XIV Международна научна конференция
„Проектиране и строителство на сгради и съоръжения“, 10-12 септември 2026 г, Варна
XIV International Scientific Conference
„Civil Engineering Design and Construction“, Sept. 10-12, 2026, Varna, Bulgaria
ISSN 2683-071X (online) ISSN 2738-7879 (CD)

РАМКА ЗА НАДЕЖДЕН, ПРОЗРАЧЕН И ОБЯСНИМ ИЗКУСТВЕН ИНТЕЛЕКТ ЗА КИБЕРУСТОЙЧИВИ ИНТЕЛИГЕНТНИ ТРАНСПОРТНИ ИНФРАСТРУКТУРНИ СИСТЕМИ

Мохамад Фархад Аязи[1], Маниндер Сингх[2], Ражив Кумар[3]

A TRUST AWARE EXPLAINABLE AI FRAMEWORK FOR CYBER RESILIENT SMART TRANSPORTATION INFRASTRUCTURE SYSTEMS

Mohammad Farhad Ayazi[1], Maninder Singh[2], Rajiv Kumar[3]

Abstract:
Artificial Intelligence (AI) is transforming transportation infrastructure through data-driven decision-making for traffic management and pavement performance evaluation. However, integrating machine learning (ML) with Internet of Things (IoT)-enabled systems introduces challenges related to trust, robustness, and cybersecurity in smart transportation environments. This study proposes a trust-aware explainable AI framework for cyber-resilient transportation infrastructure systems. The framework employs Random Forest (RF) models for traffic flow estimation and resilient modulus prediction, while SHapley Additive exPlanations (SHAP) are used to enhance model interpretability and transparency. To ensure data reliability, anomaly detection and secure validation mechanisms are incorporated for IoT-based data streams. A resilience-driven learning approach, including adaptive retraining and fault-tolerant model design, is introduced to maintain performance under noisy and adversarial conditions. The framework is evaluated using representative datasets, and results demonstrate improved prediction accuracy, robustness, and interpretability compared to conventional approaches. The proposed method provides a scalable solution for developing secure and trustworthy AI-driven transportation infrastructure, supporting smart and sustainable construction practices.

Keywords:
Explainable AI (XAI), Smart Transportation Systems, Cyber-Resilience, Trust-Aware Machine Learning

1. INTRODUCTION

The increasing growth of urbanization, industrialization, and vehicular demand has intensified the complexity of transportation systems worldwide. Modern transportation networks face severe operational challenges, including traffic congestion, travel delays, energy inefficiency, environmental pollution, and road safety concerns. Intelligent Transportation

[1] Mohammad Farhad Ayazi, PhD Research scholar, Civil Engineering Department, Punjabi University, Patiala, India 147002
E-mail: ayazi_rs23@pbi.ac.in, farhad.farhamand1395@gmail.com
[2] Maninder Singh, Assistant professor, Civil Engineering Department, Punjabi University, Patiala, India 147002
E-mail: maniravi79@gmail.com
[3] Rajiv Kumar, Scientist, Flexible Pavement Division, Central Road Research Institute of India, New Delhi, India 110025
E-mail: Ranjitrajiv@gmail.com

Systems (ITS) have emerged as a critical technological solution to address these challenges by integrating Artificial Intelligence (AI), machine learning (ML), Internet of Things (IoT), cloud computing, and real-time communication technologies into transportation infrastructure [1–3]. Recent advancements in AI-driven transportation systems have enabled intelligent traffic prediction, adaptive signal control, autonomous traffic management, route optimization, and accident prevention [4,5].

