International Journal of Science, Strategic Management and Technology
Volume **02** Issue **08** August- **2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

OPEN ACCESS

Machine Learning-Based Population-Level Risk Ranking of Space Objects Using Orbital and Debris-Environment Features

A Physics-Informed Proxy-Risk Classification Framework for Standalone Catalog Screening

Ms. Juluru Girisha Varshini,
B.Tech 3rd Year CSE, SRM University, Andhra Pradesh
girishajuluru@gmail.com

https://doi.org/10.55041/ijsmt.v2i8.069

Cite this Article: Varshini, J. G. (2026). Machine Learning-Based Population-Level Risk Ranking of Space Objects Using Orbital and Debris-Environment Features. International Journal of Science, Strategic Management and Technology, *02*(8), 1-5. https://doi.org/10.55041/ijsmt.v2i8.069

Abstract— The increasing number of cataloged resident space objects (RSOs) in low Earth orbit (LEO) has heightened concerns regarding the long-term sustainability of orbits and the potential for a self-perpetuating collisional cascade, known as the Kessler Syndrome. Existing machine learning approaches to space situational areness (SSA) are overwhelmingly reactive; they operate on pairwise Conjunction Data Messages (CDMs) issued only after a close approach has already been detected by a surveillance network, or they correct single-object orbit-propagation errors without evaluating the surrounding debris environment. In contrast, macro-scale stochastic and deterministic evolutionary models characterize population-level sustainability but cannot rank individual cataloged objects. This study addresses the resulting gap between object-level, CDM-dependent risk assessment and population-level, density-based environmental modeling by developing a pairwise-independent, physics-informed proxy-risk classification framework built entirely from standalone Two-Line Element (TLE) catalog data. A dataset of 34,136 currently orbiting payloads, debris fragments, and rocket bodies was constructed from the SATCAT and orbital element catalog, and twenty predictive features describing orbital geometry, local orbital crowding, altitude shell density, debris and rocket body neighbor density, and estimated relative velocity were engineered. A composite Overall Proxy-Risk Score, combining collision-risk and debris-cascade-risk proxies, was used to assign each object to a LOW, MEDIUM, or HIGH physics-informed proxy-risk category. Logistic Regression, Random Forest, and Gradient Boosting classifiers were evaluated using five-fold stratified cross-validation and an untouched 20% holdout test set (6,828 objects) against a majority-class dummy baseline. Logistic Regression achieved the highest holdout performance (accuracy = 0.9985, F1 = 0.9985, ROC-AUC = 0.9999), substantially exceeding the dummy baseline (accuracy = 0.4783). Feature importance analysis identified the relative velocity and local debris-neighbor density as dominant predictors. Because the proxy-risk label is a deterministic function of several of the input features, the very high classification performance is interpreted as evidence that the models recovered the structure of the proxy-risk formulation rather than as evidence of collision-prediction skill; this dependency is discussed explicitly as a central limitation. The framework is proposed as a lightweight, CDM-independent screening layer for prioritizing objects ahead of a detailed physics-based conjunction assessment, and not as a replacement for it.

Keywords— space debris, space situational awareness, orbital crowding, machine learning, proxy-risk classification, Two-Line Elements, Kessler Syndrome, collision-risk screening, Random Forest; Logistic Regression..

I. INTRODUCTION

Sustained growth in launch activity has resulted in an increasingly congested orbital environment. More than 19,000 trackable objects larger than 10 cm currently occupy low-Earth orbits, representing a cumulative mass of approximately 7,500 metric tons and originating from over 200 in-orbit breakup events [2]. Most of this population no longer serves any operational purpose; it consists of defunct payloads, spent upper stages, and fragmentation debris [2].

The long-term stability of this environment is governed by the dynamics described by Kessler in 1978. The Kessler Syndrome describes a threshold spatial density beyond which collisions between existing objects generate debris fragments that increase the probability of further collisions, producing an exponential, self-sustaining cascade capable of rendering entire orbital shells unusable [1], [15]. Although catastrophic accidental collisions have so far been rare, with the 2009 Iridium 33/Cosmos 2251 collision being the principal example [3], long-horizon evolutionary modeling indicates that LEO is already in an unstable regime, and that standard mitigation measures such as the IADC 25-year post-mission disposal guideline are necessary but insufficient to arrest population growth [4]. This has motivated growing interest in prioritization frameworks for active debris removal (ADR) and collision avoidance.

Two broad methodological traditions currently address the risk of orbital disease. The first operates at the level of individual close approaches: the Combined Space Operations Center (CSpOC) and other surveillance networks issue Conjunction Data Messages (CDMs) when a predicted miss distance falls below an operational threshold, and analytical algorithms such as Alfriend-Akella, Chan's, or Patera's method convert the encounter geometry and covariance into a probability of collision, Pc [7], [8], [9]. The second tradition models the debris environment in aggregate, using brute-force deterministic N-body propagation [14] or macro-scale stochastic, gas-kinetic approximations such as KESSYM [15] to project decadal- to centennial-scale population trends and inform space traffic policy.

Machine learning has been applied productively within both traditions — correcting TLE/SGP4 propagation error [6], regressing intermediate uncertainty parameters ahead of analytical Pc calculation [13], and classifying or regressing risk directly from CDM sequences [10], [11], [12]. In every case, however, the applicability of the model is contingent on the conjunction already having been detected and flagged. No reviewed study integrates individual, unconjuncted Keplerian elements with micro-scale local clutter descriptors — altitude-shell density, debris-neighbor density, and rocket-body-neighbor density — as direct machine-learning inputs for population-wide, standalone risk prioritization.

This study addresses this gap. It develops a pairwise-independent, physics-informed proxy-risk classification framework that operates directly on standalone TLE catalog data before any conjunction has been detected and ranks the entire tracked catalog rather than a single flagged encounter. The specific objectives of the study are to: (i) construct a population-level dataset of currently orbiting payloads, debris, and rocket bodies with engineered orbital crowding, debris density, and relative velocity features; (ii) define a transparent, physics-informed proxy risk score and a corresponding three-class risk category; (iii) train and rigorously validate interpretable classifiers to reproduce this categorization from unconjunctedcatalog features; and (iv) characterize which features drive the classification and critically examine the conditions under which such high classification performance is meaningful.

The central contribution of this study is not a claim of collision-prediction accuracy. As emphasized throughout this manuscript, the target variable is a constructed, physics-informed proxy risk category — a screening and ranking device — and the reported classification metrics describe how well the models reproduce that category, not how well they predict verified collision outcomes. The manuscript is organized as follows: Section II reviews related work; Section III describes the dataset; Section IV defines the proxy-risk framework; Section V describes the machine-learning methodology; Sections VI–VIII report cross-validation, holdout, and orbital-environment results together with feature interpretation; and Sections IX–XI discuss findings, limitations, and practical implications; Section XII concludes.

II. RELATED WORK AND LITERATURE REVIEW

This section synthesizes the state of the art in space debris modeling, TLE-based orbit propagation, physics-based conjunction assessment, and data-driven SSA, leading to the identification of the research gap that motivates the proposed framework.

A. Space Debris and the Kessler Syndrome

The expansion of space activity has raised concerns about the long-term sustainability of spaceflights. The tracked LEO population exceeds 19,000 objects larger than 10 cm [2], representing approximately 7,500 metric tons of material [2] accumulated from defunct payloads, spent rocket bodies, and fragmentation debris [2] generated by over 200 breakup events [2]. The Kessler Syndrome, first postulated in 1978, describes the point at which mutual collisions between existing objects generate fragments that trigger further collisions in a self-sustaining exponential chain reaction capable of rendering orbital shells unusable [1], [15]. Although catastrophic accidental collisions remain rare historically, with the 2009 Iridium 33/Cosmos 2251 event being the principal case [3], evolutionary projections indicate that LEO is already unstable and that standard mitigation guidelines, such as the IADC 25-year disposal rule, are necessary but insufficient [4]. This motivates the prioritization of objects for active debris removal and collision avoidance.

