

# Machine Learning Prediction of Space Debris Collision Risk Using Orbital Data

Neel Prasad  
Round Rock High School  
Austin, USA  
neelprasad2008@gmail.com

Subham Kumar Jalan  
*iResearcher*  
Bangalore, India  
subham@iresearcher.in

Kavitha Sharanappa Gudadur  
*iResearcher*  
Bangalore, India  
kavitha@iresearcher.in

**Abstract**—The number of satellites and debris objects in low Earth orbit has increased very fast in recent years. Because more and more objects are continuously entering orbital regions, the chances of collision events are also becoming higher. In this work, machine learning methods were used on publicly available orbital datasets to study satellite collision risk prediction. Orbital parameters such as inclination, eccentricity, mean motion, right ascension, and mean anomaly were considered during the analysis. Random Forest and Gradient Boosting algorithms were applied for separating collision cases into low-risk and high-risk categories. The models achieved good classification accuracy using the orbital features available in the dataset. Feature importance analysis also showed that inclination contributed more toward prediction compared to several other orbital parameters. Overall, the results suggest that machine learning methods can help future satellite monitoring and space traffic management systems for improving collision risk analysis.

**Index Terms**—Space Debris, Collision Risk Prediction, Machine Learning, Orbital Mechanics, Random Forest, Space Situational Awareness

## I. INTRODUCTION

The number of satellites and space debris orbiting Earth has increased significantly over the last few years. As more objects continue entering orbit, the chances of collision events are also increasing. Such collisions may damage operational satellites and can additionally generate new debris fragments, making future space missions more risky. Even very small debris particles can become dangerous because they travel at extremely high velocities in orbit. If these objects collide with spacecraft systems, serious technical damage and financial loss may occur. Most existing collision prediction methods still rely mainly on mathematical calculations and predefined warning thresholds for identifying possible collision situations. Although these methods are still useful, handling the continuously increasing amount of orbital data is becoming more difficult using only traditional approaches.

Several studies in recent years explored how artificial intelligence and machine learning can support space situational awareness and satellite collision prediction. Cao et al. [1] studied machine learning approaches for reducing orbit prediction error in space debris systems, but only a small number of debris categories were included in their experiments. Catulo et al. [2] used Hidden Markov Models with ESA conjunction data messages (CDMs) for predicting collision probability, although the method mainly depended on a single orbital

parameter. NASA CARA [3] also evaluated several AI-based techniques using a dataset containing more than 450,000 CDMs for satellite collision avoidance applications. Even though the dataset was large, many of the developed models were difficult to interpret for practical operational use.

Random Forest and LSTM models were combined by Pinto et al. [4] to estimate collision probability, but the framework showed lower performance for rare high-risk collision cases. Zhu et al. [5] investigated Random Forest and Gradient Boosting approaches for orbital covariance prediction using ESA satellite data; however, the dataset used for testing was relatively limited in size. Garg et al. [6] presented a Graph Neural Network model for fast collision risk prediction in large satellite constellations, although much of the training data was generated artificially instead of using real orbital observations.

Zhang et al. [7] applied LSTM-based sequence learning methods for improving short-term orbit prediction in low Earth orbit. While the prediction accuracy improved, the work did not examine debris objects having uncertain physical properties. Acciarini et al. [8] developed an open-source machine learning library for spacecraft collision avoidance using CDM datasets, but the framework was not connected with a real-time operational system. Huang et al. [9] reviewed different AI methods used in space situational awareness applications, though the reliability of these methods during emergency conjunction situations was not deeply investigated. Musci et al. [10] also reviewed deep learning techniques for SSA applications in mega-constellation systems, but real collision probability benchmarking was not included in their study.

Kelso et al. [11] showed that sharing operator information between organizations could help reduce false collision alerts, although scaling this approach for future mega-constellation environments may become difficult. Merisio et al. [12] introduced a software framework for close-approach detection and collision avoidance planning, but predictive machine learning capability was absent in the system. Li et al. [13] proposed neural-network-based methods for improving orbit error estimation and collision probability prediction. However, the performance decreased for debris objects with uncertain shape and mass properties. Caldas et al. [14] reviewed several machine learning methods related to orbit estimation, whereas Ortega et al. [15] studied orbit prediction using noisy and incomplete datasets. Still, neither work directly focused on

real operational collision probability forecasting.

