Acta Infologica,
ACIN 2025, 9 (2): 611–629

https://doi.org/10.26650/acin.1627153
Submitted: 26.01.2025
Revision Requested: 13.07.2025
Last Revision Received: 21.11.2025
Accepted: 25.11.2025

İSTANBUL UNIVERSITY PRESS

Acta Infologica

Research Article

Open Access

Meteorological and Temporal Dynamics in Urban Traffic Density: A Comparative Evaluation of Machine Learning and Neural Network Models

Büşra Özdenizci Köse [1], Farid Huseynov [1], Vildan Gülpınar Demirci [2] & Nurullah Taş [3]

[1] Gebze Technical University, Faculty of Management, Department of Management Information Systems, Kocaeli, Türkiye
[2] Aksaray University, Faculty of Economics and Administrative Sciences, Department of Economics, Aksaray, Türkiye
[3] Gebze Technical University, Faculty of Management, Department of Management, Kocaeli, Türkiye

Abstract Understanding the interplay between meteorological and temporal factors and urban traffic density is vital for effective traffic management and sustainable urban planning. This study explores these dynamics using a dataset from Istanbul, integrating traffic metrics with weather variables as temperature, dew point, wind speed, and precipitation, alongside temporal indicators, such as time of day and weekday/weekend distinctions. A multi-model approach that combines traditional regression techniques, advanced ensemble models, and neural networks was applied. Ridge and Lasso Regression provided baseline comparisons, whereas Decision Tree, KNN Regression, and Random Forest captured nonlinear relationships. Advanced ensemble models, such as LightGBM and XGBoost, employ boosting techniques to enhance accuracy. A feedforward neural network complementing ensemble methods further analyzed intricate data patterns. Performance evaluation based on MSE, MAE, RMSE, and R^2 Scores highlighted the superiority of LightGBM and Random Forest, which achieved the highest accuracy. Feature importance analysis revealed traffic-specific metrics, such as average speed, as the most significant predictors, followed by meteorological variables, such as temperature and pressure. Temporal factors, including morning and working hours, also played a crucial role in shaping traffic density. The results confirm the significant influence of weather and temporal variables on traffic density and validate the effectiveness of advanced ensemble models and neural networks in predictive traffic modeling. By focusing on Istanbul, this study highlights the value of region-specific approaches and provides a foundation for developing data-driven traffic management strategies. Future research can build on these findings by expanding the scope of the dataset and incorporating dynamic interactions to further improve prediction accuracy and applicability.

Keywords Urban Traffic Density · Meteorological Factors · Temporal Patterns · Machine Learning · Ensemble Models · Boosting Techniques · Neural Networks

Citation: Özdenizci Köse, B., Huseynov, F., Gülpınar Demirci, V. & Taş, N. (2025). Meteorological and temporal dynamics in urban traffic density: A comparative evaluation of machine learning and neural network models. *Acta Infologica*, *9*(2), 611-629. https://doi.org/10.26650/acin.1627153

Corresponding author: Büşra Özdenizci Köse busraozdenizci@gtu.edu.tr

Introduction

Traffic congestion is a pervasive challenge in urban environments, with profound implications for economic productivity, environmental sustainability, and overall quality of life. Rapid urbanization and increasing vehicle ownership worsen the issue, particularly in metropolitan areas, where demand for road infrastructure often exceeds supply. Efficient traffic flow management has become a critical priority for urban planners, policymakers, and researchers. This complex issue requires innovative approaches that consider the interplay of various contributing factors and their interrelationships (Khan & Ivan, 2023; Gao & Zhu, 2022).

Numerous factors influence traffic density, with meteorological conditions and temporal variations being the most significant. Meteorological variables, such as temperature, humidity, wind speed, and precipitation, significantly affect driving patterns and traffic flow. Adverse weather conditions can reduce visibility and road traction, resulting in slower vehicle speeds and increased congestion (Wu, J., & Zhang, 2024; Bi & Zhu, 2022). Conversely, favorable weather conditions, particularly in areas with limited public transportation infrastructure, may enhance traffic flow but can also encourage higher vehicle usage.

Temporal variations, including time of day, day of the week, and seasonal patterns, also play a pivotal role in traffic dynamics. The morning and evening rush hours typically correspond to peaks in vehicle density, driven by commuter behavior and work schedules. Specific days, such as weekends or holidays, exhibit unique traffic patterns influenced by recreational activities and altered travel preferences. Understanding these temporal factors alongside meteorological conditions provides a comprehensive framework for traffic flow variability analysis.

Effective traffic management requires a detailed understanding of these interdependencies. Predictive modeling techniques, which encompass both traditional statistical and modern machine learning approaches, offer promising tools for capturing the complex relationships among meteorological and temporal factors and their collective impact on vehicle density. These techniques enable accurate forecasts of traffic conditions by identifying patterns and dependencies that traditional methods might overlook. Such insights empower proactive decision-making, optimize urban traffic management strategies, and improve transportation system efficiency.

Understanding the impacts of weather and temporal factors on traffic density has broader implications for sustainability (Ceder, 2021; Tao et al., 2023; Borchers et al., 2024). Insights derived from such analyses can contribute to a more efficient allocation of transportation resources, reducing greenhouse gas emissions, improving air quality, and minimizing energy consumption. Furthermore, fostering data-driven traffic management strategies supports the development of resilient and environmentally sustainable cities. This research aligns with global initiatives to promote sustainable urban development while addressing traffic congestion's immediate challenges.

Building on this foundation, this study analyzes traffic density at a fixed location in Istanbul, a city that exemplifies the challenges of urban congestion. Istanbul provides a rich context for examining how meteorological and temporal factors influence traffic patterns with its dense population and unique geographic and demographic characteristics. The city serves as an ideal case study for developing predictive strategies to mitigate congestion's adverse effects.

This study utilizes a comprehensive dataset collected from a fixed location in Istanbul, integrating meteorological variables, such as temperature, humidity, and wind speed, with temporal factors, including the time of day and day of the week. The data were meticulously processed and merged to facilitate a detailed analysis of their influence on traffic density. By leveraging advanced analytical methods, this study identifies critical factors affecting traffic dynamics and develops predictive frameworks to support informed decision-making and effective traffic management strategies. Analyzing data from a single, consistent location allows for a focused examination of temporal and weather-related variations without introducing spatial variability and ensures more precise insights into the relationships between these factors.

To guide the analysis and achieve its objectives, the study addresses the following key research questions:

(1) How do meteorological factors such as temperature, humidity, and wind speed influence traffic density?

(2) How do temporal factors, including the time of day and day of the week, influence traffic density?

(3) Can predictive models effectively estimate traffic density by integrating meteorological and temporal factors?

(4) What are the implications of these predictive insights for sustainable traffic management and urban planning?

The remainder of this paper is organized as follows: Section 2 discusses the related work, provides an overview of studies, and examines the impact of meteorological and temporal factors on traffic dynamics. Section 3 describes the dataset, the preprocessing steps, and the analysis's methodological framework. Section 4 presents the results, including insights into the key factors that influence traffic density. Section 5 concludes the paper by discussing the findings, their implications for sustainable traffic management, the limitations of the study, and directions for future research.

Related Work

Understanding the relationship between weather, temporal factors, and traffic patterns has emerged as a critical area of research driven by urbanization and traffic congestion challenges. Recent advances in machine learning and deep learning have significantly improved the ability to model the complex interactions between these factors, which offer predictive insights and enable data-driven traffic management strategies.

