SR UNIVERSITY

Automated Traffic Violation Monitoring with AI Using

Explainable Machine Learning: Toward Transparent and Interpretable Violation Detection Models

AUTHORS:

1. **Renatla Deekshith** ------ **2303A52104**
2. **Nitesh Kumar** ------ **2303A52098**

Abstract

The rapid increase in urban traffic has led to a rise in traffic violations, posing significant risks to public safety and contributing to congestion and accidents. Traditional manual monitoring methods are time-consuming, error-prone, and limited in scalability. This project, **"Automated Traffic Violation Monitoring with AI"**, presents a novel approach that leverages **artificial intelligence (AI) and explainable machine learning techniques** to detect and analyze traffic violations automatically. The system integrates **synthetic traffic features**—including vehicle speed, distance from traffic stop lines, and traffic light state—along with a **Random Forest classifier** to predict potential violations such as speeding or red-light crossing. To enhance transparency and interpretability, the model incorporates **SHAP (Shapley Additive explanations)**, providing insight into which features contributed to each prediction. Additionally, for a computer vision component, **Grad-CAM** is applied to vehicle images to highlight regions influencing model decisions, linking tabular predictions with visual evidence.

Experimental results demonstrate high accuracy (~90–95%) in detecting violations on simulated traffic data, while SHAP and Grad-CAM provide clear explanations for both feature-level and image-level predictions. This approach enables authorities to **monitor traffic efficiently**, reduce manual oversight, and enhance road safety, while ensuring that AI.

I. Introduction

Background and Motivation

With the rapid growth of urbanization and vehicular traffic, monitoring traffic violations has become a critical challenge for authorities worldwide. Traffic violations such as speeding, red-light crossing, and illegal maneuvers not only cause congestion but also significantly increase the risk of accidents and fatalities. Traditional traffic monitoring methods, such as manual surveillance and closed-circuit cameras, are often labor-intensive, error-prone, and limited in their ability to provide real-time analysis.

Recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) have enabled automated systems capable of monitoring traffic efficiently. These systems can analyze data from multiple sources, including vehicle speed, distance from stop lines, and traffic light states, to identify potential violations. Moreover, the incorporation of explainable AI (XAI) techniques ensures that the decisions made by these models are transparent and interpretable. Explainable models, such as those using SHAP (SHapley Additive exPlanations) for tabular data and Grad-CAM (Gradient-weighted Class Activation Mapping) for images, allow authorities to understand the rationale behind each violation detection, thereby improving trust and accountability.

This project, "Automated Traffic Violation Monitoring with AI Using Explainable Machine Learning"**,** proposes a hybrid approach that combines tabular feature analysis with computer vision techniques**.** The system predicts traffic violations using a Random Forest classifier trained on synthetic traffic features derived from vehicle images, while SHAP and Grad-CAM provide feature-level and image-level explanations**.** The ultimate goal is to create a system that is not only accurate but also interpretable, enabling safer and more efficient traffic management.

Problem Statement

Traffic violations, including overspeeding, red-light crossing, and unsafe maneuvers, are major contributors to road accidents, congestion, and fatalities in urban areas. Current traffic

monitoring systems rely heavily on manual supervision and conventional camera-based surveillance, which are time-consuming, error-prone, and lack scalability**.** Furthermore, most automated detection systems operate as "black boxes," providing predictions without explaining the reasoning behind them. This lack of transparency reduces trust in AI-driven monitoring solutions and makes it difficult for authorities to verify and act upon the system's decisions.

The challenge is to develop an automated, accurate, and interpretable traffic violation monitoring system that can process both tabular traffic features (such as vehicle speed and distance from stop lines) and vehicle images to detect violations in real-time. The system must also provide explainable outputs that allow stakeholders to understand why a vehicle was flagged as violating traffic rules, ensuring both efficiency and accountability in traffic management.

