataset with 7,072 rows and 12 key features, such as asteroid size, velocity, and proximity to Earth. This data, processed with the pandas library, forms the foundation for predicting potential hazards and guiding planetary defense efforts.

**Data Processing**

The real-time data is processed into a DataFrame, featuring key attributes like asteroid ID, name, absolute magnitude, estimated diameter, velocity, and miss distance. This DataFrame is merged with the existing dataset. Data cleaning includes format conversion, handling missing values, and normalizing features using MinMaxScaler. To tackle class imbalance, a Generative Adversarial Network (GAN) is utilized to generate synthetic samples of hazardous asteroids. The GAN creates synthetic data through its generator and evaluates it with the discriminator, improving balance between hazardous and non- hazardous entries in the dataset.

**Monte Carlo Simulation for Asteroid Trajectories**

A Monte Carlo simulation is performed to estimate the probability of an asteroid colliding with Earth. This simulation involved generating multiple asteroid trajectories based on their initial positions and velocities. Using the gravitational force formula, the simulation calculated the asteroid's trajectory over time to determine the minimum miss distance from Earth. The outcomes of the simulation included the probability of collision, the minimum miss distance, and the standard deviation of the miss distances.

**Model Developing**

The balanced dataset was used to train and implement a range of machine learning models to predict asteroid hazard levels. A Random Forest Classifier was trained with key features, including 'Neo Reference ID', 'Absolute Magnitude H', and 'Estimated Diameter (min and max)', among others, and was fine tuned using GridSearchCV to optimize its parameters. A Support Vector Machine (SVM) was trained to provide a comparative analysis, and a Gaussian Naive Bayes model was employed for its probabilistic approach. A K-Nearest Neighbors (KNN) model, set with neighbors=5, was also trained for its simplicity and effectiveness. For handling time-series data, an LSTM model was developed, reshaped to fit sequential data requirements, and consisted of an LSTM layer followed by a dense layer with dropout to mitigate overfitting. This LSTM model used the Adam optimizer and binary cross- entropy loss, with a validation set to monitor its performance.

**Evaluation Metrics**

The models were evaluated using accuracy, precision, recall, and F1-score, with confusion matrices providing a clear summary of prediction outcomes, including true positives, true negatives, false positives, and false negatives. The Random Forest model was saved for future use. To ensure model robustness and reduce dependency on specific train test splits, 10 fold cross validation was employed, offering a comprehensive performance assessment across different data subsets.

In binary classification, the confusion matrix provides a detailed view of model performance, with components including:

- True Positive (TP): Correctly predicted positives.
- True Negative (TN): Correctly predicted negatives.
- False Positive (FP): Negatives predicted as positives.
- False Negative (FN): Positives predicted as negatives. Performance metrics derived from the confusion matrix include:
- Accuracy: Overall performance, calculated as (TP + TN) /total observations.
- Precision: Ratio of TP to predicted positives (TP + FP).
- Recall: Ratio of TP to actual positives (TP + FN).
- F1-Score: Harmonic mean of precision and recall, balancing both metrics.

**Forecasting**

The LSTM model was employed to forecast future asteroid related metrics using historical data sequences. The model architecture consisted of two LSTM layers with 50 units each, followed by a dense layer with a linear activation function. Dropout layers were incorporated between the LSTM layers to mitigate overfitting. Training utilized the Adam optimizer with a mean squared error loss function, and a validation set was used to monitor performance. The model generate predictions for future miss distances.

**Notification Alert**

The real-time hazardous asteroid detection system integrates with NASA's Near-Earth Object (NEO) API to automatically fetch daily asteroid data, including features like magnitude, diameter, velocity, and miss distance. This data is processed by a pre-trained Random Forest model to assess hazard status. When a hazardous asteroid is identified, an automated email alert system notifies stakeholders includes essential information about the asteroid such as the asteroid's name, diameter, velocity, and closest approach distance, ensuring prompt awareness and action.

## IV. Results And Analysis

The Monte Carlo simulation assessed asteroid trajectories to evaluate potential collision risks with Earth, using gravitational effects and varied initial conditions.