Early studies focused on statistical methods to explore the impact of weather on traffic patterns, laying the groundwork for more advanced modeling techniques. Keay and Simmonds (2005) investigated the effect of rainfall on traffic volume in Melbourne, revealing that daytime traffic exhibited greater sensitivity to precipitation than nighttime traffic. This finding highlights the role of temporal factors in moderating the influence of weather. Similarly, Cools et al. (2010) employed regression analysis to examine the influence of weather conditions on traffic and emphasized the importance of deriving location-specific insights to effectively inform traffic management policies.

Based on these foundational studies, Lin et al. (2022) combined random-effects regression with Random Forest models to investigate traffic volume variations under diverse weather and vacation conditions. Their work emphasized the dynamic interaction between rainfall and traffic patterns, and recommended for granular spatial-temporal modeling approaches. Bi et al. (2022) extended these findings across multiple cities using multivariate regression models and demonstrated the importance of complex weather indicators, such as temperature, wind speed, and visibility, in shaping urban traffic behavior. Collectively, these studies underscored the need to integrate meteorological data with location-specific temporal factors for traffic management insights.

Advancements in machine learning and deep learning have significantly expanded the scope and accuracy of traffic forecasting. Braz et al. (2022) evaluated the performance of LSTMs and CNNs in predicting coastal traffic flow. Their results revealed that weather variables, particularly wind and solar radiation, play a critical role in achieving accurate forecasts. Similarly, Bao et al. (2021) integrated SVR into a deep belief network, achieving superior performance compared to traditional neural networks, particularly under adverse weather conditions. Shaygan et al. (2022) conducted an extensive review of AI-based traffic prediction methods and highlighted their ability to improve network use and reduce congestion through accurate predictions. Their survey emphasized the importance of preprocessing methods and the integration of heterogeneous datasets for enhanced modeling. Abduljabbar et al. (2024) applied unidirectional and bidirectional LSTM networks to predict short-term traffic flow on arterial roads, achieving over 99% accuracy with bidirectional LSTM models. Their inclusion of weather variables, such as rainfall and temperature, improved predictive performance and provided practical insights for managing traffic under adverse conditions.

Deep learning frameworks have proven particularly effective in capturing the complex spatial-temporal dependencies inherent in traffic dynamics dependencies (Wang et al., 2022; Yin et al., 2022; Qi et al., 2023; Zhang et al., 2024). For example, Wang et al. (2022) highlighted the strengths of CNNs and RNNs in automating feature extraction and enabling high-dimensional data analysis, addressing limitations of traditional statistical approaches. Yin et al. (2022) emphasized the advantages of combining convolutional and recurrent architectures to model dynamic spatial and temporal correlations in traffic data. Qi et al. (2023) introduced a spatiotemporal graph convolutional network that integrates meteorological variables, such as temperature and visibility. Their findings demonstrated improved accuracy and stability in long-term traffic flow predictions by including external environmental factors.

In their comprehensive review, Xie et al. (2020) categorized five distinct traffic prediction methodologies: statistical, machine learning, deep learning, reinforcement learning, and transfer learning. They concluded that deep learning techniques were particularly effective for managing heterogeneous urban data, given their ability to capture intricate relationships and dependencies. Medina-Salgado et al. (2022) similarly reviewed smart urban traffic flow prediction techniques, highlighting the role of machine learning in addressing nonlinear and heterogeneous forecasting challenges.

As researchers strive to tailor predictive models to the unique characteristics of geographic and temporal contexts, localized and region-specific methodologies have gained prominence. Lin et al. (2024) employed geographically weighted regression to examine the advancing and lagging effects of weather on intercity traffic, revealing significant spatial heterogeneity. This study highlights the importance of region-specific modeling approaches to enhance prediction accuracy. Kendre et al. (2024) compared regression algorithms, including Decision Trees and SVM, and concluded that SVM offered the highest accuracy for short-term predictions based solely on weather data. Gerges et al. (2024) demonstrated the robustness of gradient boosting regression in linking meteorological variables to urban characteristics, further emphasizing the value of machine learning in localized traffic prediction. Beyond traffic volume forecasting, Lin et al. (2025) integrated Bayesian networks with deep learning to explore interactions between rainfall, traffic flow, and air pollutant concentrations, revealing critical nonlinear dependencies.

While a considerable body of international research has focused on the impact of meteorological and temporal factors on traffic dynamics, several studies have also specifically investigated traffic density prediction in the Türkiye, particularly using data from Istanbul. Aydin et al. (2021) proposed an LSTM-based deep learning model to estimate traffic density using open-access data from the Istanbul Metropolitan Municipality. Their results showed that the LSTM model outperformed classical machine learning models, including linear regression, Decision Trees, and Random Forests, in estimating regional traffic density. Akın et al. (2019) introduced a novel traffic flow forecasting model based on a clustering algorithm that used vehicle location and angular direction data without relying on spatial road information. Their findings demonstrated that effective clustering and forecasting could be achieved independently of road geometry, offering an alternative framework for traffic management. Utku (2023) developed a hybrid model combining CNN and RNN to leverage the strengths of both architectures in traffic density prediction. The model achieved high accuracy using a high-resolution hourly dataset from 2,321 locations in Istanbul, outperforming several benchmark models, including LR, RF, SVM, MLP, and standalone deep learning networks. Additionally, Taş and Müngen (2021) employed artificial neural networks and SVM to predict regional traffic intensity using meteorological and traffic data collected from 75 monitoring points across Istanbul. Their results demonstrated that integrating weather-related features significantly improved forecasting accuracy, achieving approximately 90% success in regional congestion estimation. Collectively, these studies underscore the importance of combining deep learning architectures with localized urban data to improve traffic density estimation predictive performance.

Despite significant advancements in traffic prediction methodologies, understanding the combined effects of meteorological variables and temporal factors on traffic dynamics remains a key challenge. Although individual factors, such as temperature, wind speed, and precipitation, have been extensively studied, the complex interactions between these variables and their collective influence on traffic patterns remain underexplored. Addressing this gap is crucial for developing more accurate predictive models that capture the dynamic relationships between weather and temporal factors. The integration of weather data, temporal patterns, and traffic flow dynamics into predictive models contributes to both effective traffic management and broader urban sustainability goals.

Method

This section outlines the study's methodological framework, including data collection, preprocessing, and modeling techniques.

Data Collection

The dataset used in this study consisted of 13,986 records, integrating traffic density and meteorological data from a specific region within Istanbul. Traffic density data were obtained from the *Open Data Portal of the Istanbul Metropolitan Municipality*. This dataset provides detailed metrics, such as the number of vehicles and their speeds, captured at various times within the designated region. Meteorological data specific to the same region in Istanbul were sourced from the *Weather Underground* platform. Some key weather variables are temperature (in Celsius), dew point (in Celsius), wind speed (in miles per hour), gust speed (in miles per hour), atmospheric pressure (in inches), precipitation rate (in inches per hour), and accumulated precipitation (in inches). The traffic and meteorological datasets were aligned using time stamps to ensure that each traffic record matched the corresponding weather conditions simultaneously. This merging process resulted in a unified dataset, facilitating the analysis of complex relationships among traffic volume, localized weather conditions, and temporal factors.

The final dataset included independent variables such as weather metrics (e.g., temperature, wind speed, and precipitation), temporal indicators (e.g., time of day, weekday/weekend), and traffic-specific measures (e.g., maximum, minimum, and average vehicle speeds). This comprehensive integration provided a robust foundation for exploring the factors affecting traffic density and for developing predictive models to support effective traffic management and urban planning.

Data Preprocessing

The collected data were subjected to a comprehensive preprocessing pipeline to ensure that they were clean, consistent, and suitable for machine learning analysis. In this process, the first step was to clean the data by identifying and removing outliers. Outliers were detected using statistical methods, such as the Z-score technique, which identifies values that significantly deviate from the mean. These outliers, which could distort the analysis or lead to overfitting in the predictive models, were excluded from the study. Although minimal, missing values were handled through statistical imputation. The mean or median was used to fill missing values for continuous variables depending on the data distribution. This approach ensured the completeness of the dataset while avoiding the introduction of significant bias.