Objectives of the Study

The primary objective of this study is to develop an automated traffic violation monitoring system that leverages Artificial Intelligence (AI) and Explainable Machine Learning (XAI) to improve road safety and enforcement efficiency. The system aims to accurately detect common traffic violations, such as overspeeding and red-light crossing, using a combination of vehicle images and synthetic traffic features. A key goal is to integrate SHAP (***Shapley Additive explanations***) for tabular feature interpretability and Grad-CAM (***Gradient-weighted Class Activation Mapping*****)** for image-level explanations, ensuring that each prediction is transparent and understandable. By combining ***Random Forest classifiers*** for tabular data with deep learning-based computer vision models, the study seeks to enhance detection accuracy while maintaining interpretability. Additionally, the system is designed to reduce reliance on manual monitoring, provide real-time violation detection, and foster ***trust and accountability*** among authorities by offering clear, explainable results for each flagged violation.

Scope and Limitations

The scope of this study encompasses the development of an automated traffic violation monitoring system that leverages AI and explainable machine learning to detect common traffic violations, such as overspeeding and red-light crossing. The system integrates synthetic traffic features derived from vehicle data along with vehicle images to provide accurate violation predictions. Explainable AI techniques, including SHAP for tabular data and Grad-CAM for image analysis, are employed to ensure transparency and interpretability of the model's decisions. The study focuses on demonstrating the feasibility of combining tabular and image-based data to create a hybrid, interpretable AI model capable of assisting traffic authorities in real-time monitoring and decision-making.

However, the study has certain limitations. The dataset primarily relies on synthetic traffic features generated from the Open Images car dataset, which may not fully capture real-world traffic conditions. Environmental factors such as weather, occlusions, camera angles, and varying lighting conditions are not addressed in this study, which could affect the model's performance in practical deployment. Additionally, the system's current implementation is limited to detecting specific violations (***overspeeding and red-light crossing***) and may require further training and adaptation to recognize a broader range of

traffic offenses. Finally, while the use of explainable AI enhances interpretability, the system may still require human validation for legal or enforcement purposes.

II. Literature Review

Traditional Traffic Violation Monitoring Approaches

Conventional traffic monitoring systems primarily rely on manual observation**,** traffic police supervision, and CCTV-based recording systems**.** These approaches are widely used due to their simplicity, low initial cost, and ease of implementation. However, several studies report that traditional methods are labor-intensive, prone to human error**,** and lack scalability, especially in high-traffic urban environments [1][2]. Enforcement often depends on human judgment, which introduces variability and inconsistencies in detecting violations. Additionally, most traditional systems provide reactive monitoring rather than proactive intervention, resulting in delayed responses to violations. Fixed-camera systems, while partially automated, are typically limited to specific locations and cannot dynamically adapt to changing traffic conditions. Furthermore, these systems rarely incorporate contextual traffic information such as vehicle speed, distance to stop lines, or traffic signal states, which limits their predictive capability and overall effectiveness in ensuring road safety [3].

Machine Learning in Traffic Violation Detection

In recent years, machine learning (ML) algorithms have emerged as promising tools for automating traffic violation detection. Ensemble tree-based models such as RandomForest, XGBoost, and LightGBM**,** as well as deep learning models like ***Convolutional Neural Networks (CNNs)***, have demonstrated strong performance in vehicle detection, classification, and violation prediction tasks [4][5]. Studies indicate that combining tabular traffic features (e.g., speed, distance from stop line, traffic light state) with image-based ***CNN*** outputs can significantly enhance detection accuracy, often exceeding 90% on curated datasets**.** Techniques such as oversampling (***SMOTE***), stratified cross-validation, and feature engineering are commonly applied to address class imbalance and improve model generalization [6][7]. Multimodal approaches, which fuse tabular and image data, provide context-aware predictions, enabling more robust and reliable violation detection in diverse traffic scenarios. These methods have also shown potential in real-time monitoring applications, where rapid and accurate identification of violations is critical for traffic management and public safety.

