International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

UNIFIED FRAMEWORK FOR REAL-TIME MULTI-FACTOR RISK-AWARE ROUTE OPTIMIZATION USING MACHINE LEARNING

Thilagavathi T[1*], Dr. Subashini A[2]

[1]Research Scholar, Department of Computer and Information Science, Annamalai University, Annamalai Nagar, Tamil Nadu, India.
thilagaponns@gmail.com

[2]Assistant Professor, Department of Computer Application, Government Arts College, Chidambaram, Tamil Nadu, India,
subanandh31@gmail.com

Abstract

In the era of smart transportation and increasing road traffic, traveler safety is a critical priority. Traditional navigation systems often minimize distance or time, overlooking contextual safety factors. This paper presents a unified, machine learning-based framework for multi-factor risk assessment and safe route optimization. The system integrates static parameters (vehicle type, user demographics, lighting, road type, group size) and dynamic parameters (real-time weather, traffic, surface condition, time). It gathers data from Google Maps, OpenStreetMap, OpenWeatherMap, and TomTom, processes it through a modular pipeline, and assigns risk values to route segments. Gradient Boosting, Random Forest, and Decision Tree models predict cumulative risk scores to recommend the safest route. A case study between Maragathapuram and Parvathipuram in Tamil Nadu validates the framework. The experimental results show Gradient Boosting provides the highest prediction accuracy. The system provides real-time, personalized, and data-driven travel recommendations, enhancing intelligent transportation with a focus on safety.

Index Terms – Route Optimization, Risk Assessment, Machine Learning, Regression Analysis, Intelligent Transport System, Random Forest, Gradient Boosting, Decision Tree, Safe Travel.

1 Introduction

As road-based travel and personal mobility increase, ensuring traveler safety has become a crucial focus of transportation planning and smart mobility. The World Health Organization reports road traffic injuries as the leading cause of death among those aged 5 to 29, with India alone recording over 150,000 fatalities annually. These alarming statistics highlight the urgent need to integrate safety considerations into traditional route planning, which has long prioritized time and distance.

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Most existing navigation systems optimize routes by minimizing travel time or distance. While this improves traffic efficiency, it often overlooks critical safety-related factors like road conditions, lighting, crime rates, and accident-prone areas. As a result, vulnerable users—including women, children, the elderly, and solo travelers—may be routed through high-risk zones. This limitation underscores the need for more comprehensive systems that prioritize both safety and efficiency.

Recent advances in Intelligent Transportation Systems (ITS) offer opportunities to address this challenge by incorporating real-time data for proactive risk-aware routing. Modern technologies enable continuous data acquisition from sources such as satellite navigation, traffic monitoring Application Programming Interfaces (APIs), weather services, and geographic information systems. Leveraging these, this study proposes a machine learning-driven framework for multi-factor risk assessment and safe route optimization.

The system combines static parameters (e.g., road type, user demographics, lighting) with dynamic data (e.g., traffic, weather, surface condition, time of travel), collected via APIs like Google Maps, OpenStreetMap, OpenWeatherMap, and TomTom. A web-based interface collects user inputs such as start and end points, number of travelers, demographics, and travel mode. Each route is divided into ~200-meter segments, which are analyzed individually through geocoding, spatial interpolation, and machine learning classification. Segment-level features such as locality, traffic density, weather, road surface, and lighting are used to compute safety scores, creating a detailed risk profile for each route option.

The risk assessment is automated using models like Random Forest, Gradient Boosting, and Decision Trees, trained on enriched datasets that include both static and dynamic features. The safest route is selected based on the lowest cumulative risk score, ensuring informed and secure travel decisions. To validate the framework, a real-world case study is conducted between Maragathapuram and Parvathipuram in Tamil Nadu—chosen for its varied urban and rural segments. The system dynamically updates risk scores in real-time, showcasing its potential as a scalable foundation for future smart mobility solutions.

This work bridges the gap between theoretical safety models and real-world navigation by offering a flexible, data-driven system for safe travel. It aims to reduce traffic-related injuries and fatalities while aligning with broader goals of sustainability, human-centered design, and intelligent transport infrastructure.

2 Related Work

Several studies have addressed route optimization to improve safety and efficiency. Reference [1] proposed a three-phase framework that integrates static and dynamic parameters, including weather, traffic, and crime data, to evaluate route safety. However, the approach remains rule-based, limiting scalability and adaptability to new

Received: August 07, 2025

data patterns. As [2] demonstrated, integrating spatial crash rate analysis with Random Forest and network modeling can effectively identify high-risk zones and suggest safer routes. Their work emphasized the importance of real-time validation and environmental data integration. In a related effort, reference [3] applied wavelet transforms with deep learning and Exponentially Weighted Moving Average (EWMA) filtering to improve highway traffic flow prediction.

Comprehensive reviews of crash prediction methodologies were conducted in [4] and [5], categorizing models into statistical, machine learning, and deep learning approaches. Neural networks emerged as high-performing predictors, particularly when environmental features were included. Reference [5] further noted the rise of hybrid models and emphasized the need for standardized, ethical practices in traffic prediction research. An explainable AI model was developed in [6] using SHAP values to trace accident causes and link behavior to injury severity. Meanwhile, [7] demonstrated a structured machine learning framework for predicting accident severity in Indian datasets, highlighting ML's promise in data-rich environments.

A real-time accident risk system using ensemble learning (Random Forest, XGBoost) was introduced in reference [8], featuring mobile alert integration. Reference [9] utilized convolutional neural networks (CNNs) to analyze grayscale traffic images and enhance accident severity prediction. A deep learning model compatible with the Internet of Vehicles (IoV) was presented in [10], using recurrent layers to forecast dynamic accident conditions. As [11] showed, Support Vector Machine (SVM) and logistic regression can be applied to recommend safe routes based on road type, lighting, and weather conditions. To further support environmental integration, [12] and [13] embedded real-time weather forecasting into navigation systems, enabling adaptive, safer routing.

