Explainable AI in Traffic Prediction

Han Juin Wong

June 13, 2025

Abstract

This paper presents a machine learning-based traffic prediction framework integrated with explainable artificial intelligence (XAI) to support intelligent transportation systems. Using quarter-hourly traffic data from selected New Zealand highway sites, the LightGBM model is employed to forecast traffic volumes for heavy and light vehicle classes. To enhance model interpretability, Shapley Additive Explanations (SHAP) are used to identify key features influencing predictions. Results demonstrate that site-specific and class-specific training improves performance. SHAP analysis highlights the importance of temporal features such as average past hour traffic, 1-hour trend, and time of day, which align with current traffic policies including ramp metering and dynamic lane control. The initial prototype effectively validates the use of XAI in traffic forecasting and policy relevance.

1 Introduction

Over the past couple of decades, the modes of transportation have evolved rapidly, from wagons and sailing ships to steam engines and modern automobiles. Transportation has become increasingly important as it significantly reduces travel time for people around the world. However, in modern times, the growing number of transportation services has begun to overload transportation systems, leading to negative impacts and compromising the original purpose of transportation. This is why traffic planning is essential, it plays a critical role not only in accommodating more transportation within the system but also ensuring the smooth flow of traffic. With current technological advancements, the ability to predict traffic has emerged, enabling the development of more efficient and intelligent traffic management systems.

The aim of an Intelligent Traffic Management System (ITS) is to reduce traffic congestion, enhance road safety, optimize fuel consumption and emissions, and enable real-time traffic management and traveler information [1].

However, the complexity of ITS has intensified the need for accurate traffic prediction models. [2] Traffic prediction models aim to forecast future traffic states, such as speed, flow, or congestion levels, based on historical and real-time data. [3] Reliable traffic prediction plays a crucial role in traffic management, enabling authorities to optimize signal control, congestion control, and ensure smooth traffic. [4]

Beyond operational improvements, traffic prediction results can also serve as valuable inputs for strategic traffic planning and policymaking. [2] Accurate forecasts provide significant insights into potential bottlenecks, resource allocation needs, and infrastructure development priorities. [5] As cities pursue smart transportation initiatives, the ability to integrate predictive analytics into long-term policy design becomes increasingly critical. [6]

However, traffic prediction remains a challenging task due to several key factors. First, issues related to data quality such as incomplete, missing, or biased traffic data, can significantly compromise the accuracy and reliability of prediction models [2, 4, 5]. Second, many models struggle to adapt to the dynamic nature of traffic environments or to scale effectively across large ITS, resulting in degraded performance [2, 7, 8]. Third, insufficient or missing feature selection can weaken the robustness and contextual sensitivity of traffic prediction models, limiting their applicability across varying conditions [3, 9]. Lastly, the lack of interpretability in complex models poses a significant barrier, as the inability to understand and explain how predictions are generated undermines stakeholder trust, safety assurance, and adoption, especially among policymakers [2, 6–8].

Several deep learning techniques, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Graph Neural Networks (GNNs), have been widely adopted in traffic prediction due to their ability to model complex spatio-temporal patterns [6]. However, these models often function as black boxes, their internal decision-making processes are difficult to interpret or explain due to their inherent complexity. This lack of transparency limits the trust, acceptance, and adoption of such models by stakeholders and policymakers, particularly in safety-critical domains such as traffic planning and regulation [2, 7, 8].

To address this gap, Explainable Artificial Intelligence (XAI) techniques have emerged as promising solutions. XAI is about making AI systems easier for people to understand. Many AI models work like "black boxes," where they give results without showing how or why they made those decisions. XAI helps open that box by providing simple explanations about what the AI looked at and why it reached a certain conclusion. XAI takes the output of an AI model, analyzes it, and provides interpretations that are understandable to users without a technical background. This helps users to better trust and make use of AI systems.

Figure 1: Explainable AI Process

In Figure 1, the Explainable AI process is illustrated. XAI functions as a modular framework that is integrated around or into machine learning and deep learning models. The architecture typically includes three key layers: the prediction layer, the explanation engine, and the explanation output. The prediction layer consists of machine learning or deep learning models, which are usually black-box models that are non-interpretable and optimized for accuracy rather than explainability. The explanation engine sits alongside or wraps around the prediction layer to provide interpretability. This component can take various forms, such as a surrogate model, a feature attribution engine, or a saliency map generator. A surrogate model is a simpler model trained on the outputs of the black-box model to approximate and justify its behavior. A feature attribution engine assigns scores to input features to indicate their contribution to a specific prediction. Meanwhile, a saliency map generator used images to highlight the areas of the input that are most responsible for the output. Lastly, the explanation output acts as the interface through which users interpret the model's behavior. It typically includes visual or textual representations such as plots, heatmaps, or summaries, enabling users to make informed decisions or policy changes based on the insights provided.

XAI methods aim to make model predictions transparent and understandable, enabling policymakers to better assess the rationale behind predictive results. Integrating explainability into traffic prediction models not only enhances model reliability but also strengthens the link between data-driven insights and policy interventions [6, 7]. By incorporating XAI, several key benefits emerge. First, transparency is improved, which fosters greater trust in AI-driven decisions and helps ensure safety. It also aids in detecting and correcting errors by highlighting incorrect or biased predictions. Additionally, XAI helps identify the source of such errors, allowing developers to refine models and enhance overall performance. Clear explanations of AI recommendations improve human-AI interaction, making systems more accessible to non-technical users. Moreover, XAI supports the validation of model outputs against safety protocols and facilitates compliance with regulatory standards by shedding light on the decision-making process. Finally, it provides valuable insights for policymakers, enabling more informed traffic planning and infrastructure development decisions.

Several industries have started implementing XAI into their models to explain results and support policy planning and changes. Shah et. al deployed SHAP, an XAI tool, to identify and rank features influencing job satisfaction and productivity. [10] For instance, features such as number of completed projects, feedback ratings, and salary were highlighted as strong predictors. These insights can help policymakers design targeted strategies, such as allocating more resources to employee training, improving performance feedback systems, or adjusting compensation structures to enhance employee well-being and drive workforce productivity.

Another example applied XAI to classify green energy jobs [11], where SHAP was also implemented to interpret the importance of features, helping researchers understand what drives the classification of a job as "green." As a result, occupational titles, industry labels, and geographic indicators were found to contribute most to the classification of green job categories. This allowed policymakers to pinpoint which sectors and regions are leading in green job growth, and tailor educational programs, incentives, or subsidies accordingly.

This paper explores the role of XAI in supporting traffic planning policies through traffic prediction, examines the advantages and limitations of current traffic prediction systems, and investigates how explainable AI can guide policymakers in traffic policy changes. The research questions addressed in this study are as follows:

RQ 1: What is Traffic Prediction and its key objectives?

RQ 1.1: How can traffic planning policies be guided by traffic prediction results?

RQ 2: What are the advantages and limitations of Traffic Prediction Systems in traffic planning policy guidance?

RQ 3: How can Explainable AI techniques be used for traffic prediction models?

RQ 4: How can Explainable AI results guide traffic policies for improving traffic congestion?