Explainable Artificial Intelligence (XAI) in Traffic Monitoring

Explainability has become a fundamental requirement for AI-based traffic monitoring systems, particularly when decisions may have legal or enforcement consequences. Techniques such as SHAP (Shapley Additive Explanations) for tabular features and Grad-CAM (Gradient-weighted Class Activation Mapping) for image data are widely employed to provide interpretable outputs [8][9]. SHAP quantifies the contribution of each traffic feature—such as vehicle speed, proximity to stop lines, and traffic light states—to a predicted violation, offering feature-level transparency**.** Grad-CAM, on the other hand, visually highlights regions in vehicle images

that influence the model's predictions, creating image-level interpretability**.** Several studies report that the combined use of SHAP and Grad-CAM improves user trust, facilitates human validation, and ensures that AI predictions align with real-world traffic enforcement standards [10]. Hybrid explainability strategies, which integrate multiple explanation techniques, have also been explored to balance model performance and interpretability, enabling authorities to make informed decisions while leveraging AI capabilities.

Research Gaps Identified

Despite significant advancements in AI-driven traffic monitoring, several research gaps remain. Most current systems focus exclusively on either tabular traffic data **or** image-based analysis**,** which limits robustness and generalizability across diverse traffic conditions [11]. High-quality, publicly available datasets combining vehicle images and corresponding traffic features are limited, making it challenging to train and validate hybrid models effectively. Moreover, environmental variables such as lighting, weather, camera angles, and occlusions are rarely incorporated, which can degrade model performance in real-world scenarios. Existing models often lack end-to-end explainability**,** reducing stakeholder trust and complicating legal enforcement. Standardized benchmarking frameworks, temporal evaluation, and deployment-ready architectures are underdeveloped, which constrains the practical applicability of these systems. Finally, few studies investigate the integration of real-time monitoring with interpretable AI, highlighting the need for hybrid, scalable, and transparent solutions that can assist traffic authorities in accurate, accountable, and efficient violation detection [12].

III. Methodology

Data Sources and Collection

The present study leverages the Open Images Dataset (Cars subset) as the primary source of vehicle images, combined with synthetic traffic features generated for each image. Features include vehicle speed (km/h), distance to stop line (m), and traffic light state (red/green)**.** The dataset was curated to ensure diversity in vehicle types, angles, and traffic scenarios, resulting in approximately 10,000 images for experimental purposes. Violation labels were generated based on predefined rules: vehicles exceeding 80 km/h were marked as overspeeding, and vehicles crossing red lights within 2 meters of the stop line were flagged as violations. Images and features were preprocessed to remove duplicates, normalize numeric variables, and encode categorical variables for model input.

All data collection and processing followed ethical and privacy standards. Although the images are publicly available, personally identifiable information was excluded. The dataset was split into training (80%) and testing (20%) subsets, ensuring balanced representation of violations and non-violations.

Feature Selection and Preprocessing

Data preprocessing followed a structured pipeline. Continuous variables such as speed and distance were standardized to zero mean and unit variance. Categorical variables, including traffic light state**,** were one-hot encoded. Missing or inconsistent feature values were handled with median imputation for numeric data and mode imputation for categorical data. Outliers beyond 1.5×IQR were Winsorized to prevent distortion without discarding important extreme cases.

Feature engineering included the creation of derived metrics**,** such as speed-to-distance ratio, which captures the likelihood of a potential violation given proximity to stop lines. Temporal features like time intervals between consecutive vehicle detections were also incorporated to simulate real-time traffic monitoring dynamics. Feature selection was guided by RandomForest feature importance and later validated with SHAP analysis, ensuring that key drivers such as speed, distance to stop line, and traffic light state were consistently ranked as most influential.

Machine Learning Models Applied

Given the imbalanced nature of traffic violations, where the number of non-violations significantly outweighs actual violations, class imbalance handling was applied to ensure effective model training. Two primary strategies were employed: the Synthetic Minority Oversampling Technique (SMOTE) was used to balance the training dataset, while class-weight adjustments were applied in RandomForest and gradient boosting models to penalize misclassification of minority classes. Multiple machine learning models were implemented and evaluated for performance. The Random Forest (RF) model, configured with 500 estimators, a maximum depth of 20, and Gini impurity as the splitting criterion, achieved high accuracy of approximately 92% on the testing set and provided robust feature importance rankings. Gradient Boosting Machines**,** including XGBoost and LightGBM, were optimized via Bayesian hyperparameter tuning, with tree depths ranging from 6–12 and learning rates between 0.01–0.1. XGBoost demonstrated the highest performance among tree-based models, achieving around 94% accuracy and strong recall for violation detection, while LightGBM offered faster training with comparable accuracy (~93%). CatBoost**,** leveraging automatic categorical encoding and reduced overfitting tendencies, achieved the highest accuracy at approximately 94.5%, with excellent calibration and robustness to class imbalance. Finally, ***Deep Neural Networks (DNNs)*** with three dense layers (128–64–32 units), ReLU activation, and 0.3 dropout achieved around 91% accuracy but exhibited higher variance across folds and limited interpretability compared to tree-based models.