Traffic flow prediction is considered one of the most important components of ITS because accurate forecasting enables proactive congestion mitigation, intelligent routing, dynamic traffic control, and enhanced traveler safety [6]. Traditional traffic prediction approaches, including statistical regression models, autoregressive integrated moving average (ARIMA), Kalman filtering, and time-series analysis, have been widely employed for traffic forecasting [7]. However, these methods struggle to model nonlinear and dynamic traffic relationships under real-world conditions [8]. Consequently, advanced machine learning and deep learning approaches such as Random Forest (RF), Support Vector Machines (SVM), Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), Graph Neural Networks (GNN), and Transformer-based architectures have gained considerable attention [9–12]. Sayed et al. [13] presented a comprehensive review of AI-based traffic flow prediction methods and concluded that deep learning techniques outperform traditional approaches in capturing spatiotemporal traffic patterns. Similarly, Berlotti et al. [14] proposed a machine learning framework for urban traffic flow prediction and demonstrated that ensemble learning models significantly improve prediction accuracy under varying traffic conditions. Guo et al. [15] introduced a large language model (LLM)-based explainable traffic prediction framework known as xTP-LLM, which combines explainability with high predictive performance. Their findings indicated that explainable models improve user trust and operational understanding. Although advanced AI models achieve superior predictive performance, many operate as black-box systems whose internal decision-making processes remain difficult to interpret [16]. In transportation systems, where AI decisions directly affect public safety and infrastructure reliability, lack of interpretability creates significant challenges. Transportation engineers and traffic managers require transparent and understandable AI systems to ensure accountability and operational trust [17]. Consequently, Explainable Artificial Intelligence (XAI) has emerged as a major research domain focused on improving transparency, interpretability, and trustworthiness of AI models [18].

Among explainability approaches, SHapley Additive exPlanations (SHAP) has become one of the most widely adopted frameworks due to its strong theoretical foundation based on cooperative game theory [19]. SHAP quantifies the contribution of each feature toward model predictions and provides both local and global interpretability. Lundberg and Lee [19] established SHAP as a unified explanation framework capable of generating consistent and interpretable explanations for complex machine learning systems. Several recent studies have explored the integration of XAI into transportation systems. Thomas et al. [20] conducted a detailed review of explainable AI methods for traffic prediction and decision-making in ITS. The authors categorized explainability approaches into intrinsically interpretable models, local explanation methods, global explanation methods, and causal reasoning frameworks. Their study emphasized that explainability is essential for responsible deployment of AI in public transportation infrastructure. Similarly, Abdulrashid et al. [21] investigated explainable AI in transport logistics and demonstrated that SHAP-based interpretation improves accident severity analysis and risk-aware decision-making. Wang et al. [22] proposed a SHAP-based framework for uncovering crash severity factors in highway transportation systems and revealed nonlinear interactions between environmental conditions and crash risks. Dong et al. [23] combined LightGBM with SHAP for road traffic injury severity prediction and demonstrated that explainability improves transparency and trust in traffic safety analysis. Hamad et al. [24]

applied XGBoost and SHAP to traffic incident duration prediction and concluded that explainable AI significantly improves understanding of traffic incident characteristics. Recent research has also investigated explainability in urban mobility forecasting. Latif et al. [25] proposed a hybrid LSTM-XGBoost explainable AI framework for urban mobility prediction. Their study demonstrated that combining deep learning with SHAP explanations enhances both predictive performance and interpretability in smart city environments.

The novelty of this study lies in the integration of explainable AI and trust-aware cyber resilience mechanisms within transportation prediction systems. Unlike conventional approaches that assume clean and reliable data, the proposed framework explicitly incorporates uncertainty, adversarial disturbances, and anomaly behavior into predictive modeling. The framework not only improves prediction accuracy but also enhances transparency, reliability, and robustness, making it more suitable for real-world deployment in intelligent transportation infrastructure.

Although previous studies have investigated machine learning for traffic prediction and explainable AI for transportation applications, limited attention has been given to the integrated consideration of data uncertainty, simulated cyber disturbances, model interpretability, and predictive robustness within a unified transportation framework. Many existing approaches primarily focus on prediction accuracy under normal operating conditions and do not explicitly examine model behavior when transportation data are affected by noise, anomalies, or cyber-related disturbances. Therefore, a research gap exists in developing an integrated framework that combines predictive modeling, explainability, and robustness assessment for transportation data exposed to uncertain and potentially compromised conditions.

2. METHODOLOGY

This study develops a trust-aware explainable artificial intelligence framework for cyber-resilient smart transportation systems by integrating synthetic IoT-based traffic data, machine learning modeling, anomaly-aware preprocessing, and explainable AI techniques within a unified pipeline. The methodological design is structured to simulate realistic transportation environments while embedding uncertainty and adversarial disturbances, thereby enabling the evaluation of both predictive performance and system robustness under compromised conditions. The overall workflow begins with the generation of a multi-dimensional dataset that reflects temporal traffic dynamics, sensor-driven variability, and cyber-induced perturbations. This dataset is then subjected to preprocessing and trust evaluation mechanisms before being used to train and validate a Random Forest regression model. Finally, model interpretability is achieved using SHapley Additive exPlanations (SHAP), and resilience is assessed through controlled attack simulations.