Feature engineering was essential for dataset enhancement. Categorical variables, such as the day of the week, were transformed into binary features (e.g., Monday, Tuesday), whereas temporal variables were represented through indicators capturing cyclical traffic behaviors, such as working hours (within working hours and after working hours) and time of day (AM or PM). High-level features, such as weekdays and weekends, were added to differentiate traffic patterns between workdays and non-workdays. Numerical variables were standardized using Z-score normalization to ensure consistent scales across all features and to prevent features with larger magnitudes, such as the number of vehicles or maximum speed, from disproportionately influencing the models.

The final dataset, including 12,842 records, was transformed into a structured format optimized for machine learning models. These preprocessing steps established a solid foundation for the analysis of the complex interplay between weather, time, and traffic patterns.

Distribution of Key Variables and Correlation Patterns

An exploratory data analysis was performed to better understand the characteristics and relationships within the dataset, focusing on variable distributions and interdependencies. Figure 1 illustrates the standardized distributions of key meteorological and traffic-related features using kernel density estimation (KDE). Most features exhibit non-normal distributions, particularly traffic speed indicators and precipitation variables, highlighting the dataset's variability and skewness. In particular, the number of vehicles demonstrates a heavy positive skew, indicating that traffic peaks during certain time intervals. This non-uniformity underscores the need for nonlinear and ensemble-based modeling approaches that can effectively capture complex patterns.

Figure 1

KDE of the standardized meteorological and traffic-related variables

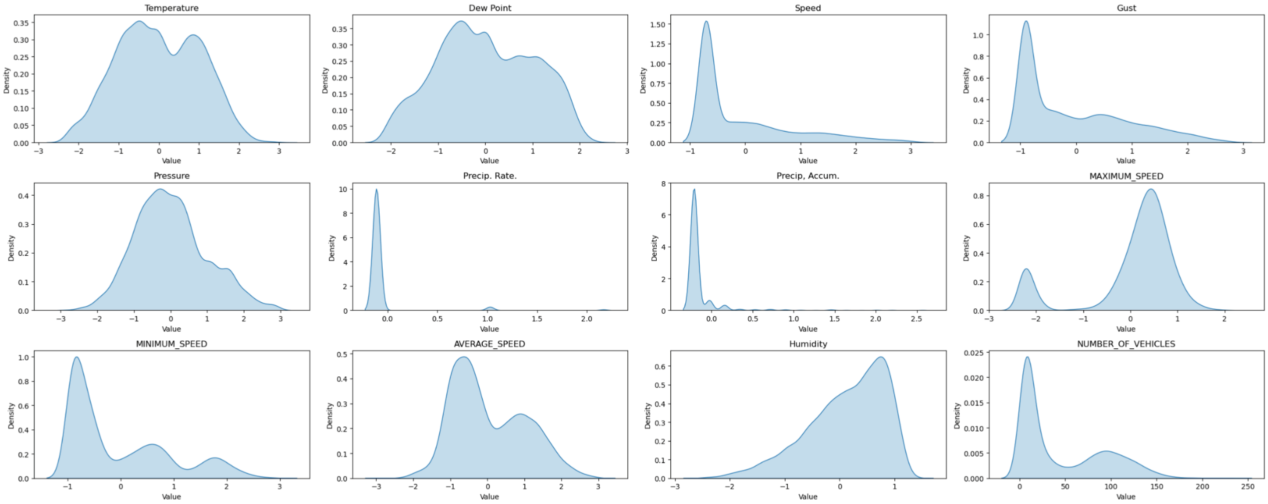


Figure 2 presents boxplots for the same variables to visualize their spread, central tendency, and presence of outliers. Traffic-related features, including vehicle speeds, display wider interquartile ranges and more extreme values compared with relatively stable meteorological variables, such as temperature and dew point. The boxplot for the number of vehicles further highlights its dispersion and outlier prevalence.

A Pearson correlation heatmap is shown in Figure 3 to identify linear dependencies among variables. As expected, strong positive correlations exist between temperature and dew point and among wind-related features, such as speed and gust. A moderate negative correlation was also observed between the number of vehicles and the average speed, indicating that increased traffic density is associated with lower speeds. However, most weather variables exhibit only weak correlations with traffic-related outcomes, underscoring the potential for nonlinear modeling approaches to better capture these interactions.

Figure 2

Boxplots of the standardized meteorological and traffic features

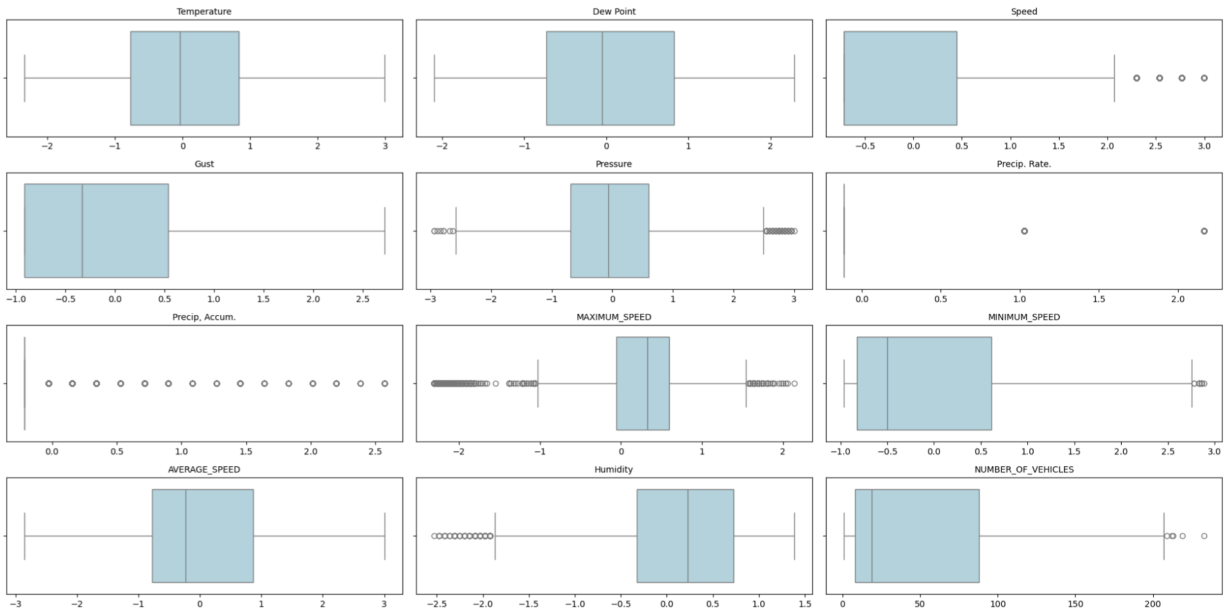
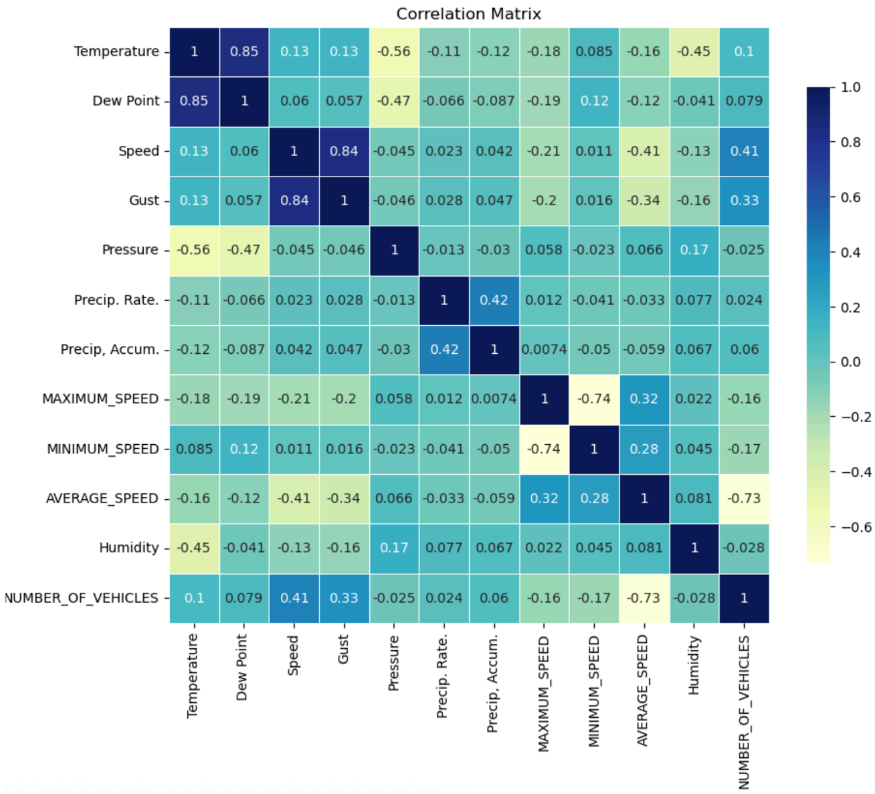