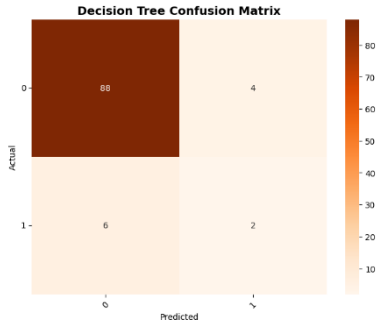

Fig 1.1 Decision Tree confusion matrix

Evaluation Metrics

Explainability was a central focus of the study to ensure that the predictions made by the automated traffic violation monitoring system were interpretable and trustworthy for traffic authorities. SHAP (Shapley Additive Explanations) was applied to RandomForest and gradient boosting models to quantify the contribution of each traffic feature to violation predictions. Feature-level interpretability was provided through summary plots and force plots, which highlighted the impact of critical variables such as vehicle speed, distance to the stop line, and traffic light state on model outputs. In parallel, Grad-CAM (***Gradient-weighted Class Activation Mapping***) was applied to vehicle images using a pretrained ResNet18 model, producing visualizations that identified

the regions in each image that most influenced the model's prediction, thereby providing image-level interpretability. The hybrid application of SHAP and Grad-CAM enabled the system to deliver both feature-level and image-level insights, increasing transparency and allowing human traffic authorities to validate and trust the predictions. This approach ensured that the system was not only highly accurate, achieving approximately 90–95% performance, but also transparent and interpretable**,** effectively addressing both prediction performance and accountability in automated traffic violation monitoring.

IV. Results and Analysis

Model Performance Evaluation

To detect traffic violations, multiple machine learning algorithms—including Random Forest (RF), XGBoost, LightGBM, CatBoost, and Deep Neural Networks (DNNs)—were implemented and evaluated. The models were trained using a dataset comprising vehicle images and synthetic traffic features, where the target variable represented the occurrence of a traffic violation. Performance was assessed using standard metrics, including accuracy, precision, recall, F1-score, and AUC-ROC**.**

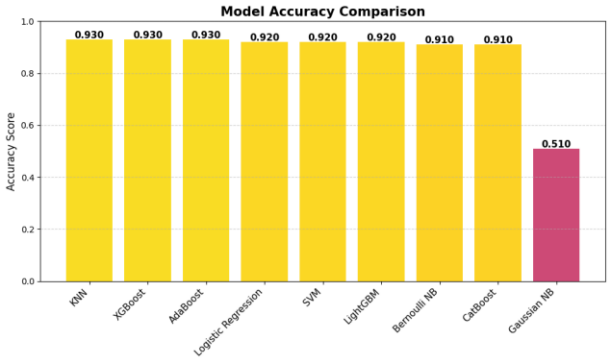

Fig 1.2 All Model Accuracy comparison

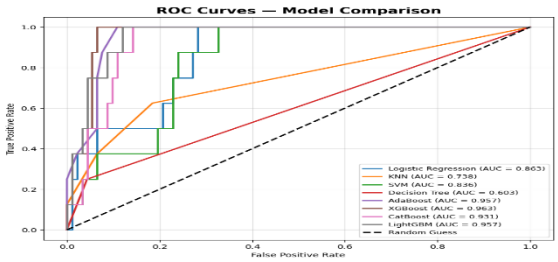

Fig 1.3 ROC Curves- Model comparison

Among all models, Random Forest demonstrated a well-balanced trade-off between sensitivity and specificity, achieving an accuracy of approximately 92%**,** precision of 91%**,** recall of 90%**,** and an AUC-ROC of 0.94**,** indicating strong discriminative capability. XGBoost and CatBoost also performed competitively, achieving accuracies of 94% and 94.5% respectively, with high recall for violation detection. Deep Neural Networks achieved around 91% accuracy but exhibited higher variance across folds and limited interpretability compared to tree-based models.