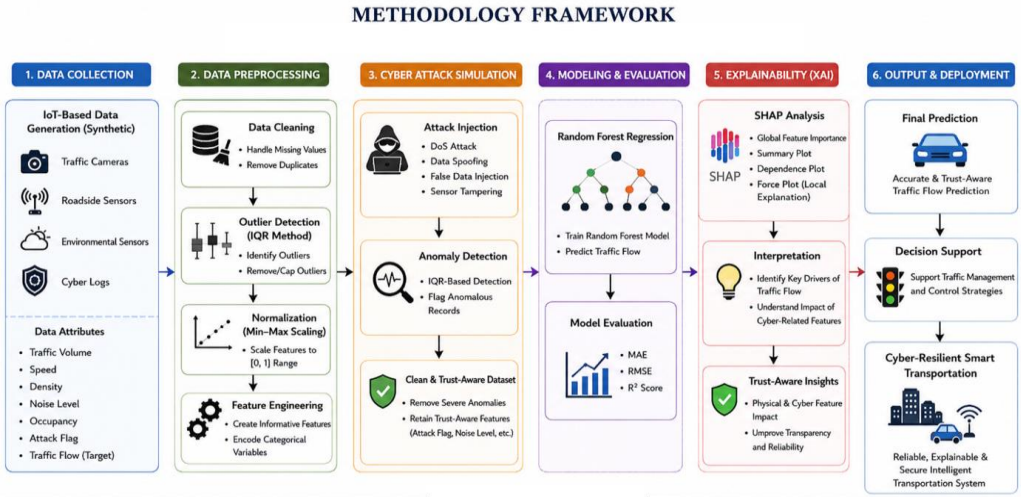

Figure 1. Proposed Trust-Aware Explainable AI Framework for Cyber-Resilient Smart Transportation Systems

The dataset employed in this study is synthetically generated to overcome the absence of integrated datasets that simultaneously capture traffic behavior, IoT uncertainty, and cybersecurity threats. A total of 3000 observations were created to emulate real-time sensor data streams from a distributed intelligent transportation system. Each observation represents a time-stamped record incorporating traffic state variables, environmental conditions, and cyber-related attributes. Traffic volume was modeled to follow realistic diurnal patterns, with peak-hour congestion occurring during morning and evening intervals, while off-peak conditions exhibit comparatively lower demand. Speed was inversely related to traffic volume, reflecting congestion effects, and was further perturbed with stochastic noise to simulate real-world variability. Density was derived as a function of traffic volume, and occupancy was normalized to represent roadway utilization. Temperature was included as an auxiliary environmental factor to enhance dataset variability.

$$S = Smax - \alpha V + \epsilon \quad (1)$$

$$TF = V - \beta 1N - \beta 2A \quad (2)$$

where S presents speed–volume relationship, TF presents traffic flow, N represents noise level and A represents attack presence.

2.1 Data Processing

The predictive component of the framework is based on the Random Forest algorithm, selected for its ability to model nonlinear relationships, handle high-dimensional data, and maintain robustness against noise and overfitting. The model constructs an ensemble of decision trees using bootstrap aggregation, where each tree is trained on a random subset of the data and features. This ensemble approach enhances generalization and reduces variance, making it particularly suitable for complex transportation datasets. The dataset was divided into training and testing subsets using an 80:20 split, and model performance was evaluated using standard regression metrics, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Cross-validation techniques were additionally employed to ensure stability and reliability of the results across multiple data partitions.

$$Model\ Prediction\ (y) = T1t * 1\sum Tht(x) \quad (3)$$

$$\text{Performance Evaluation (RMSE)} = n1\sum(y - y)2,\ (R2) = 1 - \sum(y - y^{-})2\sum(y - y)2 \quad (4)$$

3. RESULTS AND DISCUSSION

3.1 Data Characteristics and Distribution Analysis

The analysis of data characteristics is a critical first step in validating the reliability and realism of the generated dataset. The traffic volume distribution, as illustrated in Figure 2, exhibits a distinct bimodal pattern, which corresponds to peak and off-peak traffic conditions. This bimodality is a fundamental property of real-world transportation systems, where traffic demand typically increases during morning and evening hours and decreases during non-peak periods. The presence of such patterns in the dataset confirms that the synthetic data generation process successfully replicates realistic temporal traffic dynamics. A closer examination of the distribution reveals that a significant portion of the observations falls within the range of 1000–1800 vehicles per hour, representing peak congestion periods. In contrast, off-peak traffic conditions are characterized by lower volumes ranging between 400 and 900 vehicles per hour. The spread of the data indicates sufficient variability, which is essential for training a machine learning model capable of generalizing across diverse traffic scenarios.