Figure 3

Correlation heatmap of meteorological and traffic features



Model Development

After preprocessing the dataset, a multi-model approach was followed, providing a comprehensive understanding of the dataset and facilitating the development of robust predictive frameworks for analyzing traffic density in relation to weather and temporal factors. The modeling process was structured around three main groups of algorithms: traditional linear regression techniques, nonlinear regression models, and advanced machine learning approaches including ensemble methods and neural networks.

Traditional regression methods, such as Ridge Regression and Lasso Regression, were used to establish baseline performance and detect fundamental linear patterns in the data. The regularization strength (alpha), a key hyperparameter, was set to 1.0 for Ridge and 0.1 for Lasso to maintain moderate shrinkage and prevent overfitting. These models effectively addressed multicollinearity and yielded stable and interpretable coefficients.

Nonlinear regression models, including Decision Tree, K-Nearest Neighbors (KNN) Regression, and Random Forest, were applied to explore more complex, nonlinear interactions between features. For the Decision Tree model, the maximum depth was limited to 5 to prevent overfitting and maintain model generalizability. The KNN regression model used 5 neighbors to balance bias and variance in local predictions. Random Forest, configured with 100 estimators, served as an ensemble learning method that demonstrated improved prediction accuracy by aggregating outputs from multiple Decision Trees. This aggregation reduces variance and overfitting while enhancing model robustness (Dong et al., 2020; Mienye & Sun, 2022). Preliminary experimentation and validation performance consistency informed these hyperparameter settings.

Advanced ensemble models, specifically boosting algorithms such as XGBoost and LightGBM, were implemented to further improve performance by minimizing residual errors iteratively and optimizing prediction accuracy across multiple weak learners. To ensure a balance between learning capacity and generalization, the XGBoost model was configured with 100 estimators, a learning rate of 0.1, and a maximum depth of 3. LightGBM was trained using default hyperparameters with a fixed random state to maintain reproducibility. These techniques are well-known for their computational efficiency, scalability, and superior accuracy in regression tasks (Bentéjac & Martínez-Muñoz, 2021; Liu, 2025).

A feedforward Neural Network was developed to capture and analyze complex nonlinear interactions within the data. The architecture included an input layer for the selected features, followed by two hidden layers with ReLU activation functions and a single output neuron for predicting traffic density. The model was trained using the Adam optimizer. The key hyperparameters included 300 training epochs and a batch size of 32. The validation loss was monitored during training to mitigate overfitting and ensure generalizability.

Manual tuning was preferred over automated hyperparameter optimization techniques (e.g., grid search, random search, or Bayesian optimization) because of the moderate dataset size and computational resource considerations. Hyperparameter choices were informed by prior studies and refined through preliminary validation performance. Table 1 summarizes the models, their configurations, and methodological roles for clarity and ease of comparison.

Table 1

Models, hyperparameter configurations, and their methodological roles in traffic density prediction

Model Type	Algorithm	Key Hyperparameters	Purpose
Linear Models	Ridge Regression	$\alpha = 1.0$	Establish baseline and handle multicollinearity
	Lasso Regression	$\alpha = 0.1$	Selects key features and prevents overfitting
Nonlinear Models	Decision Tree	max_depth = 5	Captures nonlinear patterns; easy to interpret
	KNN	n_neighbors = 5	Local regression balances bias and variance
	Random Forest	n_estimators = 100	Stable ensemble model reduces overfitting
Boosting Models	XGBoost	n_estimators = 100, learning_rate = 0.1, max_depth = 3	Accurate and scalable boosting method
	LightGBM	Default parameters, fixed random_state	Fast and memory-efficient gradient boosting
Neural Network	Feedforward Neural Network	2 hidden layers (ReLU), 300 epochs, batch size = 32	Captures complex nonlinear patterns and avoids overfitting with validation

Evaluation and Validation of the Model

The model performance was evaluated using several metrics. Root Mean Squared Error (RMSE) quantified the average prediction error, with lower values indicating higher accuracy. Mean Absolute Error (MAE) provided a straightforward measure of the average deviation between the predicted and actual values. The R-squared (R^2) score assessed the proportion of variance in vehicle counts explained by the model, reflecting its overall goodness-of-fit.

Additionally, Symmetric Mean Absolute Percentage Error (SMAPE) was employed to evaluate the proportional accuracy. Compared with MAPE, SMAPE is more robust in the presence of extreme values and offers a more balanced percentage-based error metric by accounting for both over- and under-predictions. This makes it suitable for datasets with varying traffic volumes. The feature importance scores derived from the tree-based models further contributed to the model's interpretability by highlighting the relative influence of meteorological and temporal factors on traffic density.

The dataset was initially split into 80% for training and 20% for testing to ensure consistency across models. For models requiring internal validation, such as neural networks, an additional 20% of the training set was allocated for validation. This unified partitioning strategy allowed for robust and fair comparisons across linear, nonlinear, ensemble, and deep learning models.

In addition, all models were evaluated using 5-fold cross-validation to enhance generalizability and mitigate overfitting. This approach divides the dataset into five equally sized folds, iteratively using four folds for training and one for testing, ensuring that each data point is used for validation exactly once. The average performance across folds was recorded for the RMSE, MAE, R^2, and SMAPE metrics. This evaluation strategy provides a more stable and reliable estimate of model performance, reduces variance from single train-test splits, and enables fair comparisons across models.

All models were implemented in Python using NumPy and Pandas for data preprocessing and Scikit-learn for model implementation and evaluation. XGBoost, LightGBM, and TensorFlow were employed for gradient boosting and neural network models, respectively. Visualization tools, such as Matplotlib and Seaborn, were used to analyze feature importance, error distributions, and comparative performance metrics.

Results

Regression Analysis

The regression analysis compared the performance of several machine learning models in predicting vehicle density using meteorological and temporal features. Table 2 summarizes the evaluation metrics, including MAE, MSE, RMSE, R^2 Score, and SMAPE (%).

Ensemble methods, such as LightGBM and Random Forest, demonstrated superior performance compared with other models. Both groups achieved the lowest RMSE values (16.59 and 16.65, respectively) and the highest R^2 Scores (0.8727 and 0.8718, respectively). These results emphasize the effectiveness of ensemble methods in capturing the nonlinear and complex interactions between weather, temporal factors, and traffic density.