2 Literature Review

To address the research question, a structured search string was developed and applied across several academic databases, including IEEE, Springer, ScienceDirect, and Scopus. The formulated search string is presented below:

```
"traffic congestion" AND ("AI" OR "Artificial Intelligence" OR "Machine
Learning" OR "ML") AND "Prediction" AND ("Explainable AI" OR "XAI")
```

Criteria	**Inclusion**	**Exclusion**
Publication Type	Peer-reviewed journal articles, conference papers, systematic reviews	Non-peer-reviewed articles, blogs, web pages
Time Frame	Publications from the last 5 years (2020 onwards)	Publications before 2020
Focus	Studies evaluating AI-based adaptive traffic control systems or applying explainable AI techniques (e.g., SHAP, LIME)	Studies without experiments, models, or explainability focus
Domain	Smart cities, intelligent transportation systems, industrial traffic management	Aviation, Maritime

Table 1: Inclusion and Exclusion Criteria

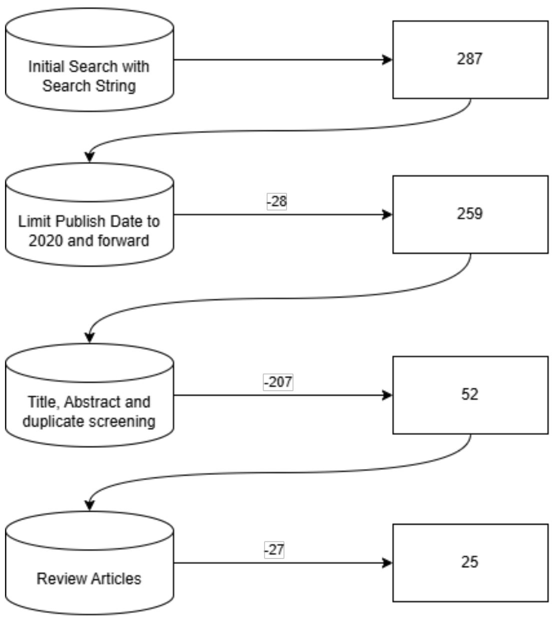


Figure 2: Study Inclusion Process

Table 1 presents the criteria considered for the literature review, while Figure 2 illustrates the study inclusion process. A total of 25 articles were selected for analysis in this paper. Section 2 is divided into five subsections, each addressing a specific topic: an overview of traffic prediction, the role of traffic prediction in traffic planning policies, the advantages and limitations of traffic prediction systems, the application of XAI in traffic prediction, and the role of XAI in guiding traffic policy decisions.

2.1 Overview of Traffic Prediction

Traffic prediction can be defined as the use of historical and real-time data to forecast future traffic states, such as vehicle speed, flow, congestion levels, and travel time. The primary objective of traffic prediction is to improve traffic management and mitigate congestion [2, 3].

According to Yin et al. traffic predictors can be categorized into five main groups. [2] The first is flow, which refers to the number of vehicles passing through a specific point over a defined period. The second is speed, which denotes the velocity at which vehicles travel, this can vary based on road type and regional speed limits. The third is demand, which involves estimating the future number of trip requests in a given area based on historical patterns, where the number of trip starts or ends represents demand at a particular time. The fourth is travel time, which calculates the time required to move between two points within a road network, typically incorporating both driving time and waiting time. Lastly, occupancy measures the proportion of road space currently occupied by vehicles and serves as a key indicator of roadway efficiency.

This section will discuss traffic prediction in 4 different section, traffic prediction in traffic planning policies, pros and cons of traffic prediction system, XAI techniques for traffic prediction and the role of XAI in guiding traffic policy changes

2.2 Traffic Prediction's Role in Traffic Planning Policies

Traffic planning policies are rules and guidelines designed to manage and improve the movement of vehicles, pedestrians, and cyclists within a specific area, such as a city or region. Grote et. al stated that the government of the country or the local government authorities have the responsibility of planning policy at the level of cities and urban areas, which includes planning for traffic. [12] Traffic prediction plays a critical role in traffic planning policies and can contribute in various ways. This section will explore these aspects in greater detail.

According to Mystakidis et at. traffic forecasting helps improve public transportation scheduling and route planning. [13] This can be achieved by predicting the number of passengers and traffic flow on public transport. Decision makers can dynamically adjust the bus frequency and routes according to the prediction to meet demand. This approach improves transit reliability, reduces passenger wait time, and encourages public transport usage.

Besides, traffic prediction models are integrated into emergency management policies to forecast congestion during pandemics or disasters [14]. These predictive results enable decision makers to identify potential bottlenecks along evacuation routes, allowing better allocation of resources, significantly reduce response times during crises, which is critical factors in life-or-death situations.

Shahid discussed several cities have incorporated traffic prediction results into policy changes and urban planning. [15] For example, Singapore deployed the Expressway Monitoring and Advisory System (EMAS), which monitors highway traffic conditions using security cameras and incident detection systems. EMAS utilizes both historical and real-time traffic data to forecast congestion levels. These predictive insights are then used to implement dynamic congestion pricing, adjusting toll rates based on traffic conditions. This approach aims to manage demand effectively and reduce congestion in real-time. Similarly, Barcelona uses traffic prediction data for urban road infrastructure planning. By analyzing predictive insights, the city is able to optimize traffic signal timings and manage parking resources. This strategy has helped reduce traffic congestion and improve sustainability.

Chen et al. introduced a dynamic lane reversal and traffic control strategy designed for autonomous vehicle environments [16]. Tested on Singapore's Ayer Rajah Expressway, this implementation demonstrated significant reductions in congestion and total travel time. This research aligns closely with short-term traffic planning policies aimed at mitigating peak-hour congestion. Its relevance is also evident in high-traffic corridors like the Auckland Harbour Bridge, where traffic volumes vary sharply between morning and evening peaks. Implementing a dynamic lane reversal strategy, supported by real-time predictive control, could enhance the effectiveness of existing lane configuration policies.

2.3 Traffic Prediction

Over the past few years, a range of machine learning and deep learning techniques have been applied to traffic prediction models.

A Spatio-Temporal Sparse Regression (ST-SR) model was introduced by Zheng et. al. as a traditional method for short-term traffic flow prediction, which fuses temporal and spatial factors into a multi-factor dictionary. [17] The authors used Least Absolute Shrinkage and Selection Operator (LASSO) to select the most relevant temporal and spatial features, combining the data and applying L1 regularization to obtain sparse coefficients for prediction. For comparison, ST-SR was evaluated against Long Short-Term Memory (LSTM) and CNN models and outperformed them across all metrics. For instance, the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) were reduced to 44.1% and 28.3% for 15-minute predictions. However, the model is linear, which improves its interpretability and prediction stability but limits its ability to capture highly nonlinear traffic patterns compared to deep learning approaches. It also lacks explainability tools, making it difficult to interpret or validate predictions, which is an important requirement for real-world traffic safety applications.

Gebre et. al. proposed a solution that uses physical laws, specifically the Lighthill-Whitham-Richards (LWR) model, as the core concept for traffic prediction. [18] They aim to predict traffic density using temporal and spatial data. They introduced a neural network called a physics-informed neural network (PINN) into the model training. Embedded with a Generative Pretrained Transformer (GPT-4), it acts as the interface between the PINN and the user. The model achieved strong performance, with a Mean Absolute Error (MAE) of 0.0188 and a Mean Squared Error (MSE) of 0.0007. However, despite the promising results, the study lacks a comparison with baseline models. Additionally, the LWR model, originally formulated in 1955, has several limitations, including assumptions of homogeneous traffic, the absence of stochastic elements, and a lack of learning capabilities.

Zhang et al. combined the Quantum Genetic Algorithm (QGA) and Learning Vector Quantization (LVQ) to create the QGA-LVQ Neural Network for traffic prediction. [19] This model aims to predict short-term traffic flow at 5-minute intervals. QGA is an optimization algorithm inspired by quantum computing and genetic algorithms, while LVQ is a supervised learning algorithm known for pattern classification. The authors first used the QGA technique to initialize a quantum population representing the LVQ weights. The optimized weights were then fed into the LVQ to perform traffic classification and prediction. When compared against the Genetic Algorithm–Backpropagation Neural Network (GA-BP) and the Wavelet Neural Network (WNN), the QGA-LVQ achieved a 25% lower MAPE value. However, the training environment utilized outdated hardware, which may have impacted the final results. Slower computational performance could have forced early stopping during model training, potentially leading to underfitting and reduced prediction accuracy. Additionally, limited processing power may have constrained model complexity and hindered thorough hyperparameter optimization. Furthermore, the novel model lacks interpretability, which is crucial for traffic safety applications.

Long Short-Term Memory (LSTM) networks were also used by Shin et. al. for traffic prediction [4]. LSTM is a type of Recurrent Neural Network designed to learn patterns from sequential data, especially when long-term dependencies are important. It is well-suited for traffic prediction because traffic data is inherently time-sequential. This paper aims to predict traffic speed 5 minutes ahead for both urban and suburban roads. One of its strengths is the use of Median Absolute Deviation (MAD) to remove outliers, along with the application of missing data correction techniques based on spatial trends, temporal trends, and historical patterns. By focusing on data preprocessing and utilizing LSTM, the model achieved a MAPE of 6.08% for urban and 4.30% for suburban roads. The proposed LSTM model was also compared with RNN and STGCN models, outperforming them with a 3–17% improvement. However, a notable limitation of the model is the lack of explainability, which is critically important in traffic prediction systems.

