Enhancing Traffic Congestion Forecasting with Explainable Deep Learning: A Framework Using LIME for Transparent Intelligent Transportation Systems

Ouhmidou Hajar, Nabou Abdellah, Elikram Moulay Ahmed
Department of Computer Science-Faculty of Science Semlalia, Cadi Ayyad University, Marrakech, Morocco

***Abstract*—Intelligent transportation systems aim to improve traffic management and road safety, manage traffic effectively, and reduce roadway system congestion. This optimally requires estimating future traffic congestion. Unfortunately, the most popular machine learning and deep learning techniques can be unsuitable for model development in this task due to interpretability challenges. This study attempts to provide a solution to this challenge by creating a tool that integrates Local Interpretable Model-agnostic Explanations (LIME) into any traffic congestion forecasting system. This tool is applied to the Metro Interstate Traffic Volume dataset, which contains samples of traffic and road system congestion along with temporal, weather, and contextual data. For global feature analysis, a Random Forest Regressor is used as a baseline model, while a neural network model is developed to predict the congestion of the traffic and road system. The neural network model achieved a congestion prediction with an R^2 score of 0.612, a mean squared error of 0.026, and a mean absolute error of 0.129. The LIME tool also provides temporal feature insights, which show that examples of weekday/holiday status reduce the sample congestion prediction for the example, while precipitation increases it. At a global level, hour of the day, day of the week, temperature, and month of the year are the dominant factors in congestion prediction. These findings illustrate the value of adding interpretability to predictive models of traffic congestion when using explainable artificial intelligence.**

***Keywords*—Explainable artificial intelligence; lime; traffic congestion forecasting; intelligent transportation systems; random forest**

I. Introduction

Accelerating urbanization and rapid population growth have presented problems for global transport infrastructure. Rapid increases in the number of cars have caused traffic congestion to be a day-to-day experience for most large cities, generating many effects ranging from economic loss of fuel to deterioration of air quality related to human well-being to lowered quality of life. To address those challenges, accurate and real-time traffic estimation was a primary objective of both researchers and transport agencies. These estimations are foundational to intelligent transport systems (ITS) for improving traffic, reducing travel time, and improving road safety [2]. Traditionally, historical and econometric models based on ARIMA and regression analyses have been dominant for traffic forecasting.

Although the output was extremely relevant for acquiring deep insight into traffic improvement, these modelling approaches, quite expectedly, do not lend themselves well to representing the complex non-linearities found in traffic because of heterogeneous and dynamically fluctuating factors such as climate and large events (festivals) and driver behaviour [3]. Recent advancements in deep learning (DL) have significantly transformed the traffic forecasting arena. The power of artificial neural models such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), and long short-term memory (LSTM) networks is demonstrated in their modelling of spatio-temporal dependencies and revealing patterns hidden in the raw traffic dataset [4]. Artificial neural models have consistently achieved state-of-the-art performance across an exceptionally wide range of applications that involve real-world processes [1]. However, along with the potential of improved accuracy brought by Deep Learning (DL) models comes the cost of non-interpretability; decisions made on top of complicated tower stacks of nonlinear transformations are opaque, and therefore DL models have also been termed "black box" models. In some applications, such as traffic management, where important decisions are made, the lack of interpretation can deter use cases because decision makers require both predictive validity and an explanation that they can trust. For example, traffic operators need to understand why a model predicted significant congestion at a certain time and location, validate the analysis, the reason for the analysis, and take appropriate action in time. To address this gap, the emerging field of Explainable Artificial Intelligence (XAI) has proposed different methods for explaining black-box models. Prominent methods such as Local Interpretable Model-agnostic Explanations (LIME) and Shapley Additive ExPlanations (SHAP) deliver post-hoc explanations for models in terms of how much the most relevant features contribute to single predictions [5],[6]. These methods provide meaningful information about model reasoning, supporting a better understanding and trust of the end user in the model's output. This revised manuscript addresses the methodological inconsistency identified during review by reframing the work as a model-agnostic explainability framework rather than as a purely deep-learning study. The predictive component may be implemented using either an ensemble model or a neural network, while LIME is used as a post-hoc explanation layer.

The main contributions are as follows:

- Evaluation framework: Propose a methodological framework for developing and evaluating interpretable traffic prediction models.
- Practical demonstration with LIME: Using a public dataset, apply LIME to explain the predictions of a congestion model and illustrate how local explanations can enhance understanding.
- Analysis of influential features: Identify and visualize both local (LIME-based) and global (feature importance-based) drivers of congestion predictions.
- Reproducible resources: Provide fully reproducible Python code to support researchers and practitioners in adapting and extending the approach.

The rest of this study is structured as follows. Section II presents related works on traffic prediction and interpretability. Section III explains the approach, starting from the description of the dataset preprocessing and predictive model, followed by LIME integration. Experimental results, visual analysis, and comparative studies are reported in Section IV, while the study is finally concluded in Section V, which describes future research directions.