B. TLE-Based Orbital Analysis and Prediction

The Two-Line Element (TLE) catalog is the sole publicly available source of orbital information for resident space objects [5], propagated using the Simplified General Perturbations-4 (SGP4) analytical model [5]. SGP4's computational efficiency is achieved at the cost of accuracy: the omission of higher-order gravitational terms, solar radiation pressure, and atmospheric density variation causes the prediction error to degrade rapidly, particularly in LEO, with an along-track error exceeding 8 km after seven days of propagation [6]. Physics-based refinements, such as differential correction, hybrid propagators, and pseudo-observation fitting, improve accuracy but remain limited by environmental-modeling complexity, sparse tracking of uncooperative debris, and the mathematical rigidity of SGP4.

C. Physics-Based Conjunction and Collision-Risk Assessment

Operational collision avoidance relies on CDMs issued by the CSpOC when a predicted close approach falls within 1 km (LEO) or 5 km (GEO) [7]. A CDM specifies the predicted state vectors and covariances at the time of the closest approach in the RTN frame [7], and analytical algorithms — principally Alfriend-Akella, Chan's, and Patera's methods — reduce the three-dimensional collision integral to a two-dimensional Gaussian evaluated on the encounter (B-plane) under simplifying independence and constant-relative-velocity assumptions [7], [8], [9]. These methods are operationally constrained by severe class imbalances in real CDM streams, irregular tracking of chaser (debris) objects, and the consequent oscillation of the predicted Pc close to the time of the closest approach [13].

D. Machine Learning for Space Situational Awareness

The growth of large LEO constellations has outpaced traditional physical-filter-based SSA, motivating machine learning applications in orbit determination, prediction, and tracking. Deep and recurrent networks have been applied to orbit determination and relative-navigation tasks, and to orbit-prediction error correction, most notably by Li et al., who used Gradient Boosting Decision Trees and CNNs to correct SGP4 along-track, cross-track, and radial errors, reducing the 14-day along-track error by more than 75% [6]. Related studies on SVMs, ANNs, and Gaussian Processes have shown similar effectiveness in mining systematic, unmodeled perturbations.

E. Machine Learning for Collision and Conjunction Risk

The release of the ESA Kelvins Collision Avoidance Challenge dataset catalyzed the direct application of machine learning to conjunction risk [10]. Sequence models (LSTM/GRU) applied to chronological CDM streams performed poorly, overfitting under severe class imbalance (approximately 3% high-risk events) and non-random sampling [10]. Two-stage classical architectures have proven to be more robust. Tulczyjew et al. combined MLP classification ($Pc \geq 10^{-6}$) with Random Forest regression of continuous Pc for flagged high-risk events using miss distance, relative velocity, and energy-dissipation-rate features [11]. Fiore demonstrated that k-nearest neighbors and boosted trees outperformed deep recurrent networks for Go/No-Go and risk-instability detection, reaching 97% classification accuracy on high-risk conjunctions [12]. Metz et al. instead predicted intermediate chaser positional uncertainties (σR, σT, σN) at TCA using Random Forest and LSTM models on Keplerian elements and solar-flux features, then processed these through the Alfriend-Akella algorithm — a hybrid design offering greater physical consistency and robustness to irregular tracking [13].

F. Space-Debris Environment and Population Modelling

Population-scale sustainability analyses rely on evolutionary simulations, which are categorized into brute-force deterministic and macro-scale stochastic models. Deterministic approaches, exemplified by Nikolaev et al. (LLNL), numerically propagate every cataloged object above 10 cm over century-scale horizons with Jacchia-Bowman 2008 drag modeling and exhaustive N × N conjunction screening, treating position as ground truth and applying the NASA Standard Breakup Model upon collision [14], [16]. These simulations are physically precise but require enormous computational resources and neglect the tracking uncertainty [14]. Stochastic models such as KESSYM instead use ideal gas-kinetic approximations to evaluate the collision probability from spatial density over 250-year horizons without trajectory tracking [15], an approach mirrored in ESA's debris-index environmental reporting (collision probability × severity) and in evidence theory (Dempster-Shafer) treatments of epistemic tracking uncertainty [17].

G. Research Gap

A distinct analytical gap separates granular, CDM-dependent risk assessments from macro-scale, density-based population modeling. Granular methods can only evaluate risk once a specific close approach has already been detected and flagged [7], [10] and are therefore incapable of proactively prioritizing standalone, un-conjuncted objects. In contrast, macro-scale models stochastically simulate population-wide sustainability and Kessler syndrome timelines while ignoring individual Keplerian elements and object-specific dynamics [15] and cannot identify which specific objects within a congested shell warrant attention. No reviewed study integrates individual object-level Keplerian parameters with micro-scale local clutter densities — local crowding, debris-neighbor density, and rocket-body-neighbor density — as direct machine-learning inputs for standalone, population-level proxy-risk ranking. This study aims to address this gap by proposing a population-level machine-learning framework that screens and ranks cataloged objects from unconjugated TLE data, independent of pairwise CDMs, using a stable, physics-informed proxy-risk category (LOW/MEDIUM/HIGH) rather than a volatile, imbalanced Pc regression target.

H. Comparative Summary of Prior Work

Study	Dataset	Method	Population-level?	Key limitation
Li et al. (2021)	TLEs + ILRS CPF ephemerides	GBDT/CNN SGP4 error corrector	No	Single-object state corrector; no covariance output
Tulczyjew et al. (2021)	ESA Kelvins CDM dataset	MLP classifier + RF regressor (2-stage)	No	Sensitive to extreme class imbalance
Acciarini et al. (2020)	Synthetic + Kelvins-calibrated encounters	Bayesian SGP4 (PyProb)	No	SGP4 fidelity; manual noise calibration
Fiore (2021)	ESA CDM database (187,118 CDMs)	LSTM/GRU/kNN/SVM	No	Regression of continuous risk unreliable; missing covariance data
Nikolaev et al. (LLNL)	2009 NORAD catalog (12,971 objects)	Deterministic N-body propagation	Yes (deterministic)	>10^7 CPU-hours; ignores tracking uncertainty
Metz et al. (2021)	ESA CDM operational dataset	RF / LSTM covariance regressor	No	Depends on chaser observation frequency
Hudson (2023)	Macro LEO population states	Monte Carlo (Palisade @Risk)	Yes (stochastic)	Bulk gas-kinetic approximation; no per-object orbits
Sánchez & Vasile (2020)	Synthetic DB1/DB2 encounters	Dempster-Shafer + ML classifier	No	Synthetic only; no ground-truth CDM validation
This study	34,136-object SATCAT/TLE catalog	LR / RF / GBM proxy-risk classifier	Yes (standalone, per-object)	Proxy label constructed from features (see Limitations)

III. DATASET AND DATA PREPARATION

The dataset was constructed entirely from publicly available orbital element and satellite catalog (SATCAT) data through a nine-stage processing pipeline implemented in Python: feature_engineering.py, neighborhood_features.py, relative_velocity. py/object_features. py, merge_satcat.py, build_current_population. py/build_debris_dataset.py, full_population_crowding.py, full_population_velocity.py, and calculate_risk_scores.py. Orbital elements were first parsed from the TLE mean elements into semi-major axis, perigee/apogee altitude, orbital period, and circular orbital velocity using standard two-body relations (Earth radius = 6378.137 km; μ = 398,600.4418 km³/s²). The object type, operational status, launch date, and decay date were then attached to the SATCAT via the NORAD_CAT_ID.

