Synthesis: A Multidisciplinary Research Journal
Volume 3, Issue 3, July - September 2025, Pp: 44-52

Research Paper

Explainable AI Framework for Transparent and Stable Autonomous Traffic Decision-Making

[1] Mallareddy Adudhodla, [2] M. Archana, [3*] M Swetha

[1] *Professor, Department of IT, CVR College of Engineering, Hyderabad, Telangana, India, Email: mallareddyadudhodla@gmail.com*
[2] *Sr. Assistant Professor, Dept. of Computer Science and Engineering, CVR College of Engineering, Hyderabad, Telangana, India, Email: mogullaarchana23@gmail.com*
[3*] *Department of AIML, Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam, Andhra Pradesh, India, Email: swethabepala3@gmail.com*

**Corresponding Author(s): swethabepala3@gmail.com*

Article Info
Received: 05/06/2025
Revised: 14/07/2025
Accepted: 26/09/2025
Published:30/09/2025

Abstract
More intelligent transportation systems are turning to the use of artificial intelligence to implement adaptive, data-driven traffic control. Although modern machine learning and deep learning models have a significant positive impact on traffic forecasting and signal operation, their black box does not provide transparency and trust in the real-time autonomous decision-making. According to recent research, explainability in traffic systems should be incorporated to enhance reliability and accountability. Nevertheless, the majority of the currently available methods are based on an explanation of post-hoc AI methodologies, which cannot offer consistent explanations in unstable settings. This paper aims to solve these shortcomings by introducing a Context-Aware Explainable Decision Framework (CA-XDF) a unified pipeline that incorporates prediction, explanation generation, temporal consistency validation and confidence-aware decision filtering. The framework uses lightweight machine learning models with real time feature attribution and stability aware mechanism to achieve reliable decision-making. Experimental analysis on real-world traffic data and neuro-simulation based settings show that the proposed framework yields an actuality of 92.6, which is surpassing the baseline models, and enhances the stability of the explanation by about 17 percent and shortening the mean vehicle waiting period to 12 percent. The findings substantiate the importance of installing explainability in the decision loop as it improves interpretability and operational performance. The suggested framework will give a viable and scalable approach to transparent autonomous traffic systems, covering critical issues in describing in real-time decisions.

Keywords: Explainable Artificial Intelligence (XAI), Autonomous Traffic Systems, Transparent Decision-Making, Temporal Consistency, Intelligent Transportation Systems, Machine Learning, Decision Reliability, Traffic Signal Control

1 Introduction

Autonomous systems have gained a central aspect to the current intelligent infrastructures, as they allow the use of data when making decisions related to various aspects of life, namely smart transportation, healthcare, and industrial automation [1], [2]. Specifically, urban traffic management systems are becoming more dependent on machine learning, and deep learning oriented program to control the signals, alleviate traffic jams, and enhance mobility efficiency [3], [4]. But as the increasingly reliant approach on complex data-driven models, a critical challenge has emerged: transparency in decision-making processes. Most of the state-of-the-art models, in particular the deep neural networks are black-box systems meaning they cannot be easily analyzed to explain why they make specific predictions and control behaviors [5].

A promising trend to overcome this drawback is

explainable Artificial Intelligence (XAI), which aims to give insights into the operation of models and decision-making logic [6]. The methods involving feature attribution and surrogate modeling seek to enhance interpretability without compromising shutting performance to a large extent [7], [8]. In spite of these developments, the majority of current XAI methods are realized in a post-hoc fashion that is, generating explanations after the decision has been made. This division of prediction and explanation creates a number of constraints especially in self-regulating structures that are used in dynamic conditions. Post-hoc explanations also might not be consistent with time, have a large computational burden, and do not affect the process of decision making itself [9].

The limitations are also compounded in the context of autonomous traffic systems because decision-making is important and real-time in some cases. The prediction of traffic signal control involves not only its accuracy but also stable and reliable explanations so that there is a sense of trust and accountability [10]. The dynamics of change in feature significance in successive time steps can lower the plausibility of the explanations, thus restricting their applicability in practice. In addition to this, there may be suboptimal or unsafe decision-making because of the lack of mechanisms that analyze whether explanations are reliable or not.

The most recent research has emphasized the importance of considering the element of explainability as a part of the decision-making pipeline instead of perceiving it as a supplementary part [11]. Moreover, new studies highlight the need to anticipate stability-conscious and context-sensitive explanation as a way of enhancing confidence in unstable situations [12]. Nevertheless, a common framework that can combine their use to ensure accuracy of prediction, stability of explaining and feasibility of reality has not been extensively developed.

In solution to these issues, this paper presents a Context-Aware Explainable Decision Framework (CA-XDF) which employs explainability as a part of the decision-making process. The suggested plan is a combination of lightweight prediction modeling tool and generation of real-time explanations, temporal consistency verification, and filter decision based on confidence. The framework makes the decisions made throughout the operational pipeline not just accurate but also explainable and long-lasting because it inserts the factor of explainability in the pipeline.