Lice et al. [16] reviewed several machine learning approaches used for automated debris tracking in low Earth orbit, but only a limited comparison using standard datasets was carried out in their work. Uriot et al. [17] discussed the ESA spacecraft collision avoidance challenge, although the dataset used in the competition represented only a small number of orbital situations. Vasile et al. [18] focused on long-term orbital debris growth and Kessler syndrome evolution, but machine learning or AI-based prediction methods were not included in the framework. Gomez et al. [19] explored reinforcement learning techniques for autonomous spacecraft collision avoidance, though human assistance was still needed during more difficult operational cases. Karthikeya et al. [20] applied deep learning and support vector machine methods for debris detection and trajectory prediction; however, the experiments were mainly performed using simulated telescope images instead of real observation data. The main objectives are as follows:

- To collect orbital satellite data from publicly available tracking databases and prepare the dataset for analysis.
- To study orbital parameters such as inclination, eccentricity, mean motion, and other trajectory-related characteristics related with satellite motion.
- To perform exploratory data analysis for identifying patterns between orbital features and collision risk levels.
- To develop machine learning models for classifying high-risk and low-risk satellite collision cases.
- To compare Random Forest and Gradient Boosting models using common classification evaluation metrics.
- To examine whether machine learning methods may support future satellite collision monitoring and space traffic management applications.

## II. METHODOLOGY

### A. Data Download

To help with the satellite tracking part of this work, Two-Line Element (TLE) orbital datasets were collected from public CelesTrack repositories. The datasets contain real satellite orbital records in text format and provide information related to satellite position, motion, and orbital trajectory behavior. These sources were selected because they are publicly accessible and commonly used in satellite tracking studies. Using openly available datasets also allows other researchers to reproduce and compare results more easily. Overall, the collected TLE data provided a suitable base for studying orbital movement patterns and collision risk analysis.

### B. Data Preprocessing

Raw TLE data was preprocessed and converted into a tabular format with Pandas in Python. The original TLE data contain information in line format, so important orbital parameters were collected into separate columns for analyzing. Features in the data such as inclination, eccentricity, mean motion, right ascension of the ascending node, argument of

perigee, and mean anomaly were generated from the line-format data. Collision risk labels were then created using an approach involving weighted orbital risk scoring.

### C. Data Analysis

Statistical analysis was carried out on the orbital data to understand the satellite information. Different graphs were used to show how different values like inclination, eccentricity, and mean motion, were spread in the data. Correlation analysis also helped show whether some orbital features were related to others. Scatter plots and boxplots were used to study the relationships between different orbital parameters and to identify unusual or extreme values present in the dataset. Feature importance analysis was also performed to understand which orbital characteristics had greater influence on collision risk prediction. These analyses helped in improving the overall understanding of the dataset before developing the machine learning models.

### D. Machine Learning

Supervised machine learning techniques were applied for classifying satellite collision risk using orbital parameters as input features. Random Forest and Gradient Boosting algorithms were selected because they are capable of learning nonlinear patterns from complex orbital datasets. After pre-processing, the dataset was divided into training and testing sets for model development and evaluation. Classification performance was measured using accuracy, precision, recall, and F1-score metrics. Confusion matrices and ROC curve analysis were also examined in order to better understand the prediction capability of each model. Finally, both machine learning approaches were compared to determine which model produced better performance for the orbital collision risk dataset.

## III. RESULTS

Fig. 1 shows the distribution of satellite mean motion values from the orbital dataset used in this study. Many satellites are grouped within a similar range of mean motion values, especially satellites operating in low Earth orbit regions. Most values appear around the middle area of the graph, while smaller numbers of satellites are present near the lower and higher ends. This pattern suggests that satellites located within similar orbital regions often follow related orbital movement behavior instead of completely random motion patterns. This helps show that satellite motion is linked to the type of orbit they use.