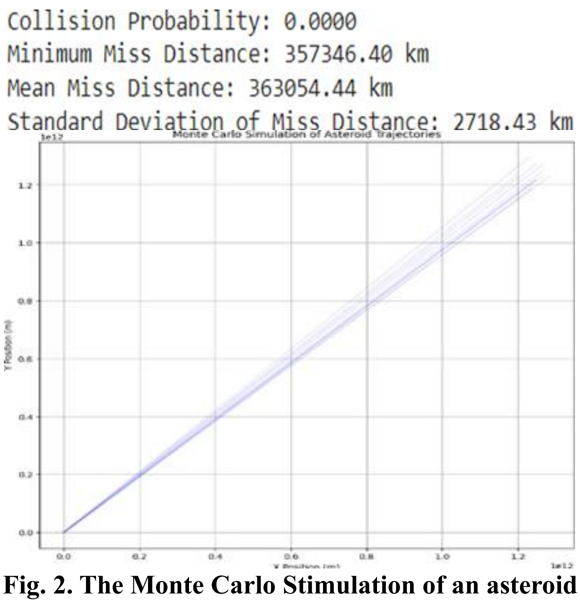


**Fig. 2. The Monte Carlo Stimulation of an asteroid**

The Fig 2 depicts the Monte Carlo simulation conducted to evaluate the potential collision risk of the asteroid (2006 CU10) with Earth. The simulation, which modeled 10 different trajectories based on varying initial conditions, showed that the asteroid has a collision probability of 0.0000, indicating no impact risk. The minimum miss distance observed was approximately 357,346.40 km, with a mean miss distance of about 363,054.44 km and a standard deviation of 2,718.43 km. The visualization of the trajectories revealed that the asteroid, moving with a significant x-component in its initial velocity, is likely to pass by Earth rather than impact it.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.94 | 0.98 | 0.96 | 1327 |
| 1 | 0.98 | 0.94 | 0.96 | 1331 |
| accuracy | | | 0.96 | 2658 |
| macro avg | 0.96 | 0.96 | 0.96 | 2658 |
| weighted avg | 0.96 | 0.96 | 0.96 | 2658 |

accuracy 96.3130173062453

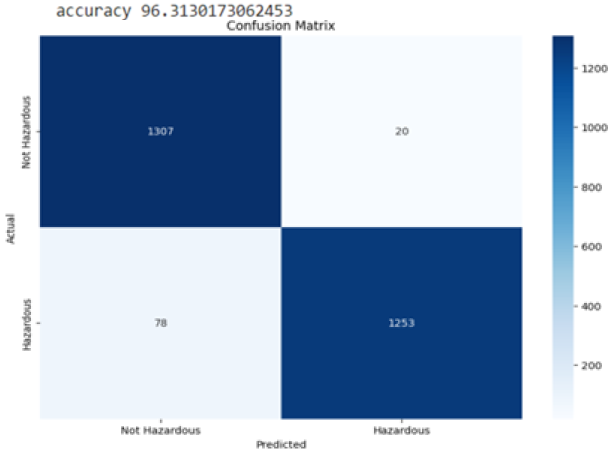


**Fig. 3. Random Forest Classifier**

The Fig 3 depicts that the Random Forest classifier achieved strong performance in predicting whether an asteroid is hazardous. With an overall accuracy of 96.31%, the model demonstrated high precision (0.94 for class 0 and 0.98 for class 1) and recall (0.98 for class 0 and 0.94 for class 1), resulting in a balanced F1-score of 0.96 for both classes. These metrics indicate that the model is equally proficient at identifying both hazardous and non-hazardous asteroids, providing reliable predictions across the test set of 2,658 instances.

SVM Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.93 | 0.75 | 0.83 | 1327 |
| 1 | 0.79 | 0.94 | 0.86 | 1331 |
| accuracy | | | 0.84 | 2658 |
| macro avg | 0.86 | 0.84 | 0.84 | 2658 |
| weighted avg | 0.86 | 0.84 | 0.84 | 2658 |