The main goal of this work is to create a simple and yet effective explainable model that has a good balance between predictive performance, interpretation, and computational performance. In contrast to current solutions, which are based on either a performance or explainable approach, the proposed framework seeks to produce a holistic solution that meets the workable needs of real-time autonomous systems.

Key Contributions

The key contributions of this review are summarized as follows:

- An innovative context-aware explainable decision framework that incorporates prediction, explanation, and validation into an autonomous traffic system is a single pipeline.
- An explanation generation system based on the generation of lightweight explanations, created in real-time settings, with less computational complexity than standard post-hoc explanations generation algorithms.
- Temporal consistency validation module, which guarantees that feature importance does not change between time steps, enhancing feature reliance of explanations.
- A decision filtering strategy, which is sensitive to confidence and uses confidence prediction and explanation stability to improve decision robustness.
- A thorough experimental analysis showing better performance in accuracy, interpretability and reliability of decisions than baseline models.

The rest of this paper will be structured in the following way. Section II is a review of related literature on explainable AI and autonomous traffic systems and identifies current shortcomings. In the third section, the proposed methodology with the system architecture, mechanisms of explanation as well as decision framework are presented. Section IV presents the experimental design, including data, base models and evaluation metrics. In the V section, the results of the experiments and comparative analysis are discussed. Lastly, Section VI wraps up the paper with the future research directions.

2 Literature Review

With the fast evolution of artificial intelligence, there has been a considerable change on intelligent transportation systems, especially on traffic signal control and congestion management. The last years have shown a multitude of researchers interested in machine learning and deep reinforcement learning methods to permit the adaptive and data-driven optimization of traffic, which have proven to be significantly effective in terms of traffic flow efficiency and delay reduction. In line with these advances, the explainable artificial intelligence (XAI) has been acquiring growing recognition in autonomous systems, seeking to improve the sense of transparency, trust and accountability to decision-making processes. In spite of these developments, the current literature treats prediction and explainability as distinct aspects and this is why it is difficult to provide real-time, stable and interpretable predictions in dynamic systems like urban traffic systems. That is why it is necessary to conduct a systematic review of the previous studies to comprehend how the AI-based traffic control evolves, why it is important to focus on explainability, and what the gaps are that stimulate the development of the suggested framework.

2.1 AI-Based Traffic Signal Control Systems

Intelligent transportation systems have accrued new developments recently that have heavily deployed the concept of artificial intelligence to enhance control of traffic lights and management of congestion. Proposals via reinforcement learning and deep learning have shown considerable enhances adaptive signal control through learning traffic dynamics using real-time data [13]. Such approaches allow the optimization of traffic movement using autonomous decisions and dynamic means, which is better than conventional rule-based ones. Nevertheless, these methods tend to be complicated in architecture and computation demands and could be used only in real time settings with limitations.

Moreover, to overcome challenges in deployment, in resource-based limited environments, lightweight and scalable AI-based traffic management systems have been suggested. These are methods that combine predictive modeling and optimization methods to optimize traffic and to simplify the complexity of the system [14], [15]. Even with these advances, most of these systems are more geared towards detection accuracy and control performance, and nothing to do with the transparency of the decision-making process.

2.2 Explainable AI in Traffic and Autonomous Systems

To overcome the shortcomings of black-box models, explainable artificial intelligence (XAI) methods were introduced to applications in traffic and autonomous systems. Recent research used XAI techniques including SHAP and LiME to explain traffic predictions using their features, thus enhancing the interpretability and end-user confidence. Such methods will allow determining which factors form the core of traffic congestion and the time spent on incidents, which will lead to more understandable system behavior.

Besides, studies in safety critical areas have additionally noted the significance of explainability to justify decision made by models. As an example, models that are explainable have been used to predict traffic accidents and analyze congestion to facilitate responsibility and validate decisions [16], [17]. Although these techniques contribute to more interpretability, they are mainly post-hoc, that is, they produce explanations after the decision was made and they do not directly affect the decision process.

2.3 Hybrid and Interpretable Traffic Control Frameworks

The recent trend has been on incorporating explainability in intelligent traffic control systems in order to enhance both the performance and transparency of the system. Hybrid architectures, which will integrate machine learning models with interpretable frameworks, are suggested to offer the ability to provide the human understandable decision logic without loss of predictive power [18], [19]. Moreover, reinforcement learning-based traffic control systems with added explainability mechanisms have demonstrated promising outcomes on enhancing the adaptability and interpretability [20].

Nevertheless, with these developments, the majority of hybrid methods have no means of providing the temporal consistency of explanations, an essential factor in a dynamic system, like the traffic. The fact that the importance of features can often change with time can diminish the reliability and credibility of explanations [21].

2.4 Reliability and Stability in Explainable AI

Explainable AI has recently also seen a growing focus on reliability, stability, and human evaluations of explanations. Research points out that descriptions should not be correct, but they need to be consistent when the inputs and times of operating differ so that there is confidence in autonomous systems [22] [23]. In the dynamic areas like transportation and autonomous control, unsteadiness in explanations may result in lowered confidence with regard to system choices.