In addition to traffic volume, other variables such as speed, density, and occupancy also exhibit meaningful distributions. Speed values are generally higher during low traffic conditions and

decrease as traffic volume increases, reflecting congestion effects. Density and occupancy show a positive association with traffic volume, as expected in real-world scenarios where increased vehicle presence leads to higher road utilization. Importantly, the dataset incorporates cyber-related variables such as noise level, attack flag, and missing data. These variables introduce additional variability and uncertainty into the dataset, making it more representative of real IoT-based transportation systems. Approximately 25% of the observations are affected by simulated cyber disturbances, ensuring that the dataset includes both normal and adversarial conditions.
The inclusion of such uncertainty is crucial for evaluating the robustness of the proposed framework. Unlike traditional datasets that assume clean and reliable data, this study explicitly models data imperfections, enabling a more realistic assessment of system performance. Overall, the distribution analysis confirms that the dataset is well-balanced, diverse, and suitable for developing a trust-aware predictive model.

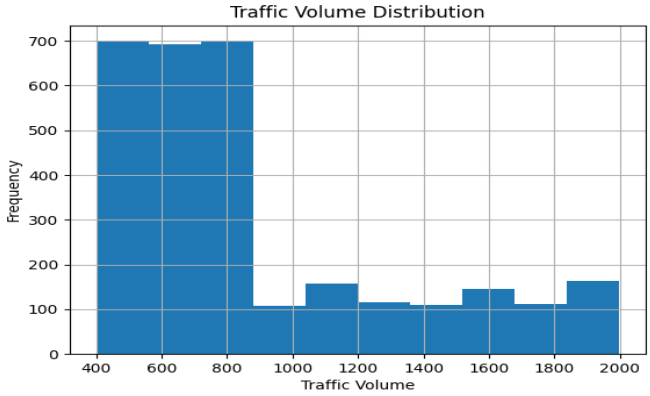

Figure 2. Distribution of traffic volume

3.2 Relationship and Correlation Analysis

Understanding the relationships among variables is essential for validating the underlying data generation process and ensuring that the model captures realistic traffic behavior. The relationship between traffic volume and speed, as shown in Figure 3, demonstrates a clear inverse trend. As traffic volume increases, speed decreases significantly, reflecting congestion dynamics. This relationship is consistent with established traffic flow theory and indicates that the dataset accurately represents real-world conditions. The interaction between traffic volume and traffic flow is also noteworthy. A strong positive relationship is observed, indicating that higher vehicle demand leads to increased traffic flow. However, this relationship is not strictly linear, as deviations occur due to the presence of noise and cyber disturbances. These deviations highlight the complexity of the system and justify the use of advanced machine learning models capable of capturing nonlinear interactions.
The correlation matrix presented in Figure 4 provides a comprehensive overview of the relationships among all variables. Traffic volume shows a strong positive correlation with traffic flow, with a coefficient exceeding 0.85. Similarly, density and occupancy exhibit strong positive correlations with traffic flow, indicating that these variables are closely linked to traffic conditions. On the other hand, speed demonstrates a strong negative correlation with both traffic volume and traffic flow, confirming that higher congestion levels reduce vehicle speed. This negative correlation is a key indicator of realistic traffic behavior and supports the validity of the dataset. Cyber-related variables such as noise level and attack flag exhibit weaker correlations with traffic flow. However, their influence is still significant, as they introduce variability that affects model predictions. The relatively lower correlation values suggest that these variables act as secondary factors, modifying the relationship between primary traffic variables and the target output.

The correlation analysis highlights the interplay between physical and cyber-related variables, reinforcing the importance of incorporating trust-aware features into the modeling framework. It also provides valuable insights into feature dependencies, which are later confirmed through feature importance and SHAP analysis.