accuracy 84.46200150489089

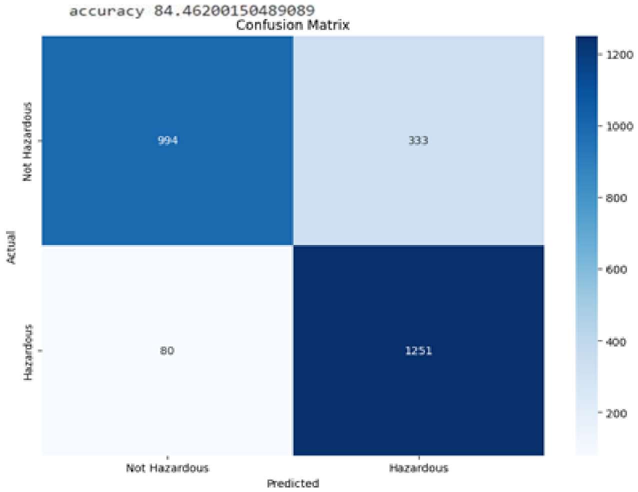


**Fig. 4. SVM Classification**

The Fig 4 depicts SVM classifier which achieved an accuracy of 84.46%. While it had a high recall for hazardous asteroids (0.94), its precision for this class was lower at 0.79, and it struggled more with non-hazardous asteroids, achieving a precision of 0.93 and a recall of 0.75. The F1-scores were

0.83 for non-hazardous and 0.86 for hazardous asteroids, indicating a less balanced performance compared to the Random Forest model.

```
Naive Bayes Classification Report:
              precision    recall  f1-score   support

           0       0.92      0.99      0.96      1327
           1       0.99      0.92      0.95      1331

    accuracy                           0.95      2658
   macro avg       0.96      0.95      0.95      2658
weighted avg       0.96      0.95      0.95      2658

accuracy 95.4477050413845
```

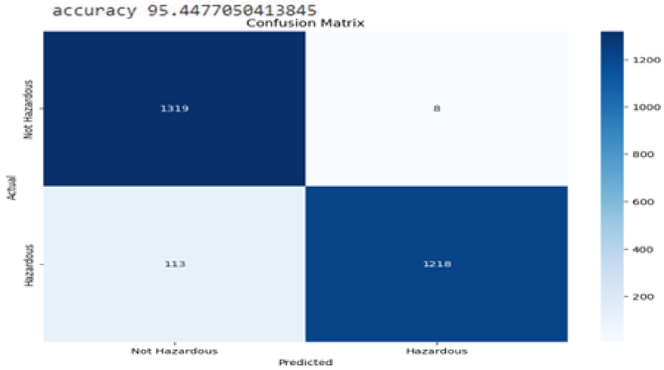


**Fig. 5. Navie Bayes Classification**

The Fig 5 depicts the Naive Bayes classifier achieved an accuracy of 95.45%. It exhibited a high precision for hazardous asteroids (0.99) but had a slightly lower recall for this class at 0.92. For non-hazardous asteroids, the precision was 0.92, and the recall was impressively high at 0.99. The F1-scores were 0.96 for non-hazardous and 0.95 for hazardous asteroids, showing a strong performance, although slightly less balanced compared to the Random Forest model, which demonstrated more consistent results across both classes.

```
KNN Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.97      0.96      1327
           1       0.97      0.94      0.96      1331

    accuracy                           0.96      2658
   macro avg       0.96      0.96      0.96      2658
weighted avg       0.96      0.96      0.96      2658

accuracy 95.71106094808127
```

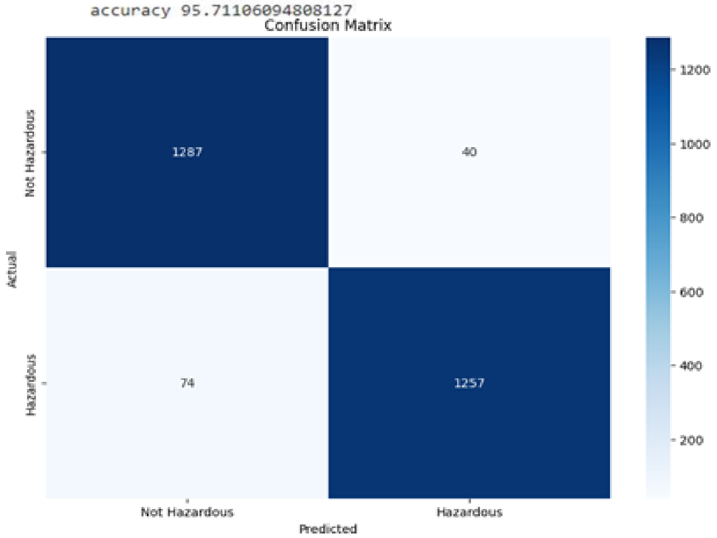


**Fig. 6. KNN Classification**

The Fig 6 represents K-Nearest Neighbors (KNN) classifier achieved an accuracy of 95.71%. It demonstrated strong performance with a precision of 0.95 and a recall of 0.97 for non-hazardous asteroids, and a precision of 0.97 and a recall of

0.94 for hazardous asteroids. The F1-scores were 0.96 for both classes, indicating a well-balanced performance. Compared to the Random Forest model, which had a slightly higher overall accuracy (96.31%), KNN performed very similarly, though the Random Forest model still showed a slight edge in terms of overall accuracy and balance across the two classes.