II. Related Work

The field of traffic forecasting has undergone a significant paradigm shift over the last decade. Early methodologies primarily relied on classical statistical frameworks and econometric models, such as ARIMA and Kalman filters. While these approaches provided foundational insights into traffic patterns, they often struggled to account for the volatile, non-linear dependencies caused by heterogeneous factors like sudden weather shifts or large-scale urban events [7]. With the rise of "Big Data," researchers transitioned toward machine learning techniques, including Support Vector Machines (SVM) and Random Forests, to better capture these complexities. However, the most transformative leap occurred with the introduction of deep learning (DL). Architectures such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) [8] networks have demonstrated an unparalleled ability to model spatio-temporal dependencies, consistently achieving state-of-the-art performance across diverse real-world datasets. Despite these gains in predictive power, the "black-box" nature of deep neural networks remains a critical bottleneck. In the context of Intelligent Transportation Systems (ITS), raw accuracy is insufficient if the underlying logic is opaque. Traffic operators [9] require "trustworthy" systems that allow them to validate predictions and perform root-cause analysis before committing to high-stakes management decisions. To bridge this accountability gap, the emerging field of Explainable Artificial Intelligence (XAI) has introduced post-hoc interpretability tools [10]. Methods such as Local Interpretable Model-agnostic Explanations (LIME) have gained prominence for their ability to provide human-understandable justifications for individual model decisions without sacrificing the performance of the underlying complex model. This research builds upon this momentum, proposing a unified framework that integrates LIME with deep learning to provide the transparency required for modern, resilient urban mobility.

For traffic and mobility applications, explainability can reveal whether congestion forecasts are driven by time of day, weather, holiday status, or other contextual factors. Visual analytics approaches have also been developed to help domain experts understand congestion-influencing factors using explainable machine learning [15]. The present work contributes to this line of research by combining local LIME explanations with global feature importance in a transparent traffic-forecasting workflow.

III. Methodology

The predictive performance and interpretability of the traffic congestion prediction framework are emphasised. The system is built on four main components. The first is the dataset. The second is data preprocessing. The third is predictive modelling. The fourth is the use of LIME for interpretability. A reliable and reproducible approach to creating traffic prediction systems is provided by these components when used together.

A. Dataset

This study uses the Metro Interstate Traffic Volume dataset from the UCI Machine Learning Repository [12], which is publicly accessible. The data set contains hourly traffic volume counts for a section of the I-94 interstate highway in Minnesota, USA. It also includes context variable information to show factors affecting traffic flow.

Table I summarizes the main variables used from the Metro Interstate Traffic Volume dataset, including temporal, weather, contextual, and traffic-volume attributes.

TABLE I. The Main Dataset Variables

Variable	Description
Holiday	Indicator of public holiday status.
Temperature	Recorded in Kelvin.
Rain and Snow	Precipitation levels in millimetres
Clouds	Percentage cloud cover.
Date time	Hourly timestamp used to derive hour, day of week, month, and year.
Weather main and Weather description	General and detailed weather conditions.
Traffic volume	Target variable, representing the number of vehicles per hour.

The presence of meteorological, time, and contextual variables makes this dataset particularly well-suited for studying real-world traffic dynamics and prediction, which is a significant advantage.

B. Data Processing

The preprocessing pipeline was revised for clarity and reproducibility. First, the timestamp was converted to a datetime format and decomposed into hour, day of week, month, and year. Second, categorical variables such as holiday and weather conditions were encoded using one-hot encoding to avoid artificial ordinal relationships. Third, the target variable was separated from the explanatory variables. Fourth, the data were

divided into training and testing subsets using an 80/20 split with a fixed random seed. Finally, scaling was applied where required by the neural-network model and by the LIME prediction wrapper.

C. Prediction Model

To demonstrate interpretability, employ a Random Forest Regressor as the predictive model. Random Forests, an ensemble of decision trees, are robust against overfitting and capable of modelling complex nonlinear relationships. While not the most advanced option compared to deep neural networks, Random Forests serve as an effective and transparent example for integrating interpretability tools such as LIME. Importantly, because LIME is model-agnostic, the same interpretability process can be extended to deep learning architectures.

A neural network regressor is also evaluated on normalized congestion values. It is implemented as a feed-forward multilayer perception operating on the engineered tabular variables. The input layer corresponds to the number of preprocessed features, hidden dense layers learn non-linear interactions, and the output layer contains a single regression neuron. The network is trained using Mean Squared Error as the loss function and monitored through validation loss and Mean Absolute Error. This formulation is consistent with the reported training and validation curves and avoids presenting the Random Forest as a deep-learning model.

The two models are not used to claim state-of-the-art performance. Instead, they demonstrate that the same explanation framework can be connected to different predictive engines.

D. Deep Learning Network Architecture for Traffic Prediction

Deep learning has emerged as a transformative approach to traffic prediction. Models such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) are achieving state-of-the-art results. Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs), which are achieving state-of-the-art results. These architectures excel due to their ability to capture spatio-temporal dependencies on traffic data, i.e., patterns that evolve over time and space. Long Short-Term Memory (LSTM) networks are well-suited to time series forecasting. By retaining information over long sequences, LSTMs can model long-term dependencies in traffic flow, which is a capability that traditional methods often lack.