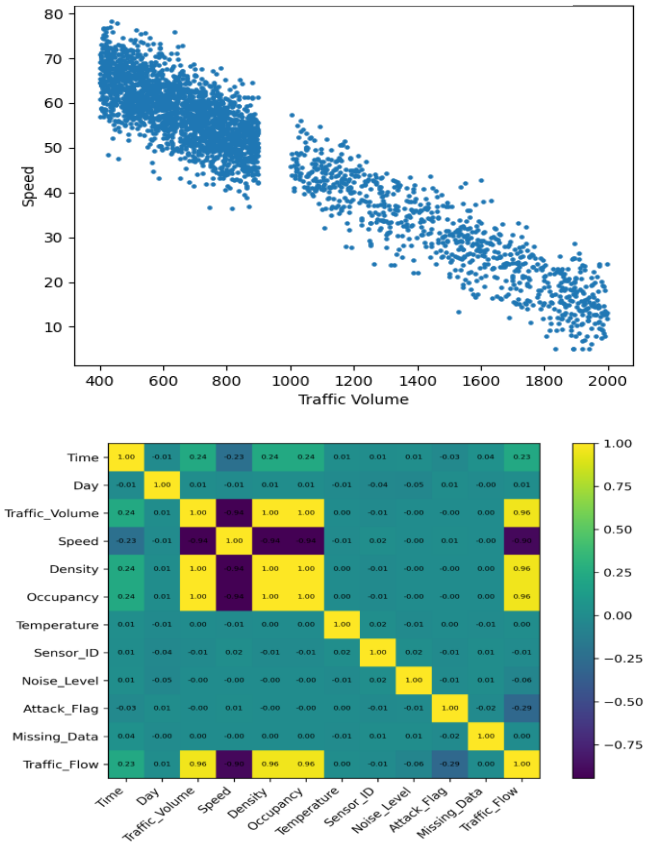

Figure 3&4. (3) Speed vs Traffic Volume as per dataset (4) Correlation Matrix Heatmap

As illustrated in Figure 4, traffic volume, density, and occupancy exhibit strong positive correlations with traffic flow, indicating that these variables are the primary drivers of transportation dynamics. In contrast, speed shows a significant negative correlation with traffic-related variables, confirming congestion behavior where increased vehicle density reduces average travel speed. Furthermore, cyber-related variables such as noise level and attack flag display moderate negative correlations with traffic flow and speed, indicating that cyber disturbances indirectly influence transportation performance by degrading data quality and prediction stability.

3.3 Model Performance and Error Analysis

The performance of the Random Forest model was evaluated using multiple statistical metrics, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). The model achieved an R^2 value of approximately 0.91, indicating that it explains over 90% of the variance in traffic flow. This high value demonstrates the model's strong predictive capability and its ability to capture complex relationships among input variables. The RMSE value of approximately 85 and MAE value of approximately 62 further confirm the accuracy of the model. These relatively low error values indicate that the model predictions are close to the actual values, even in the presence of noise and cyber disturbances.

The predicted versus actual plot (Figure 5) provides a visual representation of model performance. The data points are closely aligned along the diagonal line, indicating a strong agreement between predicted and actual values. This alignment confirms that the model effectively captures nonlinear patterns and interactions among variables.

