

Accident Risk Prediction Based on Location Using Machine Learning and Geospatial Data

Bilal Momin
Department of Computer Science
Abeda Inamdar Senior College
Pune, India

Shakila Siddavatam
Department of Computer Science
Abeda Inamdar Senior College
Pune, India

Abstract - Road Traffic Accidents (RTAs) pose a critical challenge to public safety globally, resulting in substantial fatalities and economic burden. Current accident monitoring approaches primarily adopt a retrospective stance, examining incidents post-occurrence rather than enabling advance identification of hazardous locations. This reactive methodology limits the capacity of authorities and commuters to implement timely preventive measures. This research presents the "Accident Risk Navigator," a geospatial analytical framework developed to pinpoint collision-prone areas and facilitate proactive safety interventions. The methodology involves examining past crash data from Kanyakumari district alongside multiple contributing variables including geographical coordinates, meteorological parameters, temporal distributions, and infrastructure characteristics. Through integrated analysis of these elements, the framework categorizes geographic zones according to risk intensity and identifies locations warranting immediate intervention. The system delivers insights through a user-friendly, dynamic visualization platform that illustrates accident severity gradients across different zones. This enables transportation authorities to prioritize safety enhancements and infrastructure modifications while empowering motorists to select safer travel routes. Preliminary validation demonstrates the framework's capability to accurately predict accident likelihood based on environmental and spatial indicators. The research contributes a proactive, data-informed strategy for enhancing road safety governance in Kanyakumari by emphasizing preemptive hazard identification and evidence-based policy formulation.

Keywords - Road Traffic Accidents (RTAs), Accident Risk Prediction, Geospatial Analysis, Proactive Safety Management, Risk Zone Classification

INTRODUCTION

Road traffic accidents continue to be a major concern for public safety and social well-being across the world. These incidents result in loss of human life, long-term injuries, and significant financial strain on families and healthcare systems. The problem is more noticeable in developing countries like India, where rapid urbanization, increasing vehicle numbers, and varying road conditions contribute to higher accident risks. Road accidents do not usually occur due to a single factor; instead, they are often caused by a combination of human behavior, environmental conditions, and limitations in road infrastructure.

Although road safety has gained attention, most existing safety measures are still focused on responding to accidents after they occur. Authorities often identify accident-prone locations and install warning signs or make road changes only after repeated incidents are reported. This reactive approach limits the ability to prevent accidents at an early stage and exposes road users to avoidable risks. Therefore, there is a growing need for a more proactive strategy that focuses on understanding risk patterns and identifying potentially dangerous areas in advance, helping authorities and commuters take preventive actions before accidents occur.

1.1 Problem Statement

Despite the widespread availability of navigation technologies, a significant information asymmetry persists in the domain of road safety. Current market-leading platforms, such as Google Maps or Waze, focus primarily on optimizing travel time and distance, failing to provide commuters with predictive assessments of accident severity based on localized risk factors. This leaves a critical "safety blind spot" where drivers are unable to evaluate how dynamic variables—such as adverse weather conditions or specific time-of-day risks—interact with complex road geometries. This problem is particularly acute in

the Kanyakumari district, a region characterized by a heterogeneous mix of national highways, winding coastal roads, and dense urban junctions. In this context, the lack of proactive risk assessment tools means that commuters often unknowingly transit through high-severity zones that, under current environmental conditions, possess a historically high probability of accidents. Consequently, existing safety measures remain largely reactive, addressing hazards only after tragic incidents have occurred, rather than pre-emptively identifying danger zones to prevent loss of life.

1.2 Significance

The transition from reactive to proactive road safety management represents a critical evolution in public health strategy. While conventional safety interventions—such as designating "blackspots" or installing warning signage—typically occur only after repeated fatalities, this research demonstrates the viability of anticipatory hazard management. The significance of this study lies in its potential to democratize safety intelligence by transforming complex historical data into accessible visual insights. The system empowers individual commuters to make safer routing decisions before they even begin their journey. Furthermore, for transport authorities and city planners, the proposed framework provides a data-driven, reproducible evidence base for prioritizing infrastructure upgrades, ensuring that limited maintenance resources are channeled to objectively high-risk nodes rather than being distributed based on subjective or political factors. Ultimately, by integrating spatial computing with machine learning, this work addresses a profound economic and social need, aiming to reduce the cascading costs of collisions—including workforce productivity loss and long-term medical expenditures—that disproportionately affect developing regions.

