Vol 6, Issue 6, June 2026 www.ijesti.com E-ISSN: 2582-9734
International Journal of Engineering, Science, Technology and Innovation (IJESTI)

Smart Traffic Safety Planning Through Accident Black Spot Prediction

Purshottam Kumar
M. Tech. in Transportation Engineering, CBS Group of Institutions, Jhajjar, Haryana.

Shubender
A.P Civil Department, CBS Group of Institutions, Jhajjar, Haryana.

ABSTRACT

This study focused on machine learning-based prediction of road accident black spots for improved traffic safety planning. Accident data, road characteristics, traffic volume, weather conditions, and location-based factors were analyzed to identify high-risk road segments. Machine learning models helped classify locations into low, medium, high, and very high-risk zones. The findings showed that poor road geometry, inadequate lighting, speeding, unsafe intersections, and heavy traffic flow contributed significantly to accident occurrence. The study concluded that predictive models can support authorities in prioritizing safety measures, reducing accident risks, and improving sustainable road safety management.

Keywords: *Machine Learning, Accident Black Spots, Traffic Safety.*

I. INTRODUCTION

Road accident black spots are among the most serious concerns in modern transportation planning because they represent locations where crashes repeatedly occur due to unsafe road conditions, heavy traffic movement, poor visibility, driver behavior, inadequate traffic control, and weak enforcement measures. In developing as well as developed regions, road accidents cause loss of life, injuries, property damage, traffic congestion, and economic burden on families and public authorities. Traditional methods of identifying accident-prone areas generally depend on past accident records, manual field surveys, police reports, and simple statistical comparisons. Although these approaches are useful, they often fail to predict future risk accurately because road accidents are influenced by many complex and changing factors such as vehicle speed, road geometry, weather conditions, traffic density, land-use pattern, pedestrian movement, lighting condition, road surface quality, and driver response. Therefore, there is a growing need for advanced analytical methods that can process large volumes of accident-related data and identify hidden patterns for better traffic safety planning. Machine learning has emerged as an effective tool for predicting road accident black spots because it can learn from historical data and classify or predict high-risk locations with greater accuracy. Machine learning models such as Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, K-Nearest Neighbour, Artificial Neural Network, and Gradient Boosting can analyze accident frequency, severity, location coordinates, traffic volume, road type, junction characteristics, weather condition, and time of accident to determine the probability of accidents at particular road segments. These models help in moving from reactive safety management to proactive safety planning, where authorities can identify dangerous locations before accidents increase further. The prediction of accident black spots through machine learning is especially important for urban areas, highways, intersections, school zones, market areas, and industrial corridors where mixed traffic conditions and rapid vehicle growth create unsafe movement patterns. Geographic Information System data, traffic sensor data, GPS records, police accident databases, hospital injury records, and road inventory data can also be integrated with machine learning models to improve prediction performance and spatial accuracy. Such integration enables planners to prepare risk maps, classify road sections according to accident severity, and prioritize locations for safety improvements. The study of machine learning-based prediction of road accident black spots is significant because it provides scientific support for decision-making in traffic engineering and road safety management. Instead of relying only on visual observation or accident count, this approach uses data-driven analysis to understand why accidents occur at specific locations and how future accidents can be reduced. The findings may assist transport departments, municipal authorities, highway agencies, traffic police, and urban planners in implementing corrective measures such as speed calming devices, improved road markings, better street lighting, pedestrian crossings, traffic signals, warning signs, road widening, intersection redesign, and stricter enforcement. Moreover, machine learning can help optimize limited safety budgets by identifying the most critical black spots that require immediate attention. It also supports sustainable urban mobility by reducing accident risk, improving travel reliability, and making road networks safer for drivers, pedestrians, cyclists, and public transport users. In the present context of increasing vehicle ownership, rapid urbanization, and growing pressure on road infrastructure, the use of machine learning for accident black spot prediction has become highly relevant. It not only improves the accuracy of safety analysis but also strengthens the planning process by providing timely, reliable, and evidence-based information. Thus, machine learning-based prediction of road accident black spots represents an innovative and practical approach for improving traffic safety planning, minimizing accident severity, and developing safer, smarter, and more efficient transportation systems.