LSTM Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.93 | 0.99 | 0.96 | 1327 |
| 1 | 0.99 | 0.92 | 0.95 | 1331 |
| accuracy | | | 0.95 | 2658 |
| macro avg | 0.96 | 0.95 | 0.95 | 2658 |
| weighted avg | 0.96 | 0.95 | 0.95 | 2658 |

accuracy 95.41008276899925

**Fig. 7. Fig 7 LSTM Classification**

The Fig 7 depicts LSTM classifier achieved an accuracy of 95.41%. It demonstrated strong performance with a precision of 0.93 and a recall of 0.99 for non-hazardous asteroids, and a precision of 0.99 and a recall of 0.92 for hazardous asteroids.

The F1-scores were 0.96 for non-hazardous and 0.95 for hazardous asteroids, indicating a well-balanced performance. While the LSTM model's accuracy is very close to that of the Random Forest model (96.31%), the Random Forest model still slightly outperforms LSTM, particularly in its balanced precision and recall across both classes.

| | Date | Forecasted Miss Distance (kilometers) |
|---|---|---|
| 0 | 2024-09-25 | 0.504762 |
| 1 | 2024-09-26 | 0.404691 |
| 2 | 2024-09-27 | 0.501904 |
| 3 | 2024-09-28 | 0.508086 |
| 4 | 2024-09-29 | 0.521983 |
| 5 | 2024-09-30 | 0.524797 |
| 6 | 2024-10-01 | 0.526647 |
| 7 | 2024-10-02 | 0.527188 |
| 8 | 2024-10-03 | 0.527446 |
| 9 | 2024-10-04 | 0.527535 |

**Fig. 8. Forecasting the Miss Distance (kilometer) for one of the asteroid**

The LSTM model was employed to forecast, the Fig 8 depicts the forecast of the miss distance of asteroid (2021 LL15) over a 10-day period of its close approach date, from September 25, 2024, to October 04, 2024. The forecasted miss distances suggest a gradual increase in the distance between the asteroid and Earth during this period. Initially, on September 25, 2024, the model predicts a miss distance of 0.504762 kilometers, which slightly decreases the following day to 0.404691 kilometers. Over the next several days, the predicted miss distance fluctuates, with values such as 0.508086 kilometers on September 28, 2024 and gradually increasing to 0.527535 kilometers by October 04, 2024. These fluctuations suggest that the asteroid's distance from Earth will vary slightly over this time, but overall it is expected to stay relatively close.

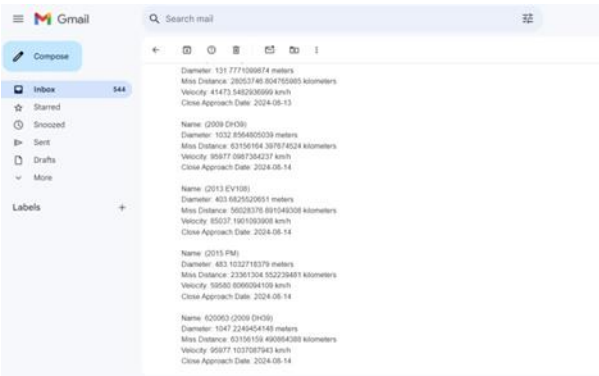


**Fig. 9. Alert Notification**

The Fig 9 depicts a real-time hazardous asteroid detection system. It integrates with NASA's Near Earth Object (NEO) API to fetch data on asteroids approaching Earth on a specific date. The code then uses a pre-trained Random Forest model to classify the asteroids based on features such as absolute magnitude, estimated diameter, relative velocity, and miss distance. When the model predicts that any of the fetched asteroids are potentially hazardous, the system generates an email alert. This email includes details such as the asteroid's name, diameter, miss distance, velocity, and close approach date. An alert send to a specified recipient via email.