Although most of traffic prediction systems utilized deep learning or machine learning techniques to achieve high predictive accuracy, they often function as "black boxes" with limited transparency. This lack of interpretability makes it difficult for stakeholder to understand the rationale behind the model's predictions. Consequently, Explainable AI has emerged as a solution to bridge this gap by providing deep insight into model behavior, enhancing trust, and supporting informed decision-making in intelligent transportation systems.

2.4 Explainable AI Techniques for Traffic Prediction

There are many Explainable AI techniques available today. The most common ones are SHapley Additive exPlanations (SHAP) and Local Interpretable Model-Agnostic Explanations(LIME). Other tools include Integrated Gradients, XAI-Toolbox, XAITK, and InterpretML. SHAP is a game theory-based approach that assigns an importance score to each feature, explaining how it contributes to the model's prediction. LIME provides explanations for individual predictions by approximating the complex model with a simpler, interpretable one. Several research papers have already discussed the use of SHAP and LIME in traffic prediction models, which will be explored in this section.

Waqas et al. used an LSTM model for traffic prediction, specifically focused on autonomous vehicles. [20] SHAP was then applied to explain the model's predictions. The aim was to predict traffic congestion levels such as blocked, mild, moderate, and severe. The dataset used consisted of autonomous vehicle sensor data, weather inputs, and historical traffic logs. The model achieved a 99.8% training accuracy and 99.1% validation accuracy, which raises concerns about potential overfitting due to the high performance. SHAP successfully identified the most influential features contributing to traffic congestion in the proposed model. These included time of day, traffic volume, weather conditions (e.g., rain, snow, cloud cover), and holiday indicators. These insights validated the model's reasoning and helped explain why congestion was predicted at certain times or routes. Although the model performs well on training data, it lacks validation results for unseen data, which makes it hard to assess how well the model generalizes to real-world scenarios.

Rather than using existing Explainable AI tools, Kong et. al. developed a custom framework called Traffexplainer. [6] They used Graph Neural Networks (GNNs) for traffic prediction and incorporated a Perturbation-Based Hierarchical Interpretation Generator as the interpretation layer. This layer includes two types of masks: a spatial mask and a temporal mask. The spatial mask learns the importance of first and second-order neighbors by perturbing the adjacency matrix at different GNN layers, while the temporal mask identifies the most influential time steps for each node by perturbing the input traffic time series.

The model aims to predict short-term traffic flow or speed. The authors compared the performance of the base GNN model with and without the interpretation layer. The version with the interpretation layer achieved a better RMSE value of 15.504, compared to 15.561 for the baseline model.

As for the interpretation results, the most influential time intervals were found to be 55–60 minutes prior, indicating that congestion patterns from nearly an hour earlier are key predictors of future traffic conditions. Another notable finding was that second-order neighbors had a greater influence than first-order ones, suggesting that roads one or two intersections away can significantly affect traffic buildup, especially in dense urban areas.

Although the proposed model provides explanations by training hierarchical masks after freezing the predictor, it does not form an end-to-end interpretable framework. The prediction and explanation processes are separated, which may introduce a disconnect between the model's actual decision-making and the generated interpretations. Nevertheless, the study demonstrates that integrating an interpretation mechanism into a model can also enhance its predictive accuracy.

Vijaya et al. used Random Forest (RF) and k-Nearest Neighbors (kNN) for model training to predict public transportation delays. [21] Although the paper focuses on predicting delays in public transit rather than road traffic congestion or flow, it remains highly relevant due to its extensive application of Explainable AI techniques. The study employed both SHAP and LIME, comparing their results across the two machine learning models.

RF is an ensemble learning method that constructs multiple decision trees and combines their outputs to improve accuracy. It trains each decision tree on a random subset of the data, each tree makes a prediction, and the final output is determined by majority vote. In contrast, kNN is a learning algorithm that makes predictions based on the distance between data points, outputting the most common class among the k nearest neighbors.

In terms of performance, the RF model outperformed kNN, achieving a Mean Absolute Error of 3.39 compared to 4.35 for kNN.

Regarding explainability, SHAP identified the most influential features as time of day, day of the week, and route type, concluding that weekday peak hours are strongly correlated with delays. Meanwhile, LIME provided case-specific insights, revealing that evening rush hour delays were primarily caused by route congestion. One of the limitation of this paper is it only focuses on train and bus delay prediction, and does not address road traffic flow, congestion levels, or dynamic routing. This limits the generalizability of findings to broader traffic systems.

The reviewed literature demonstrated that Explainable AI help bridge the gap between accurate prediction and meaningful interpretation, making AI systems more trustworthy and usable in real-world applications. Having established the value of Explainable AI in understanding traffic prediction models, the next chapter explores how these insights can extend beyond model evaluation to inform and guide traffic policy changes.

2.5 The Role of Explainable AI in Guiding Traffic Policy Changes

ITS often generate many data, enabling predictive model to predict traffic congestion, delays and traffic flow. However, for these prediction to influence policy decision significant, they must be interpretable and transparent. AI model, often called as "black-box", often little value to policy makers if the rationale behind the outputs cannot be understood or trusted. This section explores how XAI can bridge the gap by translating complex model behaviour into actionable insights that inform traffic policy decision.

Policy making in transportation industry requires accountability, traceability and stakeholder trust. Decisions such as adjusting traffic signals, rerouting vehicles, or allocation of resources must be based on clear, evidence reasoning. XAI techniques make machine learning models more transparent by showing which features influence outcomes. This clarity is crucial for building confidence in AI systems and ensuring that policies is derived from them are defensible and aligned with real-world dynamics.

Multiple studies have demonstrated how Explainable AI can provide actionable insights to guide traffic-related policy decisions. In the study involving autonomous vehicle traffic prediction [20], SHAP was used to explain the high-accuracy outputs of the LSTM model. SHAP revealed that congestion was most strongly influenced by features such as time of day, traffic volume, weather conditions, and holidays. These insights can help policymakers design dynamic routing strategies and plan emergency responses in advance. During holidays or bad weather conditions, such policies become particularly useful in ensuring that traffic is managed safely and efficiently.

Similarly, Kong et al. developed the Traffexplainer framework to interpret GNN-based traffic predictions, revealing that second-order neighboring roads and traffic patterns from 50–60 minutes prior significantly influence congestion [6]. This highlights that decision makers should not only focus on direct intersection, but also consider and monitor indirect influences of roads that are 1 or 2 steps away, because congestion and flow changes can propagate. These insights also enable policymakers to prioritize strategic infrastructure upgrades (road upgrade) in areas that may otherwise be overlooked.

In another study focused on public transit delays, SHAP and LIME were applied to Random Forest and kNN models, identifying weekday peak hours and route characteristics as major delay factors [21]. SHAP provided a global view useful for scheduling and resource planning, while LIME offered instance-specific insights, such as pinpointing evening rush-hour congestion on specific routes. By combining the result of both model, policymakers can design adaptive and targeted interventions, such as traffic signal timing adjustments, route prioritization, and congestion mitigation measures, based on transparent, data-driven evidence.

Model explanations from XAI tools can play an important role in guiding policy decisions by translating complex predictions into actionable insights. XAI outputs also enable scenario testing and impact simulations, equipping planners with the tools to evaluate potential interventions before implementation [22]. However, despite these advantages, XAI is still relatively new in the field of ITS, and its practical implementation remains limited. Thus, further research, tool development are needed to fully realize its potential in real-world ITS environments. The research questions and corresponding insights derived from the literature are summarized below:

2.6 Discussion

RQ 1: What is Traffic Prediction and its key objectives?