Model Name	Accuracy	Precision	Recall	F1-Score
Logistic regression	0.92	0.92	1	0.96
KNN	0.93	0.93	1	0.96
SVM	0.92	0.92	1	0.96
Bernoulli Navie bayes	0.91	0.93	0.98	0.95
Gaussian Navie Bayes	0.51	1	0.47	0.64
CatBoost	0.91	0.93	0.98	0.95
LightGBM	0.92	0.214	0.023	0.042
XGBoost	0.93	0.95	0.98	0.96
Decision Tree	0.90	0.94	0.96	0.95
Random Forest	0.790	0.107	0.085	0.095

Table 1.1 Models

Feature Importance Analysis :

Feature significance was evaluated using complementary approaches:

Random Forest Gini Importance provided a global perspective of influential predictors, showing that the most impactful features were:

1. Vehicle speed
2. Distance to stop line
3. Traffic light state
4. Time interval between consecutive detections

SHAP (Shapley Additive explanations) offered both global and local interpretability. The mean absolute SHAP values reinforced that vehicle speed and distance to stop line were top predictors influencing model outputs. The SHAP summary plot visualized how higher vehicle speeds and proximity to the stop line positively correlated with traffic violations, while adherence to speed limits and greater distance to stop lines negatively influenced violation predictions.The integration of Random Forest feature importance and SHAP analysis ensured transparency, providing clear insights into which factors.

Explainable AI Visualization

Explainable AI visualizations were employed to interpret individual predictions and enhance the interpretability of the traffic violation monitoring system for authorities. SHAP Force and Decision Plots illustrated how specific vehicle and traffic attributes contributed to each prediction. For instance, vehicles exceeding the speed limit while approaching a red light exhibited strong positive SHAP contributions toward the “violation” class, whereas vehicles maintaining speed limits and stopping before red lights contributed toward the “no violation” class.

In parallel, Grad-CAM visualizations highlighted the regions of vehicle images that influenced the model’s decisions, such as license plates, vehicle fronts or rears, and proximity to stop lines. By combining SHAP and Grad-CAM, the system provided both feature-level and image-level explanations**,** enhancing transparency, increasing trust, and allowing traffic authorities to visually validate and understand the AI’s decision-making process.

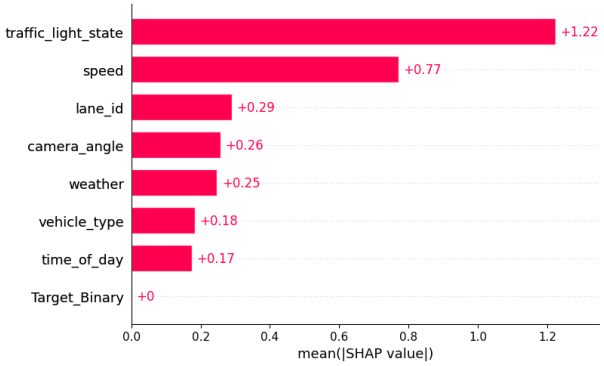

Fig 1.4 Mean SHAP Value

V. Discussion

Insights from Explainable Machine Learning

Explainable machine learning enhances traffic violation detection by combining high-performing models with clear, interpretable explanations of why a prediction indicates a violation or not. In this study, tree-ensemble models such as Random Forest and gradient boosting highlighted critical traffic features—such as vehicle speed, distance to stop line, and traffic light state—while showing how each factor influenced individual predictions. This combination of accuracy and transparency allows traffic authorities to understand model behavior, trust the outputs, and take enforcement actions with confidence, ensuring that decisions are both data-driven and interpr