Energy consumption prediction for battery electric vehicles (BEVs) was addressed in [14] using a hybrid model based on both pre-trip and real-time parameters, achieving a MAPE under 5%. Reference [15] combined estimated time of arrival (ETA) and weather data with ML algorithms to reach 92% accuracy in congestion forecasting. Boosting techniques like Light Gradient Boosting Machine (LGB), eXtreme Gradient Boosting (XGBoost), and Categorical Boosting (CatBoost) were applied in [16] to improve delivery time predictions, identifying distance, direction, and temperature as primary features. As [17] proposed, a hybrid graph and time-series model (FEN-MRMGCN) significantly improves bus arrival time predictions by including traffic and weather as external inputs.

Reference [18] introduced the Inundation Hazard Index (IHI), leveraging digital elevation models (DEM) and rainfall history to evaluate flood risk along road links, which was then integrated into routing systems for flood safety. A hybrid destination prediction framework using autoencoders and factorization machines was developed in [19], effectively managing Global Positioning System (GPS) deviations and improving

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

accuracy. As [20] showed, online learning applied to taxi-parcel routing minimizes idle time and improves efficiency by analyzing historical and real-time requests. Reference [21] presented DeepFEC, a deep learning approach combining residual and recurrent layers to accurately estimate energy usage across vehicle types.

Geospatial analysis in forest logistics was explored in [22] using GIS to assess road types and conditions, revealing that prioritizing safety increased transport costs by 15.76%. SHAP-enhanced XGBoost was used in [23] to analyze crash data from Texas, identifying median width, curb types, and shoulder structures as major factors in accident severity. In urban safety applications, [24] employed Poisson and negative binomial models to estimate accident frequency, aiding in the identification of high-risk segments. Lastly, reference [25] emphasized the role of user experience, road quality, and historical crash data in developing safe route recommendations. The 2019 Road Safety Technical Report also reaffirmed the importance of context-sensitive, localized safety strategies.

These studies highlight the growing shift toward intelligent, data-driven route safety systems using machine learning, spatial analytics, and real-time environmental data. While methods such as ensemble models, deep neural networks, and graph convolutional networks have shown strong predictive power, many frameworks remain constrained by rule-based logic, insufficient multi-source integration, and limited adaptability to dynamic travel conditions like flooding, congestion, or poor lighting.

To address these limitations, our study proposes a unified, machine learning-based risk assessment and route optimization model that incorporates both static (vehicle type, road infrastructure, demographics) and dynamic (traffic, weather, lighting, time) factors. By leveraging real-time data from APIs (Google Maps, OpenStreetMap, OpenWeatherMap, TomTom), the model computes segment-level safety scores, enabling informed, personalized route recommendations that prioritize both efficiency and safety.

3 Proposed Methodology

The methodology adopted in this research integrates static and dynamic risk parameters with real-time data collection, preprocessing, feature extraction, Risk Factor Assessment and machine learning-based risk prediction to determine the safest route for travelers. Fig. 1 illustrates the proposed methodology framework, detailing the sequential flow of data from user input to route optimization through real-time risk assessment and machine learning-based prediction.

The static and dynamic parameters are fixed in the earlier rule-based model [1]. In addition to the 14-features, a new feature, road crossing type, is added in the proposed model. The proposed methodology framework contains nine stages and described as in figure 1.

Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

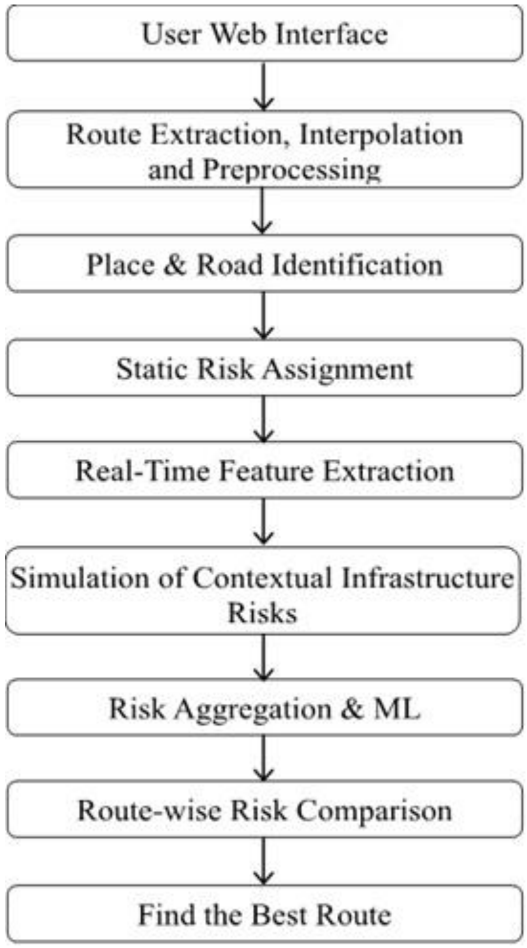


Figure 1. Methodology Framework

3.1 User Web Interface

The methodology for the user interface design revolves around capturing essential input parameters required for personalized and risk-aware route planning and illustrated in figure 2. The interface is structured to guide users in providing both geospatial data (source and destination) and demographic information (vehicle type, number of travelers, gender, and age).

Google Maps and Places APIs are integrated to allow intuitive search-based or map-click selection of locations, ensuring ease of use and accuracy. Dynamic Document Object Model (DOM) manipulation enables the form to adapt based on the number of travelers, rendering appropriate input fields in real-time. This user-friendly design ensures consistent and structured data entry to support backend risk assessment models.