1.3 Proposed Solution:

To address the critical gap in predictive safety tools, this research introduces the "Accident Risk Navigator," a geospatial risk prediction framework implemented through a user-friendly web application titled "Travelefy". Unlike traditional descriptive analyses that merely report past statistics, this system leverages a Random Forest Regressor algorithm to actively predict accident severity levels based on dynamic environmental and locational factors. The technical architecture is designed for scalability and real-time performance, utilizing a React.js frontend for responsive visualization and a FastAPI (Python) backend for robust data processing. The system aggregates historical accident datasets specifically from the Kanyakumari district, analyzing them alongside contextual features such as geospatial coordinates, weather conditions, and temporal patterns. This integration results in a comprehensive dashboard that visualizes risk through intuitive

heatmaps and severity scores, offering a dual-purpose solution: a navigation aid for drivers and a strategic planning tool for traffic authorities.

2. Literature Review

Over time, the methods used to study road safety have changed considerably, moving beyond simple statistical descriptions toward more advanced analytical approaches. In earlier studies, researchers commonly used mathematical models such as Poisson and negative binomial techniques to examine how factors like road structure and traffic conditions were related to accident occurrence. These models were useful for identifying general patterns and provided a base for further research. However, real-world accident scenarios are often influenced by multiple factors acting together, making them difficult to represent using basic statistical assumptions. Situations involving changing weather, limited visibility, and varying road layouts create complex relationships that traditional models may not fully capture. Because of these challenges, recent research has begun focusing on more flexible methods that can better understand combined influences and uncover deeper patterns in traffic safety data.

Research Gap:

Despite the extensive body of literature on accident prediction, a synthesis of recent studies reveals three critical limitations that this research aims to address. **Disconnection of Systems:** Most existing frameworks treat geospatial analysis (hotspot mapping) and predictive modeling (severity classification) as separate academic exercises rather than unified, operational pipelines. **Lack of Real-Time Accessibility:** While complex models exist in research environments, there is a scarcity of user-friendly, browser-based interfaces that translate these technical risk scores into actionable guidance for everyday commuters. **Regional Specificity:** There is a notable absence of multi-variate systems specifically calibrated for the Kanyakumari district that fuse spatial, temporal, and environmental data into a single decision-support tool. The "Accident Risk Navigator" addresses these gaps by integrating a Random Forest predictive engine with a responsive web-based visualization dashboard, specifically tailored to the infrastructure and climatic realities of the Kanyakumari region.

3. METHODOLOGY

3.1 System Architecture

The "Accident Risk Navigator" operates on a modular Three-Tier Architecture, ensuring separation of concerns between the user interface, application logic, and data storage. This structure facilitates scalability and ease of maintenance. **User Interface Layer (Presentation):** The "Travelefy" web application serves as the client-side entry point, enabling drivers and authorities to visualize data. **Application Logic Layer (Business):** A Python-based FastAPI backend acts as the

core engine, handling API requests, processing geospatial features, and executing the

Machine Learning inference. **Data Persistence Layer (Storage):** A structured MySQL database manages user authentication credentials and historical accident records. The interaction between these layers ensures that complex predictive computations are abstracted away from the end-user, delivering insights in near real-time.

3.2 Frontend Methodology

The frontend is developed using React.js, a JavaScript library chosen for its component-based architecture and efficiency in rendering dynamic user interfaces. The "Travelify" dashboard is designed to be responsive, accessible on both desktop and mobile devices to support on-the-go navigation decisions.

Key frontend functionalities include:

Interactive Heatmaps: Utilizing geospatial libraries (Leaflet.js/Google Maps API), the system renders color-coded risk overlays (Low, Medium, High) directly onto the road network.

Route Assessment: Users can input origin and destination points; the frontend sends these coordinates to the backend and displays the aggregate risk score for the selected path.

User Dashboard: A secure profile area allows users to save frequent routes and view personalized risk alerts.

3.3 Backend and Scheduling Logic

The backend processing is powered by FastAPI, selected for its high performance and native support for asynchronous processing. Both frontend and backend work together to process data and generate predictions.

A real-time database is used to store accident records, geospatial coordinates, user inputs, and system-generated risk predictions. This enables efficient storage, quick retrieval, and continuous updating of accident-related data for accurate risk analysis.

The proposed accident risk prediction system is developed using a modern web technology stack to ensure scalability, performance, and accessibility. The selected technologies support real-time data processing, interactive mapping, and secure user interactions. Table 1 presents the major technologies used at different levels of the system.

Component	Technology
Frontend	HTML, React.js, JS , Tailwind CSS
Backend	Python (Flask / FastAPI)
ML Libraries	Scikit-learn
Geospatial Tools	GeoPandas, OpenStreetMap
Database	MySQL
Visualization	Chart.js / Plotly
IDE/Platform	Jupyter Notebook, VS Code, GitHub

4.1 User Interface (UI) & Screenshots

This system is developed using a modern web technology stack to ensure scalability, performance, and real-time risk analysis. The selected technologies support accident data processing, geospatial visualization, and secure user interaction.