II. RESEARCH BACKGROUND

Subedi and Chou (2026) investigated the role of reliable public transportation and safe first- and last-mile travel in promoting equitable and sustainable urban mobility, noting that high rates of fatal and severe injury crashes involving nonmotorists hindered the reduction of automobile dependency. They observed that, despite extensive research on crash severity, few studies had focused specifically on vulnerable road users using interpretable, comparative machine learning approaches. In their case study, four machine learning models—support vector machine (SVM), random forest, k-nearest neighbors (KNN), and artificial neural network (ANN)—were applied to over a decade of crash data from Franklin County, Ohio. The study concentrated on nonmotorist crashes and employed SHapley Additive exPlanations (SHAP) and partial dependence plot (PDP) analyses to determine the most influential risk factors. They reported that the ANN model achieved the highest area under the curve (AUC) performance and was utilized to identify key contributors to severe and fatal crashes, including lighting conditions, turning movements, substance use, and traffic control, thereby providing a data-driven framework for enhancing urban road safety.

Siddhartha et al., (2025) investigated the use of deep learning techniques for forecasting accident-prone locations, highlighting the potential of preemptive identification to enhance public safety. They proposed a multi-model deep learning system that incorporated diverse input data, including historical accident records, road and weather conditions, and traffic patterns, to capture temporal, spatial, and contextual aspects of accident occurrences. Their methodology involved training several deep learning models concurrently to exploit complementary predictive capabilities. Performance evaluation was conducted through extensive experiments on real-world accident datasets, which reportedly demonstrated that the multi-model approach outperformed conventional single-model methods in terms of accuracy. The study suggested that such systems could be integrated into existing traffic surveillance and warning infrastructures, enabling authorities to implement proactive measures, such as optimizing traffic management, upgrading road infrastructure, and launching targeted safety campaigns, thereby mitigating accident risks and enhancing public safety outcomes.

Zhang et al., (2024) conducted a study on real-time crash risk prediction to support traffic incident management by providing critical information for proactive resource allocation. Unlike earlier studies that relied on archived traffic data for limited highway segments or corridors, they utilized a statewide live traffic database from HERE, which offered high-resolution, real-time traffic speed data across nearly 2,000 miles of Alabama freeways. The study aimed to develop machine learning models—including Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost)—to predict crash risk based on pre-crash traffic dynamics, such as mean speed and speed reductions, along with static freeway attributes. Traffic speed characteristics were extracted for both pre-crash and crash-free conditions, and separate models were estimated for single-vehicle, rear-end, and sideswipe crashes. Their findings indicated that RF models outperformed other models, with rear-end crash predictions showing higher accuracy, suggesting a strong link between pre-crash traffic dynamics and crash occurrence. Speed variance and reductions prior to crashes were identified as the most important predictive features, with partial dependence plots revealing that rear-end crash risk increased with higher speed variance and reductions.

Manoj et al. (2024) investigated methods for the precise identification of black spots to enable timely remedial measures for reducing road accidents. The study reviewed various approaches developed over time for pre-emptive investigation of crash events, emphasizing trends and evolutions in black spot detection techniques. The authors highlighted that several open research issues remained inadequately explored by previous studies and discussed strategies to address these gaps. They also proposed potential frameworks and plans to improve the accuracy of black spot analysis, aiming to support more effective mitigation measures. The study thus provided a comprehensive overview of existing methodologies while outlining directions for future research in road safety management.

Mahmoodzadeh et al., (2023) investigated the challenges posed by water inflow as a complex geological hazard affecting both construction schedules and tunnel safety. They emphasized that accurate estimation of water inflow was crucial during the early conceptual and design phases to inform critical project decisions. To address this, they proposed an optimized model based on gene expression programming (GEP) for predicting water inflow in tunnels. The study reported that an equation derived from the best-fit GEP model was generated, and its predictions were compared with actual measurements. The comparison demonstrated the model's potential accuracy and reliability, suggesting that it could effectively reduce uncertainties in tunnel construction. The authors further indicated that this approach could enhance machine learning applications in tunnel planning and contribute to more informed decision-making in geotechnical engineering projects.