Received: August 07, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

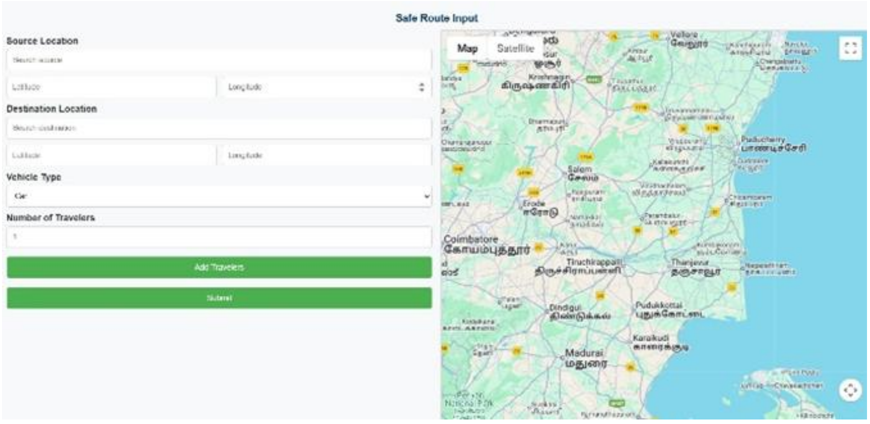


Figure 2. User Interface

3.2 Route Extraction, Interpolation and Preprocessing

In the route extraction and interpolation phase, the system uses the Google Directions API to obtain all alternate routes between the given source and destination. Each route, initially encoded as a polyline, is decoded into a sequence of GPS coordinates representing the full spatial path. To maintain uniform spatial granularity across all routes, these paths are interpolated at fixed 200-meter intervals. This ensures high-resolution analysis of road segments, curves, and intersections for effective risk assessment.

Each interpolated point is geotagged using the Google Geocoding API to extract place name and road type for risk evaluation. Data cleaning includes correcting "Unknown" or "Error" entries via forward or backward fill, removing duplicate GPS points, and inferring missing road types from nearby segments—ensuring reliable labeling for risk modeling.

3.3 Place and Road Identification

Using the Nominatim API, the place name is converted to a generalized OSM Place Type, which is mapped to a risk level as shown in table 1.

Table 1. Place type risk values

OSM Place Type	Risk Value
City	0.1
Town	0.2
Suburb	0.3
Village	0.4
Administrative	0.35
Centre	0.7
Station	0.25
Worship	0.35

Received: August 07, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Neighborhood	0.35
Park	0.25
House	0.35
Tertiary	0.5
Secondary	0.5

Each place type is assigned a risk level based on its environment—urban areas like cities are rated lower risk, while remote or undefined areas such as "Centre" carry higher risk. This helps quantify environmental safety during route risk assessment. Road types are assigned risk values based on their classification, allowing the model to assess infrastructure-related safety. High-class roads such as National Highways and State Highways are given lower risk values of around 0.2 to 0.3 due to better infrastructure and safety standards. Main roads have a risk value of 0.4, district roads 0.5, and town roads 0.6. Streets are rated at 0.7, while local roads carry the highest risk value of 0.8, reflecting their narrower design and potentially poorer maintenance. This classification helps evaluate the relative safety of each route segment.

3.4 Demographic Risk Assignment

To complement dynamic contextual risks, static user-related parameters are integrated into the route risk dataset. These represent relatively constant attributes during travel and help personalize the risk profile. Key parameters include vehicle type, gender, age, and group size, which collectively account for traveler vulnerability and behavioral risk when combined with dynamic factors.

3.4.1 User Demographic Input Extraction

Data from the web interface includes source and destination coordinates, vehicle type (e.g., Car, Van, Two-Wheeler), number of travelers, and demographic details (age, gender) for each individual. Inputs are parsed and validated—gender values normalized (e.g., 'M' to 'male') and ages converted to numeric types—for use in risk computation.

3.4.2 Vehicle Type Risk

Risk varies by vehicle due to protection and stability levels. Vehicle type risk is determined based on the protection and stability each type offers during travel. Two-wheelers are assigned the highest risk value of 0.7 due to their lower crash resistance and higher vulnerability in accidents. Cars, offering better structural protection, are assigned a moderate risk value of 0.5. Vans, which generally provide greater stability and enclosure for passengers, are assigned the lowest risk value of 0.3.

3.4.3 Gender-Based Risk Assignment

Studies [26-27] indicate that female travelers face higher risks, especially during night travel or in isolated areas. Female travelers are assigned a higher risk value of 0.9, reflecting increased vulnerability in certain travel conditions, especially during night

Received: August 07, 2025

travel or in isolated areas. Male travelers are assigned a lower risk value of 0.6. For each trip, the system calculates the average gender risk by aggregating these individual scores across all travelers in the group.

3.4.4 Age-Based Risk Assignment

Age affects risk due to varying levels of mobility and situational response. Higher risk values are assigned to children and elderly individuals, with scores of 0.9 for ages 15 and below and 0.55 for those over 60. Travelers aged 16–30 have a moderate risk value of 0.5, those aged 46–60 have a value of 0.3, and the lowest risk value of 0.2 is assigned to the 31–45 age group. The system computes the average age risk for all travelers in a group and incorporates it into the dataset.

3.4.5 Group-Risk Assignment

Group size plays a significant role in both perceived and actual safety. Solo travelers are assigned the highest risk value of 0.8 due to greater vulnerability, while a group of two has a lower risk value of 0.5. Groups of three to five travelers are considered safer, with a risk value of 0.3, and groups larger than five have the lowest risk value of 0.1.

3.5 Real-Time Feature Extraction

Real-time and structural factors significantly enhance route segment risk assessment. The system integrates dynamic parameters like weather, time, traffic, and surface quality, along with a static feature—road curvature—using live data sources.

3.5.1 Weather Risk extraction

Weather data for each route segment is obtained from the OpenWeatherMap API using GPS coordinates, and each condition is assigned a risk value to reflect its impact on travel safety. Clear weather is considered low risk with a value of 0.1, while cloudy conditions have a value of 0.3. Mist and fog increase the risk to 0.5 and 0.6, respectively. Drizzle is rated at 0.7, rain at 0.8, and thunderstorms pose the highest risk at 0.9 due to poor visibility and slippery roads. Snow also carries a high risk value of 0.7, and unknown conditions are assigned a default value of 0.5.