Yin et al. have categorized traffic prediction into speed, flow, congestion levels, and travel time using historical and real-time data. [2] Its primary objectives include improving traffic management, mitigating congestion, optimizing resource allocation, and enabling proactive control measures for better road safety and efficiency. The key objectives of traffic prediction is to improve traffic management and forecasting which highlighted in Zheng et. al. paper. [17]. It also show in Shin et. al. paper where it another main objective is to use real time data to anticipate road conditions. These papers clarify that the key objectives include real-time forecasting, reducing congestion, improving traffic flow and enabling control of traffic.

RQ 1.1: How can traffic planning policies be guided by traffic prediction results? Traffic prediction results offer actionable insights that inform traffic planning policies by identifying congestion patterns, peak usage periods, and infrastructure bottlenecks. These forecasts help governments and planners design adaptive strategies such as dynamic lane reversal, emergency routing, and better road layout. The practical application of such strategies is evident in several studies.

Mystakidis et al. [13] demonstrated how traffic prediction can support traffic policies, particularly in public transport scheduling and route planning. Similarly, Shahid [15] discussed the Singapore EMAS system, which leverages traffic prediction to implement dynamic congestion pricing, while Barcelona utilizes predictive data to optimize traffic signal timing. Chen et al. [16] proposed dynamic lane reversal policies on the Ayer Rajah Expressway, based on predictive traffic control strategies. These examples highlight how traffic prediction can benefit policymakers in designing effective traffic management policies. Other studies, including [18], [6], and [20], also explore the role of traffic prediction in informing policy decisions.

RQ 2: What are the advantages and limitations of Traffic Prediction Systems in traffic planning policy guidance?

Traffic prediction systems offer several advantages that enhance traffic planning policy. For example, Mystakidis et al. [13] showed how predictive models enable dynamic public transport scheduling, allowing authorities to adjust bus frequency and routes based on real-time demand forecasts. Shahid highlighted the Singapore EMAS system, which uses congestion predictions for dynamic toll pricing, while Barcelona optimizes traffic signal timing using forecasted flow data. [15] Chen et al. demonstrated that predictive models support dynamic lane reversal policies, significantly reducing congestion during peak hours. [16] These studies illustrate how traffic prediction can improve operational efficiency, emergency response, and urban mobility planning.

However, there is also limitations from current research. Zheng et al. pointed out that traditional models like ST-SR lack the capacity to capture non-linear traffic patterns and offer no explainability. [17] Similarly, Zhang et al. faced issues with underfitting caused by hardware constraints and lacked interpretability. [19] Shin et al. used LSTM models that performed well but provided little transparency for policy decision-making. [4] Moreover, Waqas et al. achieved high accuracy but showed signs of overfitting, and their model's generalizability was not validated on unseen data. [20] These limitations suggest that while traffic prediction models are powerful, their trustworthiness and policy applicability depend on transparency, scalability, and robust validation.

3 Research Method

This study aims to explore the role of Explainable AI in guiding traffic policy decisions by analyzing how model explainability can inform policy makers. To achieve this, three research methods were considered: Systematic Literature Review, Case Study Analysis, and Experimental modeling. Each method was evaluated based on its alignment with the research objectives, feasibility, and the level of insight it could provide into the integration of XAI in ITS.

A Systematic Literature Review involves a structured and comprehensive analysis of existing peer-reviewed studies. This method is effective for identifying common themes, research gaps, and established frameworks across different applications of XAI in traffic prediction. Saranya et. al. provides a comprehensive SLR of XAI approaches across various applications, including transportation. [23] It analyzes 91 recently published articles, categorizing them based on XAI techniques used, application domains, and evaluation methods. The review identifies trends, challenges, and future research directions in the field of XAI. SLR often offers depth understanding of research topic, support theoretical grounding. However, it often do not involve original data or testing, and cannot assess real-time applicability. While valuable at the theoretical level, it is insufficient alone for evaluating and making policy impacts. Although SLR is not directly applied in this paper, it proved useful during the writing of the literature review section. The structured approach helped streamline the workflow for identifying relevant studies, critically evaluating them, and addressing RQ1 and RQ2 by highlighting the objectives, strengths, and limitations of existing traffic prediction systems.

Case study analysis involves in-depth investigation of real-world XAI applications in traffic systems. It typically includes qualitative data like stakeholder interviews or system design documentation. For instances, this paper uses case study analysis to deploy urban traffic restriction policies in Xi'an, China [24]. The process begin with identifying stakeholders and training them in basic concepts to effectively contribute in the modeling process. Through a series of interviews and workshops, the group understood traffic behavior and policy impacts. These models are then used to simulate multiple policy scenarios, including public transport expansion, vehicle restrictions, and license plate-based driving bans. The simulation results are presented back to stakeholders for discussion.

Although case study analysis provides rich contextual detail and is valuable for real-world system deployment, it is not suitable for the focus of this study, which is to developed a model and implement with Explainable AI techniques. The paper it aims to analyze existing models and interpretability tools across various studies to derive deeper insights into how traffic prediction and XAI can inform traffic policy. Therefore, a case study approach would not align with the goal of synthesizing generalizable technical findings or evaluating model performance metrics in a measurable way.

Experimental modeling involves developing or adapting a predictive traffic model integrated with XAI tools, then analyzing the interpretability outputs to simulate how they could influence policy decisions. For instances, this paper designed a model that can anticipate traffic accidents by learning dashcam video data. The model integrates a post-hoc attention mechanism to generate visual explaination of its predictions, which are then compared to human attention maps for validation [25]. This study demonstrates the application of experimental modeling in developing and validating XAI methods for traffic safety. It generates original result, and best aligned with this research objective. However, it requires advanced technical knowledge, and the results quality is highly depend on dataset quality.

Given the goal of this paper is explainability in traffic planning for policy changes, experimental modeling was selected as the primary research method. This approach allows for a controlled environment where different model explanations can be analyzed and compared, and their implications for policy can be evaluated systematically. Moreover, it offers the flexibility to replicate scenarios and test how various policy decisions might be influenced by model-derived insights.

4 Implementation and Result

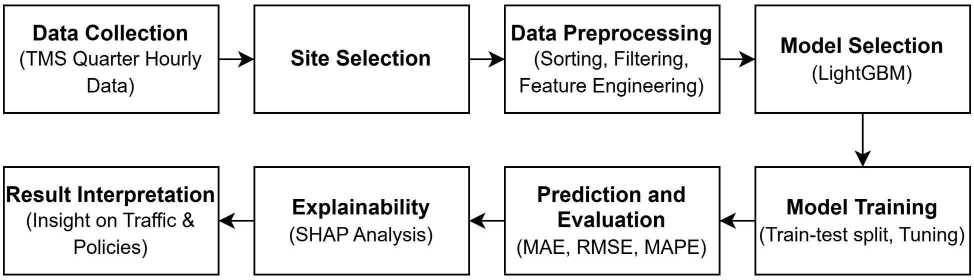


Figure 3: Workflow illustrating the implementation process of the traffic prediction study

Figure 3 outlines the overall workflow of the traffic prediction study. It summarizes the key stages, from data collection and site selection to preprocessing, model training with LightGBM, evaluation, and explainability using SHAP. Each component of the workflow is discussed in detail below.

4.1 Data Collection

To predict traffic, data was collected from the New Zealand Transport Agency's TMS Traffic Quarter-Hourly dataset, covering the period from October 2020 to January 2022. The dataset includes columns such as Start Date, Site Alias, Region Name, Site Reference, Class Weight, Site Description, Lane Number, Flow Direction, and Traffic Count. Certain columns such as Site Alias, Region Name, Site Description, and Flow Direction—were dropped as they were irrelevant to this study and to reduce dimensionality. Although the dataset spans over a year, only two months of data were used to minimize training time. At this stage, the goal is to develop an initial prototype as a proof of concept.