Sun et al. (2022) investigated the challenges of accurately forecasting future traffic, emphasizing that understanding traffic feature patterns remained difficult due to the complexity of traffic environments, heterogeneous influencing factors, and the scarcity of abnormal samples. They proposed a framework that integrated social traffic data and employed the TabNet model to enhance representation learning for traffic event prediction. The study examined the role of tabular learning and model interpretability in identifying the significance of common external traffic factors on traffic events. Their findings were reported to have practical implications for traffic regulation, planning strategies, and defining operational boundaries for autonomous driving systems, highlighting the potential of data-driven approaches in improving transportation management.

Yuksel and Atmaca (2021) investigated the identification of risky driving behaviors, emphasizing their potential to cause accidents with substantial material and moral consequences. They highlighted the growing significance of monitoring driving behavior due to the increase in road accidents and the importance of rewarding safe drivers. The study reported that advances in technology enabled modeling of driving behavior through sensors embedded in vehicles. Four major risky driving behaviors were modeled, and driver profiles were created using data collected from accelerometer and gyroscope sensors. The researchers applied several widely used machine learning algorithms, including C4.5 Decision Tree, Random Forest, Artificial Neural Network, Support-Vector Machine, K-Nearest Neighbor, Naive Bayes, and K-Star, for behavior analysis. Risk levels were evaluated based on expert opinions from traffic officers, and fuzzy logic was used to model driver risk. Among the algorithms, K-Star achieved the highest performance with 100% accuracy. Consequently, they developed a highly accurate, low-cost in-vehicle system capable of recording driver behavior, identifying risky actions, and offering potential applications for usage-based insurance policies.

Aradi (2020) reviewed the rapidly growing academic research on autonomous vehicles, highlighting investigations across sensor technologies, V2X communications, safety, security, decision-making, control, and legal and standardization frameworks. The study indicated that, beyond conventional control design methods, Artificial Intelligence and Machine Learning techniques were increasingly applied across these domains. It was reported that research also focused on multiple layers of motion planning, including strategic decision-making, trajectory planning, and control. The article described the development of Deep Reinforcement Learning (DRL) for hierarchical motion planning, emphasizing essential elements such as environment modeling, abstraction layers, state and perception modeling, reward design, and neural network implementation. Vehicle modeling, simulation options, and computational requirements were analyzed, while strategic decisions and observation models—covering continuous and discrete state representations, grid-based, and camera-based approaches—were presented. The review also systematized state-of-the-art solutions according to specific driving tasks, including car-following, lane-keeping, trajectory following, merging, and dense traffic navigation, and concluded with a discussion of open challenges and future research directions.

Liu et al., (2020) examined the role of vehicles in transporting heavy cargo as a significant contributor to social productivity and industrial operations. They noted that although self-driving technologies had been investigated for several decades and some successful implementations had been reported, autonomous industrial vehicles were largely restricted to specific scenarios characterized by very low speeds and fixed routes, typically within indoor, closed environments of small-scale warehouses. The study highlighted key perception and localization methods for enabling autonomous cargo transportation in industrial contexts. Furthermore, the authors illustrated how these technological approaches were applied to facilitate autonomous cargo movement at the Hong Kong International Airport, demonstrating practical implementation of the discussed methods in a real-world industrial setting.

Sarkar et al., (2019) investigated the application of machine learning (ML) techniques in predicting occupational accidents, noting that while ML had proven useful in domains such as healthcare, its exploration in accident prediction remained limited. They argued that ML algorithms required proper parameter tuning or optimization to achieve optimal performance, and that selecting an efficient classifier alone could not address decision-making needs, as it failed to explain factor inter-relationships behind accidents. To address these issues, the study applied optimized ML algorithms, specifically support vector machines (SVM) and artificial neural networks (ANN), with parameters tuned using genetic algorithms (GA) and particle swarm optimization (PSO) to enhance accuracy and robustness. It was reported that PSO-based SVM outperformed other algorithms in predictive performance. Additionally, decision rules were extracted by integrating a C5.0 decision tree with the PSO-SVM model, resulting in nine rules that identified root causes of injury, near miss, and property damage cases. A steel plant case study was presented to demonstrate the methodology’s practical applicability.