3.5.2 Time Risk

Travel time is converted to Indian Standard Time and categorized according to visibility and public activity levels. Night hours from 00:00 to 04:59 carry the highest risk value of 0.8, while early mornings from 05:00 to 06:59 have a risk value of 0.6. Morning hours from 07:00 to 10:59 are assigned a value of 0.3, and mid-day from 11:00 to 16:59 has the lowest risk value of 0.2. Evening travel from 17:00 to 19:59 is rated at 0.4, and late-night hours from 20:00 to 23:59 have a higher risk value of 0.7.

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.5.3 Traffic Risk Evaluation

The TomTom Traffic Flow API provides current and free-flow speeds per segment. The traffic congestion factor is computed using the (1).

$$\boldsymbol{Congestion = 1 - \left(\frac{Current\ Speed}{Free\ Flow\ Speed}\right)} \qquad (1)$$

Traffic congestion values ranging from 0.0 to 1.0 are translated into corresponding risk scores, where higher congestion indicates reduced maneuverability and a greater likelihood of accidents. Low congestion between 0.00 and 0.20 is assigned a risk value of 0.2, moderate congestion from 0.21 to 0.50 has a value of 0.4, heavier congestion between 0.51 and 0.80 is rated at 0.6, and severe congestion from 0.81 to 1.00 carries the highest risk value of 0.75. This approach ensures that real-time traffic conditions are effectively incorporated into the risk model, enabling context-aware route optimization.

3.5.4 Surface Risk Evaluation

Road surface (road condition) data is queried from OpenStreetMap using the Overpass API. A 20-meter radius is searched for surface attributes (e.g., asphalt, gravel, mud), which are mapped to risk and condition values and given in Table 2. If no surface is found, a default risk of 0.6 is applied.

Table 2. Surface risk assignment

OSM Place Type	Risk Value	Condition Label
Asphalt	0.1	Good
Paved, Concrete	0.2	Good
Paving Stones	0.3	Moderate
Compacted	0.4	Moderate
Gravel, Fine Gravel	0.5	Rough
Dirt, Ground and Pebble stone	0.6	Uneven
Grass	0.7	Damaged
Earth	0.8	Damaged
Mud, and Sand	0.9	Tough
Unknown	0.6	Default

Received: August 07, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

3.5.5 Curve Risk

Curvature risk (Road Complexity Risk) is determined by calculating the turning angle at each segment using three GPS points (*P1, P2, P3*). The geodesic distances between these points—*a* (from *P1* to *P2*), *b* (from *P2* to *P3*), and *c* (from *P1* to *P3*) —are computed using the Haversine or geodesic formula. The angle at *P2* is then derived using the Law of Cosines and given in (2).

$$\boldsymbol{\theta = cos^{-1}\left(\frac{a^2+b^2+c^2}{2ab}\right)} \qquad (2)$$

Table 3 maps angles to risk values, with sharper turns (e.g., <30°) carrying higher risk.

Table 3. Curve risk values with explanation

Turning Angle (°)	Risk Value	Explanation
< 30°	0.9	Very sharp turn, high accident risk
30° – 49°	0.8	Sharp curve
50° – 69°	0.7	Noticeable turn
70° – 99°	0.5	Gentle to moderate curve
100° – 139°	0.4	Mild bend
140° – 159°	0.3	Slight bend, low risk
≥ 160°	0.1	Almost straight, minimal risk
Error or Undefined	0.5	Default value in case of computational error

3.6 Simulation of Contextual Infrastructure Risks

In real-world travel, factors like lighting, CCTV presence, nearby public spaces, and crossing safety significantly influence route safety. Since these features are often unavailable via APIs, they are simulated using heuristics and randomization.

Each route segment includes three simulated binary features: cctv_risk (1 if CCTV is present), lighting_risk (1 if street lighting exists), and public_space_risk (1 if near parks, schools, or public buildings). These enhance the safety assessment of each segment.

Crossing risks are applied to 20–25% of segments in each route, with a randomly assigned crossing type determining the associated risk value. Highway crossings have a risk value of 0.25, state highway crossings are rated at 0.4, zebra crossings at 0.5, and

Received: August 07, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

narrow crossings at 0.7. Wider roads such as highways generally offer lower crossing safety, while designated crossings like zebra crossings provide relatively safer passage. Segments not selected for crossing are assigned a risk value of 0.0, and only numeric values are stored in the dataset for further analysis.

4 Implementation

4.1 Experimental setup

The experimental setup automates safe route generation using a web-based Flask interface to collect user inputs such as coordinates, vehicle type, and demographics. Submitted data is stored in a CSV file for processing. The backend comprises five Python modules: the first uses the Google Maps API for route interpolation and adds weather, time, and curve risks; the second maps vehicle, gender, age, and group size to risk scores; the third queries OpenStreetMap for surface types; the fourth integrates TomTom Traffic data; and the fifth injects contextual and crossing risks.

To enhance efficiency, the system uses ThreadPoolExecutor for parallel API calls across weather, traffic, and surface queries. This multithreading reduces latency and supports real-time processing while handling API rate limits. The entire pipeline is triggered via a POST request and outputs a risk-annotated dataset and an interactive map for visualizing safe route options.

4.2 Risk factor calculation

The risk factor computation in this study builds upon the framework in [1], originally using ten static and four dynamic features. We enhance this model by adding road crossing risk as an eleventh static parameter, enabling more precise assessment of pedestrian hazards, especially in high-speed or poorly marked areas. Mathematical formulations are presented to calculate static and dynamic risks at the segment level, which are then aggregated to determine the safest route. Each feasible route R_j (where $1 \leq j \leq m$) is broken into uniform 200-meter segments S_i (where $1 \leq i \leq N$), where composite risk scores—integrating demographics, road infrastructure, environmental, and traffic data—are computed to guide machine learning-based route safety prediction.