Fig. 1 presents the overall workflow adopted in this study, starting from traffic data collection and verification, followed by data curation, feature extraction, analytical processing, prediction, and validation. This framework highlights the complete pipeline used to transform raw traffic-related data into interpretable forecasting output.

A typical LSTM-based setup receives sequential traffic data and learns to pick up on repeated patterns and trends in the data, which are then used to predict future traffic. Unlike the old ways of coping with prohibitive amounts of feature engineering, these networks learn relevant representations from raw data automatically. Such an ability to unveil hidden patterns in large data renders deep neural networks excellent instruments for intelligent transport systems (ITS), where timely and precise traffic prediction is important to improve road safety, reduce congestion, and optimize urban mobility.

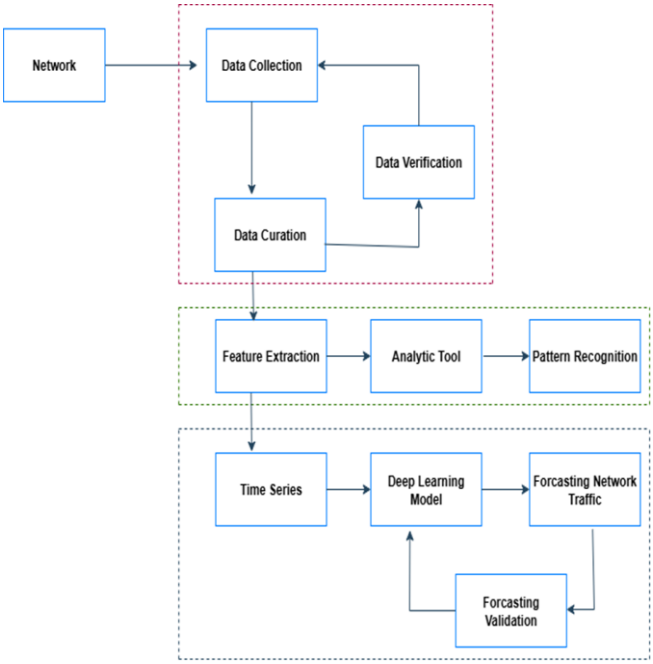

Fig. 1. End-to-end framework for traffic data collection, analysis, and prediction.

E. Lime Architecture

The primary idea behind LIME (Local Interpretable Model-agnostic Explanations) is to increase the transparency of complex models by creating interpretable, intuitive approximations for nearby individual predictions [11]. Unlike explaining the entire model globally, LIME addresses only a local neighborhood around a specific instance. The process begins by perturbing the input data point to create a set of nearby samples. The original model then makes predictions on the perturbed instances. From these predictions, LIME derives a sparse, interpretable model of ten times a linear regressor wherein the contribution of each sample is weighted by its proximity to the original instance.

F. Mathematical Problem Formulation

The primary objective of traffic congestion forecasting is to map a high-dimensional set of historical and contextual features to a continuous target variable representing future traffic volume. Formally, define the traffic state at a specific time interval t as $x_t \in R$ the task is to construct a predictive mapping function $f: x \rightarrow y$, where the input space $\mathcal{X}$ is composed of a multidimensional feature vector $\mathcal{V}_t$. This vector incorporates temporal attributes (hour, day, month), meteorological data (temperature, precipitation), and contextual indicators such as holiday status.

The forecasting model aims to minimize the discrepancy between the predicted volume $\hat{y}_t$ and the ground truth observation obtained from the Metro Interstate dataset. This is

achieved by optimizing a loss function, specifically the Mean Squared Error (MSE), defined as:

$$MSE = \frac{1}{n} \setminus \sum_{i=1} (y_i - \hat{y}i)^2$$

To ensure transparency, augment this predictive model with a local surrogate explanation model g . For a specific traffic instance x, the explanation $\xi(x)$ is derived by solving the following optimization problem:

$$\xi(x) = argmin_{g \in \mathbb{G}} \mathcal{L}(f, g, \pi x) + \Omega(g)$$

In this formulation, $L(f, g, \pi x)$ represents the fidelity loss, which measures how accurately the interpretable model g approximates the complex "black-box" model f within the local neighbourhood πx . The term $\Omega_{(g)}$denotes the complexity of the explanation, ensuring that the resulting insights remain human-interpretable. By minimizing this joint objective, the framework identifies the specific features such as heavy rain or peak commute hours that exert the most significant influence on a given congestion forecast [13].

The Fig. 2 Effect of LIME original prediction explained by fake data around the input instance fitted with a locally weighted interpretable model. While the example is shown for image prediction, applying this approach in the work to tabular traffic data can concisely explain individual congestion forecasts.

