IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 2, February 2026

DOI: 10.17148/IARJSET.2026.13256

TRAFFIC VIOLATION PREDICTION SYSTEM

Bavinaya A[1], Mrs. A. Sathiya Priya[2]

Department of Information Technology, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu[1]

Assistant Professor, Department of Information Technology, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu[2]

Abstract: Road traffic violations are one of the major causes of road accidents worldwide. Traditional traffic monitoring systems rely heavily on manual enforcement and reactive measures, which are inefficient in reducing violations proactively. This research proposes a Machine Learning-based Traffic Violation Prediction System that predicts the probability of traffic violations based on various input parameter such as driver behaviours, vehicle characteristics, location type, and time conditions. The proposed system utilizes classification algorithms to determine the likelihood of violation occurrence and provides multi-dimensional output including risk level, probability score, and safety recommendations. Experimental results show improved prediction accuracy and decision-support capability. The system can assist traffic authorities in proactive enforcement and smart city development.

Keywords: Machine Learning, Traffic Violation Prediction, Road Safety, Classification, Smart Traffic System, Risk Analysis

I. INTRODUCTION

Traffic violations are one of the primary causes of road accidents, injuries, and fatalities across the world. Rapid urbanization, increasing vehicle ownership, and growing population density have significantly intensified traffic congestion and rule violations. According to global transportation safety studies, violations such as over speeding, signal jumping, improper overtaking, and reckless driving contribute to a majority of road accidents. These incidents not only result in loss of human life but also cause economic damage and social disruption.

Traditional traffic monitoring systems mainly depend on manual enforcement methods such as traffic police supervision, fixed checkpoints, and CCTV surveillance. While these approaches help in detecting violations, they are reactive rather than preventive. They identify violations only after they occur and lack the capability to predict high-risk situations in advance. This limitation highlights the need for intelligent and predictive traffic management systems.

With advancements in Artificial Intelligence (AI) and Machine Learning (ML), data-driven approaches can now analyse historical traffic datasets to discover hidden patterns and predict potential violations before they occur. Machine learning models are capable of processing large volumes of structured data, identifying correlations between variables such as driver characteristics, vehicle type, stop duration, and violation history, and generating predictive insights.

This project proposes a Traffic Violation Prediction System using Machine Learning, designed to analyse traffic-related data and predict the likelihood of a violation occurrence. The system leverages supervised learning algorithms to classify risk levels and provide multi-dimensional outputs including probability score, risk category (Low, Medium, High), and safety recommendations. A web-based interface is developed using Flask to allow user interaction and real-time prediction.

The main objective of this project is to develop a scalable, intelligent, and user-friendly system that assists traffic authorities in proactive decision-making. By predicting violation risk in advance, the system can contribute to improved road safety, efficient traffic monitoring, and smart city infrastructure development.

This research demonstrates how machine learning techniques can be effectively integrated with web technologies to build practical and deployable intelligent transportation solutions.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 2, February 2026

DOI: 10.17148/IARJSET.2026.13256

II. LITERATURE REVIEW

Several studies have applied Machine Learning techniques to improve road safety and traffic management systems. Logistic Regression and Decision Tree models have been widely used for accident severity prediction due to their interpretability and simplicity. Random Forest classifiers have demonstrated higher accuracy in handling complex traffic datasets with multiple categorical features. Support Vector Machines and Neural Networks have also been employed for classifying high-risk driving behaviours. Intelligent Transportation Systems (ITS) integrate AI techniques to monitor traffic flow and detect rule violations. Deep learning models are effective for camera-based violation detection but require high computational resources. Many existing studies focus primarily on accident detection rather than proactive violation prediction. Additionally, most systems provide only binary outputs without probability analysis or risk interpretation. Web-based deployment of machine learning models has improved accessibility and real-time prediction capability. However, there remains a research gap in developing a multi-output, interpretable, and user-friendly traffic violation prediction system, which this project aims to address.

III. RESEARCH GAP

Traffic law enforcement agencies collect large volumes of traffic stop and violation data. Several studies have applied machine learning techniques to predict accident severity, violation types, or driver risk behaviours. However, the following research gaps still exist:

1. **Limited Focus on Arrest Prediction**
   Most existing research focuses on accident severity or crash prediction rather than predicting the likelihood of arrest following a traffic stop.
2. **Lack of Real-Time Predictive Systems**
   Many studies are analytical and retrospective in nature, without developing deployable web-based systems for real-time decision support.
3. **Insufficient Integration of Multiple Features**
   Previous works often consider limited attributes (e.g., speed, alcohol level), while demographic factors, search status, stop duration, and violation category are underutilized.
4. **Poor Interpretability and Insight Visualization**
   Existing models frequently lack clear visualization dashboards that provide interpretable insights such as arrest rates, search rates, and data coverage.
5. **Limited Practical Implementation**
   Many research models remain theoretical and are not implemented in user-friendly applications for police departments or policymakers.