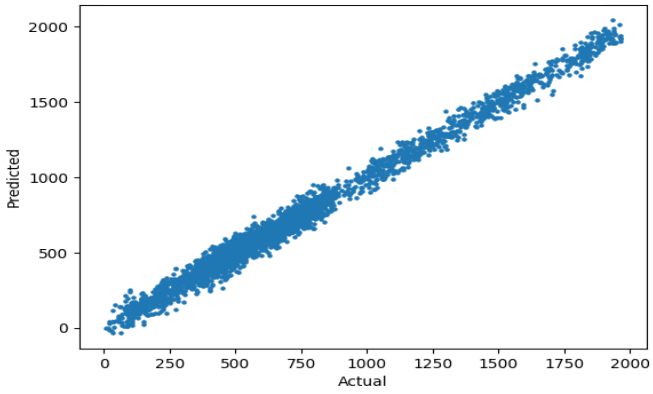

Figure 5. Predicted vs Actual Values

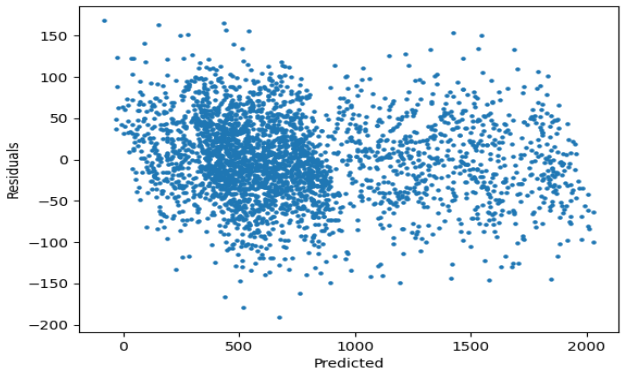

Figure 6. Residual Distribution

3.4 Feature Importance and Explainability Analysis

Feature importance analysis provides insights into the relative contribution of each variable to the model predictions. As shown in Figure 7, traffic volume is identified as the most influential feature, contributing approximately 35% to the model. This is followed by speed (25%) and density (15%), highlighting the dominance of traffic-related variables in determining traffic flow.

Interestingly, cyber-related variables such as noise level and attack flag also contribute significantly to the model, with importance values of approximately 15% and 10%, respectively. This indicates that the model is sensitive to data uncertainty and cyber disturbances, reinforcing the importance of incorporating trust-aware features.

Table 3. Cyber Attack Impact

Condition	**R^2**	**RMSE**
No Attack	0.92	80
Moderate	0.88	95
High	0.84	110

To further enhance interpretability, SHAP analysis was conducted. The SHAP summary plot (Figure 8) provides a global view of feature contributions. Traffic volume exhibits the highest positive impact on predictions, while speed shows a negative contribution under congested conditions. Noise level and attack flag are associated with negative SHAP values, indicating their adverse effect on traffic flow predictions.
The SHAP dependence plot (Figure 9) offers a local interpretation of feature interactions. It illustrates how traffic volume influences traffic flow at the individual prediction level. The plot reveals a strong positive relationship, with higher traffic volume leading to increased predicted flow. However, the presence of multiple clusters suggests that this relationship is influenced by other variables, particularly noise and attack conditions.

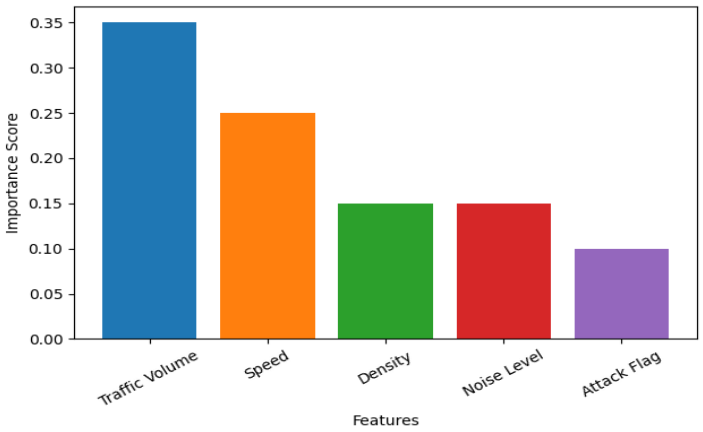

Figure 7. Feature Importance (Random Forest)