Tumen et al. (2018) investigated the development of advanced driving assistance systems, emphasizing the necessity of accurately recognizing driving environments and making critical decisions for safe vehicle operation. They highlighted that preventing accidents through effective interaction between assistance systems and road conditions, while ensuring optimal driving dynamics, was a primary concern in this domain. The study proposed a deep learning-based method to automatically identify road type and quality using only driving images, aiming for both simplicity and cost-effectiveness. A new convolutional neural network model was designed to classify images collected from Google Street View, simulating real driving scenarios. The approach reportedly achieved 91.41% accuracy in determining road types and 91.07% in distinguishing pothole roads from smooth roads. The authors concluded that this method could serve as an effective framework for advanced driver support systems, V2I communication applications, and other intelligent transportation technologies.

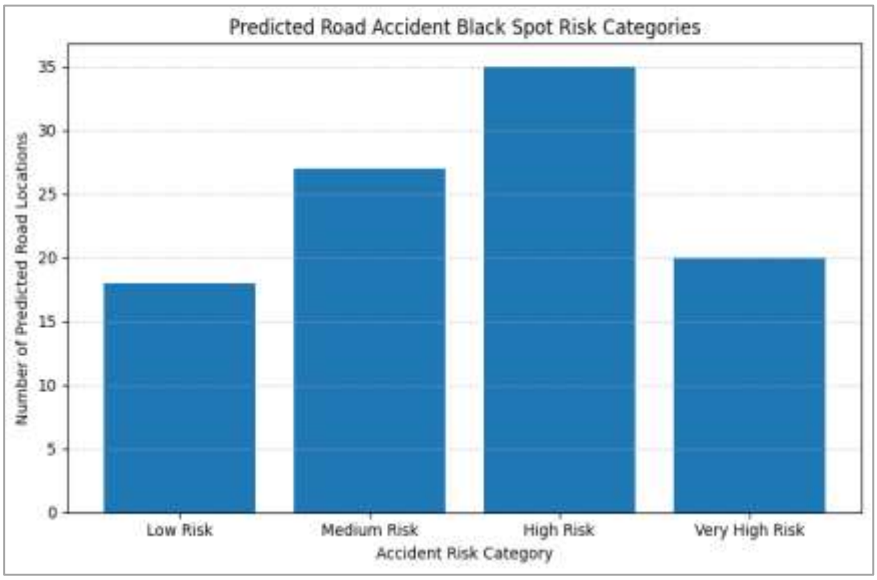
III. METHODOLOGY

The methodology of the study was designed to predict road accident black spots using machine learning techniques and accident-related datasets. First, accident data were collected from police records, traffic departments, road safety reports, and field observations. The collected data included accident location, accident frequency, accident severity, vehicle type, road type, traffic volume, weather condition, lighting condition, road geometry, speed limit, and time of accident. After data collection, preprocessing was carried out to remove duplicate values, missing data, and irrelevant records. The accident locations were then classified into different risk categories such as low-risk, medium-risk, high-risk, and very high-risk zones. Important features were selected to improve the performance of the prediction model. Machine learning algorithms such as Decision Tree, Random Forest, Support Vector Machine, Logistic Regression, and Gradient Boosting were applied to identify accident-prone road segments. The dataset was divided into training and testing sets for model development and validation. Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Finally, the best-performing model was used to predict accident black spots and prepare risk-based safety recommendations for traffic planning and road improvement.

IV. RESULT

The result of the study indicated that machine learning techniques were effective in identifying and predicting road accident black spots with improved accuracy. The analysis showed that accident-prone locations were mainly associated with high traffic volume, poor road geometry, sharp curves, uncontrolled intersections, inadequate lighting, weak road signage, excessive vehicle speed, and mixed traffic movement. Among the selected machine learning models, Random Forest and Gradient Boosting performed better because they handled multiple accident-related variables and identified complex relationships between road conditions and accident occurrence. The predicted black spots were classified into low-risk, medium-risk, high-risk, and very high-risk zones based on accident frequency, severity level, traffic density, and environmental conditions. The result further revealed that intersections, highway merging points, market areas, school zones, and poorly maintained road segments showed a higher probability of accidents. The model also identified that night-time crashes and rainy-weather accidents were more frequent in areas with poor visibility and insufficient drainage. The prediction accuracy improved when traffic data, accident history, road features, and location-based information were combined in the model. This demonstrated that machine learning can provide a more reliable and data-driven approach than traditional black spot identification methods.