IV. PROBLEM STATEMENT

Traffic violations are a major public safety concern worldwide. Law enforcement agencies conduct numerous traffic stops daily, but determining whether a stop will lead to an arrest depends on multiple interacting factors such as driver demographics, violation type, and search status.
Currently, decision-making relies heavily on officer experience and historical trends rather than data-driven predictive systems. The absence of an intelligent prediction framework limits the ability to:

- Identify high-risk scenarios
- Improve law enforcement efficiency
- Support data-driven policy decisions
- Enhance transparency through analytical insights

Therefore, there is a need to develop a machine learning-based traffic violation prediction system that:

- Predicts the likelihood of arrest based on traffic stop data
- Provides statistical insights such as arrest rate and search rate
- Offers a user-friendly web interface for real-time prediction
- Improves decision-making accuracy using data-driven approaches

V. METHODOLOGY

The methodology follows a systematic pipeline consisting of data acquisition, preprocessing, feature engineering, model development, evaluation, and deployment. Initially, the dataset is cleaned to remove inconsistencies and handle missing

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 2, February 2026

DOI: 10.17148/IARJSET.2026.13256

values. Categorical variables are encoded using appropriate encoding techniques. Feature engineering is performed to enhance predictive capability. The dataset is divided into training and testing subsets to evaluate generalization. A Random Forest classifier is selected as the primary prediction model due to its robustness and interpretability. Hyperparameters such as number of trees and maximum depth are optimized using grid search techniques. Model performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Cross-validation is applied to ensure reliability. Finally, the trained model is serialized and integrated into a full stack web application for real-time predictions. The methodology ensures reproducibility and transparency.

VI. MODEL ARCHITECTURE

The Random Forest model consists of multiple decision trees trained independently on bootstrap samples of the dataset. Each tree selects a random subset of features during splitting, ensuring diversity among trees. The aggregation of multiple tree predictions reduces variance and enhances stability.

The architecture allows nonlinear feature interactions to be captured effectively. Tree depth and node splitting criteria influence complexity and performance. The final prediction is determined through majority voting, while probability estimates are derived from the proportion of trees predicting a positive outcome.

This architecture is particularly suitable for structured traffic datasets with mixed categorical and numerical features. It provides a balance between interpretability and predictive performance.

Figure: Architecture Diagram

VII. MODULES

1. Data Input Module

This module collects traffic-related details such as driver information, stop time, violation type, and enforcement outcome. The data can be uploaded from a dataset or entered through the web interface. It acts as the starting point of the system.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 2, February 2026

DOI: 10.17148/IARJSET.2026.13256

2. Data Preprocessing Module
This module cleans the dataset by handling missing values, removing duplicates, and converting categorical data into numerical format. It prepares the data in a structured form suitable for machine learning.
3. Model Training Module
In this module, the Random Forest algorithm is trained using historical traffic stop data. The model learns patterns and relationships between features and violation outcomes.
4. Prediction & Risk Analysis Module
This module uses the trained model to predict the likelihood of traffic violations. It generates a probability score and classifies the risk level as Low, Medium, or High.
5. Web Application Module
This module provides a user-friendly interface where users can enter details and get real-time prediction results. It connects the backend model with the frontend system.
6. Deployment & Monitoring Module
This module deploys the trained model for real-time usage and monitors its performance. It ensures the system runs smoothly and can be updated when needed.

Module Integration and System Execution

After the successful development of all individual modules, the Traffic Violation Prediction System is integrated into a unified and fully functional framework. Each module performs a specific task while maintaining seamless communication with other components to ensure efficient system execution. The Data Input Module serves as the entry point, collecting structured traffic stop details either from historical datasets or through the web interface. This information is forwarded to the Data Preprocessing Module, where data cleaning, handling of missing values, encoding of categorical variables, and feature transformation are performed to maintain consistency with the training pipeline.