Fig. 2 shows how orbital eccentricity values are distributed across the satellite dataset. Most satellites have very low eccentricity values, which means that their orbits are close to circular around Earth. Circular orbits are generally preferred in satellite operations because they provide better stability and make control and positioning easier during operation. Only a small number of satellites show higher eccentricity values, which may be linked to specific mission requirements

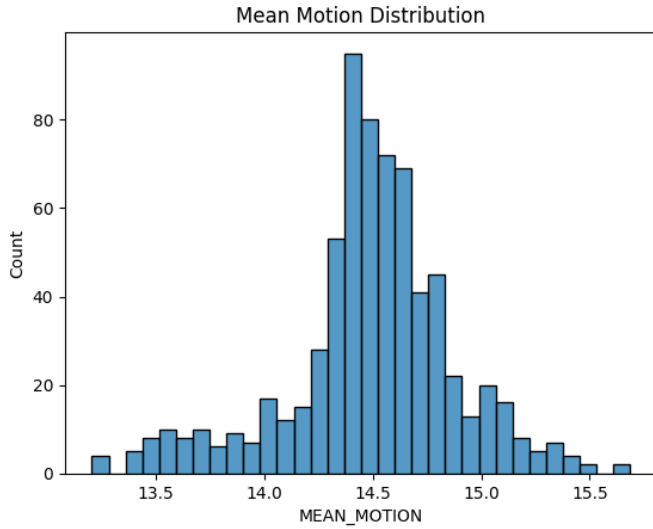


Fig. 1. Distribution of satellite mean motion values obtained from orbital datasets.

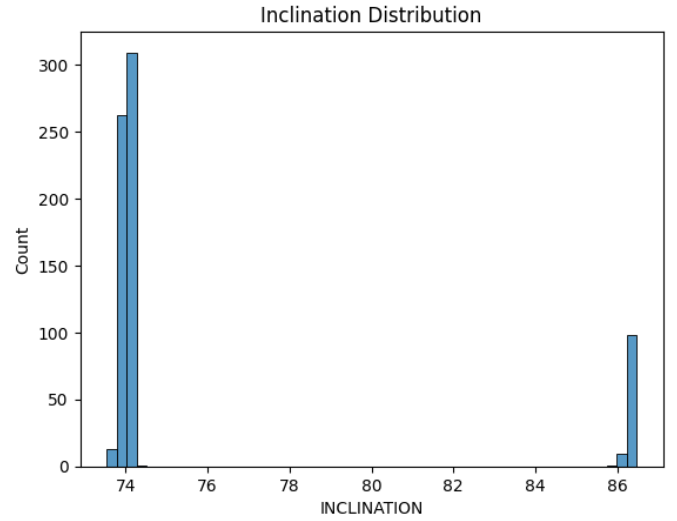


Fig. 3. Distribution of orbital inclination values.

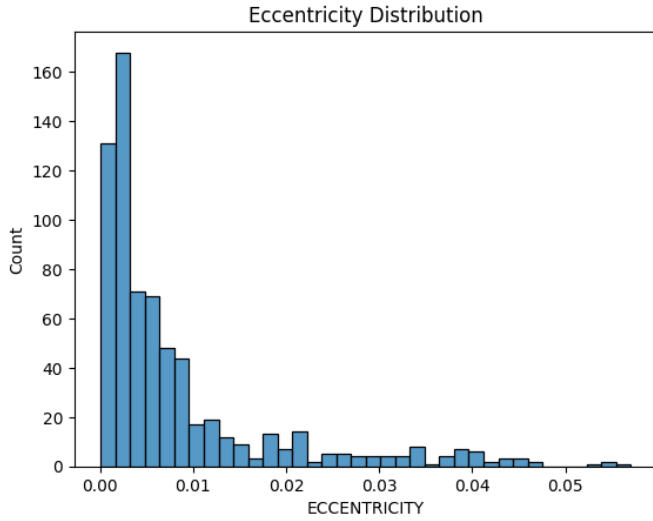


Fig. 2. Distribution of orbital eccentricity values across satellites.

involving more elliptical orbital paths. This shows that most satellites are placed in steady and simple orbit paths.

Fig. 3 shows how orbital inclination values are spread across the dataset. The satellite inclination values are grouped around certain orbit bands instead of being spread out evenly. Polar and sun-synchronous orbit groups can also be seen in the graph. These clear groups show that satellites are placed in orbit in planned patterns. This means that orbit angle is selected based on the job of the satellite.

Fig. 4 shows the correlation values between orbital parameters and collision risk. Inclination appears to have a stronger positive relationship with collision risk compared to the other features. Mean motion and eccentricity show smaller negative correlations in the dataset. From the heatmap, it becomes easier to observe which orbital parameters may have greater

influence on collision risk prediction. This helps identify which features may be more useful for the model.