In contrast, traditional regression models, including Ridge Regression and Lasso Regression, showed limited predictive capability, with RMSE values exceeding 28.83 and R^2 Scores around 0.615. Although these models provide interpretable insights, their inability to model nonlinear relationships restricts their utility in this context.

Nonlinear models, such as Decision Tree and XGBoost, performed significantly better than traditional methods. Decision Tree achieved an RMSE of 19.92 and an R^2 Score of 0.8164, while XGBoost achieved an RMSE of 17.06 and an R^2 Score of 0.8653. These models effectively capture the interactions between features while maintaining interpretability.

In addition to traditional performance metrics, the SMAPE was used to assess the proportional accuracy across models. LightGBM and Random Forest again yielded the lowest SMAPE values (33.48% and 31.46%, respectively), highlighting their consistency in absolute error and relative performance across varying traffic volumes. Conversely, the Ridge and Lasso regressions exhibited the highest SMAPE values, indicating limited proportional accuracy and further reinforcing the superiority of ensemble-based approaches.

Figures 4(a) and 4(b) show the residual (actual - predicted) distributions for the Ridge Regression and LightGBM models, respectively, aggregated across the cross-validation folds. The LightGBM model yields residuals that are tightly clustered around zero, indicating that the predicted values closely align with the actual traffic density observations. In contrast, the Ridge Regression model shows a wider spread in residuals, reflecting greater deviations between predicted and actual values and highlighting its limited capacity to capture complex, nonlinear patterns in the data.

Overall, the findings highlight the value of advanced ensemble methods and gradient boosting techniques, such as LightGBM and XGBoost, in developing robust traffic density predictive frameworks. These models provide a strong foundation for improving traffic management strategies and urban planning efforts.

Table 2

Average performance metrics of regression models

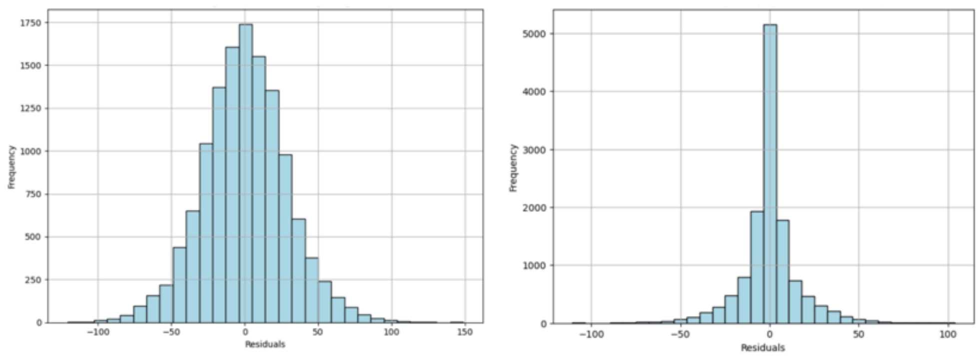
Model	MAE	MSE	RMSE	R^2 Score	SMAPE (%)
Ridge Regression	22.41	831.45	28.83	0.6155	78.97
Lasso Regression	22.45	831.83	28.84	0.6152	78.84
Decision Tree	12.39	396.88	19.92	0.8164	37.95
KNN Regression	14.29	526.46	22.94	0.7566	41.65
Random Forest	10.13	277.39	16.65	0.8718	30.92
XGBoost	10.56	341.27	17.06	0.8653	33.13
LightGBM	10.35	275.23	16.59	0.8727	33.48

As shown in Table 3 and Figure 5, the average feature importance scores derived from Random Forest, LightGBM, and XGBoost models highlight the contribution of individual features to traffic density predictions. While traffic-related metrics, such as average speed, minimum speed, and maximum speed, consistently ranked among the top predictors due to their direct relationship with vehicle flow, meteorological and temporal features also demonstrated a notable impact.

Among meteorological variables, temperature and pressure demonstrated significant importance, particularly in LightGBM, highlighting their role in capturing weather-driven fluctuations in traffic patterns. Additional environmental indicators, such as humidity, dew point, and gush, provided complementary insights into how atmospheric conditions shape driving behavior. Precipitation variables such as PrecipRate and PrecipAccum had lower importance scores, likely due to the limited occurrence of severe weather in the dataset.

Figure 4

Histograms of residual distribution aggregated across folds



(a) Ridge Regreession (b) LightGBM

Temporal variables, such as Is_Summer, Within_WorkingHour, and Is_AM, were also influential, effectively modeling seasonal traffic surges, working-hour dynamics, and daily commuting patterns. For instance, Is_Summer was especially important in XGBoost, showing that seasonal effects were captured differently across model architectures. These findings emphasize that both meteorological and temporal features contribute to understanding and predicting traffic density beyond core traffic metrics. Importantly, these feature rankings reflect the average results obtained across 5-fold cross-validation, ensuring that the observed importance patterns are consistent and generalizable.

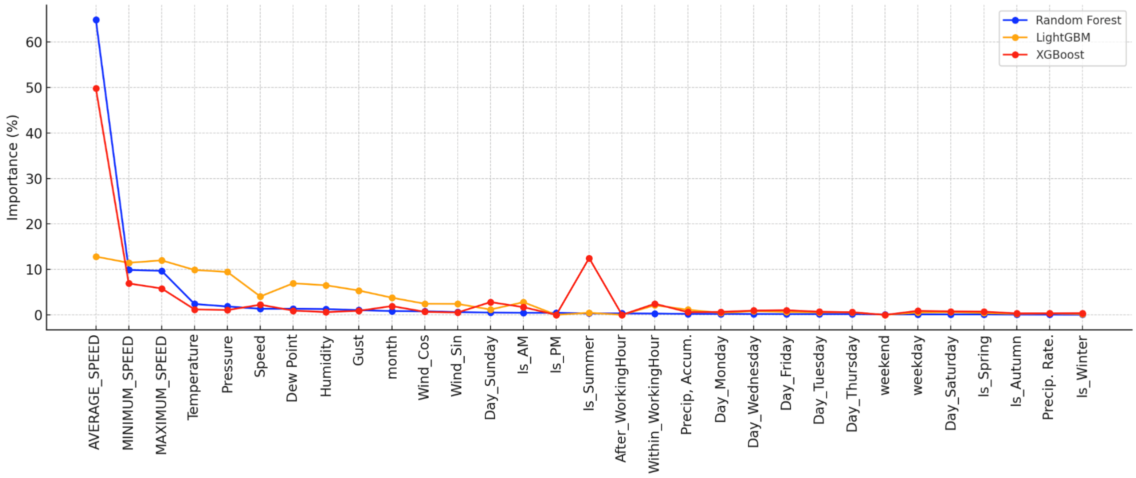
Table 3

Average Feature Importance Analysis of the LightGBM, XGBoost, and Random Forest Models

Feature	LightGBM (%)	XGBoost (%)	Random Forest (%)
AVERAGE_SPEED	12.82	49.8	64.9
MINIMUM_SPEED	11.45	6.91	9.9
MAXIMUM_SPEED	11.99	5.81	9.69
Temperature	9.88	1.21	2.39
Pressure	9.44	1.06	1.85
Dew Point	6.94	0.94	1.34
Humidity	6.51	0.61	1.29
Gust	5.35	0.9	1.04
Speed	4.05	2.23	1.34
month	3.78	1.93	0.85
Wind_Cos	2.45	0.68	0.77
Wind_Sin	2.42	0.52	0.62
Day_Sunday	1.17	2.81	0.51
Is_AM	2.79	1.7	0.45
Is_PM	0.0	0.0	0.44
Is_Summer	0.47	12.49	0.32
After_WorkingHour	0.0	0.0	0.32
Within_WorkingHour	2.13	2.45	0.3
Precip, Accum.	1.13	0.6	0.22
Day_Monday	0.43	0.65	0.21
Day_Wednesday	0.82	0.95	0.2
Day_Friday	0.65	1.03	0.17
Day_Tuesday	0.55	0.69	0.16
Day_Thursday	0.53	0.59	0.16
weekend	0.0	0.0	0.1
weekday	0.51	0.89	0.1
Day_Saturday	0.6	0.76	0.09
Is_Spring	0.41	0.71	0.09
Is_Autumn	0.23	0.33	0.07
Precip. Rate.	0.33	0.35	0.06
Is_Winter	0.19	0.38	0.05