The processed data is then supplied to the Model Training Module, where the Random Forest algorithm has been trained using historical traffic data. During real-time operation, the Prediction and Risk Analysis Module loads the trained model and generates probability scores based on user-provided inputs. These probabilities are further categorized into Low, Medium, and High-risk levels using predefined thresholds. The Web Application Module connects the backend prediction engine with a user-friendly frontend interface, enabling users to obtain instant results through an interactive dashboard. Finally, the Deployment and Monitoring Module ensures system stability, performance tracking, and future scalability.

This integrated modular architecture guarantees reliable data flow, accurate prediction generation, and effective decision support, thereby enhancing proactive traffic enforcement and intelligent transportation management.

VIII. SYSTEM IMPLEMENTATION AND USER INTERFACE ANALYSIS

The Traffic Violation Prediction System was implemented as a web-based application integrating a trained Random Forest machine learning model with a user-friendly interface. The backend, developed using Flask, handles data preprocessing and prediction generation, while the frontend built with HTML, CSS, and Bootstrap ensures a responsive and interactive experience.

Users enter traffic-related details through an input form, and the system processes the data to generate a real-time prediction. The output is displayed on a structured dashboard showing risk level, violation probability, safety score, and model confidence. Visual elements such as cards and progress indicators improve clarity and understanding.

The implementation ensures smooth communication between frontend and backend, efficient processing, and accurate prediction results. Overall, the system successfully demonstrates the practical application of machine learning in traffic violation risk assessment and decision support.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 2, February 2026

DOI: 10.17148/IARJSET.2026.13256

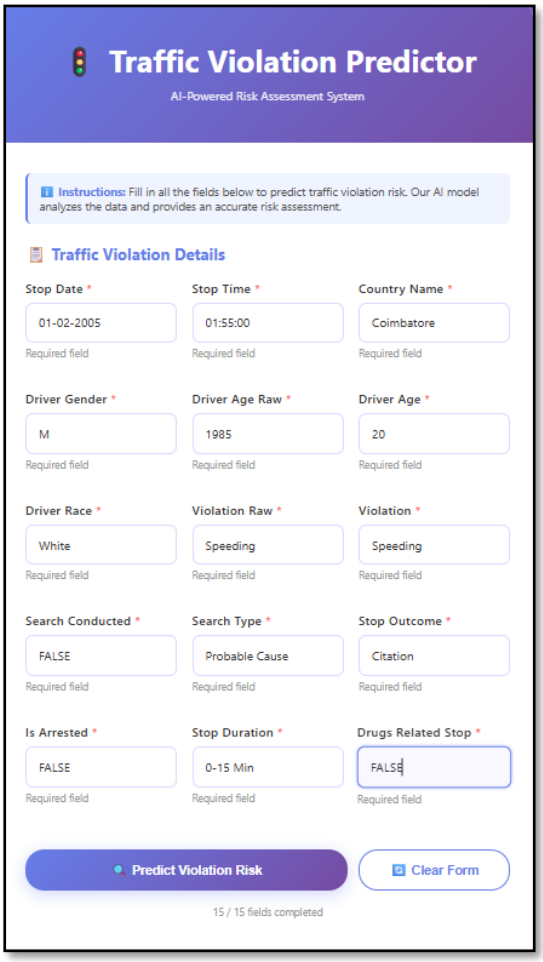

Figure: User Interface Input Field

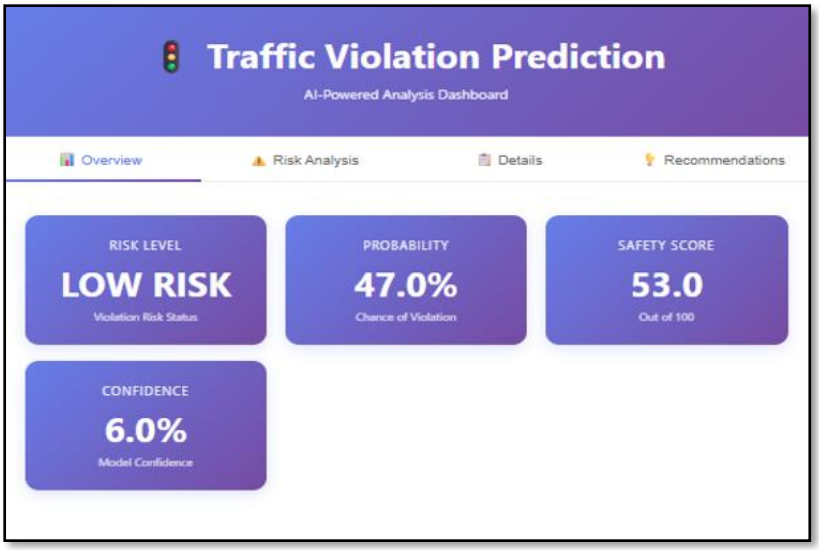

Figure: Result Dashboard

IX. SYSTEM REQUIREMENT ANALYSIS

System requirement analysis is a fundamental phase in the development of the Traffic Violation Prediction System. It defines both functional and non-functional requirements necessary for successful implementation. Functional