A. Population Selection

The analysis population was restricted to objects that (a) are currently orbiting — defined as having no decay date, or a decay date on or after the reference date — and (b) belong to one of the three physically meaningful SATCAT object-type codes: payload (PAY), debris (DEB), or rocket body (R/B). Objects lacking a complete set of orbital parameters (period, inclination, apogee, and perigee) were excluded. This filtering, implemented in build_current_population.py, yielded the final analysis catalog of 34,136 objects, composed as follows.

TABLE II

DATASET COMPOSITION.

Category	**Count**	**Share of catalog**
Payloads (PAY)	19,587	57.4%
Debris (DEB)	12,309	36.1%
Rocket bodies (R/B)	2,240	6.6%
Total objects	34,136	100.0%
Development / training set (80%)	27,308	—
Final untouched test set (20%)	6,828	—

B. Engineered Feature Groups

For every object, the pipeline computed the following categories of features using a pairwise altitude/inclination neighborhood search over the full 34,136-object population (neighbourhood_features. py, recomputed on the current population in full_population_crowding.py):

- Orbital characteristics: mean altitude, altitude range (apogee − perigee), inclination, orbital period, apogee, perigee, and circular orbital velocities.
- Local orbital crowding: neighbor counts within 25 km/1°, 50 km/1°, 100 km/1°, and 50 km/5° altitude–inclination windows; count of objects sharing the same 100-km altitude shell; and a local orbital density measure (50 km/1° neighbor count normalized by shell width).
- Debris environment: number of debris-classified objects and number of rocket-body-classified objects within a 50 km/1° neighborhood of each object.
- Relative velocity: an estimated local relative-velocity potential (full_population_velocity.py), computed from the law-of-cosines combination of each object's orbital velocity with the velocities of neighbors within 50 km/5°, using the inclination difference as a simplified proxy for the relative-velocity-vector angle, and the count of velocity-neighbors contributing to that estimate.
- Object type: Binary indicators for payload, debris, and rocket body classification encoded from the SATCAT-merged OBJECT_TYPE field.

The dataset used for machine learning retained 45 variables (identifiers, orbital elements, engineered crowding/velocity features, three proxy risk scores, and categorical labels). Of these, 20 variables — excluding identifiers, the proxy-risk scores themselves, and the risk category — were retained as the predictive feature set (Section V, Table IV), specifically to avoid direct data leakage from the constructed target back into the predictors used for reproduction.

IV. PHYSICS-INFORMED PROXY-RISK FRAMEWORK

No ground-truth collision outcome exists for the vast majority of cataloged objects, and calibrated collision probabilities require object-specific covariance information that is unavailable outside the active CDM streams. To enable population-wide screening in the absence of this information, a transparent, physics-informed proxy risk score was constructed directly from the engineered orbital, crowding, and velocity features described in Section III, following the exact formulation implemented in full_population_crowding.py,

International Journal of Science, Strategic Management and Technology
Volume 02 Issue 08 August- 2026 | ISSN: 3108-1762 (Online) | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

full_population_velocity.py, and calculate_risk_scores.py. This proxy score is explicitly a ranking and screening device, not a calibrated probability of a collision.

A. Orbital Crowding Score

Each crowding component (25 km/1°, 50 km/1°, 100 km/1°, and 50 km/5° neighbor counts and same-shell object count) was min–max normalized across the population and combined as:

$$ORBITAL\_CROWDING\_SCORE = 0.20 \cdot N_{(25,1°)} + 0.30 \cdot N_{(50,1°)} + 0.20 \cdot N_{(100,1°)} + 0.15 \cdot N_{(50,5°)} + 0.15 \cdot S \quad (1)$$

where $N_{(\cdot,\cdot)}$ denotes the normalized neighbor count for the indicated altitude/inclination window and S is the normalized count of objects in the same 100-km altitude shell (this final population weighting, from full_population_crowding.py, supersedes an earlier exploratory weighting used in neighborhood_features.py during initial TLE-only feature development).

B. Relative Velocity Score

An encounter speed potential was estimated for each object from the law-of-cosines combination of orbital velocity vectors, using the inclination difference as a simplified proxy for the angle between velocity vectors:

$$V_r = \sqrt{v_1^2 + v_2^2 - 2 \cdot v_1 \cdot v_2 \cdot \cos\theta} \quad (2)$$

computed against all neighbors within 50 km altitude and 5° inclination, with the maximum resulting value retained as RELATIVE_VELOCITY_KM_S. This value was min–max normalized to RELATIVE_VELOCITY_SCORE and combined with orbital crowding as follows:

$$VELOCITY\_CROWDING\_SCORE = 0.60 \cdot RELATIVE\_VELOCITY\_SCORE + 0.40 \cdot ORBITAL\_CROWDING\_SCORE \quad (3)$$

C. Debris Crowding Score

The count of debris-classified neighbors within 50 km/1° (DEBRIS_NEIGHBORS_50KM_1DEG) was min–max normalized to produce the DEBRIS_CROWDING_SCORE.

D. Collision-Risk and Debris-Cascade-Risk Proxies

The Collision-Risk proxy weighs orbital crowding, relative velocity, velocity-weighted crowding, and debris crowding:

$$COLLISION\_RISK\_SCORE = 0.40 \cdot ORBITAL\_CROWDING\_SCORE + 0.30 \cdot RELATIVE\_VELOCITY\_SCORE + 0.20 \cdot VELOCITY\_CROWDING\_SCORE + 0.10 \cdot DEBRIS\_CROWDING\_SCORE \quad (4)$$

The object-hazard term reflects the debris-generation potential associated with the object type:

$$OBJECT\_HAZARD\_SCORE = 0.70 \cdot IS\_DEBRIS + 0.30 \cdot IS\_ROCKET\_BODY \quad (5)$$

The Debris-Cascade-Risk proxy weighs orbital crowding, debris crowding, relative velocity, and object hazard.

$$DEBRIS\_CASCADE\_RISK\_SCORE = 0.35 \cdot ORBITAL\_CROWDING\_SCORE + 0.30 \cdot DEBRIS\_CROWDING\_SCORE + 0.20 \cdot RELATIVE\_VELOCITY\_SCORE + 0.15 \cdot OBJECT\_HAZARD\_SCORE \quad (6)$$

E. Overall Proxy-Risk Score and Category

$$OVERALL\_RISK\_SCORE = 0.60 \cdot COLLISION\_RISK\_SCORE + 0.40 \cdot DEBRIS\_CASCADE\_RISK\_SCORE \quad \text{(clipped to } [0, 1]) \quad (7)$$

The overall score is mapped to a categorical proxy risk label by fixed thresholds, implemented identically for the overall, collision, and cascade scores:

TABLE III

PROXY-RISK CLASSIFICATION THRESHOLDS (CALCULATE_RISK_SCORES.PY).

Category	Overall risk score range
LOW	score < 0.25
MEDIUM	0.25 ≤ score < 0.50
HIGH	0.50 ≤ score < 0.75
CRITICAL (defined, not observed)	score ≥ 0.75

The CRITICAL category is defined in the code but was not populated in the analyzedcatalog. The maximum observed OVERALL_RISK_SCORE across all 34,136 objects is approximately 0.71 (Fig. 9), below the 0.75 CRITICAL threshold. The realized classification task is therefore effectively a three-class task (LOW/MEDIUM/HIGH); this is noted explicitly in Section XI-C and reflected in the machine learning target definition (Section V), which restricts the modeled classes to LOW, MEDIUM, and HIGH.