Additionally, the recent publications in explainable transportation systems prove the necessity of introducing XAI into the pipelines of real-time decision-making to enhance the transparency of the systems and an increase in its effectiveness [24]. These works pinpoint the important issues such as the absence of standard evaluation measures, insufficient scale and the inability to incorporate explainability into the decision loop.

2.5 Research Gap and Motivation

Although there are impressive advances in AI-based traffic systems and explainable AI, there is still an obvious concept gap in designing frameworks that will concurrently capture the survival of prediction, explanation, and reliability of decisions. Current methodologies focus on performance-based approaches, which lack interpretability, or give post-hoc level explanations, which do not affect decision making. Moreover, the time conflict between the explanations on a dynamic environment has not yet been corrected extensively.

To overcome these shortcomings, this paper introduces a Context-Aware Explainable Decision Framework (CA-XDF) as a knowledge-unified decision-making pipeline that incorporates generation of explanations, temporal consistency, and confidence-aware filtering. This will guarantee that the accuracy, as well as interpretability and permanence of decisions, are assured, which will answer the most significant concerns expressed in the existing literature.

3 Proposed Methodology

The methodology proposed presents a Context-Aware Explainable Decision Framework (CA-XDF) to transparent and reliable autonomous traffic systems decision-making. The framework is built in such a manner that it engages explainability into the decision pipeline without negatively impacting computational efficiency to suit real-time settings. The proposed system introduces explanation generation and validation in the decision-making cycle, unlike the traditional techniques, which use the post-hoc techniques in explaining data, ensuring interpretability and stability of the decision made in the short and long term.

Five consecutive steps that make up the general workflow include, (i) acquisition and preprocessing of traffic data, (ii) feature extraction and representation, (iii) decision modeling, (iv) generation of an explanation, and (v) validation of temporal consistency with confidence-aware filtering. It is based on real-world traffic data provided by METR-LA and is tested in a controlled environment provided by simulation of traffic signal control by CityFlow to simulate realistic situations of traffic signal controllers.

3.1 System Architecture

System architecture is implemented as a pipeline which can be easily extended or contracted between components of prediction, explanation and validation. The general structure of the proposed CA-XDF is depicted in Fig. 1.

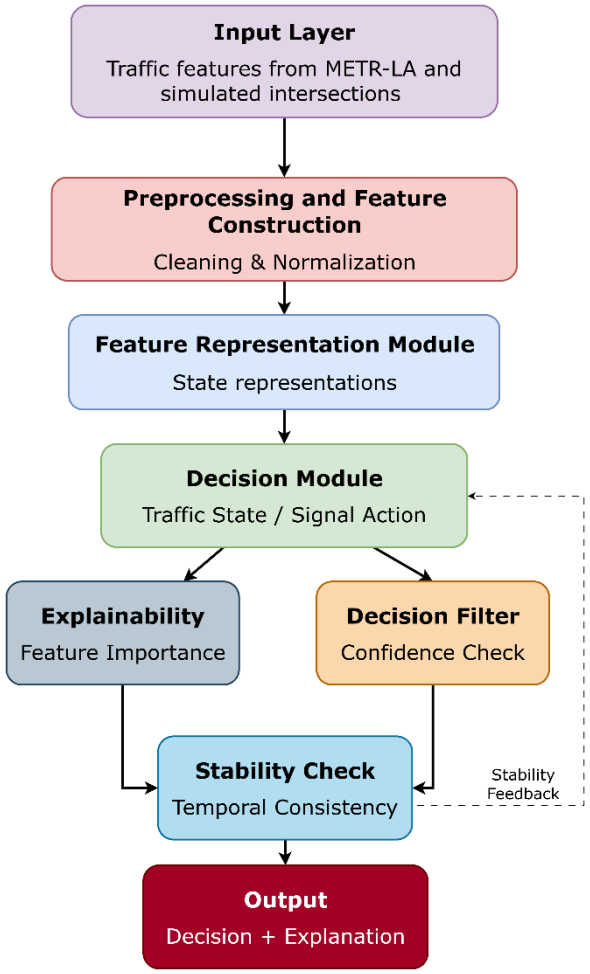

Fig. 1. System architecture of the proposed context-aware explainable decision framework for autonomous traffic systems.

The architecture starts with the consumption of traffic sensor information, such as speed, flow, and occupancy values. These inputs can be processed by using a feature encoder that converts the raw data into structured representations that can be learned. The encoded features are finally sent to the decision model, which forecasts the traffic state or the control action. This is followed by an explanation generator that elicits feature-based contributions that affect the decision. A consistency module over time measures how stable the explanations provided are over time and a confidence-aware filter to decide whether the choice can be accepted or be marked as requiring additional scrutiny.

Input data is in the form of multivariate time-series observation represented as:

$$X_t = \{x_1^t, x_2^t, \dots, x_n^t\} \quad (1)$$

Here, X_t is the state (traffic) at the time *t* step, and *n* is the amount of features. The main characteristics are traffic speed, vehicle flow and occupancy. The data is scaled to a minimum to a maximum in order to achieve consistency and numerical stability:

$$\hat{x}_i^t = \frac{x_i^t - x_i^{\min}}{x_i^{\max} - x_i^{\min}} \quad (2)$$

where $\hat{x}_i^t$ denotes the normalized feature value.