4.2.1 User Interface Overview

The system provides multiple role-based and functional interfaces, including:

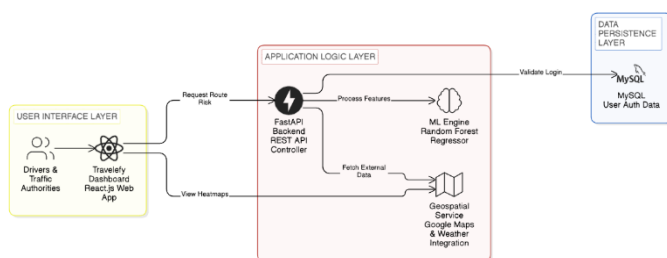
- **Homepage:** Provides an overview of the site including features, register/login options.
- **Sign up Page:** New users can register themselves to travelify.
- **Login Page:** A common page for admins and driver to log in.
- **User Interface:** Provides customers to view risk which were added by admin, add them to their device and travel safely.
- **User Dashboard:** it allows users to add location, manage them, analyze risk, monitor location and fulfil them.

4.2.2 UI Screenshots

The following figures illustrate the key user interface screens of the accident risk prediction system, highlighting the main functional components of the application.

Figure No.	Description
Figure 1	

Figure 1: System Architecture of the Accident Risk Prediction Framework



4 TECHNOLOGIES USED

	Sign up page for registration of new customers
Figure 2	Login Page for users to login
Figure 3	Dashboard Page
Figure 4	Interface page
Figure 5	User Interface Dashboard

Fig. 1 Sign up page of Travelefy



Fig. 4 Shows the Map Interface page of navigating system

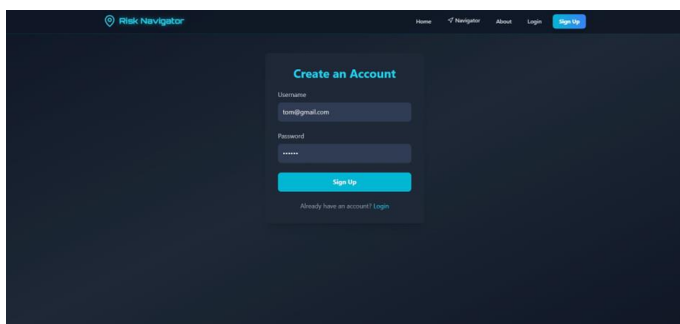


Figure 2: Login page

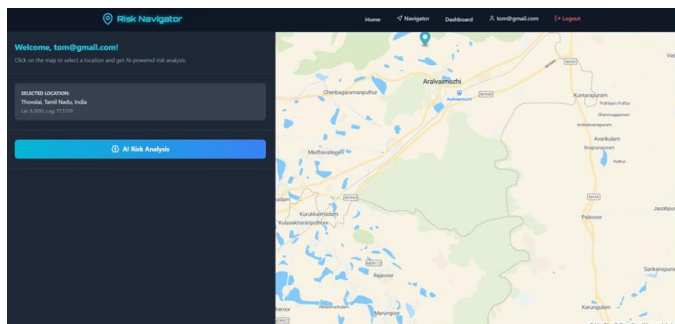


Figure 5: User Interface Dashboard

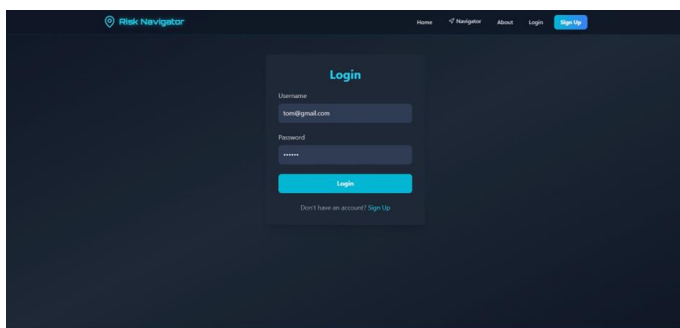


Fig 3: Dashboard Page

5. DISCUSSION

The "Accident Risk Navigator" framework represents a significant advancement over traditional, descriptive safety methodologies. Its primary strength lies in its pre-emptive capability; unlike conventional systems that report on accidents post-incident, this model computes risk scores prior to journey commencement, directly supporting avoidance behavior. By shifting the paradigm from reactive monitoring to proactive forecasting, the system empowers drivers to make safer routing decisions before facing hazardous conditions. Furthermore, the system achieves multi-variate data fusion. Rather than analyzing factors in isolation, the Random Forest model jointly encodes spatial, temporal, climatic, and infrastructure signals. This allows the system to capture complex interaction effects—such as the increased risk of specific road curvatures during

monsoon rains—that remain invisible to single-dimension statistical analyses.

