Geojournal of Tourism and Geosites
ISSN 2065-0817, E-ISSN 2065-1198

Year XIX, vol. 67, no. 3, 2026, p.1936-1945
DOI 10.30892/gtg.67310-1822

AN INTEGRATED ML AND GEOSPATIAL FRAMEWORK FOR CONTEXT-AWARE SHORT-TERM CONGESTION FORECASTING IN TOURISM-HEAVY URBAN MOBILITY NODE

Enas ELSHEBLI [1], Ferenc ERDŐS [1*], Varadarajan VIJAYAKUMAR [2], Anwar KHAN [3], Bayboboeva FIRUZA [4]

[1] Széchenyi István University, Faculty of Informatics and Electrical Engineering, Department of Informatics, Győr, Hungary; elshebli.enas.abd.almajeed@hallgato.sze.hu (E.E.); erdosf@sze.hu (F.E.)
[2] University of Technology Sydney, Swiss School of Business and Management, Australia, Geneva, Switzerland; Vijayakumar.Varadarajan@uts.edu.au (V.V.)
[3] University of Peshawar, Department of Electronics, Peshawar, KPK, Pakistan; arkhan@uop.edu.pk (A.K.)
[4] Tashkent State University of Economics, Department of Economical and Financial Security, Tashkent, Uzbekistan; f.bayboboeva@tsue.uz (B.F.N.)

Citation: Elshebli, E., Erdős, F., Vijayakumar, V., Khan, A., & Firuza, B. (2026). An integrated ML and geospatial framework for context-aware short-term congestion forecasting in tourism-heavy urban mobility node. *Geojournal of Tourism and Geosites*, 67(3), 1936–1945. https://doi.org/10.30892/gtg.67310-1822

Abstract: Tourism-heavy urban mobility nodes often experience short, intense congestion spikes where visitor surges interact with commuter-related traffic. Despite growing interest in smart mobility management, few studies integrate real-time computer vision sensing with uncertainty-aware machine learning to support operational decision-making at tourism-intensive urban nodes. This study aims to develop and validate a real-time, uncertainty-aware traffic prediction framework that provides actionable, short-horizon (15-minute) congestion intelligence in such areas. Town Quay (Southampton, UK) is used as a case study. Three months of continuous webcam footage (July–September 2025) were processed with a YOLO + SORT pipeline to generate 15-minute vehicle counts. Counts were enriched with contextual features capturing weather (clear/foggy/rainy), day–night cycle, and day type (business vs. leisure). CatBoost and LightGBM were trained using a chronological split (first two months for training, third month for testing). Uncertainty quantification combined conformal prediction intervals with Gaussian Mixture Model (GMM). Traffic volumes varied systematically with context, with the highest average loads occurring during clear-weather daytime business periods. The framework achieved MAE = 43.62, R^2 = 0.902, and CCC = 0.94. These findings demonstrate that integrating computer-vision traffic sensing with contextual feature engineering and uncertainty-aware learning yields reliable, interpretable 15-minute congestion forecasts. This enables proactive, risk-informed interventions (e.g., port-arrival coordination, adaptive traffic control) that help protect the physical carrying capacity and heritage-setting quality of tourism-intensive areas while maintaining everyday urban mobility.

Keywords: traffic prediction, tourism mobility, urban mobility node, geospatial analysis, machine learning, intelligent transport systems, urban sustainability

* * * * * *

INTRODUCTION

Tourism presents a range of challenges for urban environments, including traffic, environmental degradation, and strain on local infrastructure and services (Kibiro & Mwangi, 2025). In many tourist destinations, vehicle congestion can spike dramatically during peak seasons or local events. These surge traffic flows not only slow down vehicle movement and create long queues but also disrupt the daily lives of local residents. This challenge is increasingly documented in mobility-and-tourism studies (Nyong et al., 2024; Artemyev et al., 2025). Ghazali et al. (2019) found a significant correlation between traffic congestion and reduced quality of life in Seri Kembangan, Malaysia, citing frustration, time loss, and increased health risks due to pollution. Within tourism geography, mobility is a central component of destination management, as the efficiency of tourist movement directly affects visitor experience and the spatial distribution of tourism flows. Congestion in touristic areas is therefore not only a transportation problem but also a spatial management challenge. This is particularly evident in historic urban centres, where the form and size of inherited road networks cannot accommodate the passenger flows generated by modern travel and tourism demand, so that mobility constraints become an obstacle to both local development and tourism (Chelabi et al., 2025).