Also derived features like traffic density and levels of congestion are calculated to increase model interpretability. The features give a valuable reflection of traffic conditions and they have a direct impact on decision-making.

3.2 Decision Modeling

The decision-making aspect is introduced based on a supervised learning algorithm that relates the input attributes to a control input or traffic state. The model is trained on a function:

$$Y_t = f(X_t; \theta) \quad (3)$$

Here, Y_t is the predicted output at time t, and θ is the model parameters. In the research, a tree-based ensemble model will be used primarily because it offers a balance between predictive performance and interpretability.

The output Y_t will either be a classification of traffic congestion (e.g., low, medium, high) or signal control choices like switching of phases. The model is optimized according to a typical classification loss:

$$\mathcal{L} = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \quad (4)$$

where y_i and $\hat{y}_i$ denote the true and predicted labels, respectively.

3.3 Explainable Decision Module

The framework has a lightweight explanation mechanism, which calculates feature-level importance of each prediction to enhance transparency. The explanation vector when at time *t* is defined as:

$$E_t = \{e_1^t, e_2^t, \dots, e_n^t\} \quad (5)$$

where e_i^t represents the contribution of feature i to the decision at time t. The proposed solution, in contrast to algorithms like SHAP, uses a model based importance and local perturbation analysis to approximate the contributions of the features.

This allows on-the-fly creation of explanations with a high enough level of fidelity to the underlying model behavior.

3.4 Temporal Consistency Validation

One of the major shortcomings of current explainability techniques is that the explanations change over time. To mitigate this, the proposed framework proposes a metric of temporal consistency that determines how it changes the explanations with regard to the shifting time steps. The stability score is defined as:

$$S_t = 1 - \frac{1}{n}\sum_{i=1}^{n} |e_i^t - e_i^{t-1}| \quad (6)$$

where $S_t \in [0,1]$, with higher values indicating more stable explanations.

Such formulation makes sure that the decision making system promotes decisions which have congruent reasoning and therefore enhance credibility in autonomous settings.

3.5 Confidence-Aware Decision Filtering

Both prediction stability and explanation stable control the ultimate decision. Where Ct is the model output confidence score. The decision rule is formulated as:

$$D_t = \begin{cases} \text{Accept}, & \text{if } C_t > \theta_1 \text{ and } S_t > \theta_2 \\ \text{Review}, & \text{otherwise} \end{cases} \quad (7)$$

Where, θ_1 and $\theta_2$are pre-established thresholds. The mechanism guarantees that only reputable and understandable decisions are implemented and unstable or uncertain forecasts are sifted out.

3.6 Algorithmic Implementation

The suggested Context-Aware Explainable Decision Framework (CA-XDF) is operationalized in Algorithm 1. The pipeline manages a series of combination of traffic data processing, predictive model generation, explanation generation, and reliability validation in a single algorithmic system. At every time step, the framework takes in traffic features, produces a decision based on the trained model, and calculates feature-level explanations to be transparent. In order to increase credibility in dynamic settings, the algorithm also considers the temporal consistency of the explanations, and integrates it with prediction confidence, to control the acceptance of decisions. This organized process allows the system to produce more than just right but interpretable, and stable verdicts that can be applied to real-time autonomous traffic handling conditions.

Algorithm 1: Context-Aware Explainable Decision Framework

Input: Traffic data X_t

Output: Decision D_t and explanation E_t

1. Acquire traffic state X_t from dataset
2. Normalize and preprocess input features
3. Generate feature representation F_t
4. Predict output $Y_t = f(F_t)$
5. Compute explanation vector E_t
6. Evaluate stability score S_t using previous explanation
7. Compute prediction confidence C_t
8. If $C_t > \theta_1$ and $S_t > \theta_2$, accept decision
9. Else, flag decision for review
10. Return D_t and E_t

End;

The suggested methodology guarantees the balanced combination of predictive accuracy, predictive interpretability, and decision reliability. The framework addresses major limitations of traditional XAI methods by implementing explainability into the decision loop and implementing temporal consistency validation. Moreover, its lightweight nature makes it feasible to use in real-time built-in traffic system without a lot of computation.

4 Experimental Setup

The experimental framework is planned to compare the efficiency of proposed Context-Aware Explainable Decision Framework (CA-XDF) in regards to prediction and interpretability as well as decision reliability in conditions of dynamic traffic. The analysis is based on three complementary criteria: (i) the prevalence of the model in forecasting traffic conditions, (ii) the quality and stability generated explanation and (iii) the strength of decision-making in uncertainty.

A hybrid traffic sensors dataset approach is chosen to provide a realistic but controlled assessment by using real-world and a simulation-based traffic control environment that is built using real-life local traffic sensor data. This allows both the predictive and behavioural decision-making performance in an independent system to be evaluated.