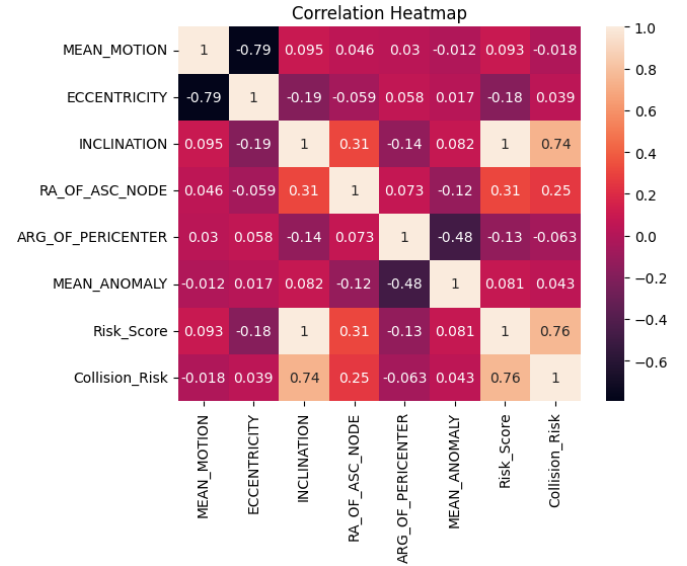


Fig. 4. Correlation heatmap showing relationships among orbital parameters and collision risk.

Fig. 5 presents the scatter plot between orbital inclination and mean motion for different collision risk categories. Distinct groupings of collision risk classes can be observed in several orbital regions of the graph. Certain regions contain a larger concentration of high-risk objects compared to others, which suggests that some orbital conditions may contribute more toward possible collision events. The distribution pattern in the scatter plot also indicates that inclination and mean motion are useful features for separating collision risk classes within the dataset. This shows that some orbit areas may need closer tracking.

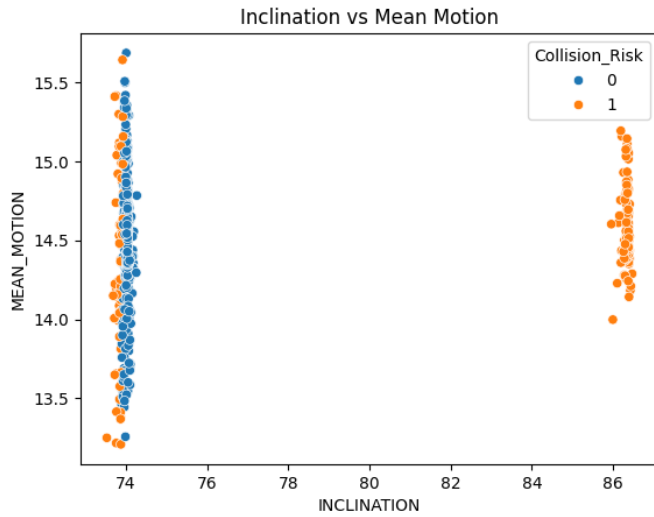


Fig. 5. Scatter plot comparing inclination and mean motion with collision risk classes.

Fig. 6 shows the eccentricity distribution for different collision risk groups using a boxplot. The high-risk group contains a wider spread of eccentricity values compared with the low-risk category, which indicates larger variation in orbital paths. A few outlier points are also visible in the graph and may represent satellites having less common orbital behavior. From the observed trend, eccentricity appears to contribute toward collision risk in certain orbital situations. This suggests that unusual orbit shapes may increase risk in some cases.

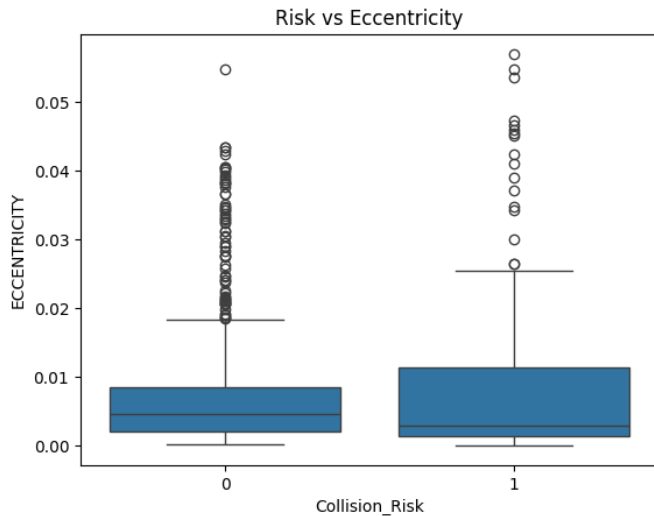


Fig. 6. Boxplot illustrating eccentricity distribution across collision risk classes.