## V. Conclusion

As the need for early detection and mitigation of potential threats from near-Earth objects grows, this study introduces a comprehensive approach to predicting hazardous asteroids using advanced machine learning techniques. With the increasing urgency to protect Earth from potential impact events, accurately classifying and forecasting hazardous asteroids is essential. The study involves collecting and processing both real-time and historical asteroid data, extracting key features, and training various machine learning models to predict hazardous asteroids where Random Forest Classifier obtained strong performance with accuracy of 96.31%. Monte Carlo simulations are employed to estimate the collision risks by generating multiple asteroid trajectories, assessing the minimum miss distance, and calculating the probability of collision. Additionally, the study incorporates a forecasting module using Long Short-Term Memory (LSTM) networks to predict future asteroid behavior based on historical and simulated data. To address class imbalance in the dataset, a Generative Adversarial Network (GAN) is utilized to create synthetic samples for hazardous asteroids, enhancing the balance between hazardous and non-hazardous entries. The system also includes a notification alert mechanism that sends emails to relevant stakeholders when a potentially hazardous asteroid is detected, ensuring timely and effective response to potential threats. Future improvements could explore integrating multisensor data for enhanced asteroid tracking and predictions.

## References

[1] Surekha, Nemmaluri Purna, Preethi Nanjundan, And Shahid Bashir. ”Prediction Of Hazardous Asteroids Using Machine Learning.” In 2024 International Conference On Emerging Systems And Intelligent Comput- Ing (ESIC), Pp. 169-174. IEEE, 2024.

[2] Bahel, Vedant, Pratik Bhongade, Jagrity Sharma, Samiksha Shukla, And Mahendra Gaikwad. ”Supervised Classification For Analysis And Detection Of Potentially Hazardous Asteroid.” In 2021 International Conference On Computational Intelligence And Computing Applications (ICCICA), Pp. 1-4. IEEE, 2021.

[3] Sandeep, Avala, Panta Jeevan Reddy, And T. Madhu Perkin. ”Planetary Defense Using Machine Learning Early Warning Systems For Hazardous Asteroids.” In 2024 International Conference On Advances In Data Engineering And Intelligent Computing Systems (ADICS), Pp. 1-6. IEEE, 2024.

[4] Bhavsar, Rushir, Nilesh Kumar Jadav, Umesh Bodkhe, Rajesh Gupta, Sudeep Tanwar, Gulshan Sharma, Pitshou N. Bokoro, And Ravi Sharma. ”Classification Of Potentially Hazardous Asteroids Using Supervised Quan- Tum Machine Learning.” IEEE Access (2023).

[5] Si, Anish. ”Hazardous Asteroid Classification Through Various Machine Learning Techniques.” International Research Journal Of Engineering And Technology 7, No. 3 (2020): 5388-5390.

[6] Sharma, Tushar, Shamneesh Sharma, Ajay Sharma, Aman Kumar, Arun Malik, And Abhimanyu Sharma. ”Asteroid Hazard Prediction Using Machine Learning: A Comparative Analysis Of Different Algorithms.” In 2023 3rd International Conference On Advancement In Electronics & Communication Engineering (AECE), Pp. 673-677. IEEE, 2023.

[7] Chomette, Gregoire, Lorien Wheeler, And Donovan Mathias. ”Machine Learning For The Prediction Of Local Asteroid Damages.” Acta Astronautica 219 (2024): 250-263.

[8] Malakouti, Seyed Matin, Mohammad Bagher Menhaj, And Amir Abol- Fazl Suratgar. ”Machine Learning Techniques For Classifying Dangerous Asteroids.” Methodsx 11 (2023): 102337.

[9] Pasko, Vadym. ”Prediction Of Orbital Parameters For Undiscovered Po- Tentially Hazardous Asteroids Using Machine Learning.” In Stardust Final Conference: Advances In Asteroids And Space Debris Engineering And Science, Pp. 45-65. Springer International Publishing, 2018.

[10] Prasad, LVR Chaitanya, Thadisina Anjali Shivani Reddy, And Bhu- Vaneshwari Kashi. ”Asteroid Detection Using Machine Learning Algo- Rithm.” Communications Of The Byurakan Astrophysical Observatory 67 (2020): 329-334.