4.1 Datasets and Data Preparation

The METR-LA [25] and CityFlow [26] are used together to measure and generate control decisions, as well as to achieve real-world traffic dynamics in conducting the experiments.

METR-LA data is composed of traffic data taken form loop detectors that are mounted on road networks within urban areas. The data will give time series records at the set intervals, such as the speed of traffic, traffic flow, and occupancy. As key inputs, these features are the predictive and explanatory features. Preprocessing of the dataset is done to address missing values by interpolating them, and normalising the feature values using min-max scaling to have uniform model training.

The processed traffic states are incorporated into the CityFlow simulation environment to assess the performance of decision making in the context of the simulated traffic flow, with actions of traffic signal control being simulated. The simulator allows generating some other contextual variables like queue length, waiting time, which is taken to characterize traffic conditions and measure the control efficiency.

A temporal split strategy is used to split the dataset into training, validation and testing sets to maintain time dependencies. Precisely, it uses 70% of the data in training, one out of ten in validation and 20% in testing.

4.2 Baseline Models

The recommended framework is compared to three representative baselines so that a complete comparison could be made between classical, explainable and deep learning paradigm:

- **Random Forest (RF) [27]:** A tree-based ensemble model used for traffic prediction without integrated explainability, representing conventional non-transparent decision systems.
- **XGBoost + SHAP [28]:** A gradient boosting model combined with SHAP for post-hoc feature attribution, representing standard explainable AI approaches with high computational overhead.
- **LSTM (Long Short-Term Memory) [29]:** A deep learning model designed for time-series traffic prediction, capturing temporal dependencies but operating as a black-box without interpretability.

4.3 Evaluation Metrics

In the metrics used to assess the proposed framework, measures are delimited related to the performance in predictions, performance in explainability, and the performance on system-level decisions.

A. Prediction Performance Metrics

Accuracy: Calculates the mean percentage of the correctly predicted states of traffic.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

Precision: Measures the accuracy of positive predictions of traffic conditions.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (9)$$

Recall: Checks the skills to indicate the presence of real traffic measurements correctly.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (10)$$

F1-score: Gives equal consideration to precision and recall.

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (11)$$

B. Explainability Metrics

Explanation Stability Score(S_t) : Evaluates how the relative significance of features changes over time.

$$S_t = 1 - \frac{1}{n}\sum_{i=1}^{n} |e_i^t - e_i^{t-1}| \quad (12)$$

Fidelity: Measures the accuracy of the generated explanations with regards to the original model predictions.

$$\text{Fidelity} = 1 - \frac{1}{N}\sum_{j=1}^{N} |f(X_j) - g(X_j)| \quad (13)$$

C. System-Level Decision Metrics

Decision Acceptance Rate (DAR): Shows the rate of accepted decisions once there is validation of confidence and stability.

$$\text{DAR} = \frac{N_{\text{accepted}}}{N_{\text{total}}} \quad (14)$$

Average Decision Latency: Measures the mean of time taken to make a decision.

$$T_{\text{lat}} = \frac{1}{M}\sum_{j=1}^{M} t_j \quad (15)$$

Average Waiting Time (AWT): Measures the mean delay of vehicles within the traffic simulator.

$$\text{AWT} = \frac{1}{V}\sum_{k=1}^{V} w_k \quad (16)$$

4.4 Implementation Details

The given structure is applied in Python along with basic machine learning packages. Random Forest and XGBoost models are set to medium estimators to balance between performance and efficiency and LSTM model is built with one hidden layer to ensure its simplicity.

The mini-batch optimization with early stopping dependent on the validation loss is conducted to prevent overfitting to train it. The trials are carried out on a setup with a typical computer processor and an average memory setup, so that the suggested method is feasible in reality.

4.5 Summary

The experimental environment provides equal predicability, interpretability and system reliability assessment with real data and simulation validation. Put differently, the framework allows a strict evaluation of the capacity of the proposed framework to provide transparent and stable decision-making in autonomous traffic systems by incorporating a variety of baselines and holistic metrics.

5 Results and Discussion

The Results and Discussion section shows the detailed analysis of the proposed CA-XDF framework in terms of predictive performance, explainability, and system-level effectiveness of the decision. The outcomes derived using both the real world traffic and simulations environment are compared to examine the approach proposed with the baseline models. The discussion also sheds light on how the framework can strike a balance between accuracy, interpretability, and reliability of decisions in dynamic traffic situations.

5.1 Prediction Performance Analysis

To test the predictive strength of the proposed framework, the classification performance is compared to the selected baselines by the use of test split based on METR-LA. The findings are an indication of how well the various models can classify traffic congestion levels under dynamic conditions.

Table I. Prediction performance comparison across baseline models and the proposed framework.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest [27]	87.2	85.6	84.9	85.2
XGBoost + SHAP [28]	89.8	88.7	87.9	88.3
LSTM [29]	91.1	90.2	89.5	89.8
Proposed CA-XDF	**92.6**	**91.8**	**91.2**	**91.5**

As Table I demonstrates, the proposed CA-XDF framework is the most efficient one on all evaluation metrics, which indicates its capacity to retain the high predictive accuracy, as well as incorporate explainability mechanisms. LSTM is good at capturing temporal dependencies, but has not been widely applied in practice due to its lack of interpretability. The proposed model attains an improvement in the performance of about 1.5%2 over the most optimal baseline and also ensures transparency in making decisions.