It is important to plainly state what this framework is and is not. It is a transparent, reproducible, physics-informed screening score built from measurable orbital environment characteristics, intended to rank objects by their local crowding, encounter speed potential, and debris/object-type hazard. It is not a calibrated probability of collision and has not been validated against verified historical conjunction or collision outcomes. This distinction is maintained throughout this manuscript.

V. MACHINE LEARNING METHODOLOGY

The machine-learning task is defined as multi-class classification of RISK_CATEGORY (restricted to the observed classes LOW, MEDIUM, HIGH) from 20 predictive features (Table IV), implemented in ml_risk_prediction.py (five-fold cross-validation) and step13_final_test.py (final holdout evaluation). The three proxy-risk scores (COLLISION_RISK_SCORE, DEBRIS_CASCADE_RISK_SCORE, OVERALL_RISK_SCORE) and the RISK_CATEGORY label itself were explicitly excluded from the feature set, as they were the outputs of the scoring formula in Section IV and would constitute direct target leakage.

TABLE IV

MACHINE LEARNING PREDICTIVE FEATURE SET (20 FEATURES).

Group	Features
Orbital characteristics	MEAN_ALTITUDE_KM, ALTITUDE_RANGE_KM, INCLINATION, PERIOD, APOGEE, PERIGEE, ORBITAL_VELOCITY_KM_S
Local orbital crowding	NEIGHBORS_25KM_1DEG, NEIGHBORS_50KM_1DEG, NEIGHBORS_100KM_1DEG, NEIGHBORS_50KM_5DEG, OBJECTS_IN_ALTITUDE_SHELL, LOCAL_ORBITAL_DENSITY
Debris environment	DEBRIS_NEIGHBORS_50KM_1DEG, ROCKET_BODY_NEIGHBORS_50KM_1DEG
Relative velocity	RELATIVE_VELOCITY_KM_S, VELOCITY_NEIGHBOR_COUNT
Object type	IS_PAYLOAD, IS_DEBRIS, IS_ROCKET_BODY

A. Preprocessing

All feature columns were coerced to numeric type; infinite values arising from normalization edge cases were replaced with missing values; and the remaining missing values were imputed with the column median (zero if the median itself was undefined). Logistic Regression was trained within a pipeline including standard (zero-mean, unit-variance) feature scaling; Random Forest and Gradient Boosting were trained on unscaled features, consistent with their scale-invariant, tree-based decision rules.

B. Data Splitting and Validation

Two complementary validation protocols were used in this study. First (ml_risk_prediction.py), five-fold stratified cross-validation (shuffled, random_state = 42) was applied to the full 34,136-object dataset to compare the candidate models under repeated resampling. Second (step13_final_test.py), an 80/20 stratified train–test split (random_state = 42) reserved 6,828 objects (20%) as a final, untouched test set; the remaining 27,308 objects (80%) were used for five-fold stratified cross-validation model selection, after which each candidate model was refitted on the full development set and evaluated exactly once on the held-out test set. The two protocols are independent; the holdout test objects were never used for any model-selection decision.

C. Models

- Dummy Baseline — predicts the most frequent class (MEDIUM) for every object, establishing the performance floor.
- Logistic Regression — multinomial logistic regression with class_weight = 'balanced' and max_iter = 3000, trained on standardized features.
- Random Forest — 300 trees, max_depth = 15, min_samples_leaf = 2, class_weight = '' balanced, random_state = 42.
- Gradient Boosting — 200 estimators, learning_rate = 0.05, max_depth = 5, and random_state = 42.

No hyperparameter values beyond those specified directly in the source code were used or inferred in this study.

D. Evaluation Metrics

Models were compared using weighted-average accuracy, precision, recall, F1-score, and one-vs-rest weighted ROC-AUC, computed identically for cross-validation folds and the final holdout set. Weighted averaging accounts for the class imbalance between the LOW, MEDIUM, and HIGH categories (Table VIII).

VI. CROSS-VALIDATION RESULTS

Table V reports the mean five-fold stratified cross-validation performance across the full 34,136-object dataset (ml_risk_prediction.py).

TABLE V
FIVE-FOLD STRATIFIED CROSS-VALIDATION RESULTS (MEAN VALUES).

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.998008	0.998013	0.998008	0.998008	0.999991
Gradient Boosting	0.996572	0.996576	0.996572	0.996573	0.999974
Random Forest	0.995635	0.995644	0.995635	0.995635	0.999957
Dummy Baseline	0.478293	—	—	0.309497	0.500000

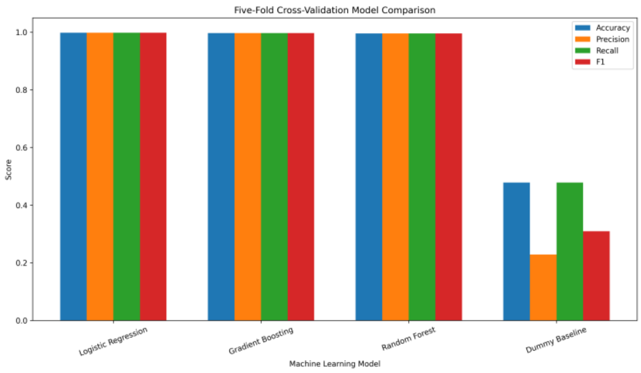

Fig. 1. Five-fold cross-validation model comparison across accuracy, precision, recall and F1-score. The three trained classifiers clustered near 1.0 on every metric, and the dummy baseline was markedly lower.

All three trained classifiers substantially outperformed the dummy baseline on every metric and were closely clustered with one another, with Logistic Regression marginally ahead. The magnitude and uniformity of this advantage, rather than any specific model's superiority, is the primary finding requiring interpretation and is addressed directly in Sections X and XI.

VII. FINAL TEST RESULTS

Following cross-validation-based model selection on the 27,308-object development set, each model was refitted on the full development set and evaluated once on the untouched 6,828-object test set (step13_final_test.py). The results are presented in Table VI.

TABLE VI

FINAL UNTOUCHED HOLDOUT TEST PERFORMANCE (N = 6,828).

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	0.998535	0.998538	0.998535	0.998536	0.999994
Gradient Boosting	0.995313	0.995326	0.995313	0.995313	0.999959
Random Forest	0.995021	0.995023	0.995021	0.995020	0.999945
Dummy Baseline	0.478325	0.228794	0.478325	0.309532	0.500000

Logistic Regression was selected by cross-validation as the best-performing model and achieved the highest holdout accuracy (0.998535) and F1-score (0.998536) of the three trained classifiers. The confusion matrix for the selected model in the holdout set is shown in Fig. 2.

Fig. 2. Final holdout confusion matrix, Logistic Regression (LOW = 1,428 objects; MEDIUM =

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **08** August- **2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

3,266; HIGH = 2,134). Off-diagonal counts were minimal: one LOW object predicted MEDIUM, six MEDIUM objects predicted LOW, and three MEDIUM objects predicted HIGH.

All ten misclassified objects fell immediately adjacent to a classification threshold (0.25 or 0.50) in the underlying OVERALL_RISK_SCORE, consistent with the model recovering a threshold rule rather than making a substantive category error. The corresponding one-vs-rest ROC curves (Fig. 3) are visually indistinguishable from a perfect classifier for all three models.