Figure 5

Comparison of feature importance across models



Neural Network Analysis

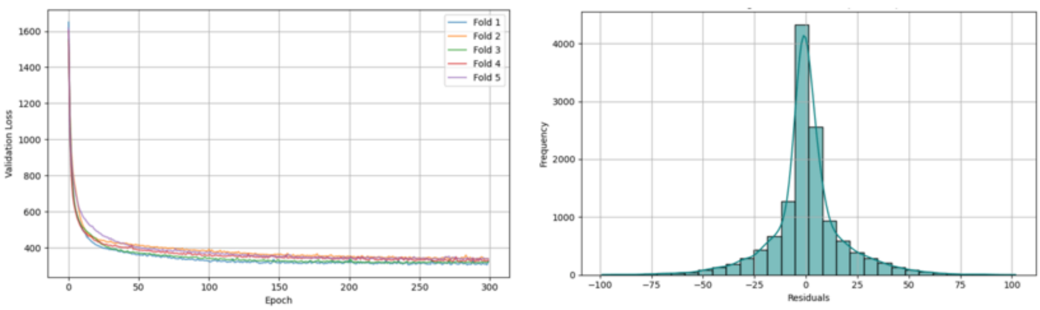
To further investigate the nonlinear relationships and interactions among features, a feedforward neural network was implemented. This approach was designed to complement the findings from traditional and ensemble models by leveraging the ability of deep learning frameworks to uncover intricate patterns within the data.

Several neural network architectures were evaluated by varying the number of hidden layers and neurons using a consistent set of 30 meteorological and temporal input features. The best-performing configuration consisted of two hidden layers with 16 and 8 neurons and demonstrated strong predictive performance. On average, the model achieved an RMSE of 17.96, an R^2 score of 0.8508, a SMAPE of 36.91%, and an EVS (Explained Variance Score) of 0.8509 across the 5-fold cross-validation. These results confirm the robustness and effectiveness of the model in capturing complex nonlinear relationships within the data.

Figure 6(a) shows that the validation loss curves for each fold show consistent convergence trends, with the model reaching a stable minimum after approximately 200 epochs. This reflects strong generalizability and the absence of overfitting. Figure 6(b) illustrates the distribution of residuals (prediction errors) aggregated across all folds. Most errors are centered around zero with a slightly right-skewed distribution, indicating the model's overall reliability and limited bias.

Figure 6

Evaluation of the performance of the neural network model



(a) Validation loss curves across epochs and folds

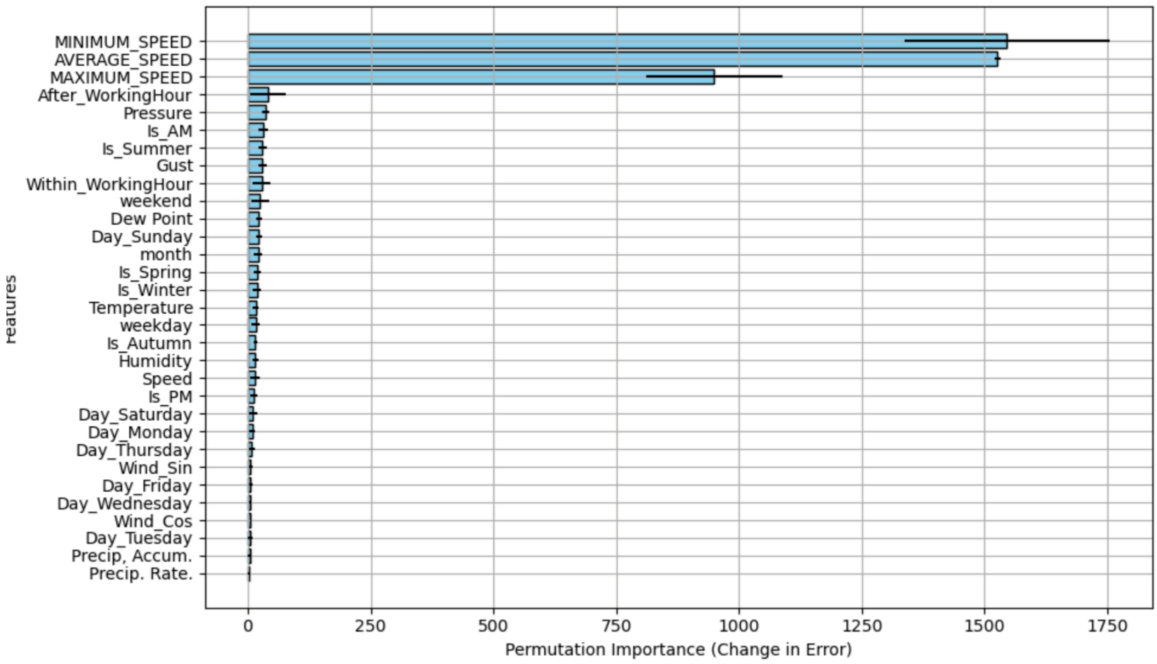
(b) Residual histogram of the neural network model

Permutation importance analysis was employed to further interpret the neural network's behavior and quantify the impact of each feature on prediction performance. This model-agnostic approach randomly permutes one feature at a time and measures the change in RMSE. A larger increase in error indicates the feature's greater importance in generating accurate predictions.

As shown in Figure 7, speed-related features (i.e., minimum, average, and maximum speed) had the highest importance, as expected. However, the results also indicated the influence of several temporal and contextual factors. In particular, After_WorkingHour and Is_AM were among the most important non-speed features, indicating their role in capturing working-hour-related daily traffic patterns. Weather-related variables, such as pressure and gush, showed moderate effects, suggesting that short-term weather changes can have an impact on traffic flow. Time-based indicators, such as Within_WorkingHour, weekday/weekend, and Day_Sunday, also contributed, highlighting regular weekly traffic trends. In comparison, seasonal variables (i.e., Is_Summer, Is_Winter, Is_Spring, and Is_Autumn) and precipitation-related features had relatively low importance. This may be due to the limited variation in these features within the dataset. Overall, the permutation importance results show that the neural network captures patterns not only from the main traffic indicators but also from the time and weather context, thereby helping improve traffic density predictions.

Figure 7

Permutation importance of features in neural network predictions



Discussion and Conclusion

Understanding the intricate relationships between meteorological and temporal factors and traffic density is pivotal for advancing urban traffic management strategies and sustainable city planning. This study investigated these dynamics using a combination of traditional regression models, advanced ensemble techniques, and neural networks. The findings provide insights into the predictive capabilities of different modeling approaches and emphasize the significance of weather and temporal variables in shaping traffic patterns.

Model Comparison

The analysis underscores the superior performance of advanced ensemble methods, particularly LightGBM and Random Forest, in traffic density modeling. These models achieved the lowest RMSE values (16.59 and 16.65, respectively) and the highest R^2 Scores (0.8727 and 0.8718, respectively). Their ability to capture complex, nonlinear relationships between variables highlights their suitability for traffic density prediction in diverse and dynamic contexts. Neural networks also demonstrate competitive performance, achieving an RMSE of 17.96 and an R^2 Score of 0.8508, which underscores their capacity to model intricate interactions and complement findings from ensemble methods.