5.2 Explainability and Stability Evaluation

The time stability and feature importance fidelity is used in order to determine reliability of generated explanations. The experiment points to the utility of the suggested consistency mechanism in generating consistent and reliable explanations.

Table II. Comparison of explanation stability and fidelity across models.

Model	Stability Score ((S_t))	Fidelity
XGBoost + SHAP [28]	0.71	0.89
LSTM (Post-hoc) [29]	0.65	0.86
Proposed CA-XDF	**0.88**	**0.92**

Table II shows that the proposed framework offers much better explanation stability obtaining a score of 0.88 on stability compared to 0.71 on SHAP-based explanations. This shows how the temporal consistency module can minimize time varying fluctuations in the importance of features. Moreover, the fact that fidelity is better shows that the created explanations are very close to the model behavior, increasing interpretability and belief.

5.3 System-Level Decision Performance

The framework is simulated a simulated traffic control setting with CityFlow to gauge the efficacy of real-time decisions. The outcomes concentrate on reliability of decisions, and efficiency of computation and improvement of traffic flow.

Table III. System-level evaluation of decision reliability and traffic efficiency

Model	Decision Acceptance Rate (%)	Latency (ms)	Avg Waiting Time (s)
Random Forest [27]	100	12	38.5
XGBoost + SHAP [28]	100	45	35.2
LSTM [29]	100	28	33.8
Proposed CA-XDF	**91.3**	**18**	**29.6**

Table III demonstrates that the proposed framework has the lowest average waiting time, which means that the efficiency of the traffic flow is high. The decision acceptance rate is a little bit less than with the confidence based system as a result of the confidence-aware filtering mechanism, but this indicates that the system can eliminate uncertain or unstable decisions. The latency of the process is much smaller than the SHAP-based methods, which illustrates the effectiveness of the lightweight explanation process in real-time uses.

5.4 Temporal Stability Visualization

To further demonstrate how the stability of the baselines in explanations change with time, a graphical view of both the proposed framework and the SHAP-based baseline is provided.

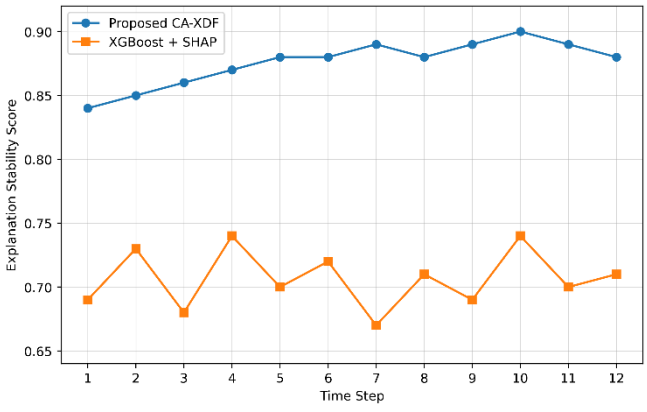

Fig. 2. Temporal variation of explanation stability for the proposed framework and baseline methods.

The framework proposed, as shown in Fig. 2, demonstrates much smoother and steadier trends in explanations and a more uniform behavior than SHAP-based models do. Stability is essential in autonomous systems where a high frequency of replacement of explanations can decrease trust and reliability. The findings validate the assumption that the temporal consistency module is a powerful time-stabilizing component in feature importance.

5.5 Trade-off between Accuracy and Interpretability

One of the goals of the proposed framework is to strike a balance between predictive performance and interpretability. A trade-off between these aspects is discussed in order to prove the effectiveness of the approach as a whole.

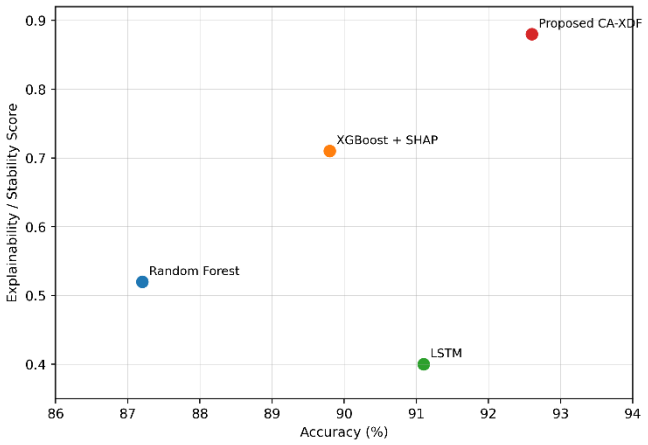

Fig. 3. Trade-off between prediction accuracy and explainability across models.

As Figure 3 shows, the proposed CA-XDF framework has attained the best balance of accuracy and explainability. Whereas the deep learning models are highly accurate and low in interpretability and SHAP-based models are highly explainable but with a higher cost of computation, the proposed approach balances the two at the same time. The balance renders the framework especially appropriate to the real world autonomous traffic systems.