Site	Site Description
01N29424	Auckland Harbour Bridge - Classifier Site No 1 - SB
01639008	SH16 Southwestern On Ramp WB (Link)
01N39422	SH1 Northern Busway On Ramp SB
02020017	SH20 Hillsborough Rd Off Ramp to Dominion Rd On Ramp SB
02020002	SH20 Lambie Dr Off Ramp to Cavendish Dr On Ramp SB

Table 2: Site Table

4.2 Site Selection

Table 2 shows the selected sites considered for this study. In total, five highways are analyzed in this research: Auckland Harbour Bridge, SH16 Southwestern On-Ramp (Westbound), SH1 Northern Busway On-Ramp (Southbound), SH20 Hillsborough Road Off-Ramp to Dominion Road On-Ramp (Southbound), and SH20 Lambie Drive Off-Ramp to Cavendish Drive On-Ramp (Southbound). These five highways were chosen because they are among the most heavily trafficked in Auckland. Furthermore, various traffic reduction methods, such as dynamic lanes, bus lanes, and ramp signals have been implemented at these sites, making them suitable for evaluating the explainability aspect of traffic prediction models.

4.3 Data Preprocessing

Data preprocessing is an important step in any model training pipeline. Since the model is sensitive to time series data, the dataset is first sorted chronologically and then filtered to include only the five selected sites. Additional features are added to enhance model training, including: average traffic in the past hour, minimum past hour traffic, hourly volatility, 30-minute lag, hour of the day, 15-minute lag, previous day same time traffic, one-hour trend, maximum past hour traffic, and day of the week. These features are also used in the explainability analysis to assess their relevance in traffic prediction. Finally, the dataset is split into 70% for training and 30% for testing.

4.4 Model Selection

The chosen model for this study is LightGBM, a machine learning algorithm that builds decision trees using a technique called gradient boosting. Key features of LightGBM include high efficiency and speed, native support for missing values, built-in categorical feature handling, and no requirement for feature scaling. This model is well-suited to the study as it handles tabular data effectively, captures non-linear relationships (which are typical in traffic patterns), and supports categorical and missing features. These capabilities make it ideal for developing an initial prototype as a proof of concept due to its fast training performance.

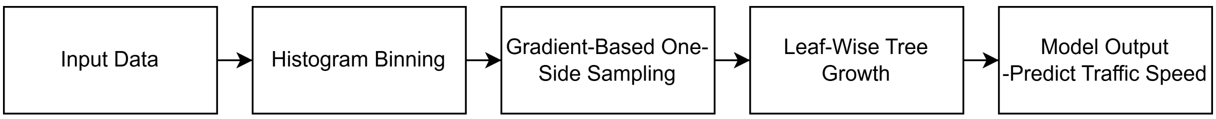


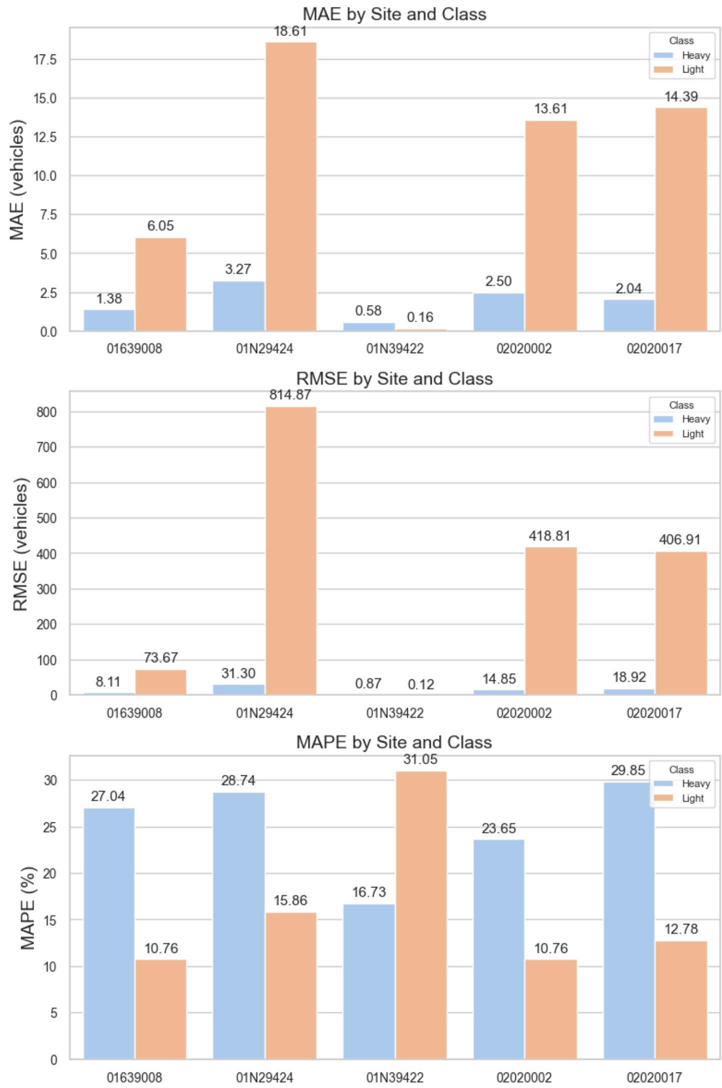
Figure 4: Workflow of LightBGM

LightGBM is built on the principle of gradient boosting, where models are trained sequentially, and each new model attempts to correct the errors made by the previous one. This is achieved by minimizing a loss function through the addition of weak learners. Unlike traditional level-wise tree growth algorithms, LightGBM uses a leaf-wise tree growth strategy. It selects the leaf with the highest loss reduction and splits it first, resulting in deeper trees and improved accuracy.

Based on figure 4, the model begins by converting continuous variables into bins to reduce memory usage and speed up computation. It then applies Gradient-based One-Side Sampling (GOSS) to prioritize high-impact data points while randomly sampling the rest. After training, the model outputs traffic predictions.

4.5 Prediction and Evaluation

For the evaluation metrics, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) were chosen to assess model performance. MAE measures the average magnitude of errors in a set of predictions, providing a straightforward interpretation in the same units as the target variable. RMSE penalizes larger errors more heavily, making it sensitive to outliers and useful for capturing overall prediction accuracy. MAPE expresses the error as a percentage, offering interpretability across different scales. The model was trained separately for each site and each vehicle class (heavy and light) because training on the entire dataset produced poor results. By segmenting the model based on site and class weight, the training error was significantly reduced and performance improved.



Site	Site Description
01639008	SH16 Southwestern On Ramp WB (Link)
01N29424	Auckland Harbour Bridge - Classifier Site No 1 - SB
01N39422	SH1 Northern Busway On Ramp SB
02020002	SH20 Lambie Dr Off Ramp to Cavendish Dr On Ramp SB
02020017	SH20 Hillsborough Rd Off Ramp to Dominion Rd On Ramp SB

Figure 5: Evaluation of MAE, RMSE, and MAPE Across Sites and Vehicle Classes

As shown in figure 5, MAE, RMSE, and MAPE highlighted significant variation in model performance across different sites and vehicle classes. In general, predictions for heavy vehicles yield lower errors compared to light vehicles, particularly for RMSE and MAE. The best performing model overall is observed at the SH1 Northern Busway On Ramp SB (01N39422) for the heavy class, with a MAE of 0.58, RMSE of 0.87, and a moderate MAPE of 16.73%, indicating highly accurate predictions. In contrast, the worst performing model appears at the Auckland Harbour Bridge (01N29424) for the light class, where the RMSE reaches 814.87 and MAE is 18.61, suggesting large deviations between predicted and actual values. These discrepancies reflect the diverse traffic dynamics across sites and vehicle classes, emphasizing the importance of model segmentation. The results confirm that training models separately by site and class weight significantly improves prediction accuracy and helps reveal localized traffic behavior patterns on Auckland's major highways.