In contrast, traditional regression models, such as Ridge and Lasso, have limited predictive capacity. These models provide interpretability and simplicity but fail to account for the nonlinear relationships essential for capturing the traffic density dynamics' complexities. Decision Tree and XGBoost models provide a balance between interpretability and predictive accuracy and achieve better results than traditional methods. However, their performance is slightly lower than that of ensemble techniques such as LightGBM and Random Forest.

Feature importance analysis provided critical insights into the factors influencing traffic density, enhancing the predictive models' interpretability. Traffic-specific variables consistently ranked highest across ensemble methods, underscoring their direct association with vehicle flow. Meteorological factors, including temperature and pressure, also demonstrated significant importance in capturing variations caused by changing weather conditions. Temporal factors, such as Is_AM and Within_WorkingHour, revealed patterns related to peak traffic hours, further emphasizing their role in understanding daily traffic dynamics. These findings emphasize the importance of incorporating traffic, weather, and temporal feature types into comprehensive predictive frameworks.

Overall, ensemble methods, such as LightGBM and Random Forest, delivered high predictive accuracy and transparent insights into feature contributions, enhancing model interpretability. By capturing deeper, nonlinear dependencies, neural networks further enriched the modeling landscape. Together, these methods effectively addressed the research objectives, providing robust, explainable, and actionable solutions for the prediction and management of urban traffic density.

Addressing Research Questions

This study effectively answered the research questions by demonstrating the intricate relationships between meteorological and temporal factors and traffic density through advanced modeling techniques.

The findings confirmed the first research question, "How do meteorological factors influence traffic density?" by demonstrating that meteorological variables, such as temperature, pressure, and humidity, significantly affect traffic patterns. These factors provide critical insights into how environmental conditions affect driving behaviors and traffic flow. While extreme weather conditions lead to noticeable alterations in vehicular patterns, precipitation metrics showed lower importance, likely due to the limited number of adverse weather instances in the dataset. Nonetheless, they remain relevant for understanding specific traffic scenarios under unfavorable conditions.

The second research question, "How do temporal factors influence traffic density?" was addressed by analyzing features such as Is_AM, Within_WorkingHour, and Is_Summer. These variables captured critical temporal dynamics, including daily commuting peaks during the morning hours and increased vehicle density during the working hours. Seasonal patterns, represented by Is_Summer, highlight the impact of tourism and recreational activities on traffic density. Although the differentiation between weekday and weekend traffic was less pronounced, the temporal indicators demonstrated the significance of time-related variations in understanding and predicting traffic patterns.

The third research question, “Can predictive models effectively estimate traffic density by integrating meteorological and temporal factors?” was confirmed through the evaluation of various machine learning models. Advanced ensemble methods, such as LightGBM and Random Forest, exhibited superior predictive performance, achieving the lowest RMSE values and highest R^2 scores across experiments. Neural networks also delivered competitive results, reinforcing their ability to model complex and nonlinear relationships within the data. These findings confirm the effectiveness of integrating meteorological and temporal features in traffic density prediction and highlight the practical potential of such models for real-world traffic management and urban planning applications.

Implications of the Study

This section discusses the implications of the study, addressing the fourth research question: “What are the implications of these predictive insights for sustainable traffic management and urban planning?”

This study contributes to the theoretical understanding of the complex interplay between meteorological and temporal factors and their collective influence on urban traffic density. The findings validate and extend prior research that emphasizes the significance of integrating weather and temporal variables in predictive modeling by focusing on a region-specific context. Previous studies, such as those by Lin et al. (2022) and Bi et al. (2022), highlighted the impact of weather conditions, such as temperature, wind speed, and precipitation, on traffic patterns. This research builds on these insights by confirming their relevance in Istanbul’s unique urban and meteorological context and providing evidence of their dynamic interactions with temporal factors. This study illustrates how such localized dynamics can offer actionable insights for sustainable traffic management.

The application of advanced ensemble models and neural networks underscores the strengths of machine learning and deep learning frameworks in capturing nonlinear relationships. Shaygan et al. (2022) and Wang et al. (2022) emphasized the advantages of these models in managing high-dimensional data and modeling complex dependencies. This research reinforces these theoretical advancements by demonstrating that ensemble methods, such as LightGBM and Random Forest can outperform traditional models while offering interpretable insights. This study’s comparative evaluation illustrates the relative strengths of different approaches and highlights the ability of neural networks to complement ensemble methods in handling intricate interactions. These findings affirm the potential of these models for real-time traffic forecasting and adaptive urban planning.

This study further supports the importance of localized modeling, as emphasized by Lin et al. (2024) and Kendre et al. (2024). The focus on Istanbul exemplifies the importance of tailoring predictive models to specific geographic and temporal contexts. By integrating region-specific weather patterns and urban characteristics, this study enhances the accuracy and relevance of traffic density predictions and provides a replicable framework for other urban regions facing similar challenges. This localized approach offers a theoretical framework for future research in other urban settings and reinforces the need for context-aware modeling.

Feature importance analysis provides additional theoretical insights into traffic density modeling. Consistent with the findings of Braz et al. (2022) and Gerges et al. (2024), traffic-specific variables, such as average and minimum speeds, emerged as the most significant predictors, underscoring their direct influence on traffic flow. Meteorological variables, such as temperature and pressure, also demonstrated substantial importance, particularly in capturing variations caused by extreme weather. Temporal factors, such as morning hours and working periods, provide insights into cyclical traffic behaviors, emphasizing the importance of integrating daily and seasonal temporal patterns into predictive models. These findings establish the multifaceted nature of traffic density and the need for comprehensive feature inclusion in predictive frameworks.

This study addresses a critical gap in the literature on the interaction effects of weather and temporal variables on traffic dynamics. While previous research (e.g., Qi et al., 2023; Lin et al., 2025) acknowledged the underexplored nature of these interactions, this study directly examines their combined influence and offers valuable robust evidence of their significance. This theoretical contribution bridges an important research gap and encourages the incorporation of multifactorial approaches in future studies.

In summary, the study aligns with broader discussions in the field by extending theoretical frameworks on traffic modeling, emphasizing localized and context-aware approaches, and validating the utility of advanced machine learning models. These contributions fill existing gaps in the literature and provide a foundation for future research aiming to enhance urban traffic systems and promote sustainable city planning.

Limitations of the Research

This study has several limitations that should be considered. The analysis focused on a specific region within Istanbul, limiting the generalizability of the findings to other urban contexts with varying geographic, infrastructural, or cultural characteristics. Additionally, the dataset primarily captured typical weather and restricted insights into the impact of extreme events, such as severe storms or heavy snowfall. Relying on historical traffic and weather data excluded real-time variations caused by unexpected factors, such as accidents or road closures, which underscores the need for future studies to integrate real-time analytics. Although advanced machine learning and deep learning models were employed, their trade-offs between accuracy and interpretability pose challenges for stakeholders who require explainable insights for policy and planning. Addressing these limitations in future research could enhance the robustness, applicability, and relevance of traffic density models.