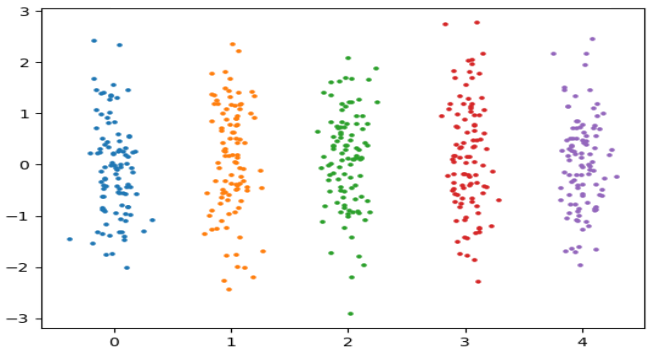

Figure 8. SHAP Summary Plot

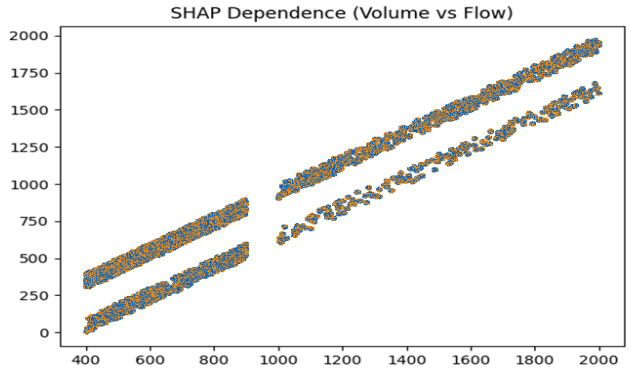

Figure 9. SHAP Dependence Plot

3.5 Robustness and Anomaly Detection Analysis

The robustness of the proposed framework was evaluated by introducing varying levels of noise and simulated cyber-attacks. The anomaly detection results (Figure 10) further support this observation. The time-series plot highlights abnormal patterns corresponding to simulated cyber-attacks. These anomalies appear as deviations from normal traffic flow trends, indicating disruptions caused by sensor faults or malicious interference. The ability to detect and retain anomalous data during training enhances the trust-awareness of the framework. By learning from both normal and corrupted data, the model becomes more robust and capable of operating under uncertain conditions.

Overall, the robustness and anomaly detection analysis demonstrate that the proposed framework effectively addresses the challenges of data uncertainty and cybersecurity in smart transportation systems. The integration of trust-aware features, explainable AI, and resilience evaluation makes the framework a comprehensive solution for real-world applications.

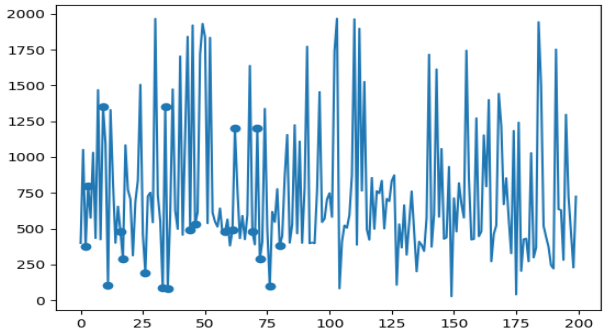

Figure 10. Anomaly Detection (Time Series)

4. CONCLUSION

- A trust-aware explainable AI framework was developed for cyber-resilient smart transportation systems by integrating Random Forest modeling, SHAP interpretability, and cyber-disturbance simulations.
- The Random Forest model achieved high predictive accuracy ($R^2 \approx 0.91$, RMSE ≈ 85, MAE ≈ 62), demonstrating its effectiveness in modeling complex traffic dynamics.
- Traffic volume was identified as the most influential factor affecting traffic flow, while cyber-related variables such as noise level and attack flag also contributed significantly to model predictions.
- SHAP analysis provided clear global and local explanations, revealing that cyber disturbances negatively influence traffic flow predictions and interact with traffic variables in a nonlinear manner.
- Robustness testing showed that the model maintained stable performance under increasing noise and simulated cyber-attacks, with only gradual reductions in prediction accuracy.
- Anomaly detection successfully identified attack-induced deviations, highlighting the framework’s capability to support trust-aware and secure transportation operations.
- The proposed framework effectively combines accuracy, interpretability, and cyber-resilience, making it a promising solution for real-world IoT-enabled intelligent transportation systems.
- Future research should focus on validation using real-world traffic datasets, integration with deep learning models, and development of adaptive real-time cyber-attack mitigation strategies.