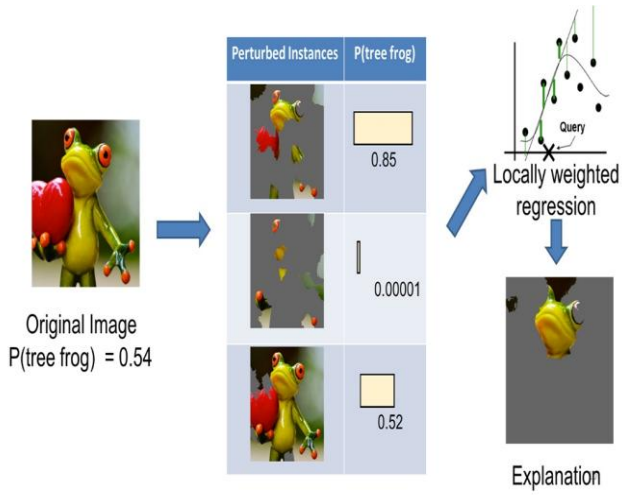

Fig. 2. Local Interpretable Model-Agnostic Explanations (LIME) applied to image predictions.

The coefficients of the surrogate model act as explanations, indicating which of the features contributed most to the prediction of the original model for that sample. In this way, LIME provides a localized, human-interpretable view of the reasoning of a "black box" model. Rather than giving an overall view, it illuminates decision-making in specific settings, giving practitioners valuable insight into why the model predicted the outcome in question.

G. Lime Integration

To address the "black box" nature of predictive models, integrate Local Interpretable Model-agnostic Explanations (LIME) into the framework. LIME provides local human-understandable explanations for individual predictions through the following procedure:

- Initialization: A LimeTabularExplainer instance is built using the training data, feature names, and the regression mode.
- Instance Selection: A representative instance from the test set is selected for explanation, allowing us to examine model reasoning for a specific traffic scenario.
- Prediction Function Wrapper: A custom predict fn raw function is defined. This wrapper takes raw feature inputs, scales them using the previously fitted StandardScaler, and then passes them to the trained DNN model for prediction.

This ensures that LIME interacts with the model in its expected scaled input space while generating explanations based on raw feature values.

IV. Explanation Generation

LIME perturbs the selected instance and evaluates the model's predictions on these variations. It then fits a simple interpretable model (e.g., a local linear model) to approximate the complex model's behaviour in the vicinity of the instance.

Through this process, LIME highlights the relative contribution of input features (e.g., weather, time of day, holidays) to a given prediction. These explanations make the model's reasoning more transparent, bridging the gap between predictive performance and interpretability in traffic management systems. The experimental evaluation highlights both the predictive performance of the model and the interpretability provided through LIME.

In addition to presenting standard performance metrics, include visual explanations that illustrate how different features contribute to traffic congestion predictions at both the local and global levels.

A. Model Evaluation

The revised evaluation separates raw-volume results from normalized-congestion results to avoid misleading comparisons. The Random Forest result is reported on raw traffic-volume units, whereas the neural-network result is reported on normalized congestion values. Therefore, the numerical values of the errors should not be compared directly unless the same target scale, preprocessing, and split are used.

TABLE II. Reported Model Performance

Model	Target Scale	Metrics reported	Interpretation
Random Forest Regressor	Raw traffic volume	MSE= 189,619.42	Baseline model and source of global feature importance.
Neural network regressor	Normalized congestion	R^2=0.612; MSE=0.026 MAE=0.129	Predictive model used for the reported training curves and local LIME explanations.

Table II presents the reported performance metrics for the Random Forest Regressor and the neural network regressor, while clarifying the target scale used for each model.

The neural network explains approximately 61.2% of the variance in normalized congestion values. Fig. 3 and 4 show that training and validation errors decrease over the training epochs, which indicates model convergence without a severe divergence between training and validation curves.

Fig. 3 shows the Training & Validation loss curves from model training. The gradual decrease in the curves indicates that the model learned the traffic pattern progressively. The training and the validation loss are rather similar (721 out of 884), indicating that the model converged well without any overfitting or very little evidence.

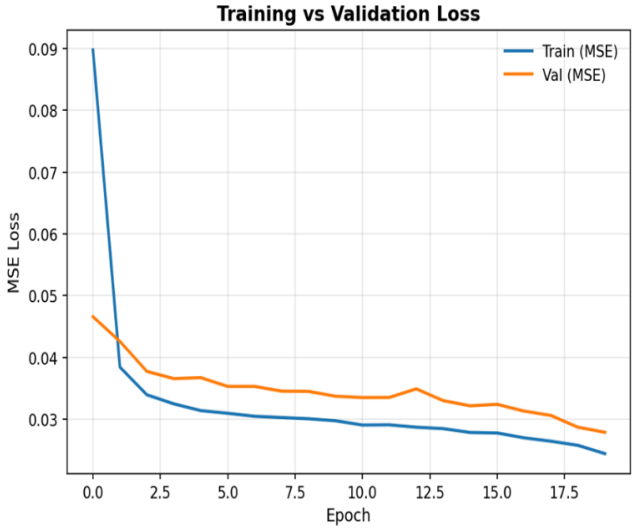

Fig. 3. Training and validation loss curves showing model convergence without overfitting.