The Static Risk Factor (SRF_i) comprises eleven parameters related to fixed personal and environmental conditions—namely vehicle type, gender, age, group size, place type, road type, curve complexity, public space availability, lighting infrastructure, CCTV presence and road crossing type. This is formulated in (3).

$$SRF_i = \frac{1}{11} \sum_{k=1}^{11} P_{ik} \qquad (3)$$

where P_{ik} represents the static risk value of the k^{th} parameter at segment S_i. The Dynamic Risk Factor (DRF_i) captures real-time and contextual risks derived from weather conditions, traffic congestion, road surface type, and time of day. It is calculated as in (4).

Received: August 07, 2025

Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

$$DRF_i = \frac{1}{4}\sum_{k=1}^{4} D_{ir} \quad (4)$$

where D_{ir} denotes the dynamic risk value of the r^{th} parameter at segment S_i. The Final Risk Score (RF_i) for each segment is computed by averaging the static and dynamic components and provided in (5).

$$RF_i = \frac{SRF_i + DRF_i}{2} \quad (5)$$

To assess the overall safety of a route R_j, the segment-wise risk scores are aggregated and averaged, which is given in (6).

$$R_j = \frac{1}{N}\sum_{i=1}^{N} RF_i \quad (6)$$

The safest route (SR) is then determined by selecting the route with the lowest overall risk score among all feasible paths and given in (7).

$$SR = arg \min_{1 \le j \le m} \left(R_j\right) \quad (7)$$

4.3 Machine Learning Models

The machine learning models were implemented using the Scikit-learn library in Python, with additional support from Pandas, Matplotlib, and Seaborn for data processing and visualization. Three supervised learning algorithms—Random Forest, Gradient Boosting, and Decision Tree—were applied to assess the classification and regression performance using metrics such as accuracy, precision, F1-score, confusion matrix, and AUC-ROC.

5 Results and Discussion

This study compared the predictive performance of the proposed framework using real-time route data against an earlier rule-based model based on synthetic data. The system computes static and dynamic risks for each segment, averages them to obtain a final risk score, and identifies the optimal route with the lowest overall risk. A total of 1,425 segments from three alternative routes were analyzed after integrating all risk factors. The dataset was split 80:20 for training and testing. Gradient Boosting, Random Forest, and Decision Tree regression models were used to predict final risk scores and support safe route selection.

5.1 Route Map Visualization

The map visualization process starts when users input source and destination coordinates via a web form (see Figure 2). These coordinates are used to query the Google Directions API, which returns alternate routes. Each route is decoded into GPS coordinates using the polyline library. Routes are then plotted in different colors, with midpoint markers displaying the route number and distance for easy comparison. The resulting map is shown in Figure 3.

Received: August 07, 2025

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

This visual representation enhances user understanding and supports informed decision-making by combining spatial layout with key route metrics. Route 1 (green) spans 51.86 km, Route 2 (red) covers 57.73 km, and Route 3 (purple) is the longest at 79.5 km. Midpoint labels on the map allow users to easily compare route lengths and orientations.

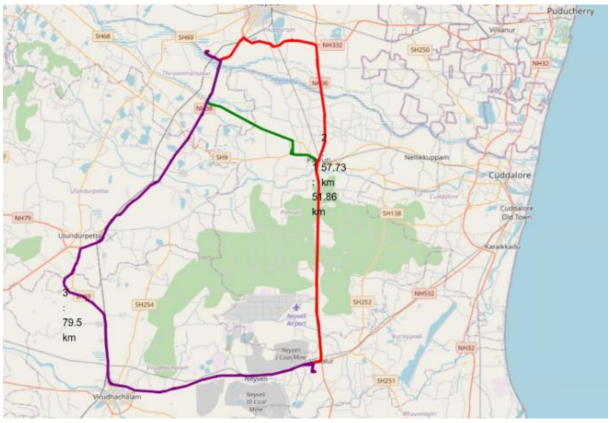


Figure 3. Route Map Visualization

5.2 Classification Metrics Analysis

A comparative analysis of Gradient Boosting, Random Forest, and Decision Tree models was conducted using key classification metrics: accuracy, precision, recall, and F1-score (Table 4, Figure 4). All models were evaluated on the same test set containing 154 high-risk segments, ensuring consistency and statistical validity in the results.

Table 4. Comparison of Classification Metrics

Metrics	Gradient Boosting	Random Forest	Decision Tree
Accuracy	98.95%	96.49%	98.25%
Precision	98.09%	96.15%	99.34%
Recall	100.00%	97.40%	97.40%
F1 Score	99.04%	96.77%	98.36%
Support	154	154	154

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

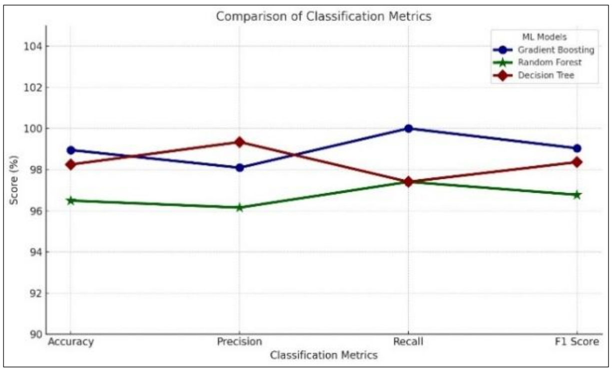


Figure 4. Visualization of Classification Metrics

The "Support" row reflects the total number of actual high-risk segments used to compute these scores. Among the models, Gradient Boosting excelled with the highest F1-score and perfect recall, showing its superior capability to detect high-risk segments without missing any true positives.

5.3 Regression Metrics Analysis

The regression performance metrics in Table 5 and Figure 5 compare the predictive accuracy of Gradient Boosting, Random Forest, and Decision Tree models in estimating final route risk scores.