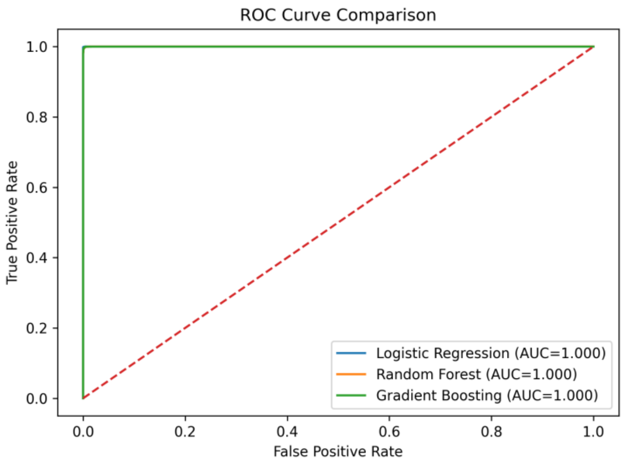

Fig. 3. One-vs-rest ROC curves for Logistic Regression, Random Forest, and Gradient Boosting (holdout set). All three curves are effectively coincident with the top-left corner (AUC ≈ 1.000), and the diagonal reference line indicates random-chance performance.

CRITICAL FRAMING: The accuracy figures above (e.g., 99.85%) describe how well the classifiers reproduce the constructed physics-informed proxy-risk category defined in Section IV. They are not, and must not be interpreted as, measures of accuracy in predicting real collisions, verified conjunction outcomes, or operational collision probabilities. No historical collision or verified conjunction ground truth was available for or used in this study.

VIII. ORBITAL ENVIRONMENT ANALYSIS

This section characterizes the analyzedcatalog and resulting proxy risk distribution, independent of the classification task.

A. Population Composition and Altitude Distribution

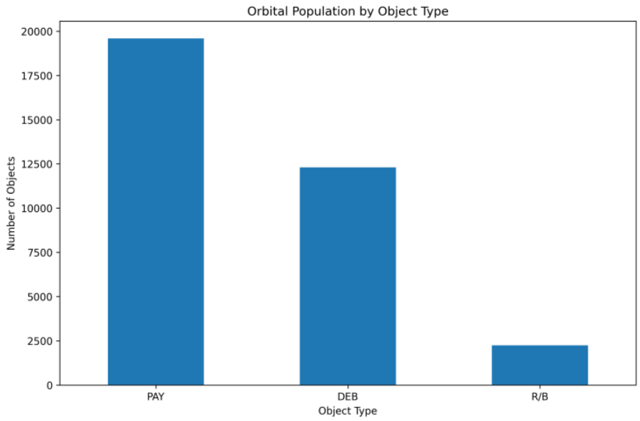

Fig. 4. Orbital population by object type. Payloads (19,587) are the largest category, followed by debris (12,309) and rocket bodies (2,240).

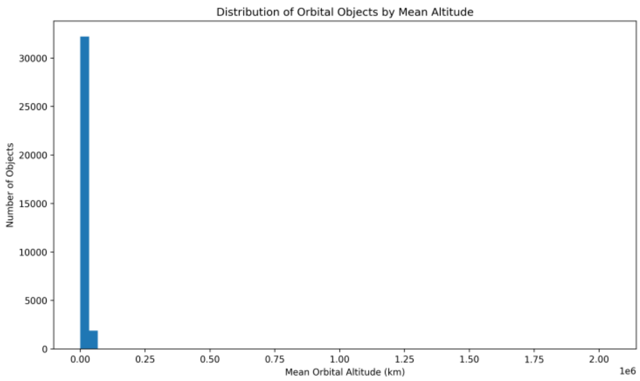

Fig. 5. Distribution of orbital objects by mean altitude. The population is heavily concentrated below roughly 2,000 km, consistent with the dominance of LEO in the tracked catalog, with sparse secondary concentrations near GEO altitude (~35,786 km).

B. Debris Concentration by Altitude

Table VII lists the most debris-dense 100-km altitude shells identified by the neighborhood analysis (debris_altitude_regions.csv / Fig. 6).

TABLE VII

DEBRIS CONCENTRATION BY ALTITUDE SHELL (TOP FOUR SHELLS).

Altitude shell (lower bound)	Debris object count
800 km	2,501
700 km	2,151
900 km	1,229
600 km	1,023

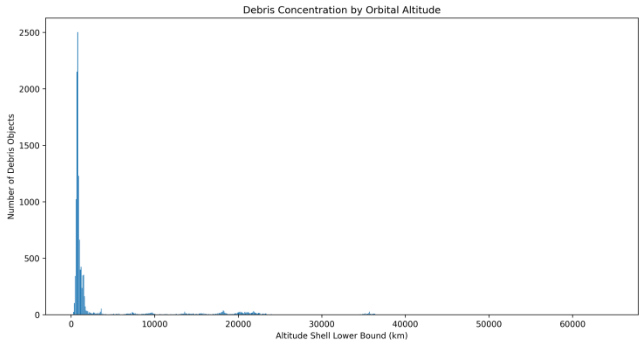

Fig. 6. Debris concentration by orbital altitude shell. A pronounced peak occurs in the 700–900 km band, a historically debris-dense region associated with past fragmentation events.