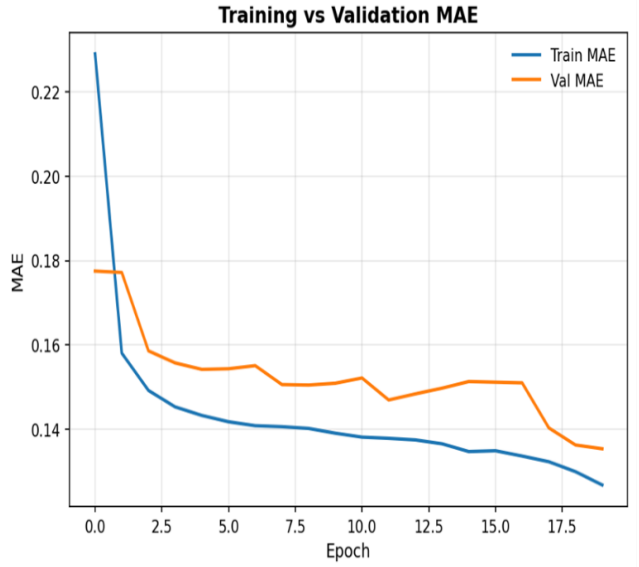

Fig. 4. Mean Absolute Error progression during training, demonstrating consistent improvement.

Fig. 4 shows the evolution of the Mean Absolute Error (MAE) during training and validation. The decreasing trend of both curves confirms that the prediction error was reduced over successive epochs. This behavior indicates that the model improved its forecasting capability as the training process progressed.

B. Prediction Accuracy Analysis

The predicted-versus-actual plot in Fig. 5 shows that predictions generally follow the diagonal reference direction, although dispersion remains visible. This confirms that the model captures a meaningful part of the congestion dynamics but also indicates that additional features, temporal lags, or more advanced architecture could improve prediction quality.

It's comparing the predicted congestion values to what was observed in the test set. The point distribution distributed around the diagonal reference line of the graph reveals that the model was able to reconstruct a general trend in traffic congestion. On the other hand, the dispersion of some points also indicates that due to those traffic conditions, there are corresponding prediction errors that require other interpretability tools to use, such as LIME.

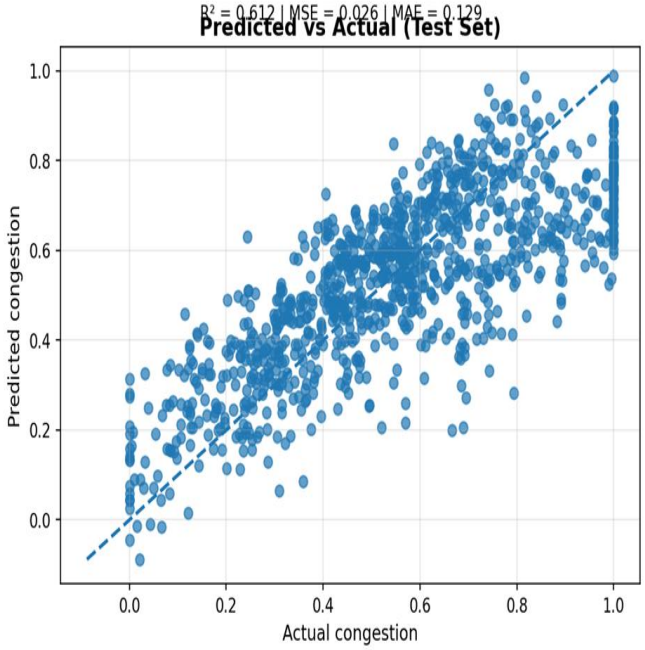

Fig. 5. Predicted vs. actual congestion values.

C. Residual Analysis

Residual diagnostics are presented in Fig. 6 and 7. The residual histogram is centered near zero, which suggests that the model does not show a strong average bias. The residuals versus predicted plot does not show a single linear systematic pattern, although the spread indicates that prediction errors vary across congestion levels. These diagnostics should be retained because they provide more information than aggregate metrics alone.

Fig. 6 shows the histogram of residuals (the gap between actual and predicted congestion values. Finding that the residuals mainly lie close to zero, indicating there is little global bias in the model. The nearly symmetric histogram also confirms that there are roughly as many overestimations as underestimations in the test set.