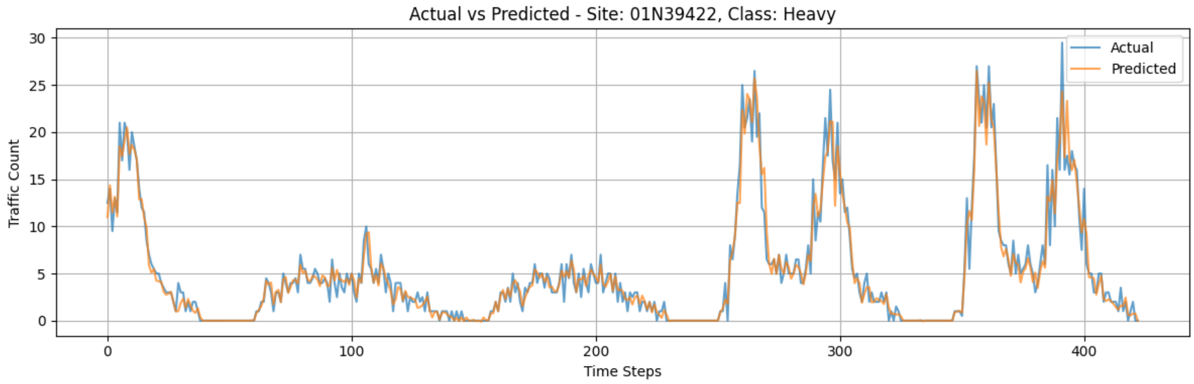


Figure 6: Actual vs predicted traffic counts for site 01N39422 in the heavy vehicle class.

Figure 6 presents a time series comparison between the actual and predicted traffic counts for the heavy vehicle class at site 01N39422 (SH1 Northern Busway On Ramp SB). The x-axis represents sequential time steps, while the y-axis indicates the number of vehicles detected in each time interval. From the visual trend, the predicted values (orange line) closely follow the actual observed values (blue line), including during both low and high traffic periods. In particular, the model captures key fluctuations and peak patterns with minimal lag or deviation. The alignment of both lines indicates that the model effectively generalizes the traffic flow dynamics for this site and class.

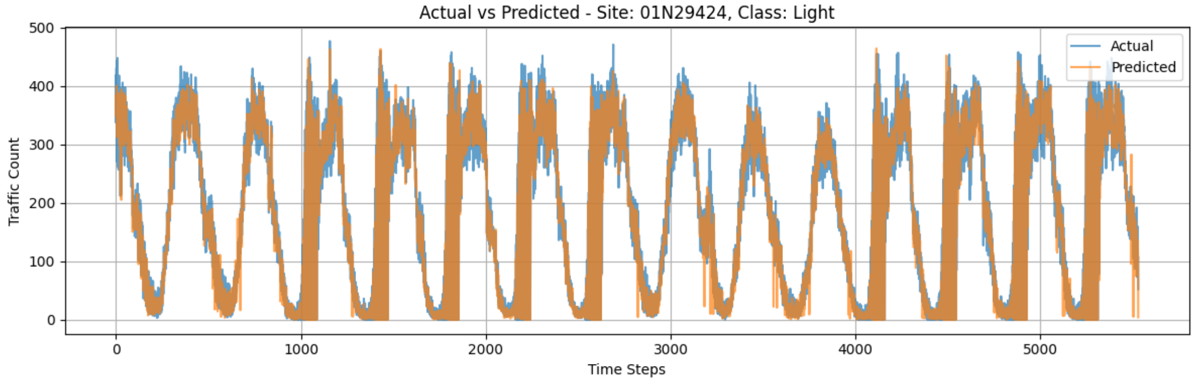


Figure 7: Actual vs predicted traffic counts for site 01N29424 in the light vehicle class.

Figure 7 illustrates the comparison between actual and predicted traffic counts for the light vehicle class at site 01N29424 (Auckland Harbour Bridge). The time series reveals a clear wave pattern, likely reflecting daily or peak-hour traffic behavior. While the predicted values (orange line) generally follow the overall trend and periodicity of the actual counts (blue line), there are several regions with notable deviation, particularly during peak times. The model tends to underestimate or overshoot during high-volume periods. This visual observation aligns with the quantitative evaluation, where this site and class combination recorded the highest RMSE (814.87) and MAE (18.61) in the study. The results suggest that the model struggles to capture the high variability and rapid spikes in traffic volume at this busy site, making it the worst-performing configuration overall. Potential reasons could include irregular patterns, extreme values, or insufficient feature representation for the light vehicle class at this location.

4.6 Explainability Analysis

For explainability, SHAP (SHapley Additive exPlanations) is used to interpret the model's predictions. SHAP assigns each feature an importance value for a particular prediction by computing the contribution of each feature towards the output, based on cooperative game theory. This allows for a clear understanding of how each input feature influences the model's predictions, making it possible to evaluate the effectiveness of different traffic policies and gain insights into the key factors affecting traffic flow.

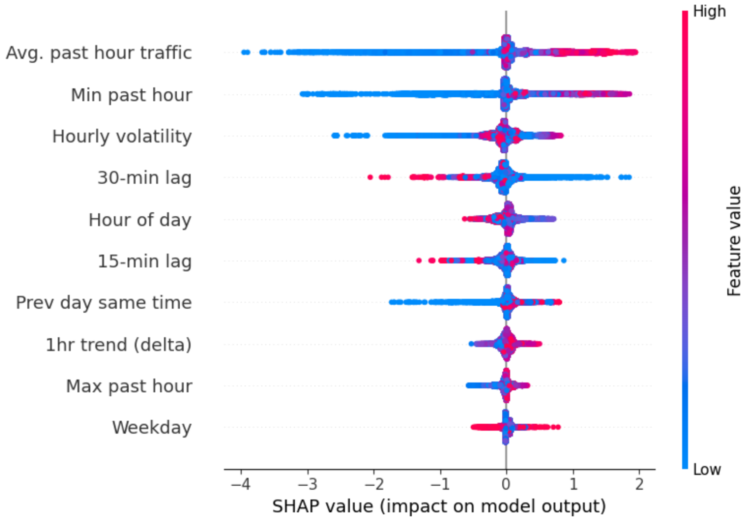


Figure 8: SHAP summary plot showing the impact of each feature on the model's output

The SHAP summary plot in figure 8 visualizes the contribution of each feature to the model predictions. Each row represents a feature, and each dot corresponds to a single prediction instance. The horizontal axis shows the SHAP value, which quantifies how much a feature pushes the prediction higher (positive values) or lower (negative values). Features are ranked by importance, with those having the greatest average impact appearing at the top. The color of each dot reflects the actual feature value, red for high values and blue for low values, allowing the reader to see whether high or low values of a feature increase or decrease the prediction. As the plot shows the three features that contribute the most for the model are average past hour traffic, minimum past hour traffic and hourly volatility.

The average past hour traffic represents the mean vehicle count over the previous 60 minutes, providing a smoothed view of recent traffic conditions. This feature is important for identifying traffic momentum, filtering out short-term fluctuations, and helping the model understand general flow trends. From a policy perspective, it can be used to detect ongoing congestion and inform proactive traffic management strategies such as adaptive signal control, rerouting, or ramp metering.

The minimum past hour traffic captures the lowest vehicle count in the last 60 minutes. This feature is particularly useful for identifying sudden traffic drops, which may result from low demand or unexpected incidents. By detecting these interruptions, the model can support policy decisions related to incident detection and flexible lane allocation during off-peak or disrupted periods.

The hourly volatility measures the change in traffic volume compared to one hour earlier, indicating whether traffic is increasing or decreasing. This trend is essential for capturing the acceleration or deceleration of traffic flow, offering insight into emerging congestion or clearing conditions. In practice, this feature can support real-time traffic interventions, such as triggering early warnings or updating dynamic message signs to reflect upcoming traffic buildup or relief.

This plot helps interpret not only which features matter most but also how their values influence the model's output, making it a powerful tool for understanding model behavior and supporting data-driven traffic policy decisions.

4.7 New Zealand Traffic Policies

To support the interpretation of the SHAP plot, several implemented traffic policies in New Zealand can be discussed in relation to the model feature importance. New Zealand has implemented several traffic management strategies, particularly in Auckland, to mitigate congestion and improve traffic flow. Several policies is discussed below.

Dyanmic lane control that used on the Auckland Harbour Bridge, the center lane is reversed based on peak hour directioanlity using real time traffic data.

Ramp metering is a traffic signals that installed on almost all Auckland motorway on-ramps to regulate the rate of vehicle entry to prevent downstream congestion.

High-Occupancy Vehicle (HOV) and bus lanes is deployed on several motorways such as Northern Busway and SH20 to promote public transport and carpooling. It also improves bus efficiency.