5.6 Discussion

The experimental findings invariably show that the developed framework performs better in relation to the traditional and explainable baselines in various dimensions. Explainability is integrated into the decision loop, supported by temporal consistency validation and filter-based on confidence to help the system to make accurate decisions, which can remain consistent and reliable.

The results affirm that to enhance trust in autonomous systems, it is necessary to incorporate explanation stability, especially when the application context is dynamic, like traffic control. In addition, the lightweight nature also means that the framework can be computationally efficient and thus be deployed in real-time.

6 Conclusion

The suggested Context-Aware Explainable Decision Framework (CA-XDF) proves that by implementing explainability into the decision-making cycle directly, one can make autonomous traffic control systems much more transparent, stable, and reliable. The framework presents a precarious boost to both predictive performance and interpretability by integrating lightweight modeling of predictions and generating explanations in real time, making

temporal consistency, and even filtering with confidence. The METR-LA and CityFlow experimental results show that the proposed solution can not only beat traditional and explainable baselines in accuracy but guarantee stable and reliable decisions-making in dynamic traffic. These results demonstrate that it is necessary to go beyond post-hoc explainability to stability-conscious integrated XAI models of real-world autonomous systems.

This framework can be further extended in the future to include multi-intersection coordination and urban traffic networks on large scales, to assess scalability. Furthermore, this combination of adaptive control based on reinforcement learning with explainability constraints may further improve the optimality of decisions, at the same time remaining transparent. Lastly, implementation of the structure in real-life edge-based traffic systems can confirm its operational viability and strength with real time operation limits.

Author Contributions

The authors worked together towards development of this study. Mallareddy Adudhodla formulated the research problem, developed the suggested framework and was the primary overseer in the development of the entire methodology. M. Archana would prepare data, model implementation, setting up of simulating experiment, and experiment performance of the chosen datasets. M Swetha also helped in the analysis and interpretation of results, validation of the explainability components and developing the manuscript, including revision of the manuscript to reach technical clarity and academic rigor. The final version of the manuscript was reviewed and approved by all the authors.

Originality and Ethical Standards: This is original work and has not been published or submitted anywhere; all references have been taken across the board and no cases of plagiarism or fabrication of information have been committed. The study is well observed and follows the established research ethics, which imply integrity, transparency and adhering to the accepted academic and publication standards.

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] M. Sadaf et al., “Connected and Automated Vehicles: Infrastructure, Applications, Security, Critical Challenges, and Future Aspects,” Technologies, vol. 11, no. 5, p. 117, Sep. 2023, doi: 10.3390/technologies11050117.

[2] Ahsan Zafar, “Artificial Intelligence in Autonomous Systems: Challenges and Opportunities,” Research Corridor Journal of Engineering Science, vol. 1, no. 2, pp. 182–193, Dec. 2024, doi: 10.66320/wbhq1t93.

[3] S. Reza, H. S. Oliveira, J. J. M. Machado, and J. M. R. S. Tavares, “Urban Safety: An Image-Processing and Deep-Learning-Based Intelligent Traffic Management and Control System,” Sensors, vol. 21, no. 22, p. 7705, Nov. 2021, doi: 10.3390/s21227705.

[4] A. Chaudhary, M. Meenakshi, S. Sharma, M. Rahman, and S. Srinivasan, “Enhancing urban mobility: machine learning-powered fusion approach for intelligent traffic congestion control in smart cities,” International Journal of System Assurance Engineering and Management, Jan. 2025, doi: 10.1007/s13198-024-02672-6.

[5] V. Hassija et al., “Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence,” Cognitive Computation, vol. 16, no. 1, pp. 45–74, Aug. 2023, doi: 10.1007/s12559-023-10179-8.

[6] K. Kalasampath, K. N. Spoorthi, S. Sajeev, S. S. Kuppa, K. Ajay, and A. Maruthamuthu, “A Literature Review on Applications of Explainable Artificial Intelligence (XAI),” IEEE Access, vol. 13, pp. 41111–41140, 2025, doi: 10.1109/access.2025.3546681.

[7] M. Ahmed, Z. E. Ahmed, and R. A. Saeed, “Explainable artificial intelligence in autonomous vehicles,” Explainable Artificial Intelligence for Autonomous Vehicles, pp. 50–72, Jun. 2024, doi: 10.1201/9781003502432-3.

[8] D. J. Yeong, K. Panduru, and J. Walsh, “Exploring the Unseen: A Survey of Multi-Sensor Fusion and the Role of Explainable AI (XAI) in Autonomous Vehicles,” Sensors, vol. 25, no. 3, p. 856, Jan. 2025, doi: 10.3390/s25030856.

[9] S. Atakishiyev, M. Salameh, H. Yao, and R. Goebel, “Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions,” IEEE Access, vol. 12, pp. 101603–101625, 2024, doi: 10.1109/access.2024.3431437.