Table 5. Regression Metrics of all Models

Metrics	Gradient Boosting	Random Forest	Decision Tree
MAE	0.00196	0.003375	0.004115
MSE	0.000007	0.000041	0.000065
R2 Score	0.997167	0.983915	0.974152

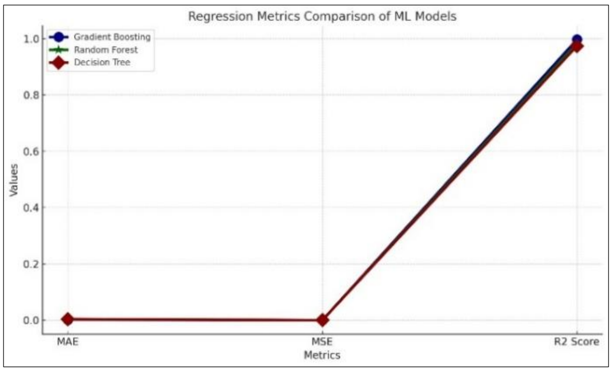


Figure 5. Visualization of Regression Metrics

Gradient Boosting achieves the best performance across all metrics—with the lowest MAE and MSE and the highest R^2 score—indicating its strong capability to accurately

Received: August 07, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

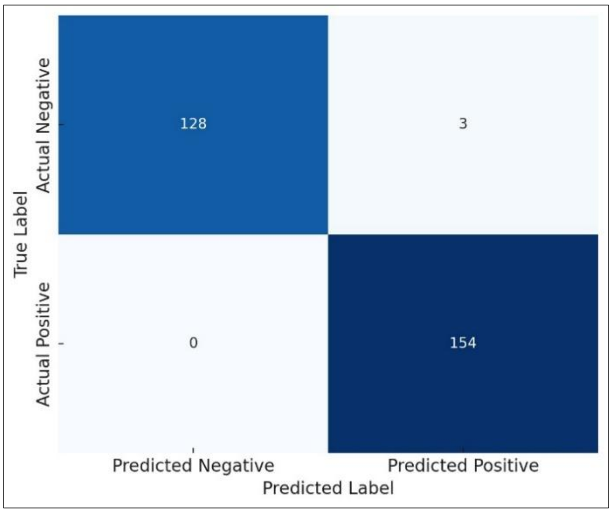
predict route risk with minimal error and high variance explanation. Random Forest and Decision Tree also perform well but show slightly lower predictive precision.

5.4 Confusion Matrix and ROC-AUC Analysis

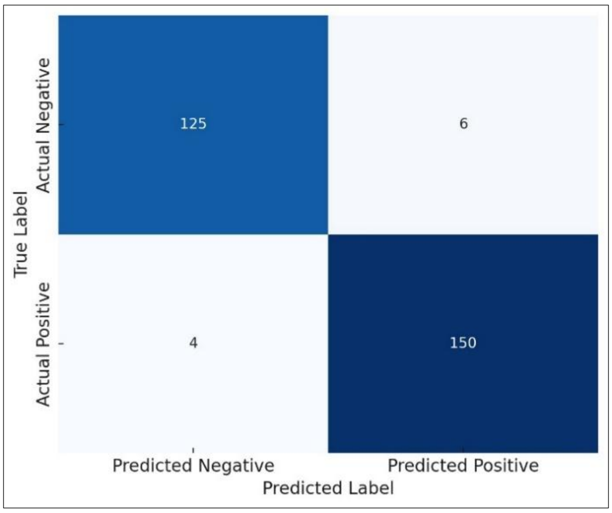
The confusion matrix for Gradient Boosting, Random Forest, and Decision Tree models provides deeper insight into their classification performance and shown in Table 6 and visualized in Figure 6.

Table 6. Confusion Matrix Analysis

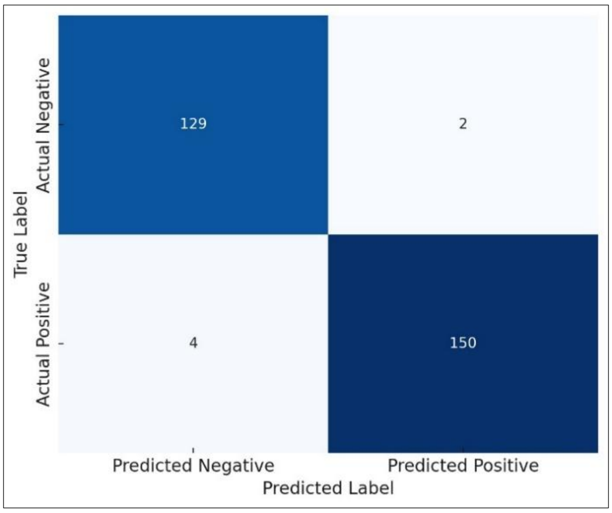
Model	True Negatives (TN)	False Positives (FP)	False Negatives (FN)	True Positives (TP)
Gradient Boosting	128	3	0	154
Random Forest	125	6	4	150
Decision Tree	129	2	4	150



(A)



(B)



(C)

Figure 6. Confusion Matrix Analysis: (A) Gradient Boosting, (B) Random Forest and (C) Decision Tree

Figures 6.(A) to 6.(C) show the confusion matrices for Gradient Boosting, Random Forest, and Decision Tree models. Gradient Boosting (Figure 6.a) achieves near-perfect classification with no false negatives and only three false positives, ensuring excellent recall and precision. Random Forest (Figure 6.b) misclassifies slightly more, with four false negatives and six false positives, slightly reducing both recall and precision. Decision Tree (Figure 6.c) has the fewest false positives (two) but shares the same false negatives (four) as Random Forest, indicating better precision but similar recall. Figure 7 presents the ROC curves, illustrating each model's ability to distinguish between high- and low-risk segments across all thresholds, with Gradient Boosting showing superior separability.