Dynamic signage system is used to informs drivers of estimated travel times for key destination and traffic conditions.

Event-Based Management where temporary restriction and diversion applied during major event such as concert, sports event, road construction.

Theses polices rely on traffic sensor networks, loop detectors and real time feeds to operate effectively.

4.8 Result Interpretation

Dynamic Lane Control, as seen on the Auckland Harbour Bridge, is dependent on understanding real-time traffic directionality. The SHAP results highlight features like hour of day, average past hour traffic, and hourly volatility as influential. From a policy making perspective, these features suggest that this systems could be further enhanced using real-time data from in-road loop detectors and CCTV-based vehicle counting. A detection frequency of 5–15 minutes would allow for proactive lane configuration before congestion builds up.

Ramp Metering is tied to short-term traffic fluctuations. SHAP features such as 15-minute lag, 30-minute lag, and 1-hour trend can be integrated into metering algorithms. By combining data from roadside sensors and GPS traffic feeds, ramp meters can dynamically adjust signal timings in real time. This would help smooth vehicle entry and reduce bottlenecks, especially during peak morning and evening hours.

The previous day same time feature helps identify recurring traffic patterns, reinforcing the relevance of event-based traffic management plans. These include the deployment of high-occupancy vehicle (HOV) lanes and bus lanes, which are common on highway such as the Northern Busway and SH20, aiming to prioritize public transport and reduce single-occupancy vehicle usage during peak hours. The model ability to forecast congestion could also inform dynamic lane assignments, such as opening bus, priority lanes during high congestion periods, and closing them during low traffic hours to maximize efficiency.

Additionally, several highways in New Zealand feature dynamic signage that displays estimated travel times to key destinations. This ties to the hourly volatility feature from the SHAP plot, which captures fluctuations in traffic flow and helps forecast delays. Such real-time feedback mechanisms could be further enhanced by predictive traffic models to support smarter routing and reduce network-wide congestion.

4.9 Performance Evaluation with Top 3 SHAP Features

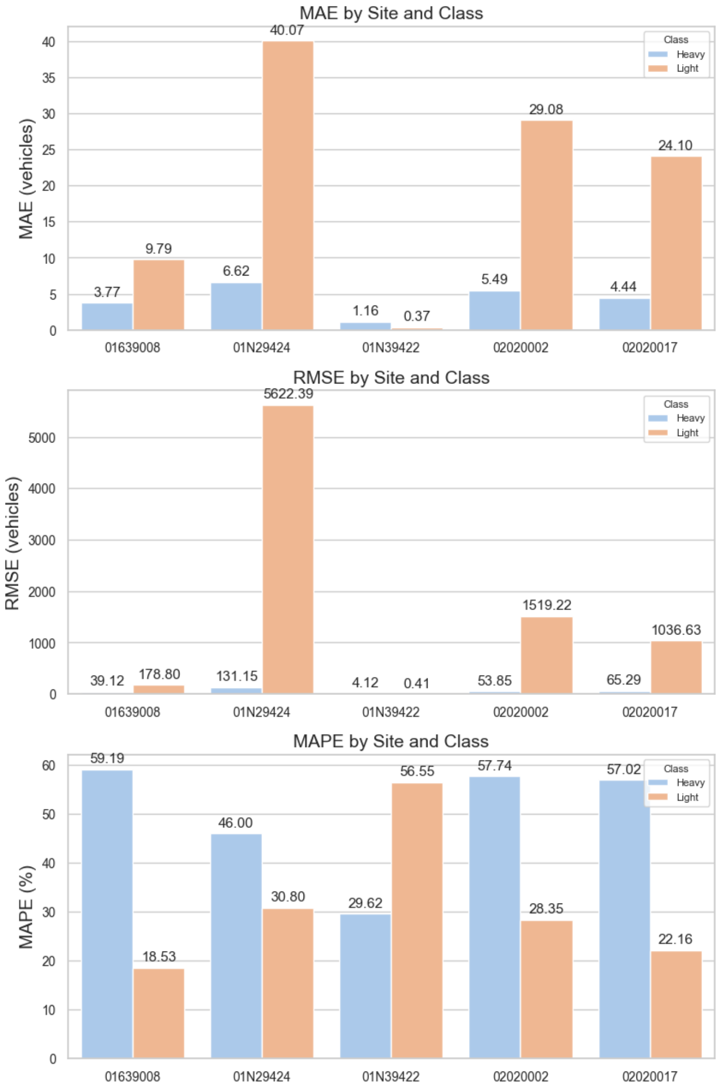


Figure 9: Model result based on top 3 SHAP features

To evaluate the prediction model of only using the top SHAP-ranked features, a simplified LightGBM model was trained using only the top 3 inputs. However, performance across all three metrics, MAE, RMSE, and MAPE is bad significantly as shown in figure 9. This demonstrates that although these features are globally important, they are insufficient for accurate predictions when using alone. By limiting the model to only 3 features, it strip away complementary information that help reduce prediction noise. Traffic prediction is a complex equation and requires a diverse set of features to capture site specific, temporal, and contextual patterns. The poor results confirm that SHAP values should only guide model interpretation rather than using as a feature selection method.

5 Conclusion

This paper successfully addressed the research questions outlined earlier. Through a combination of literature review and experimental modeling, it demonstrated not only the fundamentals of traffic prediction but also how Explainable AI can be integrated into predictive models. The primary objective was not only to forecast traffic flow but also to make the model decision-making process interpretable and transparent. SHAP analysis provided valuable insights into the most influential features driving model predictions, such as hour of day, average past hour traffic, and hourly volatility. These insights align closely with existing traffic policies in New Zealand, such as ramp metering, dynamic lane control, and adaptive signal systems.

5.1 Linking Research Questions to Results

RQ 3: How can Explainable AI techniques be used for traffic prediction models?

This was addressed through the integration of SHAP with the LightGBM model. SHAP provided both global and local interpretability, offering a clear understanding of the behavior of the model.

Globally, SHAP identified the most influential features contributing to congestion predictions, such as hour of day, average past hour traffic, and hourly volatility. Locally, it explained each individual predictions by quantifying how each feature pushed the model output higher or lower. This dual layer interpretability helped understanding black-box AI model such as LightGBM model and demonstrated that XAI technique can make complex machine learning model to be more transparent and trustworthy.

RQ 4: How can Explainable AI results guide traffic policies for improving traffic congestion?

This question was explored by linking the SHAP plot most important feature to reald world traffic policies that currently implemented in New Zealand.

For example, Dynamic Lane Control, that deployed on the Auckland Harbour Bridge, depends on understanding traffic directionality in real time. SHAP identified features like hour of day, average past hour traffic, and hourly volatility reflect precisely the types of conditions that inform lane reversal decisions. Similarly, Ramp Metering policies are tied to short-term traffic fluctuations. SHAP features such as 15-minute lag, 30-minute lag, and 1-hour traffic trends capture the types of dynamic congestion build-ups that metering systems aim to prevent.

Overall, these examples illustrate how XAI techniques such as SHAP can inform traffic policy design by translating model outputs into actionable insights. This provides a foundation for decision making in real time traffic management systems.

5.2 Limitations

This research does have certain limitations. Only a subset of available traffic monitoring sites was analyzed, and the dataset was limited to two months to allow for timely prototype development. This constraint affects the generalizability of the findings. Additionally, the model was trained separately for each site and vehicle class, without capturing the spatial or cross-site dependencies that could be crucial for city-wide traffic optimization. Besides, external features like weather, events, and incidents were not included, and no real-world system validation or stakeholder feedback was involved. Moreover, while SHAP helped interpret the model, it was also shown that relying solely on the top SHAP-ranked features for prediction resulted in poor performance

5.3 Future Work

To address the above limitations, future work should expand the dataset to include a broader range of traffic sites and extend the time period of analysis to improve the generalizability of the findings.

To overcome the current model site isolated training, future models could integrate spatial and temporal dependencies using graph-based models or spatio temporaldeep learning frameworks such as Graph Convolutional Network (GCN). This would enable more holistic and coordinated traffic prediction across a city-wide network.