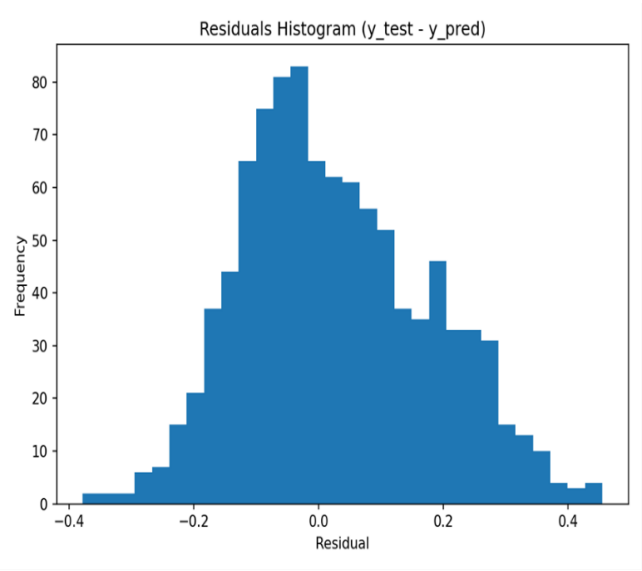

Fig. 6. Histogram of residuals showing approximately normal distribution centered at zero.

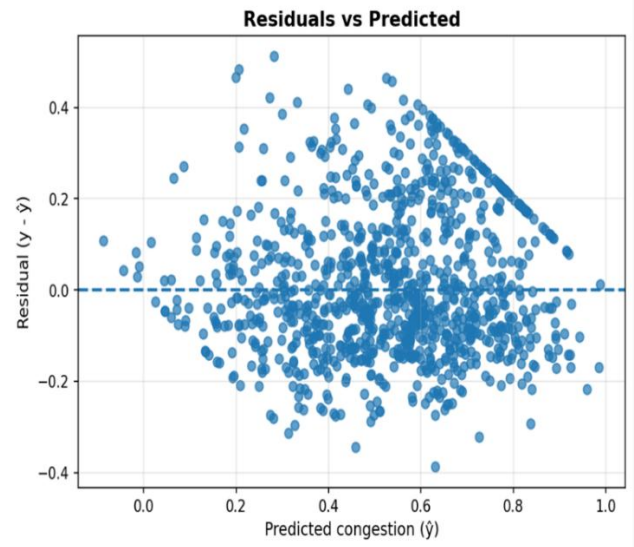

Fig. 7. Residuals vs. predicted values showing a homoscedastic pattern.

Fig. 7 shows the distribution of residuals with respect to the predicted congestion values. The absence of a clear systematic trend suggests that the model does not strongly overestimate or underestimate congestion within a specific prediction range. This supports the reliability of the prediction results, although some dispersion remains visible for medium and high predicted congestion values.

D. Local Explanations with LIME

To gain insight into individual predictions, LIME was applied to selected test instances. LIME produces bar plots that highlight the relative influence of features on a given prediction, with the bar's length and direction indicating the strength and sign of the contribution. Positive values increase the predicted traffic volume, while negative values decrease it.

LIME analysis provides crucial insights into feature importance for individual predictions:

Key Interpretability Findings:

- Day of Week (Dominant Factor): Shows the strongest negative contribution (-0.30), indicating weekends or specific days significantly reduce congestion prediction.
- Holiday Status: Secondary negative contributor (-0.15), confirming reduced traffic during holidays.
- Precipitation: Positive contribution (+0.08), suggesting increased congestion during rainy conditions.
- Temperature and Hour: Moderate influences on prediction, with temperature showing a slight negative impact and hour showing a minimal positive contribution.

Fig. 8 presents the local LIME explanation for the selected representative test instance. It shows which variables increase or decrease the predicted congestion value, thereby making the individual model prediction more transparent.

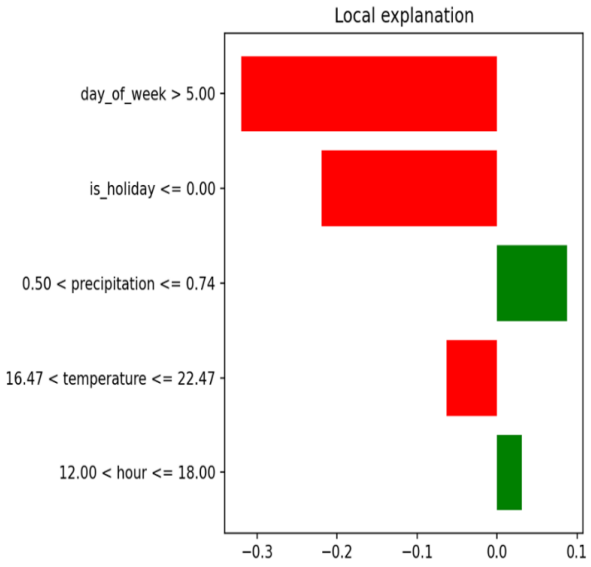

Fig. 8. LIME explanation for a representative test instance showing feature contributions.

Feature Interpretation;

- Red bars: Features decreasing congestion prediction
- Green bars: Features increasing congestion prediction
- Bar length: Magnitude of feature impact on this specific prediction

This LIME analysis demonstrates that temporal factors (day of week, holidays) are primary drivers of congestion prediction, while environmental factors (precipitation, temperature) provide secondary but meaningful contributions.

These local explanations make the model's decision-making process transparent, allowing practitioners to trace back a prediction to specific contributing factors.