C. Altitude–Inclination Structure and Risk Category

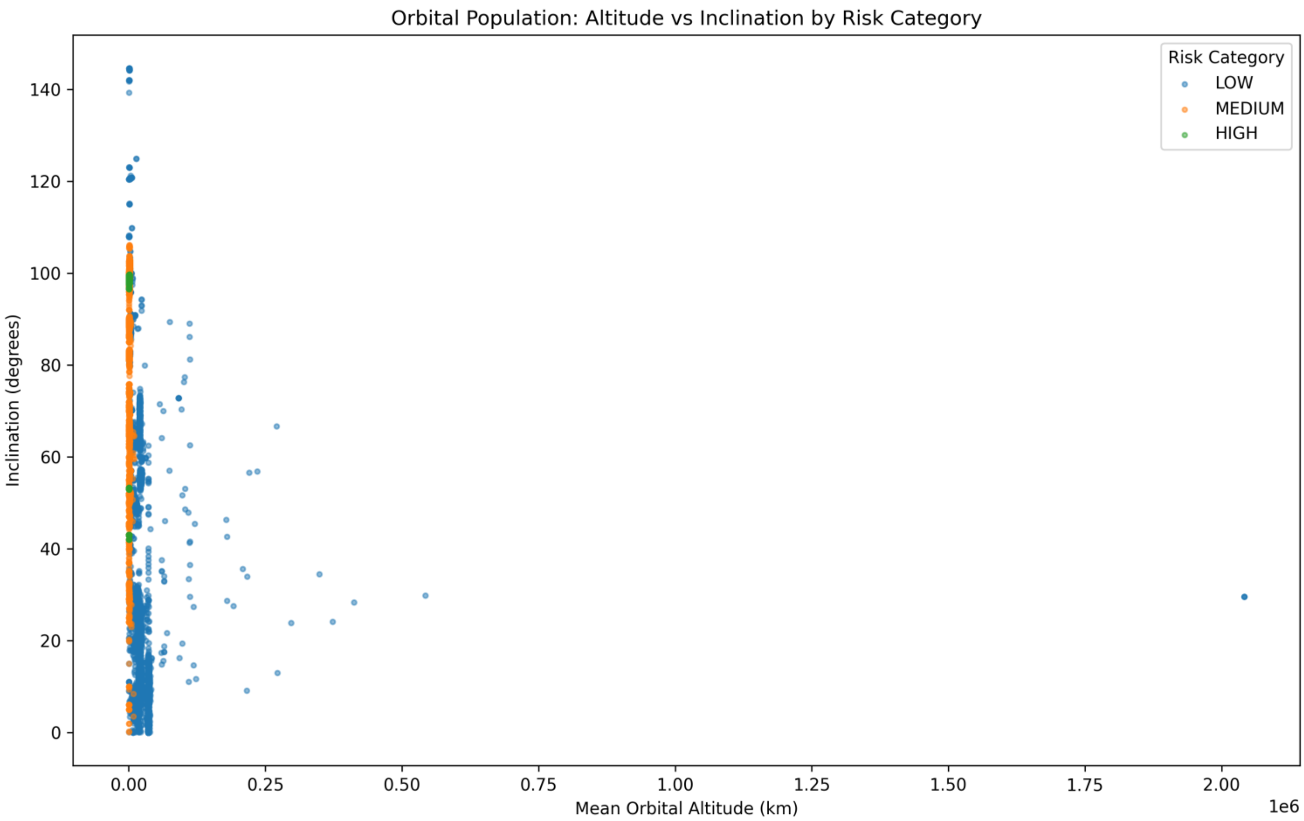

Fig. 7. Altitude vs. inclination, colored by proxy-risk category. HIGH-risk objects (green) are concentrated near common LEO inclinations (~53°, ~97–99° sun-synchronous, ~28.5°) at low altitude, where crowding and relative-velocity potential are greatest.

D. Risk Category and Score Distributions

TABLE VIII

PROXY-RISK CATEGORY DISTRIBUTION AND MEAN SCORES BY OBJECT TYPE.

Category / Object type	**Count or mean score**
LOW	7,138 objects
MEDIUM	16,327 objects
HIGH	10,671 objects
Mean OVERALL_RISK_SCORE — Payloads (PAY)	0.4464
Mean OVERALL_RISK_SCORE — Debris (DEB)	0.3767
Mean OVERALL_RISK_SCORE — Rocket bodies (R/B)	0.2204

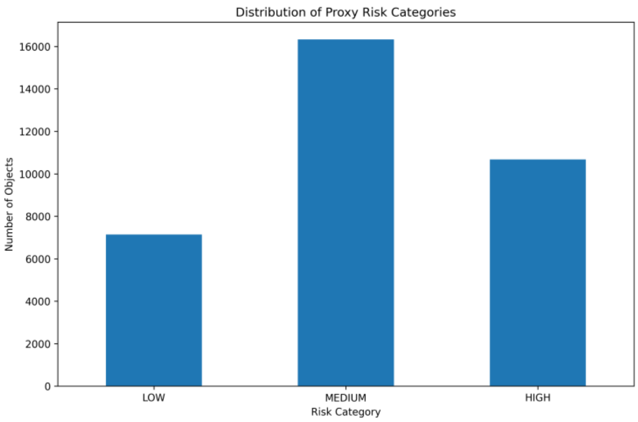

Fig. 8. Distribution of proxy-risk categories across the 34,136-object catalog. MEDIUM is the modal category; HIGH exceeds LOW in object count.

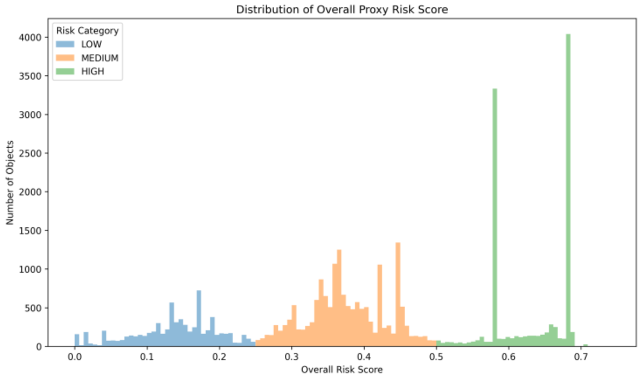

Fig. 9. Distribution of the continuous OVERALL_RISK_SCORE, colored by category. The score is multimodal, and its empirical maximum (~0.71) falls short of the 0.75 CRITICAL threshold, so no object is categorized as CRITICAL.

The finding that payloads carry the highest mean overall proxy-risk score, ahead of debris and rocket bodies, follows directly from the scoring formulation: payloads are disproportionately represented in densely populated operational shells (Sections IV-A–IV-B), which drives the crowding and relative-velocity components of the score, whereas the object-hazard term in the cascade-risk proxy (Section IV-D) only elevates debris and rocket-body scores, and carries a comparatively small weight (0.15 within the cascade proxy, itself weighted at 0.40 in the overall score). This is a property of the constructed proxy and should not be read as an empirical finding that payloads are physically more dangerous than debris.

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **08** August- **2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

IX. FEATURE INTERPRETATION

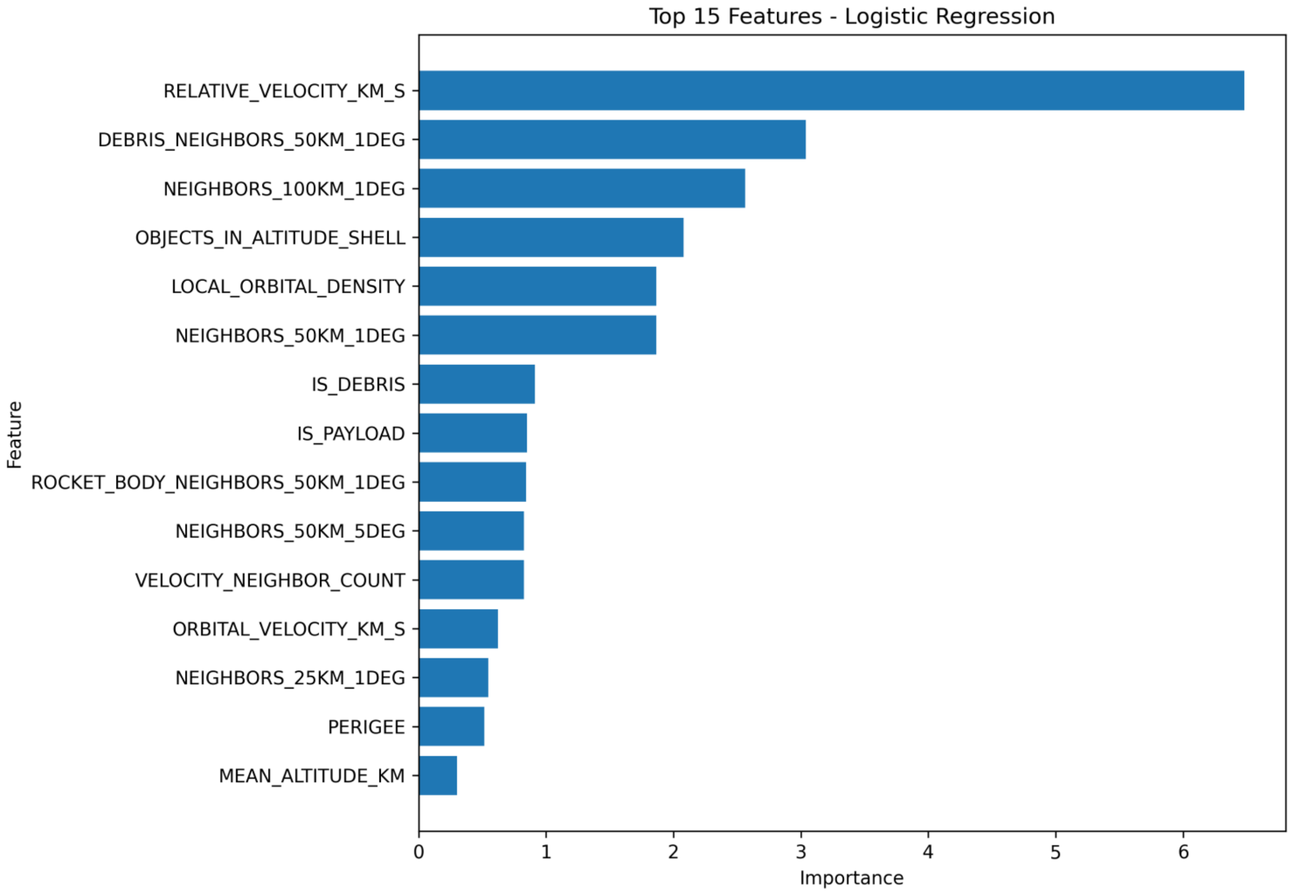

Fig. 10. Top features by absolute Logistic Regression coefficient magnitude. Approximate values estimated visually from the generated figure; exact coefficients should be sourced from the underlying model-output file when available.