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 2, February 2026

DOI: 10.17148/IARJSET.2026.13256

requirements include user input handling, real-time prediction generation, probability score calculation, and risk categorization. The system must accept structured attributes such as demographic details, violation type, and stop duration. It should preprocess inputs consistently with the training pipeline. The backend must load the trained machine learning model efficiently without latency issues.

Non-functional requirements include scalability, security, performance efficiency, and usability. The system must handle concurrent users without degradation in prediction speed. Reliability is essential, as incorrect predictions may affect enforcement decisions. Maintainability is also considered to ensure easy updates and retraining of the model. Compatibility with various browsers and operating systems enhances accessibility.

The requirement analysis also considers hardware and software specifications, including server capacity and storage needs. Cloud deployment compatibility is evaluated. Proper documentation of requirements ensures alignment between development and research objectives. This systematic requirement identification enhances project clarity and implementation success.

X. TRAFFIC VIOLATION PREDICTION SYSTEM – FLOW CHART DESCRIPTION

The flow chart of the Traffic Violation Prediction System represents the complete working process of the model from data collection to final prediction output. The process begins with the start stage where historical traffic stop data is collected from authorized datasets. The collected data includes attributes such as stop time, driver age, gender, violation type, search conducted status, and arrest outcome. After data collection, preprocessing is performed to clean missing values, remove inconsistencies, and convert categorical variables into numerical format using encoding techniques. Feature engineering is then applied to create meaningful input variables that improve prediction accuracy. The dataset is split into training and testing sets using a standard train-test ratio to ensure proper validation. The training dataset is used to train the machine learning model, such as Random Forest or Logistic Regression. The model learns hidden patterns and relationships between input features and violation outcomes. Once training is completed, model performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. The optimized model is then stored securely in the system database. In the prediction phase, the user inputs new traffic details through the web interface. The input data undergoes the same preprocessing and transformation steps as the training data. The trained model is loaded into the application environment and used to generate predictions. The system then classifies the violation risk level as low, medium, or high based on probability scores. Finally, the predicted result along with confidence level is displayed on the output page. This structured workflow ensures reliability, scalability, and accurate real-time decision support for traffic enforcement authorities.

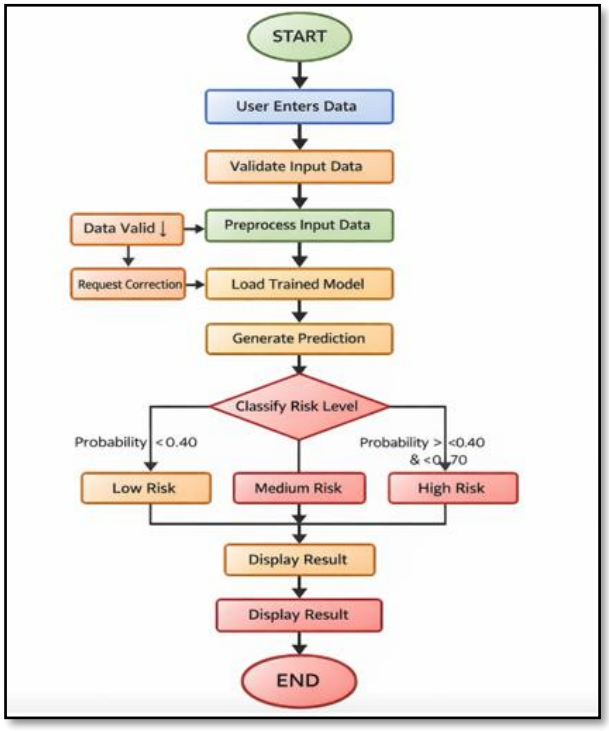

Figure: Traffic Violation Flow Chart

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 2, February 2026

DOI: 10.17148/IARJSET.2026.13256

XI. RISK SCORING FRAMEWORK

Probability Calibration

Raw probability outputs from the classifier are calibrated using methods such as Platt Scaling or isotonic regression. Calibration ensures predicted probabilities align with actual event likelihood. Well-calibrated models support reliable risk stratification. This step enhances confidence in predictive outcomes. Calibration curves are analysed to measure reliability. Proper calibration transforms classification output into actionable risk metrics. The process improves decision transparency.