Fig. 9 gives a more detailed local explanation for a single prediction with an estimated congestion value of 0.12. The result

confirms that LIME can identify the specific variables responsible for a given model output. In this instance, weekend-related temporal effects reduce the predicted congestion, whereas precipitation and afternoon timing slightly increase the forecasted value.

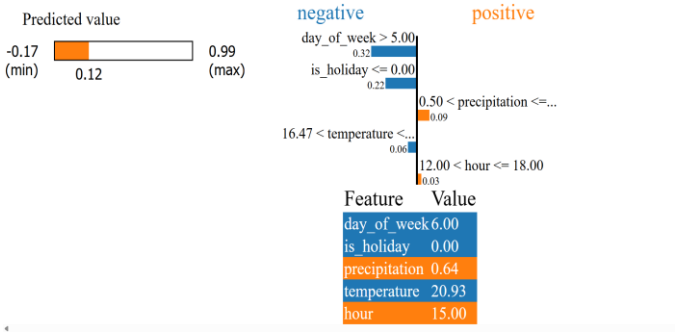

Fig. 9. LIME explanation of a single prediction (0.12), showing the local positive and negative contributions of temporal and weather-related features.

Analysis of the LIME explanation for a specific instance reveals a local prediction of 0.12, falling between a minimum value of -0.17 and a maximum value of 0.99. Factors with a significant negative impact on this prediction include the day of the week being a weekend (day 6.00, contributing negatively at 0.32) and the absence of a public holiday (isholiday 0.00, with a negative contribution of 0.22). A temperature above 16.47 (specifically 20.93) also has a slight negative effect (0.06). On the other hand, precipitation between 0.50 and a higher value (specifically 0.64) has a notable positive influence (0.09), as does the time of day between 12:00 p.m. and 6:00 p.m. (specifically 3:00 p.m.), with a positive contribution of 0.03. These results highlight the importance of weather conditions and timing in determining the model's prediction for this instance.

E. Global Feature Importance

Beyond individual predictions, it is also important to capture the global influence of features across the entire dataset. In Random Forests, feature importance is derived from the average reduction in variance contributed by each variable across all trees in the ensemble.

Fig. 10 presents the top 10 global feature importance rankings obtained from the Random Forest model. This global analysis complements the local LIME explanations by identifying the variables that most consistently influence congestion predictions across the dataset.

The top ten features identified included hour, temperature, and month, all of which consistently played a dominant role in shaping traffic patterns. This global feature importance ranking was visualized in a bar chart, showing which variables the model relies on most when generating predictions.

Together, the local explanations from LIME and the global feature importance analysis provide a complementary perspective. While LIME sheds light on why the model predicts a specific outcome in each context, feature importance summarizes the broader, dataset-wide drivers of congestion. This dual perspective not only improves transparency but also supports more informed decision-making in traffic management applications.

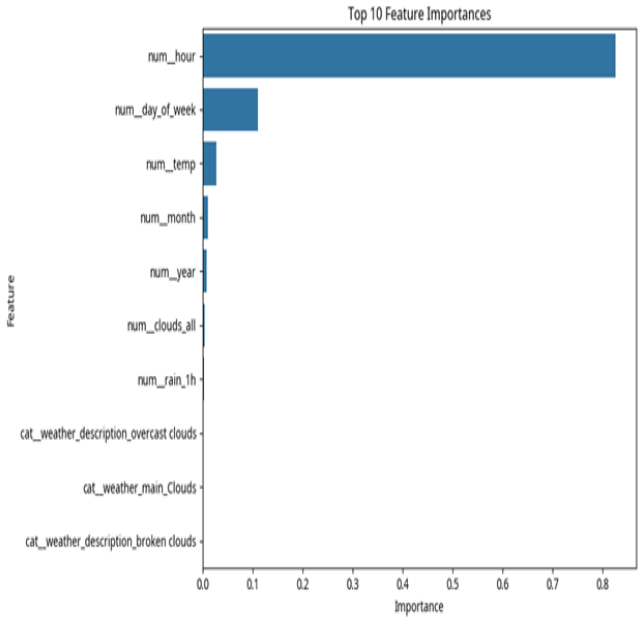

Fig. 10. Feature importance analysis for network traffic forecasting.

F. Positioning Against Related Work

Table III positions the results against representative studies. Direct numerical comparison is possible only when the same dataset, target scale, preprocessing pipeline, and temporal split are used.

TABLE III. COMPARISON WITH REPRESENTATIVE RELATED STUDIES

Study	Model or scope	Relation to this work
Lv et al. [7]	Deep learning for traffic-flow prediction using large-scale traffic data.	Supports the use of deep learning for non-linear traffic dynamics but uses a different dataset and setting.
Ma et al. [8]	LSTM model for traffic-speed prediction.	Shows the value of recurrent models for long-term temporal dependencies.
Polson and Sokolov [9]	Deep-learning architecture for short-term traffic-flow prediction.	Demonstrates the ability of deep models to capture traffic regime changes.
Li et al. [7]	DCRNN for graph-based traffic forecasting.	Addresses road-network spatio-temporal dependencies not modeled in the present single-station tabular setup.
Pranolo et al. [14]	LSTM and Bi-LSTM on the Metro Interstate dataset.	Provides a dataset-specific reference point; however, comparison requires identical preprocessing and splits.