TABLE IX

APPROXIMATE FEATURE IMPORTANCE, LOGISTIC REGRESSION (READ FROM FIG. 10).

Feature	**Approx. importance**
RELATIVE_VELOCITY_KM_S	≈ 6.5
DEBRIS_NEIGHBORS_50KM_1DEG	≈ 3.0
NEIGHBORS_100KM_1DEG	≈ 2.5
OBJECTS_IN_ALTITUDE_SHELL	≈ 2.1
LOCAL_ORBITAL_DENSITY	≈ 1.9
NEIGHBORS_50KM_1DEG	≈ 1.9
IS_DEBRIS / IS_PAYLOAD / ROCKET_BODY_NEIGHBORS_50KM_1DEG	≈ 0.85–0.9
ORBITAL_VELOCITY_KM_S / NEIGHBORS_25KM_1DEG / PERIGEE / MEAN_ALTITUDE_KM	≈ 0.3–0.6 (lowest ranked)

RELATIVE_VELOCITY_KM_S is, by a wide margin, the most influential predictor, followed by the debris- and general-crowding neighbor counts. This ranking mirrors the weighting structure of the proxy-risk formula itself (Section IV): RELATIVE_VELOCITY_SCORE enters the collision-risk proxy directly at weight 0.30 and indirectly (via VELOCITY_CROWDING_SCORE) at a further 0.20 × 0.60 = 0.12, while the crowding-derived neighbor counts feed ORBITAL_CROWDING_SCORE, which contributes to both proxy components. Conversely, features with the weakest direct algebraic link to the proxy formula — MEAN_ALTITUDE_KM, PERIGEE, and raw ORBITAL_VELOCITY_KM_S — rank lowest in importance. This correspondence is expected given the deterministic construction of the target (Section IV) and is discussed as a central limitation in Section XI rather than as an independent empirical discovery.

Correlation with OVERALL_RISK_SCORE (risk_correlations.csv) is consistent with this pattern: crowding, altitude-shell density, and relative-velocity features correlate most strongly with the proxy score, as expected given that they are direct algebraic inputs to it. Correlation values in this context describe internal consistency of the constructed scoring formula, not an independently observed physical relationship, and are not interpreted as evidence of causation.

X. DISCUSSION

The results demonstrate that a physics-informed proxy-risk category, once defined from measurable orbital-crowding, debris-neighbor, and relative-velocity characteristics, can be reproduced with very high fidelity by interpretable classifiers operating solely on standalone, un-conjuncted TLE-derived features — without reference to any CDM, covariance, or externally flagged conjunction event. This is consistent with the study's positioning relative to the literature reviewed in Section II: whereas CDM-dependent machine-learning approaches [10], [11], [13] can only act once a close approach has already been detected, and macro-scale stochastic models [15] cannot resolve individual objects, the present framework operates continuously across the full tracked catalog and produces an object-level ranking without requiring either input.

Local orbital crowding and estimated relative velocity emerge as the dominant drivers of proxy-risk classification (Section IX), which is physically intuitive: high relative-velocity potential combined with high local object density are the two conditions classical conjunction-risk theory associates with elevated collision likelihood [7], [8]. The concentration of HIGH-risk objects near common LEO operational inclinations (~53°, ~97–99°, ~28.5°) at low altitude (Fig. 7), and the pronounced debris concentration in the 700–900 km shells (Fig. 6, Table VII), are consistent with widely documented consequences of historical fragmentation events in these regions [2].

Among the three trained classifiers, Logistic Regression achieved marginally higher cross-validated and holdout performance than Random Forest and Gradient Boosting (Tables V–VI). Given that the target is a smooth, approximately linear-in-its-normalized-inputs combination of the predictors (Section IV), this is unsurprising: a linear model is well matched to a target constructed from a weighted linear sum, whereas the tree-based ensembles must approximate that structure through axis-aligned splits. The near-identical performance across all three models, rather than any one architecture's superiority, is the more informative observation — it indicates that the classification task, as constructed, is close to fully separable given the available features, regardless of model family.

Interpreted appropriately, the principal contribution of these results is methodological: the study demonstrates that engineered, un-conjuncted, catalog-derived crowding and relative-velocity features carry enough structure to support a stable, three-tier proxy-risk ranking of the entire tracked population — the kind of standalone, pairwise-independent screening layer identified as absent from the reviewed literature (Section II-G). Section XI discusses, in detail, why the specific numerical accuracy figures should not be read as evidence of collision-prediction skill, and Section XII frames how such a screening layer could plausibly be used alongside, rather than instead of, existing physics-based conjunction assessment.

XI. LIMITATIONS

A. Proxy Labels, Not Verified Collision Outcomes

The RISK_CATEGORY target is a constructed, physics-informed proxy defined by the study's own scoring formula (Section IV), not a label derived from verified historical collisions, near-miss events, or operational CDM outcomes. No such ground-truth collision or conjunction dataset was available to or used in this study. Consequently, none of the classification metrics reported in Sections VI–VII measure real-world collision-prediction accuracy.

B. Target–Feature Dependency and Its Effect on Reported Performance

This is the most important limitation of the study and is stated explicitly. Every predictive feature used for machine learning (Table IV) either directly or indirectly determines the components of the OVERALL_RISK_SCORE that defines RISK_CATEGORY (Section IV): the neighbor-count and altitude-shell features determine ORBITAL_CROWDING_SCORE; DEBRIS_NEIGHBORS_50KM_1DEG determines DEBRIS_CROWDING_SCORE; RELATIVE_VELOCITY_KM_S determines RELATIVE_VELOCITY_SCORE; and IS_DEBRIS/IS_ROCKET_BODY determine OBJECT_HAZARD_SCORE. Although the three proxy-risk scores and the RISK_CATEGORY label were explicitly excluded from the feature set to prevent direct leakage (Section V), their deterministic algebraic ingredients were retained as predictors. Because the scoring formula is a fixed, noise-free, weighted linear/threshold function of these same ingredients, the classification task substantially reduces to recovering a known function from its own inputs, rather than learning an independent statistical relationship between orbital characteristics and an externally measured risk outcome. This mechanism is the most plausible explanation for the very high accuracy, F1, and ROC-AUC values reported in Sections VI–VII, and is corroborated by the feature-importance ranking in Section IX, which closely mirrors the scoring formula's own weighting structure. The reported metrics should therefore be understood as a demonstration that the models can recover the internal structure of the proxy-risk formulation, not as independent evidence of predictive skill against an external criterion.

C. CRITICAL Category Not Observed

The proxy-risk formulation defines a fourth category, CRITICAL (score ≥ 0.75), that does not occur in the analyzedcatalog (maximum observed score ≈ 0.71; Fig. 9). The classification task is therefore effectively three-class in practice, despite four categories being defined in code.