Peer Review	Externally peer-reviewed.
Peer Review	Conception/Design of Study- B.Ö.K., F.H., V.G.D., N.T.; Data Acquisition- B.Ö.K., F.H., V.G.D., N.T.; Data Analysis/Interpretation- B.Ö.K., F.H., V.G.D., N.T.; Drafting Manuscript- B.Ö.K., F.H., V.G.D., N.T.; Critical Revision of Manuscript- B.Ö.K., F.H., V.G.D., N.T.; Final Approval and Accountability- B.Ö.K., F.H., V.G.D., N.T.
Conflict of Interest	The authors have no conflict of interest to declare.
Grant Support	This work was supported by Research Fund of the Gebze Technical University. Project Number: 2024-A-113-13

Author Details

Büşra Özdenizci Köse

[1] Gebze Technical University, Faculty of Management, Department of Management Information Systems, Kocaeli, Türkiye

0000-0002-8414-5252 busraozdenizci@gtu.edu.tr

Farid Huseynov

[1] Gebze Technical University, Faculty of Management, Department of Management Information Systems, Kocaeli, Türkiye

0000-0002-9936-0596

Vildan Gülpınar Demirci

[2] Aksaray University, Faculty of Economics and Administrative Sciences, Department of Economics, Aksaray, Türkiye

0000-0002-8824-5154

Nurullah Taş

[3] Gebze Technical University, Faculty of Management, Department of Management, Kocaeli, Türkiye

0000-0001-6221-0204

References

Abduljabbar, R., Dia, H., & Liyanage, S. (2024). Machine Learning Models for Traffic Prediction on Arterial Roads Using Traffic Features and Weather Information. *Applied Sciences*, *14*(23), 11047.

Akın, M., Sağıroğlu, Ş., & Değirmenci, A. (2019, November). *Traffic flow forecasting model with density based clustering algorithm*. In 2019 1st International Informatics and Software Engineering Conference (UBMYK) (pp. 1-5). IEEE.

Aydın, S., Taşyürek, M., & Öztürk, C. (2021). Traffic density estimation using machine learning methods. *Journal of Artificial Intelligence and Data Science*, *1*(2), 136–143.

Bao, X., Jiang, D., Yang, X., & Wang, H. (2021). An improved deep belief network for traffic prediction considering weather factors. *Alexandria Engineering Journal*, *60*(1), 413-420.

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, *54*, 1937-1967.

Bi, H., Ye, Z., & Zhu, H. (2022). Data-driven analysis of weather impacts on urban traffic conditions at the city level. *Urban Climate*, *41*, 101065.

Borchers, T., Wittowsky, D., & Fernandes, R. A. S. (2024). *A Comprehensive Survey and Future Directions on Optimizing Sustainable Urban Mobility*. IEEE Access.

Braz, F. J., Ferreira, J., Gonçalves, F., Weege, K., Almeida, J., Baldo, F., & Gonçalves, P. (2022). Road traffic forecast based on meteorological information through deep learning methods. *Sensors*, *22*(12), 4485.

Ceder, A. (2021). Urban mobility and public transport: future perspectives and review. *International Journal of Urban Sciences*, *25*(4), 455-479.

Cetin Tas, I., & Mungen, A. A. (2021). Prediction of Regional Traffic Intensity with Artificial Neural Networks and Support Vector Machines. *Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi*, *8*(15), 378-390.

Cools, M., Moons, E., & Wets, G. (2010). Assessing the impact of weather on traffic intensity. *Weather, Climate, and Society*, *2*(1), 60-68.

Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, *14*, 241-258.

Gao, Y., & Zhu, J. (2022). Characteristics, Impacts and trends of urban transportation. *Encyclopedia*, *2*(2), 1168-1182.

Gerges, F., Llaguno-Munitxa, M., Zondlo, M. A., Boufadel, M. C., & Bou-Zeid, E. (2024). Weather and the City: Machine Learning for Predicting and Attributing Fine Scale Air Quality to Meteorological and Urban Determinants. *Environmental Science & Technology*, *58*(14), 6313-6325.

Keay, K., & Simmonds, I. (2005). The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. *Accident Analysis & Prevention*, *37*(1), 109-124.

Kendre, H., Karmarkar, M., Karadbhajne, H., Kamtalwar, N., & Marathe, A. (2024, February). *Traffic Volume Prediction Based on Weather Parameters*. In 2024 IEEE International Conference for Women in Innovation, Technology & Entrepreneurship (ICWITE) (pp. 124-129). IEEE.

Khan, A. F., & Ivan, P. (2023). Integrating Machine Learning and Deep Learning in Smart Cities for Enhanced Traffic Congestion Management: An Empirical Review. *Journal of Urban Development and Management*, *2*(4), 211-221.

Lin, P., He, Y., Pei, M., & Yang, R. (2022). Data-driven spatial-temporal analysis of highway traffic volume considering weather and festival impacts. *Travel Behavior and Society*, *29*, 95-112.

Lin, P., Hong, Y., He, Y., & Pei, M. (2024). Advancing and lagging effects of weather conditions on intercity traffic volume: A geographically weighted regression analysis in the Guangdong-Hong Kong-Macao Greater Bay Area. *International Journal of Transportation Science and Technology*, *13*, 58-76.

Lin, Y. C., Lin, Y. T., Chen, C. R., & Lai, C. Y. (2025). Meteorological and traffic effects on air pollutants using Bayesian networks and deep learning. *Journal of Environmental Sciences*, *152*, 54-70.

Liu, Z. L. (2025). Ensemble learning. In *Artificial Intelligence for Engineers: Basics and Implementations* (pp. 221-242). Cham: Springer Nature Switzerland.

Medina-Salgado, B., Sánchez-DelaCruz, E., Pozos-Parra, P., & Sierra, J. E. (2022). Urban traffic flow prediction techniques: A review. *Sustainable Computing: Informatics and Systems*, *35*, 100739.

Mienye, I. D., & Sun, Y. (2022). A survey of ensemble learning: Concepts, algorithms, applications, and prospects. *IEEE Access*, *10*, 99129-99149.

Qi, X., Mei, G., Tu, J., Xi, N., & Piccialli, F. (2022). A deep learning approach for long-term traffic flow prediction with multifactor fusion using spatiotemporal graph convolutional network. *IEEE Transactions on Intelligent Transportation Systems*, *24*(8), 8687-8700.

Shaygan, M., Meese, C., Li, W., Zhao, X. G., & Nejad, M. (2022). Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities. *Transportation Research Part C: Emerging Technologies*, 145, 103921.

Tao, X., Cheng, L., Zhang, R., Chan, W. K., Chao, H., & Qin, J. (2023). Towards Green Innovation in Smart Cities: Leveraging Traffic Flow Prediction with Machine Learning Algorithms for Sustainable Transportation Systems. *Sustainability*, *16* (1), 251.

Theofilatos, A., & Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis & Prevention*, *72*, 244-256.

Utku, A. (2023). Deep Learning Based an Efficient Hybrid Model for Urban Traffic Prediction. *Bilişim Teknolojileri Dergisi*, *16*(2), 107-117.

Wang, S., Cao, J., & Philip, S. Y. (2020). Deep learning for spatio-temporal data mining: A survey. *IEEE Transactions on Knowledge and Data Engineering Transactions on Knowledge and Data Engineering*, *34*(8), 3681-3700.

Wu, J., & Zhang, C. (2024). *The role of temperature and rainfall in traffic congestion: Evidence from 98 Chinese cities*. Weather, Climate, and Society.

Xie, P., Li, T., Liu, J., Du, S., Yang, X., & Zhang, J. (2020). Urban flow prediction from spatiotemporal data using machine learning: A survey. *Information Fusion*, *59*, 1-12.

Yin, X., Wu, G., Wei, J., Shen, Y., Qi, H., & Yin, B. (2022). Deep learning on traffic prediction: Methods, analysis, and future directions. *IEEE Transactions on Intelligent Transportation Systems*, *23*(6), 4927-4943.

Zhang, W., Pan, W., Zhu, X., Yang, C., Du, J., & Yin, J. (2024). Identification of Traffic Flow Spatio-Temporal Patterns and Their Associated Weather Factors: A Case Study in the Terminal Airspace of Hong Kong. *Aerospace*, *11*(7), 531.