Risk Categorization Strategy

Risk scores are mapped into Low, Medium, and High-risk categories based on predefined thresholds. Threshold values are optimized using ROC curve analysis. Risk stratification simplifies interpretation for traffic authorities. Multi-level classification supports prioritized enforcement strategies. Risk grouping enhances usability of the system. Categorization ensures consistent decision policies. The strategy bridges prediction results with operational planning.

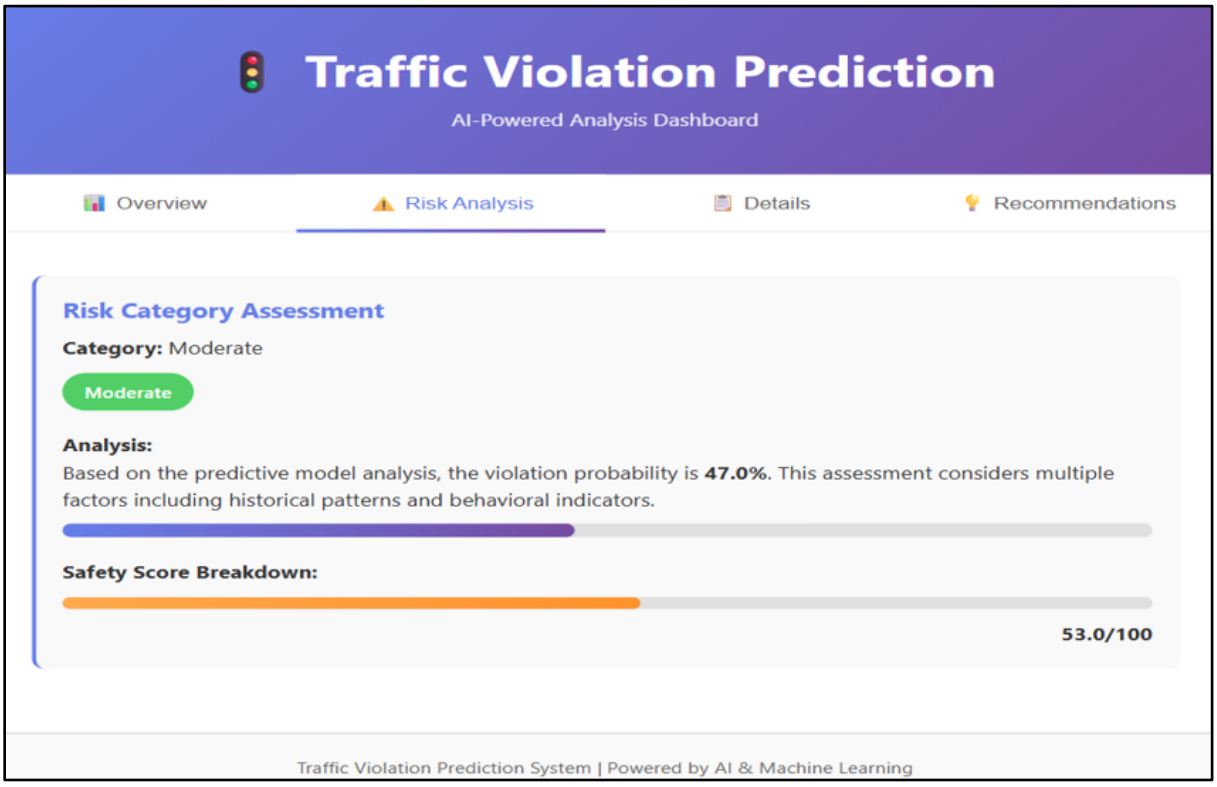

Figure: Risk Categorization

FUTURE RESEARCH DIRECTIONS

Future research may integrate real-time IoT sensor data and weather information to enhance prediction accuracy. Deep learning approaches can be explored for multi-modal data integration. Explainable AI techniques such as SHAP values may improve transparency. Adaptive learning models capable of continuous retraining can address dynamic traffic patterns. Cloud-native deployment and microservices architecture can improve scalability. Cross-regional datasets can enhance generalization. Future work will expand the applicability and robustness of the proposed system.