D. Simplified Neighborhood and Relative-Velocity Definitions

Neighborhoods are defined using fixed Euclidean thresholds in mean altitude and inclination (e.g., 50 km / 1°), which do not account for right ascension of the ascending node, argument of perigee, or true anomaly, and therefore may group objects that are never simultaneously close in three-dimensional space, or miss objects that are close in 3-D space but differ in these unconsidered elements. The relative-velocity estimate uses inclination difference as a proxy for the full three-dimensional angle between velocity vectors and assumes circular-orbit velocity magnitudes; it is a screening approximation and not an operational relative-velocity or miss-distance calculation.

E. Catalog Completeness and Measurement Uncertainty

The analysis is limited to cataloged, trackable objects in SATCAT/TLE data; the much larger population of untracked sub-10-cm fragments is not represented. TLE-derived mean elements carry their own propagation and fitting uncertainty, and no covariance information is available at the TLE level, unlike operational CDMs.

F. No Independent Historical Validation

The framework has not been validated against independent, historical conjunction or collision datasets (such as the ESA Kelvins CDM archive referenced in Section II-E), and no such validation was performed as part of this study.

XII. SPACE-SUSTAINABILITY AND PRACTICAL IMPLICATIONS

Notwithstanding the limitations in Section XI, the engineered crowding, debris-density, and relative-velocity features, and the resulting proxy-risk ranking, may have practical value as a lightweight, standalone screening layer positioned ahead of detailed operational conjunction assessment:

Large tracked catalog→ ML proxy-risk screening → Prioritized subset of objects → Detailed physics-based conjunction assessment (CDM / Pc)

In this role, the framework would not replace CDM-based conjunction analysis or Pc calculation, but could help direct limited tracking, propagation, and analyst attention toward objects sitting in persistently crowded, high-relative-velocity regions of the catalog — complementing, rather than substituting for, the reactive CDM-based pipeline discussed in Section II-C. Any operational use would require the historical validation described in Section XI-F before the proxy-risk category could be treated as more than an internally consistent screening heuristic.

XIII. CONCLUSION

This study developed a pairwise-independent, physics-informed proxy-risk classification framework for the standalone screening of cataloged space objects, addressing a gap identified in the literature between CDM-dependent, object-level collision-risk assessment and macro-scale, density-based population sustainability modeling. Using a 34,136-object catalog of currently orbiting payloads, debris, and rocket bodies, twenty features describing orbital geometry, local crowding, debris- and rocket-body-neighbor density, and estimated relative velocity were engineered and used to reproduce a three-tier (LOW/MEDIUM/HIGH) proxy-risk category derived from a transparent, weighted scoring formula. Logistic Regression, Random Forest, and Gradient Boosting classifiers all reproduced this category with very high fidelity (holdout accuracy up to 0.9985), against a Dummy baseline of 0.4783, with relative velocity and local debris/orbital crowding identified as the dominant predictors.

The scientific significance of this work lies in demonstrating that un-conjuncted, catalog-derived crowding and relative-velocity features carry sufficient structure to support a stable, population-wide risk ranking without reliance on pairwise CDMs — the kind of standalone screening layer the reviewed literature does not address. At the same time, because the proxy-risk target is a deterministic function of several of the predictive features, the reported classification accuracy reflects the models' ability to recover that constructed formula rather than independent predictive skill against a verified, external collision or conjunction outcome; this limitation is central to interpreting the results and has been stated explicitly rather than minimized.

Future work should pursue: (i) incorporation of historical conjunction-event or CDM-derived labels to replace or validate the proxy-risk target against verified outcomes; (ii) inclusion of covariance information where available; (iii) higher-fidelity relative-state calculation using full three-dimensional orbital-element differences and numerical propagation rather than the simplified inclination-difference approximation used here; (iv) temporal orbit propagation to evaluate how proxy-risk rankings evolve over time as orbits precess and decay; and (v) independent operational validation against real-world conjunction and collision-avoidance records.

REFERENCES

[1] D. J. Kessler and B. G. Cour-Palais, "Collision frequency of artificial satellites: The creation of a debris belt," J. Geophys. Res., vol. 83, no. A6, pp. 2637-2646, 1978.

[2] European Space Agency (ESA) Space Debris Office, "ESA Space Environment Report," ESA/ESOC, Darmstadt, Germany. [Online]. Available: https://www.sdo.esoc.esa.int/environment_report/Space_Environment_Report_latest.pdf

[3] T. S. Kelso, "Analysis of the Iridium 33-Cosmos 2251 collision," in Proc. Advanced Maui Optical and Space Surveillance Technologies Conf. (AMOS), Maui, HI, USA, 2009.

[4] Inter-Agency Space Debris Coordination Committee (IADC), "IADC Space Debris Mitigation Guidelines," IADC-02-01, Rev. 2, Mar. 2020.

[5] D. A. Vallado, P. Crawford, R. Hujsak, and T. S. Kelso, "Revisiting Spacetrack Report #3," in Proc. AIAA/AAS Astrodynamics Specialist Conf., Keystone, CO, USA, Aug. 2006, AIAA 2006-6753.

[6] B. Li, Y. Zhang, J. Huang, and J. Sang, "Improved orbit predictions using two-line elements through error pattern mining and transferring," Acta Astronaut., vol. 188, pp. 405-415, 2021.

[7] M. R. Akella and K. T. Alfriend, "Probability of collision between space objects," J. Guid., Control, Dyn., vol. 23, no. 5, pp. 769-772, 2000.

[8] R. P. Patera, "General method for calculating satellite collision probability," J. Guid., Control, Dyn., vol. 24, no. 4, pp. 716-722, 2001.

[9] F. K. Chan, Spacecraft Collision Probability. El Segundo, CA, USA: Aerospace Press, 2008.

[10] T. Uriot, D. Izzo, L. F. Simoes, R. Abay, N. Einecke, S. Rebhan, J. Martinez-Heras, F. Letizia, J. Siminski, and K. Merz, "Spacecraft collision avoidance challenge: Design and results of a machine learning competition," Astrodynamics, vol. 6, pp. 121-140, 2022.

[11] L. Tulczyjew, M. Myller, M. Kawulok, D. Kostrzewa, and J. Nalepa, "Predicting risk of satellite collisions using machine learning," J. Space Safety Eng., vol. 8, no. 4, pp. 339-344, 2021.

[12] N. Boscolo Fiore, "Machine learning based satellite collision avoidance strategy," M.S. thesis, Dept. Aerosp. Sci. Technol., Politecnico di Milano, Milan, Italy, 2021.

[13] S. Metz, H. Simon, and F. Letizia, "Implementation and comparison of data-based methods for collision avoidance in satellite operations," in Proc. 8th European Conf. Space Debris (ECSD), Darmstadt, Germany, 2021.

[14] S. Nikolaev, D. Phillion, H. K. Springer, W. deVries, M. Jiang, A. Pertica, J. Henderson, M. Horsley, and S. Olivier, "Brute force modeling of the Kessler syndrome," in Proc. Advanced Maui Optical and Space Surveillance Technologies Conf. (AMOS), Maui, HI, USA, 2012.

[15] J. Hudson, "KESSYM: A stochastic orbital debris model for evaluation of Kessler Syndrome risks and mitigations," J. Student Res., 2023.

[16] N. L. Johnson, P. H. Krisko, J.-C. Liou, and P. D. Anz-Meador, "NASA's new breakup model of EVOLVE 4.0," Adv. Space Res., vol. 28, no. 9, pp. 1377-1384, 2001.

[17] L. Sanchez, M. Vasile, S. Sanvido, K. Merz, and C. Taillan, "On the use of machine learning and evidence theory to improve collision risk management," Acta Astronaut., vol. 181, pp. 694-706, 2021.