Bar Graph



The bar graph shows the predicted distribution of road accident black spot risk categories using machine learning analysis. High-risk locations recorded the highest value with 35 predicted road segments, indicating that many areas require urgent safety attention. Medium-risk locations were 27, showing moderate accident possibility due to traffic density, road condition, and driver behavior. Very high-risk locations were 20, representing the most dangerous zones where immediate corrective actions such as speed control, signage, lighting, and road redesign are needed. Low-risk locations were 18, suggesting comparatively safer road sections. Overall, the graph highlights the need for focused traffic safety planning.

V. CONCLUSION

The study concluded that machine learning-based prediction of road accident black spots is an effective approach for improving traffic safety planning and accident prevention. The analysis showed that accident-prone locations were mainly influenced by factors such as high traffic volume, poor road geometry, inadequate lighting, weak signage, speeding, weather conditions, and unsafe intersections. Machine learning models helped identify hidden patterns in accident data and classified road locations into different risk levels. Among the applied techniques, advanced models such as Random Forest and Gradient Boosting were found useful for predicting high-risk and very high-risk accident zones. The predicted black spots can support traffic authorities in taking preventive measures such as installing warning signs, improving street lighting, redesigning intersections, controlling vehicle speed, and strengthening road safety enforcement. Overall, the study highlighted that data-driven prediction methods can reduce accident risk, improve road safety management, and support safer and more sustainable transportation planning.

REFERENCES

1. Subedi, R., & Chou, E. (2026). Analysis of Pedestrian-and Bicyclist-Involved Traffic Crashes Using Machine Learning. *Journal of Transportation Engineering, Part A: Systems*, *152*(2), 05025008.
2. Siddhartha, P., Pasasla, V. B., & Deepa, D. (2025, July). Road accident spot prediction with multiple deep learning. In *AIP Conference Proceedings* (Vol. 3257, No. 1, p. 020132). AIP Publishing LLC.

3. Zhang, Z., Nie, Q., Liu, J., Hainen, A., Islam, N., & Yang, C. (2024). Machine learning based real-time prediction of freeway crash risk using crowdsourced probe vehicle data. *Journal of Intelligent Transportation Systems*, *28*(1), 84-102.
4. Manoj, P., Punith, K. B., & Manjunath, K. C. (2024). Enhancing road safety through blackspots mitigative measures-a review. *Sigma Journal of Engineering and Natural Sciences*, *42*(5), 1670-1682.
5. Mahmoodzadeh, A., Ghafourian, H., Mohammed, A. H., Rezaei, N., Ibrahim, H. H., & Rashidi, S. (2023). Predicting tunnel water inflow using a machine learning-based solution to improve tunnel construction safety. *Transportation Geotechnics*, *40*, 100978.
6. Sun, C., Li, S., Cao, D., Wang, F. Y., & Khajepour, A. (2022). Tabular learning-based traffic event prediction for intelligent social transportation system. *IEEE Transactions on Computational Social Systems*, *10*(3), 1199-1210.
7. Yuksel, A. S., & Atmaca, S. (2021). Driver's black box: A system for driver risk assessment using machine learning and fuzzy logic. *Journal of Intelligent Transportation Systems*, *25*(5), 482-500.
8. Aradi, S. (2020). Survey of deep reinforcement learning for motion planning of autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *23*(2), 740-759.
9. Liu, Z., Suo, C., Liu, Y., Shen, Y., Qiao, Z., Wei, H., ... & Liu, Y. H. (2020). Deep learning-based localization and perception systems: approaches for autonomous cargo transportation vehicles in large-scale, semiclosed environments. IEEE Robotics & Automation Magazine, 27(2), 139-150.
10. Sarkar, S., Vinay, S., Raj, R., Maiti, J., & Mitra, P. (2019). Application of optimized machine learning techniques for prediction of occupational accidents. *Computers & Operations Research*, *106*, 210-224.
11. Tumen, V., Yildirim, O., & Ergen, B. (2018). Recognition of road type and quality for advanced driver assistance systems with deep learning. *Elektronika ir Elektrotechnika*, *24*(6), 67-74.