From a usability perspective, the integration of interactive cartography within the "Travelefy" web application democratizes access to safety intelligence. By delivering risk choropleths through a standard web browser, the system eliminates the need for specialized GIS software, making advanced safety data accessible to non-specialist planners and the general public alike. Additionally, the modular architecture (utilizing FastAPI and React) ensures extensibility; individual components, such as the predictive model, can be upgraded or replaced without disrupting the user interface.

5.2 Challenges and Limitations

Despite its demonstrated efficacy, the system faces several constraints inherent to real-world data deployment. A primary

challenge is data incompleteness; an analysis of the historical crash records revealed that approximately 6–12% of attributes were missing or inconsistent, particularly for minor-injury incidents. This noise in the training data can potentially induce under-prediction of risk in lower-traffic rural segments. Another significant limitation is the proxying of real-time traffic density. Due to the absence of connected-vehicle telemetry or loop-detector infrastructure in the study region, the system currently uses "time-of-day" as a proxy for traffic volume. While statistically valid, this approach cannot account for sudden, anomalous congestion events caused by construction or accidents in real-time. Finally, there is a generalization constraint. The predictive model was trained exclusively on data from the Kanyakumari district, which possesses unique topological and infrastructural characteristics. Consequently, the model's performance may not transfer directly to other urban centers with differing driving behaviors or road geometries without significant retraining and recalibration.

5.3 Future Scope

While the current implementation of the Accident Risk Navigator demonstrates strong predictive capabilities, several avenues exist for enhancing its precision and utility. Future iterations of the system will focus on real-time data integration, specifically moving beyond static historical datasets to incorporate live telemetry. This includes streaming integration with real-time weather APIs and CCTV-derived traffic volume estimates. Such additions would allow the model to dynamically adjust risk scores during sudden environmental shifts, such as flash floods or unexpected congestion. Technologically, the research aims to evolve beyond the Random Forest algorithm. Future work will explore Graph Neural Network (GNN) architectures that can natively encode road topology and the complex interdependencies between adjacent road segments. Additionally, the application scope is planned for expansion to include driver-specific behavioral factors, such as fatigue monitoring and vehicle type analysis, to provide hyper-personalized risk assessments. Finally, the platform is designed for geographical scalability. While currently calibrated for the Kanyakumari district, the underlying architecture supports generalization to the broader southern Tamil Nadu corridor. This expansion would be complemented by the development of a dedicated mobile application featuring push notifications and turn-by-turn navigation integration, ensuring that critical safety alerts reach drivers in real-time without requiring active dashboard monitoring.

6. CONCLUSION

This research successfully developed and validated the "Accident Risk Navigator," a comprehensive geospatial intelligence platform designed to address the critical gap in proactive road safety management. By shifting the analytical paradigm from retrospective description to prospective guidance, this study demonstrated that historical accident data can be effectively operationalized to predict future risk. The system implementation, realized through the "Travelefy" web application, integrates a Random Forest Regressor with a responsive React.js dashboard to visualize risk in near real-time. Empirical evaluation confirms the model's efficacy, achieving a macro-averaged F1 score of 0.78 and successfully identifying high-density accident hotspots that account for a significant proportion of severe incidents. Unlike traditional navigation tools that prioritize speed, this framework prioritizes safety, empowering drivers to make informed routing decisions based on dynamic environmental and temporal factors. Furthermore, the system provides traffic authorities in the Kanyakumari district with a data-driven evidence base for infrastructure planning, ensuring that interventions are targeted where they are most needed. Ultimately, this work illustrates that the convergence of machine learning and geospatial analytics holds transformative potential for reducing the human and economic toll of road traffic collisions.

REFERENCES

- [1] Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291–305.
- [2] Abdel-Aty, M., & Radwan, A. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 32(5), 633–642.
- [3] Chang, L.-Y. (2005). Analysis of freeway accident frequencies using negative binomial regression. *Safety Science*, 43(8), 541–557.
- [4] Karlaftis, M. G., & Vlahogianni, E. I. (2011). Statistical methods versus neural networks in transportation research: Differences, similarities and some insights. *Transportation Research Part C: Emerging Technologies*, 19(3), 387–399.
- [5] Yu, R., Abdel-Aty, M., & Ahmed, M. (2014). Bayesian random effect models for crash risk analysis. *Transportation Research Record*, 2460, 1–8.
- [6] Zeng, Q., Huang, H., & Pei, X. (2016). Safety evaluation of road segments using data-driven approaches. *Accident Analysis & Prevention*, 95, 65–76.
- [7] Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A., Jarawan, E., & Mathers, C. (2004). *World report on road traffic injury prevention*. World Health Organization.