Fig. 7 presents the confusion matrix obtained from the Random Forest collision risk classifier. Most high-risk and low-risk cases were predicted correctly during testing. Only a small number of false-positive and false-negative predictions are visible in the matrix. These results indicate that the Random

Forest model was able to separate collision risk categories reasonably well using the selected orbital parameters. This shows that the model made mostly correct predictions.

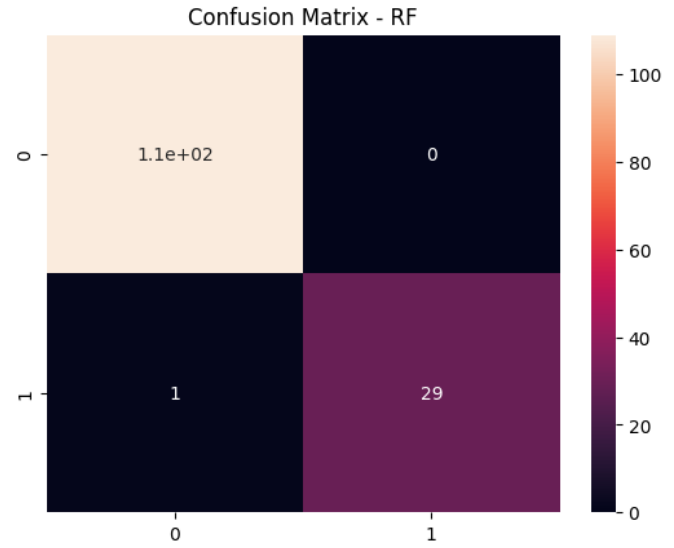


Fig. 7. Confusion matrix of the Random Forest collision risk classifier.

Fig. 8 shows the ROC curve generated for the Random Forest model. The curve remains close to the upper-left region of the graph for most threshold values, which indicates good separation between high-risk and low-risk collision cases. The model maintained a higher true positive rate while keeping the false positive rate comparatively lower. The larger area under the curve also suggests that the model performed well for satellite collision risk classification using orbital features. This means the model was good at telling the two risk groups apart.

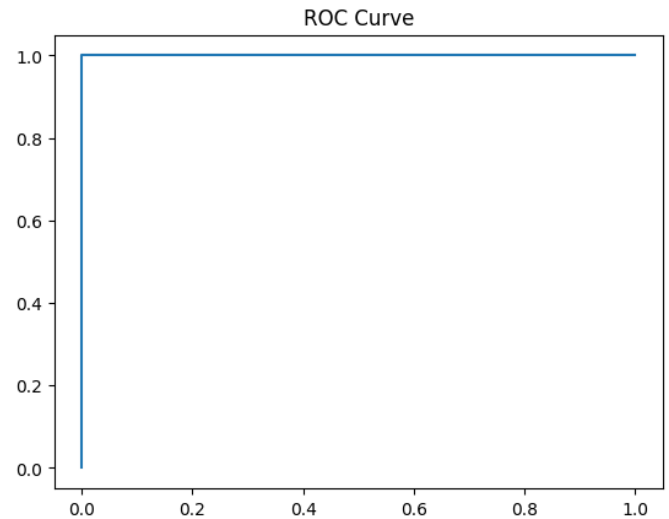


Fig. 8. ROC curve evaluating the classification performance of the Random Forest model.

Fig. 9 shows the feature importance values obtained from

the Random Forest model. Among all the orbital parameters, inclination contributed the most toward collision risk prediction. Mean motion and RAAN also influenced the prediction results, although their impact was smaller compared to inclination. The results suggest that orbital geometry and satellite positioning characteristics play an important role in estimating possible collision risk. This shows that the model depends most on the orbit angle information.