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

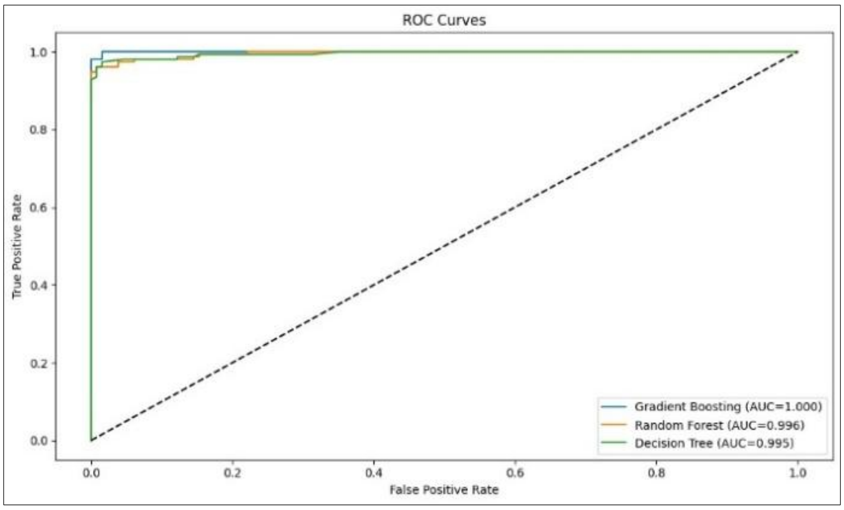


Figure 7. ROC Plot

The Gradient Boosting model achieves a near-perfect AUC of 0.9997, indicating exceptional classification accuracy with minimal false positives. Its ROC curve nearly touches the top-left corner, signifying ideal performance. Random Forest (AUC = 0.9956) and Decision Tree (AUC = 0.9955) also show strong discrimination ability but fall slightly behind. All models perform significantly better than random classification (AUC = 0.5), with Gradient Boosting clearly emerging as the most reliable for route safety prediction.

5.5 Risk Factor Comparison

The static, dynamic and final risk factors are determined using the equations discussed in section 4.2 and the three machine learning models generate the final risks. The results are shown in table 7 and visualized in Figure 8 for final risk factors.

Table 7. Final Risk Scores

Route No	Final Risk Score			
	Rule Based Model	Gradient Boosting	Random Forest	Decision Tree
1	0.374189	0.374164	0.373437	0.373917
2	0.374002	0.373884	0.373216	0.37375
3	0.367006	0.367018	0.366848	0.366907

Received: August 07, 2025

International Journal of Applied Mathematics

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

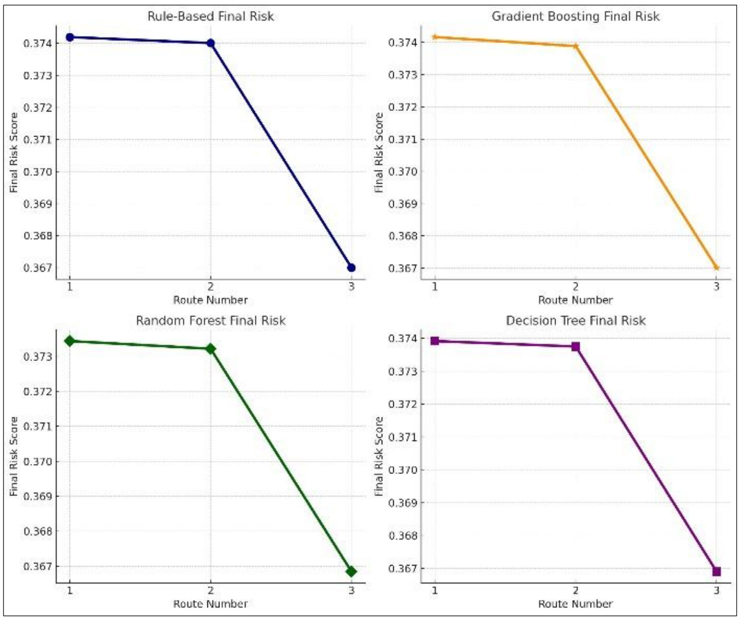


Figure 8. Comparison of Final Risk Scores of all models

All models, including the rule-based approach, consistently identify Route 3 as the safest and Route 1 as the riskiest. While the rule-based model relies on average risk values, machine learning models enhance prediction accuracy through learned patterns. Gradient Boosting aligns closely with rule-based scores but offers higher precision, Random Forest predicts slightly lower risks due to ensemble averaging, and Decision Tree may overfit. Despite its longer distance (79.5 km), Route 3's lower risk reflects safer conditions compared to the shorter, high-risk Route 1.

6 Conclusion And Future Work

The model integrates real-time features with contextually simulated parameters across feasible routes between the given source and destination. It employs three machine learning algorithms—Gradient Boosting, Random Forest, and Decision Tree—on the generated dataset to perform comprehensive route risk prediction. The Gradient Boosting model demonstrated superior performance across all evaluation metrics, achieving the lowest MAE (0.00196), MSE (0.000007), and the highest R^2 score (0.9972), with a perfect AUC of 1.000 and zero false negatives. Its predictions closely align with the rule-based approach while offering greater precision, affirming its reliability for route risk prediction. Route 3, despite being the longest (79.5 km), showed the lowest risk, highlighting the model's strength in identifying safer travel paths through favorable conditions.

For future enhancement, integrating contextual data such as crime reports and accident histories could improve predictive accuracy in urban settings. This would enable the system to evolve into a real-time, context-aware routing advisor that enhances safety and supports smart city mobility planning.

Received: August 07, 2025

Volume 38 No. 5s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

References

[1] T. Thilagavathi and A. Subashini. (2024). *Multi-factor risk assessment and route optimization for safe human travel*. International Journal of Advanced Computer Science and Applications, 15 (11), 426–435.