While SHAP provided useful insights, additional XAI techniques such as LIME, or feature interaction plots could be used to validate SHAP results and provide multiple perspectives on interpretability.

In terms of deployment, large language models could be explored to automatically translate SHAP based insights into human readable policy suggestions. This would bridge the gap between technical outputs and non-technical decision makers.

Lastly, to validate the model in real world scenarios, the pipeline could be integrated with traffic simulation environments such as SUMO. This would allow testing policy interventions based on model predictions and explainability outputs, thus strengthening the case for live deployment in intelligent transport systems.

References

[1] R. A. Khalil, Z. Safelnasr, N. Yemane, M. Kedir, A. Shafiqurrahman, and S. N, "Advanced learning technologies for intelligent transportation systems: Prospects and challenges," *IEEE Open Journal of Vehicular Technology*, vol. 5, pp. 397–427, 2024, DOI: https://doi.org/10.1109/OJVT.2024.3369691.

[2] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi, and B. Yin, "Deep learning on traffic prediction: Methods, analysis, and future directions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 4927–4943, 2022, DOI: https://doi.org/10.1109/TITS.2021.3054840.

[3] C. S. Priya and F. S. Francis, "Optimizing traffic speed prediction using a multi-objective genetic algorithm-enhanced rnn for intelligent transportation systems," *IEEE Access*, vol. 13, pp. 35 688–35 706, 2025, DOI: https://doi.org/10.1109/ACCESS.2025.3544864.

[4] D. H. Shin, K. Chung, and R. C. Park, "Prediction of traffic congestion based on lstm through correction of missing temporal and spatial data," *IEEE Access*, vol. 8, pp. 150 784–150 796, 2020, DOI: https://doi.org/10.1109/ACCESS.2020.3016469.

[5] R. Zhou, H. Chen, H. Chen, E. Liu, and S. Jiang, "Research on traffic situation analysis for urban road network through spatiotemporal data mining: A case study of xi'an, china," *IEEE Access*, vol. 9, pp. 75 553–75 567, 2021, DOI: https://doi.org/10.1109/ACCESS.2021.3082188.

[6] L. Kong, H. Yang, W. Li, Y. Zhang, J. Guan, and S. Zhou, "Traffexplainer: A framework toward gnn-based interpretable traffic prediction," *IEEE Transactions on Artificial Intelligence*, vol. 6, no. 3, pp. 559–573, 2025, DOI: https://doi.org/10.1109/TAI.2024.3459857.

[7] D. Rowan, H. He, F. Hui, A. Yasir, and Q. Mohammed, "A systematic review of machine learning-based microscopic traffic flow models and simulations," *Communications in Transportation Research*, vol. 5, p. 100164, 2025, DOI: https://doi.org/10.1016/j.commtr.2025.100164.

[8] S. Yang, B. Yang, Z. Zeng, and Z. Kang, "Causal inference multi-agent reinforcement learning for traffic signal control," *Information Fusion*, vol. 94, pp. 243–256, 2023, DOI: https://doi.org/10.1016/j.inffus.2023.02.009.

[9] X. Qi, X. Qin, Z. Jia, M. Lin, and Y. Liu, "Ncae and elm based enhanced ensemble optimized model for traffic flow forecasting," *IEEE Access*, vol. 8, pp. 200 486–200 499, 2020, DOI: https://doi.org/10.1109/ACCESS.2020.3034763.

[10] A. Shah, A. Ghugharawala, M. Patel, V. Patel, N. Rathore, and R. R. Naik, "Impact of ai and ml on employee job satisfaction and performance," in *2024 4th International Conference on Sustainable Expert Systems (ICSES)*, 2024, Conference Proceedings, pp. 1183–1188, DOI: https://doi.org/10.1109/ICSES63445.2024.10763153.

[11] H. Chen and C. M. Mason, "Explainable ai (xai) for constructing a lexicon for classifying green energy jobs: A comparative analysis of occupation, industry, and location composition with traditional energy jobs," *IEEE Access*, vol. 12, pp. 142 709–142 720, 2024, DOI: https://doi.org/10.1109/ACCESS.2024.3430317.

[12] M. Grote, B. Waterson, and F. Rudolph, "The impact of strategic transport policies on future urban traffic management systems," *Transport Policy*, vol. 110, pp. 402–414, 2021, DOI: https://doi.org/10.1016/j.tranpol.2021.06.017.

[13] A. Mystakidis, P. Koukaras, and C. Tjortjis, “Advances in traffic congestion prediction: An overview of emerging techniques and methods,” 2025,” DOI: https://doi.org/10.3390/smartcities8010025.

[14] Y. Zhu, Q. Wu, and N. Xiao, “Research on highway traffic flow prediction model and decision-making method,” *Scientific Reports*, vol. 12, no. 1, p. 19919, 2022, DOI: https://doi.org/10.1038/s41598-022-24469-y.

[15] M. Shahid, “Real-time traffic management solutions through urban data platforms,” 2024.

[16] S. Chen, H. Wang, and Q. Meng, “An optimal dynamic lane reversal and traffic control strategy for autonomous vehicles,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3804–3815, 2022, DOI: https://doi.org/10.1109/TITS.2021.3074011.

[17] Z. Zheng, L. Shi, L. Sun, and J. Du, “Short-term traffic flow prediction based on sparse regression and spatio-temporal data fusion,” *IEEE Access*, vol. 8, pp. 142 111–142 119, 2020, DOI: https://doi.org/10.1109/ACCESS.2020.3013010.

[18] T. S. Gebre, L. Beni, E. T. Wasehun, and F. E. Dorbu, “Ai-integrated traffic information system: A synergistic approach of physics informed neural network and gpt-4 for traffic estimation and real-time assistance,” *IEEE Access*, vol. 12, pp. 65 869–65 882, 2024, DOI: https://doi.org/10.1109/ACCESS.2024.3399094.

[19] F. Zhang, T. Y. Wu, Y. Wang, R. Xiong, G. Ding, P. Mei, and L. Liu, “Application of quantum genetic optimization of lvq neural network in smart city traffic network prediction,” *IEEE Access*, vol. 8, pp. 104 555–104 564, 2020, DOI: https://doi.org/10.1109/ACCESS.2020.2999608.

[20] M. Waqas, S. Abbas, U. Farooq, M. A. Khan, M. Ahmad, and N. Mahmood, “Autonomous vehicles congestion model: A transparent lstm-based prediction model corporate with explainable artificial intelligence (eai),” *Egyptian Informatics Journal*, vol. 28, p. 100582, 2024, DOI: https://doi.org/10.1016/j.eij.2024.100582.

[21] A. Vijaya, S. Bhattarai, L. S. Angreani, and H. Wicaksono, “Enhancing transparency in public transportation delay predictions with shap and lime,” in *IEEE International Conference on Industrial Engineering and Engineering Management*, 2024, Conference Proceedings, pp. 1285–1289, export Date: 05 April 2025; Cited By: 0, DOI: https://doi.org/10.1109/IEEM62345.2024.10857000.

[22] S. Khan, T. M. Ghazal, T. Alyas, M. Waqas, M. A. Raza, O. Ali, M. A. Khan, and S. Abbas, “Towards transparent traffic solutions: Reinforcement learning and explainable ai for traffic congestion,” *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 1, 2025, DOI: https://doi.org/10.14569/ijacsa.2025.0160150.

[23] S. A and S. R, “A systematic review of explainable artificial intelligence models and applications: Recent developments and future trends,” *Decision Analytics Journal*, vol. 7, p. 100230, 2023, DOI: https://doi.org/10.1016/j.dajour.2023.100230.

[24] J. Shi, X. Guo, and X. Hu, “Engaging stakeholders in urban traffic restriction policy assessment using system dynamics: The case study of xi’an city, china,” *Sustainability*, vol. 11, p. 3930, 2019, DOI: https://doi.org/10.3390/su11143930.

[25] M. M. Karim, Y. Li, and R. Qin, “Towards explainable artificial intelligence (xai) for early anticipation of traffic accidents,” *National Academy of Sciences*, 2022, DOI: https://doi.org/10.48550/arXiv.2108.00273.