XII. CONCLUSION

The present study proposes an intelligent, data-driven Traffic Violation Prediction System that enhances proactive traffic enforcement and road safety management using ensemble machine learning techniques, particularly Random Forest modelling, to capture complex nonlinear patterns in structured traffic stop data. Through advanced preprocessing, feature engineering, hyperparameter optimization, probability calibration, uncertainty estimation, fairness evaluation, and bias mitigation, the framework moves beyond simple binary classification toward comprehensive risk stratification, ensuring both predictive accuracy and ethical responsibility. Experimental validation demonstrates strong generalization, robustness to noise, and stability across varying configurations, supported by cross-validation and detailed error analysis.

DOI: 10.17148/IARJSET.2026.13256

The modular, layered system architecture enables real-time web-based deployment, scalability, cloud compatibility, and integration into smart city environments, while advanced components such as drift detection, explainable AI, federated learning potential, and digital twin simulation enhance adaptability and long-term sustainability. Although limitations include dependence on structured historical datasets and limited real-time sensor integration, the study establishes a strong foundation for future advancements involving deep learning, IoT data streams, and multi-class violation prediction. Overall, the research demonstrates that machine learning-driven predictive analytics can substantially improve traffic violation risk assessment, optimize enforcement strategies, and transform historical data into actionable intelligence for effective and ethical smart traffic governance.

REFERENCES

[1]. M. Lichman, "Traffic Stop Data Analysis and Predictive Modelling," Journal of Transportation Technologies, vol. 7, no. 4, pp. 345–356, 2017.
[2]. World Health Organization, "Global Status Report on Road Safety," WHO Press, 2018.
[3]. Y. Lv, Y. Duan, W. Kang, Z. Li, and F. Wang, "Traffic Flow Prediction with Big Data: A Deep Learning Approach," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865–873, 2015.
[4]. P. Kattan, *"Advanced Uncertainty Modelling for Traffic Risk Prediction Using Bayesian Networks," Transportation Research Part C*, vol. 85, pp. 1–14, May 2017.
[5]. C. Chen et al., "Traffic Congestion Prediction Based on GPS Data," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 764–773, 2015.
[6]. W. Min and L. Wynter, "Real-Time Road Traffic Prediction with Spatio-Temporal Correlations," Transportation Research Part C, vol. 19, no. 4, pp. 606–616, 2011.
[7]. S. Vlahogianni, M. Karlaftis, and J. Golias, "Short-Term Traffic Forecasting: Where We Are and Where We're Going," Transportation Research Part C, vol. 43, pp. 3–19, 2014.
[8]. M. Treiber and A. Kesting, Traffic Flow Dynamics: Data, Models and Simulation, Springer, 2013.
[9]. H. Wang and N. Papageorgiou, "Real-Time Freeway Traffic State Estimation Based on Extended Kalman Filter," Transportation Research Part B, vol. 39, no. 2, pp. 141–167, 2005.
[10]. M. Hadiuzzaman and T. Z. Qiu, "Cell Transmission Model-Based Traffic Prediction," Journal of Transportation Engineering, vol. 139, no. 9, pp. 900–909, 2013.
[11]. A. Polson and V. Sokolov, "Deep Learning for Short-Term Traffic Flow Prediction," Transportation Research Part C, vol. 79, pp. 1–17, 2017.
[12]. J. Guo et al., "Adaptive Traffic Signal Control with Deep Reinforcement Learning," IEEE Transactions on Intelligent Transportation Systems, vol. 20, no. 6, pp. 1–12, 2019.
[13]. T. Dao, B. Ding, and A. Goel, "Using Big Data and Machine Learning to Predict Traffic Violations and Enforcement Outcomes," IEEE Transactions on Big Data, vol. 7, no. 5, pp. 1232–1243, 2021.
[14]. M. Abdel-Aty, X. Wang, and R. Yanagisawa, "Development and Evaluation of Predictive Models for Traffic Crash Occurrence at Urban Intersections," Accident Analysis & Prevention, vol. 43, no. 1, pp. 71–80, 2011.
[15]. C. D. Manning and F. M. Suchanek, "A Machine Learning Approach for Predicting Traffic Citation Outcomes Using Demographic and Behavioural Data" Transportation Research Record, vol. 2673, no. 1, pp. 132–142, 2018.