[2] Y. Berhanu, D. Schröder, B. T. Wodajo, & E. Alemayehu. (2024). *Machine learning for predictions of road traffic accidents and spatial network analysis for safe routing on accident- and congestion-prone road network*s. Results in Engineering, 23, 102737.

[3] F. Harrou, A. Zeroual, F. Kadri, & Y. Sun. (2024). *Enhancing road traffic flow prediction with improved deep learning using wavelet transforms*. Results in Engineering, 23, 102342.

[4] P. B. Silva, M. Andrade, & S. Ferreira. (2020). *Machine learning applied to road safety modeling: A systematic literature review*. Journal of Traffic and Transportation Engineering (English Edition), 7(6), 775–790.

[5] A. B. Z. Chai, B. T. Lau, M. K. T. Tee, & C. McCarthy. (2024). *Enhancing road safety with machine learning: Current advances and future directions in accident prediction using non-visual data*. Engineering Applications of Artificial Intelligence, 137, 109086.

[6] I. Abdulrashid, R. Z. Farahani, S. Mammadov, M. Khalafalla, & W.-C. Chiang. (2024). *Explainable artificial intelligence in transport logistics: Risk analysis for road accidents. Transportation Research Part E*: Logistics and Transportation Review, 186, 103563.

[7] H. Khanum, R. Kulkarni, A. Garg, & M. I. Faheem. (2024). *Enhancing road safety in India: A predictive analysis using machine learning algorithm for accident severity modeling*. In Civil Engineering. IntechOpen.

[8] A. Bhat, T. Talati, S. Vaidya, & S. Dhage. (2022). *Safe routes recommendation for drivers by real-time prediction of accident risk scores*. 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT), 1–6.

[9] M. Zheng, et al. (2019). *Traffic accident's severity prediction: A deep-learning approach-based CNN network*. IEEE Access, 7, 39897–39910.

[10] D.-J. Lin, M.-Y. Chen, H.-S. Chiang, & P. K. Sharma. (2022). *Intelligent traffic accident prediction model for Internet of Vehicles with deep learning approach*. IEEE Transactions on Intelligent Transportation Systems, 23(3), 2340–2349.

[11] L. Indraja, & D. Suneetha. (2023). *Safe path prediction using machine learning*. International Journal for Research in Applied Science and Engineering Technology (IJRASET), 11(7), 1771–1773.

[12] P. Litzinger, G. Navratil, Å. Sivertun, & D. Meier. (2012). *Using weather information to improve route planning*. Lecture Notes in Geoinformation and Cartography.

[13] E. B. Setiawan, G. Putera Husein, & A. Setiyadi. (2023). *Travel Route*

Received: August 07, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Recommendation System Based on Weather Prediction and Geolocation Technology. 2023 International Conference on Informatics Engineering, Science & Technology (INCITEST), 1–7.

[14] J. Wang, I. Besselink, & H. Nijmeijer. (2018). *Battery electric vehicle energy consumption prediction for a trip based on route information. Proceedings of the Institution of Mechanical Engineers, Part D*: Journal of Automobile Engineering, 232, 1528–1542.

[15] N. Zafar, & I. Ul Haq. (2020). *Traffic congestion prediction based on Estimated Time of Arrival*. PLoS ONE, 15.

[16] B. Kodikara, A. Thibotuwa, H. N. Perera, & P. T. Gamage. (2022). *Comparing the behaviour of ensemble algorithms for route optimization in last-mile delivery considering the weather condition and holiday effect*. 6th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI), 1–6.

[17] T. Qiu, C. T. Lam, B. Liu, B. K. Ng, X. Yuan, & S. Im. (2025). *FEN-MRMGCN: A frontend-enhanced network based on multi-relational modeling GCN for bus arrival time predictio*n. IEEE Access, 13, 5296–5307.

[18] J. Kim, J. Lee, W. H. Lee, & K. Yu. (2010). *Proposal for an Inundation Hazard Index of Road Links for Safer Routing Services in Car Navigation Systems*. ETRI Journal, 32(3), 430–439. https://doi.org/10.4218/etrij.10.0109.0478

[19] Y. He, Z. Jia, M. Hu, G. Zhang, & H. Dong. (2023). *The hybrid trip destination prediction model of vehicles based on autoencoder and high-order interaction features*. IEEE Transactions on Intelligent Transportation Systems, 24, 8443–8451.

[20] S. N. Van, N. V. Thi Hong, D. P. Quang, H. N. Xuan, B. Babaki, & A. Dries. (2022). *Novel online routing algorithms for smart people-parcel taxi sharing services*. ETRI Journal, 44(2), 220–231. https://doi.org/10.4218/etrij.2021-0406

[21] S. Elmi, & K. Tan. (2021). *DeepFEC: Energy consumption prediction under real-world driving conditions for smart cities*. Proceedings of the Web Conference 2021.

[22] A. E. Akay. (2020). *Determination of the safest route for logging trucks based on road types and conditions*. Environmental Sciences Proceedings.

[23] C. Cheng, S. Chen, Y. Ma, F. Qiao, & Z. Xie. (2024). *Crash severity prediction and interpretation for road determinants based on a hybrid method*. Journal of Transportation Safety & Security, 17, 30–56.

[24] G. Gianfranco, S. Soddu, & P. Fadda. (2017). *An accident prediction model for urban road networks*. Journal of Transportation Safety & Security, 10(4), 387–405.

[25] *Road Safety Evaluations Based on Human Factors Method* - Technical Report. (2019).

[26] K. S. Sandyal, Anisha, B. S. Sejal, L. R. Gupta and A. Bansal (2025). *Women Safety Platform with Safe Route Prediction Using Crime Data*. 2025 3rd

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2025, pp. 1855-1861, doi: 10.1109/IDCIOT64235.2025.10914699.

[27] D. Bura, M. Singh and P. Nandal (2019). *Predicting Secure and Safe Route for Women using Google Maps*. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 103-108, doi: 10.1109/COMITCon.2019.8862173.

Received: August 07, 2025