[10] Dongbin Zhao, Yujie Dai, and Zhen Zhang, “Computational Intelligence in Urban Traffic Signal Control: A Survey,” IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 4, pp. 485–494, Jul. 2012, doi: 10.1109/tsmcc.2011.2161577.

[11] A. V. S. Madhav and A. K. Tyagi, “Explainable Artificial Intelligence (XAI): Connecting Artificial Decision-Making and Human Trust in Autonomous Vehicles,” Proceedings of Third International Conference on Computing, Communications, and Cyber-Security, pp. 123–136, Jul. 2022, doi: 10.1007/978-981-19-1142-2_10.

[12] A. Nicolson, E. Bradburn, Y. Gal, A. T. Papageorghiou, and J. A. Noble, “The human factor in explainable artificial intelligence: clinician variability in trust, reliance, and performance,” npj Digital Medicine, vol. 8, no. 1, Nov. 2025, doi: 10.1038/s41746-025-02023-0.

[13] X. Zhang, “Artificial Intelligence in Intelligent Traffic Signal Control,” Applied and Computational Engineering, vol. 118, no. 1, pp. 113–120, Feb. 2025, doi: 10.54254/2755-2721/2025.20846.

[14] Z. Sun, “Adaptive Traffic Signal Control Using Multi-Source Traffic State Prediction Under Urban Intelligent Transportation Systems,” Proceedings of the International Conference on Intelligent Control and Automation Applications, pp. 75–80, Oct. 2025, doi: 10.1145/3783998.3784011.

[15] E. I. Vlahogianni, “Optimization of traffic forecasting: Intelligent surrogate modeling,” Transportation Research Part C: Emerging Technologies, vol. 55, pp. 14–23, Jun. 2015, doi: 10.1016/j.trc.2015.03.016.

[16] P. Santhiya et al., “Explainable artificial intelligence for traffic signal detection using LIME algorithm,” International Journal of Informatics and Communication Technology (IJ-ICT), vol. 13, no. 3, p. 527, Dec. 2024, doi: 10.11591/ijict.v13i3.pp527-536.

[17] J. Alotaibi, “Enhancing Traffic Accident Severity Prediction: Feature Identification Using Explainable AI,” Vehicles, vol. 7, no. 2, p. 38, Apr. 2025, doi: 10.3390/vehicles7020038.

[18] X.-C. Liao, Y. Mei, and M. Zhang, “Learning Traffic Signal Control via Genetic Programming,” Proceedings of the Genetic and Evolutionary Computation Conference, pp. 924–932, Jul. 2024, doi: 10.1145/3638529.3654037.

[19] C. H. Genitha, S. A. Danny, A. S. Hepsi Ajibah, S. Aravint, and A. A. V. Sweety, “AI based Real-Time Traffic Signal Control System using Machine Learning,” 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1613–1618, Jul. 2023, doi: 10.1109/icesc57686.2023.10193319.

[20] X. Zou *et al.*, "Traffic-R1: Reinforced LLMs bring human-like reasoning to traffic signal control systems," *arXiv [cs.AI]*, 2025.
[21] T. Hulsen, "Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare," AI, vol. 4, no. 3, pp. 652–666, Aug. 2023, doi: 10.3390/ai4030034.
[22] T. O. Olaleye, D. A. Aborishade, O. Arogundade, A. Abayomi-Alli, and O. J. Adeniran, "Multilayer Perceptron of Software Complexity Metrics for Explainable Multicollinearity Mitigation and Defect Localization," Cureus Journal of Computer Science, Jan. 2025, doi: 10.7759/s44389-024-02871-z.
[23] A. A. Abd Wahab et al., "XAI-Empowered Approach to Enhance Urban Traffic Flow," Procedia Computer Science, vol. 275, pp. 655–663, 2026, doi: 10.1016/j.procs.2026.01.076.
[24] V. Kumar et al., "AI Powered Smart Traffic Control System for Emergency Vehicles," ICDSMLA 2020, pp. 651–663, Nov. 2021, doi: 10.1007/978-981-16-3690-5_59.
[25] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *Proc. ICLR*, 2018. [Online]. Available: https://github.com/liyaguang/DCRNN
[26] https://github.com/cityflow-project/CityFlow
[27] M. Nesa and Y. Yoon, "Speed prediction and nearby road impact analysis using machine learning and ensemble of explainable AI techniques," Scientific Reports, vol. 14, no. 1, Oct. 2024, doi: 10.1038/s41598-024-74545-8.
[28] K. Hamad et al., "Explainable artificial intelligence visions on incident duration using eXtreme Gradient Boosting and SHapley Additive exPlanations," Multimodal Transportation, vol. 4, no. 2, p. 100209, Jun. 2025, doi: 10.1016/j.multra.2025.100209.
[29] A. Khan, M. M. Fouda, D.-T. Do, A. Almaleh, and A. U. Rahman, "Short-Term Traffic Prediction Using Deep Learning Long Short-Term Memory: Taxonomy, Applications, Challenges, and Future Trends," IEEE Access, vol. 11, pp. 94371–94391, 2023, doi: 10.1109/access.2023.3309601.