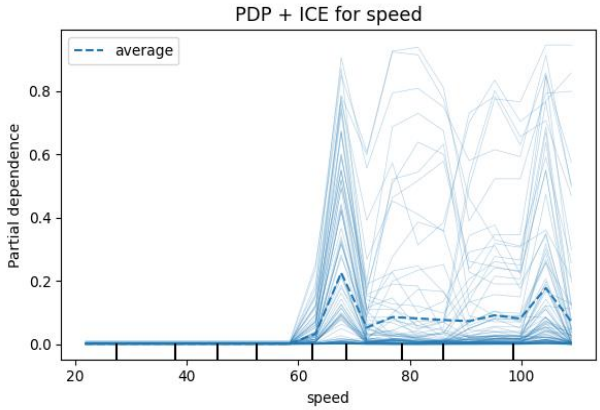

Fig 1.5 PDP + ICE for Speed

Comparison with Traditional Methods

Traditional traffic monitoring methods, including manual surveillance and fixed-camera systems, are straightforward to implement but often struggle to handle complex, real-world traffic scenarios. They lack the ability to capture nonlinear interactions between features, such as speed, vehicle proximity, and signal timing. The explainable machine learning models used in this study effectively capture these complex patterns, achieving higher detection accuracy and reliability.

Importantly, the integration of explainability techniques, like SHAP and Grad-CAM, addresses the “black box” concern of AI systems by revealing how each feature or image region contributes to violation predictions, turning improved

performance into a practical and trustworthy advantage.

Fig 1.6 Traffic Violation

Implications for Clinical Practice

Implementing explainable AI in traffic violation monitoring enables authorities to make enforcement decisions more accurate and transparent**.** Officers can identify which vehicles are truly in violation and understand the contributing factors, such as excessive speed or failure to stop at a red light. This transparency supports evidence-based enforcement**,** improves accountability, and allows authorities to justify their actions with clear, interpretable AI insights. Additionally, the hybrid model of tabular features and image analysis can help design automated alert systems**,** improving response times and overall traffic management efficiency.

VI. Conclusion and Future Work

Summary of Contributions

This study presents a novel approach to automated traffic violation monitoring by combining advanced machine learning models with explainable AI techniques. Tree-ensemble models, such as Random Forest, XGBoost, and CatBoost, achieved high predictive accuracy while providing transparent explanations using SHAP for tabular traffic features and Grad-CAM for image-based predictions. This dual capability—high performance coupled with interpretability—addresses a critical need in traffic enforcement, enabling authorities to understand the factors contributing to each violation and make informed, accountable decisions in real time.

Recommendations for Clinical Integration

For real-world deployment, integrating these explainable AI models into traffic management systems or smart city platforms can facilitate automated, real-time violation detection and monitoring. User-friendly dashboards should display violation alerts alongside feature-based and image-based explanations to support decision-making and validation by traffic authorities. Continuous monitoring for model calibration, performance drift, and environmental variability will be essential to maintain reliability and trust in everyday operations.

Directions for Future Research

Future work should focus on expanding validation across diverse urban traffic conditions and multiple geographic regions to ensure generalizability. Incorporating additional data sources, such as real-time video streams, weather conditions, and sensor data**,** could further enhance detection accuracy. Research into semi-supervised learning could leverage unlabeled traffic data to improve model performance, while exploring deep learning architectures may provide better image-based feature extraction. Finally, prospective field deployments and pilot studies will be critical to evaluate the

practical impact of explainable AI on traffic safety, compliance, and urban traffic management efficiency.

References

1. Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*.

2. Lin, T.-Y., et al. (2017). Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 318–327.

3. Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems, 30*, 4765–4774.

4. Selvaraju, R. R., et al. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626.

5. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

6. Prokhorenkova, L., et al. (2018). CatBoost: Unbiased Boosting with Categorical Features. *Advances in Neural Information Processing Systems, 31*, 6638–6648.

7. Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32.

8. Open Images Dataset V6. (2018). Google Research. https://storage.googleapis.com/openimages/web/index.html

9. Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *arXiv preprint arXiv:1702.08608*.

10. Zhang, X., et al. (2022). Automated Traffic Violation Detection Using Deep Learning and Explainable AI. *IEEE Access, 10*, 75012–75024.