G. Results and Discussion

The experimental results demonstrate that the deep learning framework effectively captures the complex dynamics of urban traffic, achieving an R^2 score of 0.612 and a Mean Squared Error (MSE) of 0.026. While these quantitative metrics confirm the model's predictive reliability, the integration of LIME provides a critical qualitative layer that moves beyond "black-box" forecasting. By decomposing individual predictions, identifying those temporal features, specifically the "Day of the Week," exerted the most significant influence on congestion levels, showing a strong negative contribution of -0.30. This suggests

that the model successfully learned the reduced commercial and commuter activity typical of weekends, allowing for more nuanced weekend-specific forecasts.

Furthermore, the model's sensitivity to environmental factors such as the positive contribution of precipitation (+0.08) indicates an accurate alignment with real-world traffic physics, where adverse weather typically reduces free-flow speeds and increases density. Unlike traditional global importance metrics, which provide a static view of the dataset, the local LIME explanations revealed that at specific peak hours, weather conditions can become the dominant driver of error if not properly accounted for. This dual-perspective analysis, combining global feature rankings with instance-specific explanations, equips traffic managers with the "reasoning" behind a forecast. Ultimately, this transparency transforms the AI from a mere prediction tool into a decision-support system that can be validated and trusted by human operators in real-time Intelligent Transportation Systems (ITS).

V. Conclusion and Future Work

This study successfully establishes a framework for integrating deep learning with Explainable AI (XAI) to resolve the "black-box" limitations of modern traffic forecasting. By developing a model that achieves a robust R^2 score of 0.612 demonstrates that high-fidelity predictive performance can coexist with technical transparency.

The integration of LIME provides a dual-layered interpretability that is essential for real-world deployment. On a local level, it empowers traffic managers to validate individual forecasts by identifying specific drivers such as the -0.30 impact of weekend temporal factors or the +0.08 influence of precipitation, allowing for targeted, data-driven interventions. Globally, the analysis confirms that variables such as the hour of the day and temperature remain the most consistent predictors of urban mobility patterns.

Ultimately, this research transitions AI from a mere forecasting tool into a trustworthy decision-support system. By providing the "reasoning" behind the numbers, this framework fosters the institutional trust required to implement safer, more efficient, and sustainable smart city infrastructures. Future work will focus on scaling these explanations to real-time dynamics and exploring the generalizability of this framework across more complex graph-based road networks.

Acknowledgment

The authors would like to thank the Faculty of Sciences Semlalia at Cadi Ayyad University for providing the research facilities and computational resources necessary to conduct this study. Also, thanks to Nabou Abdellah and Elikram Moulay Ahmed for their insightful feedback on the explainability framework and the application of LIME. Additionally, express gratitude to the anonymous reviewers for their constructive comments that improved the quality of this manuscript.

References

[1] M. Adnan and S. Islam, "A review on traffic prediction using machine learning and deep learning techniques," IEEE Access, 2019.

[2] Ribeiro, M. T. (n.d.). LIME - Local Interpretable Model-Agnostic Explanations. https://homes.cs.washington.edu/~marcotcr/blog/lime/

[3] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017 *

[4] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865–873, 2015, doi: 10.1109/TITS.2014.2345663.

[5] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," Transportation Research Part C: Emerging Technologies, vol. 54, pp. 187–197, 2015, doi: 10.1016/j.trc.2015.03.014.

[6] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," Advances in Neural Information Processing Systems, vol. 30, 2017.

[7] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 2, pp. 865–873, 2015.

[8] X. Ma, Z. Tao, Y. Wang, H. Yu, and Y. Wang, "Long short-term memory neural network for traffic speed prediction," Transportation Research Part C: Emerging Technologies, vol. 54, pp. 187–197, 2015.

[9] N. G. Polson and V. O. Sokolov, "Deep learning for traffic flow prediction," Transportation Research Part C: Emerging Technologies, vol. 79, pp. 1–17, 2017.

[10] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.

[11] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv:1702.08608, 2017.

[12] Hogue, J. (2019). Metro interstate traffic volume [Dataset]. In UC Irvine. https://doi.org/10.24432/c5x60b.

[13] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[14] A. Pranolo, S. Saifullah, A. B. U. Putra, R. Drzewski, and A. P. Wibawa, "Urban traffic volume prediction using LSTM and Bi-LSTM: Performance evaluation on the Metro Interstate dataset," ILKOM Jurnal Ilmiah, vol. 17, no. 3, pp. 227–240, 2025, doi: 10.33096/ilkom.v17i3.3001.227-240.

[15] X. Chen, J. Zhang, H. Wang, and X. Song, "TCEVis: Visual analytics of traffic congestion influencing factors based on explainable machine learning," Visual Informatics, vol. 8, no. 1, pp. 56–66, 2024, doi: 10.1016/j.visinf.2023.11.003.