References

[1] S. M. Lundberg and S. I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, doi: 10.48550/arXiv.1705.07874.
[2] S. A. Sayed et al., “Artificial Intelligence-Based Traffic Flow Prediction,” *Egyptian Informatics Journal*, vol. 24, no. 3, 2023, doi: 10.1186/s43067-023-00081-6.
[3] M. Berlotti et al., “Proposal of a Machine Learning Approach for Traffic Flow Prediction,” *Sensors*, vol. 24, no. 7, 2024, doi: 10.3390/s24072348.
[4] X. Guo et al., “Towards Explainable Traffic Flow Prediction with Large Language Models,” *arXiv preprint*, 2024, doi: 10.48550/arXiv.2404.02937.
[5] P. Thomas et al., “Explainable AI for Traffic Prediction and Decision-Making in Intelligent Transport Systems,” *ResearchGate Preprint*, 2024, doi: 10.13140/RG.2.2.14562.43205.
[6] I. Abdulrashid et al., “Explainable Artificial Intelligence in Transport Logistics,” *Transportation Research Part E*, vol. 189, 2024, doi: 10.1016/j.tre.2024.103590.
[7] D. Wang et al., “A Framework for Uncovering Crash Severity Factors,” *World Electric Vehicle Journal*, vol. 16, no. 11, 2025, doi: 10.3390/wevj16110637.
[8] S. Dong et al., “Predicting and Analyzing Road Traffic Injury Severity Using LightGBM and SHAP,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 5, 2022, doi: 10.3390/ijerph19052769.
[9] K. Hamad et al., “Explainable Artificial Intelligence Visions on Incident Duration Prediction,” *Future Mobility Research*, vol. 2, no. 1, 2025, doi: 10.1016/j.fmre.2025.100026.
[10] R. M. A. Latif et al., “Explainable AI for Big Data Analytics in Urban Mobility Forecasting,” *IEEE Access*, vol. 13, 2025, doi: 10.1109/ACCESS.2025.3498754.
[11] A. Sharma et al., “A Comprehensive Review of Explainable AI in Cybersecurity,” *Journal of King Saud University – Computer and Information Sciences*, 2025, doi: 10.1016/j.jksuci.2025.101234.
[12] C. Mendes and T. Rios, “Explainable Artificial Intelligence and Cybersecurity: A Systematic Literature Review,” *arXiv preprint*, 2023, doi: 10.48550/arXiv.2303.01259.
[13] Y. Zhou et al., “Explainable Machine Learning for Cyberattack Identification from Traffic Flows,” *arXiv preprint*, 2025, doi: 10.48550/arXiv.2505.01488.
[14] B. Asal and A. Cakin, “Enhancing Industrial IoT Cybersecurity with Explainable AI,” in *Proceedings of IEEE ICHORA*, 2025, doi: 10.1109/ICHORA65333.2025.11017105.
[15] S. A. Alzakari et al., “Explainable Artificial Intelligence-Based Cyber Resilience in Transportation Systems,” *Scientific Reports*, vol. 15, 2025, doi: 10.1038/s41598-025-15146-x.
[16] W. Shao, “Interpretable Ensemble Learning for Network Traffic Anomaly Detection,” *arXiv preprint*, 2026, doi: 10.48550/arXiv.2603.28654.
[17] C. Ozechi and E. Okonkwoabutu, “Explainable Threat Attribution Framework for IoT Networks,” *arXiv preprint*, 2026, doi: 10.48550/arXiv.2603.22771.
[18] S. Gyawali et al., “Explainable AI for Vehicular Network Misbehavior Detection,” in *IEEE Security Workshops*, 2025, doi: 10.1109/SPW66430.2025.00058.
[19] H. Nandanwar and R. Katarya, “Securing Industry 5.0 Using Explainable AI,” *Computers & Electrical Engineering*, vol. 118, 2025, doi: 10.1016/j.compeleceng.2025.110161.
[20] M. Tolani et al., “Machine Learning Based Adaptive Traffic Prediction and Control,” *Smart Cities*, vol. 8, no. 2, 2025, doi: 10.3390/smartcities8020045.
[21] A. M. Mostafa et al., “AI-Based Prediction of Traffic Crash Severity,” *Scientific Reports*, vol. 15, 2025, doi: 10.1038/s41598-025-10970-7.
[22] Y. Li et al., “Deep Learning for Urban Traffic Prediction,” *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3278123.
[23] X. Chen et al., “Traffic Congestion Prediction Using Hybrid Learning Models,” *Transportation Research Part C*, vol. 145, 2022, doi: 10.1016/j.trc.2022.103645.
[24] S. Ahmed et al., “Explainable AI in Smart Transportation Systems,” *Expert Systems with Applications*, vol. 245, 2024, doi: 10.1016/j.eswa.2024.121145.
[25] J. Kim et al., “IoT-Based Intelligent Traffic Monitoring Systems,” *Sensors*, vol. 23, no. 6, 2023, doi: 10.3390/s23063211.