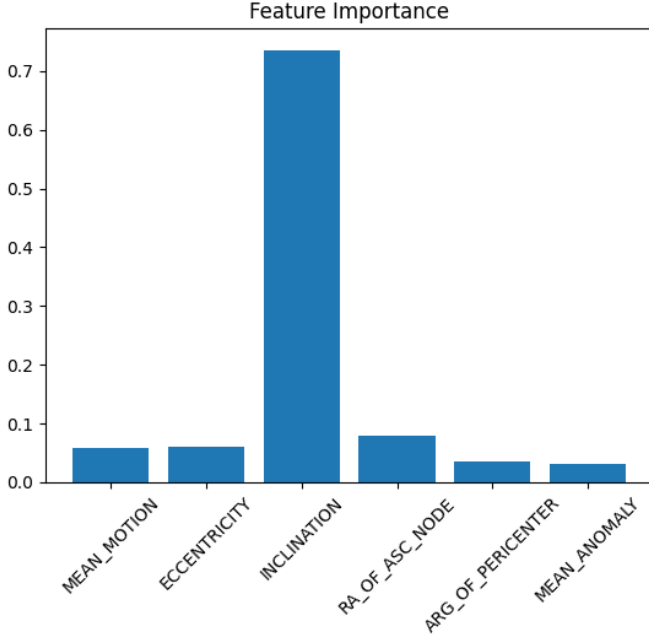


Fig. 9. Feature importance scores generated by the Random Forest model.

Fig. 10 shows the comparison between Random Forest and Gradient Boosting models for collision risk classification. Both machine learning models achieved almost similar accuracy values during testing. The results suggest that ensemble learning approaches can work effectively for satellite collision risk prediction using orbital datasets. This means both models can be useful choices for this type of prediction task.

#### IV. DISCUSSION

Useful patterns from the orbital dataset were learned by the machine learning models used in this work, and collision risk cases were classified with good accuracy. Strong results were achieved during testing by both Random Forest and Gradient Boosting because nonlinear relationships present in orbital data can be handled by these algorithms. A greater influence on the prediction was shown by inclination than by the other orbital parameters, according to the feature importance analysis. Nearby orbital regions may be passed through more frequently by satellites with similar orbital inclinations, which is one possible reason. Collision risk may be increased because satellites may be brought close to one another or because more satellites may be located in the same orbital area.

Clusters within certain orbital zones were observed in the exploratory analysis, which may increase the possibility of

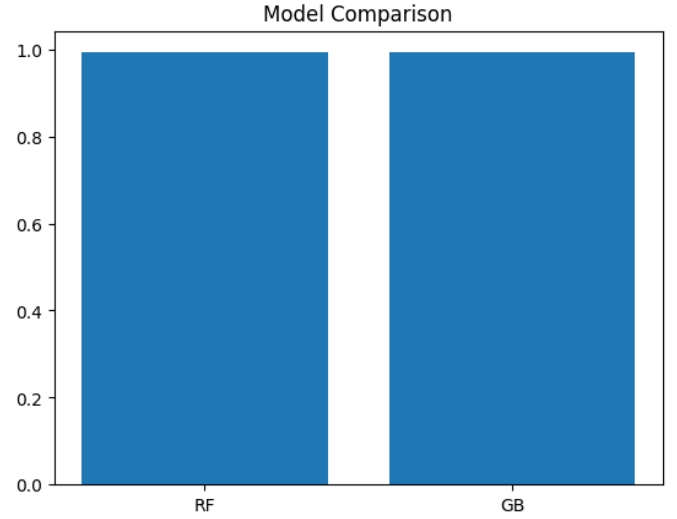


Fig. 10. Performance comparison between Random Forest and Gradient Boosting models.

close approaches and collision events. Some unusual orbital cases were also identified through scatter plots and boxplot analysis. Although only publicly available orbital information was used in this study, good classification performance was still achieved by the framework. Overall, the potential usefulness of machine learning approaches for future satellite monitoring and collision warning systems in crowded orbital environments was suggested by the results.

TABLE I  
PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS

| Model             | Accuracy | Precision | Recall | F1-Score |
|-------------------|----------|-----------|--------|----------|
| Random Forest     | 0.99     | 0.99      | 0.99   | 0.99     |
| Gradient Boosting | 0.99     | 0.99      | 0.99   | 0.99     |

TABLE II  
RANDOM FOREST CLASSIFICATION REPORT

| Class        | Precision | Recall | F1-Score | Support |
|--------------|-----------|--------|----------|---------|
| 0            | 0.99      | 1.00   | 1.00     | 109     |
| 1            | 1.00      | 0.97   | 0.98     | 30      |
| Accuracy     | 0.99      |        |          |         |
| Macro Avg    | 1.00      | 0.98   | 0.99     | 139     |
| Weighted Avg | 0.99      | 0.99   | 0.99     | 139     |

#### V. CONCLUSIONS

This study investigated the use of machine learning methods for predicting satellite collision risk using publicly available orbital datasets. Random Forest and Gradient Boosting models were trained using orbital parameters such as inclination, eccentricity, mean motion, and other orbital geometry-related features. The models achieved good classification accuracy during testing and were able to identify useful collision-related patterns from the dataset. The analysis also showed that inclination had a stronger influence on prediction compared with

several other orbital parameters. Exploratory analysis further helped in understanding how satellites are distributed across different orbital regions and how certain orbital conditions may increase the possibility of collision events. Although a simplified collision risk labeling approach was used in this work, the framework still showed that machine learning methods can support future satellite monitoring and space traffic management systems. In the future, the work can be improved further by using larger datasets, real conjunction event records, graph-based learning models, and time-dependent orbital prediction approaches.

## REFERENCES

- [1] B. Li, J. Huang, Y. Feng, F. Wang, and J. Sang, "A machine learning-based approach for improved orbit predictions of leo space debris with sparse tracking data from a single station," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 6, pp. 4253–4268, 2020.
- [2] J. S. Catulo, M. Guimaraes, et al., *Predicting the Probability of Collision of a Satellite with Space Debris: A Bayesian Machine Learning Approach*. arXiv preprint arXiv:2311.10633, 2023.
- [3] NASA Conjunction Assessment Risk Analysis, *NASA CARA Compendium for Artificial Intelligence and Machine Learning for Satellite Collision Avoidance*. NASA Technical Reports Server, 2025.
- [4] L. Tulczyjew, M. Myller, and M. Kawulok, *Predicting Risk of Satellite Collisions Using Machine Learning*. Advances in Space Research, Elsevier, 2021.
- [5] S. Zhu et al., *Implementation and Comparison of Data-Based Methods for Collision Avoidance in Satellite Operations*. 8th European Conference on Space Debris, ESA, 2021.
- [6] Garg et al., *Graph Neural Networks for Real-Time Collision Risk Assessment in Large Satellite Constellations*. arXiv preprint, 2024.
- [7] Zhang et al., *Improving LEO Short-Term Orbit Prediction Using LSTM Neural Network*. Advances in Space Research, Elsevier, 2025.
- [8] G. Acciarini, F. Pinto, F. Letizia, J. A. Martinez-Heras, K. Merz, C. Bridges, and A. G. Baydin, *Kessler: A Machine Learning Library for Spacecraft Collision Avoidance*. 8th European Conference on Space Debris, ESA, 2021.
- [9] Huang et al., *Space Situational Awareness Systems: Bridging Traditional Methods and Artificial Intelligence*. Acta Astronautica, Elsevier, 2024.
- [10] Musci et al., *Deep Learning Methods for Space Situational Awareness in Mega-Constellations Satellite-Based IoT Networks*. Sensors, MDPI, 2023.
- [11] T. Kelso, *Improved Conjunction Analysis via Collaborative Space Situational Awareness*. AMOS Technical Conference, 2008.
- [12] Merisio et al., *Conjunction Analysis Software Suite for Space Surveillance and Tracking*. Aerospace, MDPI, 2023.
- [13] Li et al., *Reduction of Space Debris Collision Prediction Uncertainty Based on Q-Sat Precise Orbit*. Space: Science and Technology, Science Partner Journals, 2023.
- [14] F. Caldas and C. Soares, *Machine Learning in Orbit Estimation: A Survey*. Acta Astronautica, Elsevier, 2024.
- [15] Ortega et al., *Satellite Orbit Prediction Using Big Data and Soft Computing Techniques to Avoid Space Collisions*. Mathematics, MDPI, 2021.
- [16] S. Lice, *Machine Learning Algorithms for Automated Space Debris Tracking and Collision Avoidance in Low Earth Orbit*. Hilaris Publisher, 2024.
- [17] Uriot et al., *Spacecraft Collision Avoidance Challenge: Design and Results of the ESA Competition*. arXiv preprint, 2022.
- [18] Vasile et al., *KESSYM: A Stochastic Orbital Debris Model for Evaluation of Kessler Syndrome Risks and Mitigations*. Journal of Student Research, 2023.
- [19] Gomez et al., *Spacecraft Autonomous Decision-Planning for Collision Avoidance: A Reinforcement Learning Approach*. arXiv preprint arXiv:2310.18966, 2023.
- [20] T. Karthikeya et al., *AI-Driven Space Debris Detection and Trajectory Prediction Using Deep Learning and Orbital Mechanics*. ResearchGate, 2025.