Tourism-intensive urban mobility nodes are transport-network nodes (e.g., gateways/hubs/destinations) where tourist flows concentrate and where tourism peaks can amplify congestion and interact with commuter traffic (Lohmann & Pearce, 2010; Cui & Zhang, 2025). Jo et al. (2016) reported that tourists who experienced more than two hours of congestion were

[*] Corresponding author

Copyright: © 2026 by the authors.This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0) License

more likely to support vehicle bans in touristic zones based on 440 survey responses. Their findings highlight how the spatiality of congestion, arising from specific street layouts, attraction clustering, and pedestrian–vehicle interactions, can shape tourist behavior and attitudes toward mobility policies. This underscores the need to understand congestion as a spatial process embedded in urban morphology rather than as an isolated flow problem. Further studies emphasize that traffic congestion in tourist cities significantly affects tourists' satisfaction. For example, Chong (2016) showed that foreign tourists perceived traffic congestion negatively, which influenced their travel behavior, where many chose destinations within walking distance or avoided congested areas altogether. This behavioral shift highlights how congestion can affect tourist experiences and how it can potentially restructure tourism flows. Moreover, traffic congestion significantly influences tourist consumer behavior, particularly in destinations experiencing rapid tourism growth.

Putra et al. (2025) examined how emotional regulation associated with Stoic principles and the traditional Balinese philosophy of Sad Ripu relates to tourists' responses to traffic congestion. Based on structured questionnaires completed by 200 tourists in Bali, the study found significant relationships among emotional self-regulation, perceptions of traffic congestion, and tourist consumer behaviour. Traffic congestion mediated the relationships between Stoic principles and tourist consumer behaviour and between Sad Ripu control and tourist consumer behaviour. Tourists reporting stronger emotional control also reported less congestion-related stress and more positive tourism experiences, suggesting that congestion affects not only travel conditions but also tourists' emotional responses and behavioural decisions.

To address these challenges, researchers have applied a range of systematic approaches. For instance, Gao et al. (2021) used regression models to demonstrate how different types of tourism land use (such as accommodations and shopping areas) affect traffic patterns, mostly during peak tourism periods. Their findings support targeted land-use planning to manage congestion. Likewise, in their bibliometric analysis, Krabokoukis & Polyzos (2023) found that interdisciplinary collaboration and data-driven strategies are essential for managing tourism-related mobility sustainably. By applying these approaches, cities can mitigate the negative impacts of tourism on traffic while enhancing the quality of life for residents and the experience for visitors. These studies reveal that tourism-related traffic congestion is not only a planning challenge but also a technological challenge. As cities grow and tourist flows intensify, traditional traffic management systems often struggle to keep up with the fast-changing demands of urban mobility. Kulkarni (2024) highlights that, as urban populations rise, conventional traffic management approaches often fail to alleviate congestion effectively. This has led cities to adopt Intelligent Traffic Management Systems (ITMS), which use adaptive technologies to respond to real-time traffic conditions, improving flow and reducing emissions.

To monitor and manage urban mobility, major cities around the world have adopted traffic cameras and CCTV systems. However, these systems are often used for surveillance or retrospective analysis, such as reviewing accidents or traffic violations, rather than as proactive analytical tools for traffic management Fedorov et al. (2019). This reactive approach limits their potential to improve real-time decision-making and congestion mitigation. On the other hand, the advancement of artificial intelligence (AI) is transforming these systems into contextually aware, real-time traffic management tools, an innovation particularly relevant for tourism-heavy urban mobility nodes where seasonal surges and multimodal flows create complex mobility patterns. The integration of AI-driven analytics enhances both efficiency and safety in urban mobility. For instance, Khule et al. (2025) presented a real-time traffic monitoring system using CCTV cameras and the deep learning model You Only Look Once (YOLO), which integrates live dashboards and automated signal adjustments to overcome the limitations of traditional deterministic systems. In their study, Shahi et al. (2024) propose a smart camera network that uses AI and real-time data combination to dynamically manage traffic, contrasting with older systems that rely only on post-event footage review.

Recent studies highlight the transformative impact of AI in evolving traffic systems from reactive mechanisms to predictive and adaptive frameworks. For example, Talaat et al. (2025) employed YOLOv11 for real-time vehicle detection, achieving a mean Average Precision of 92.4%, enabling adaptive signal control and rerouting. Additionally, El Mokhi et al. (2025) examined AI techniques such as fuzzy logic and deep reinforcement learning. Their findings revealed that these techniques can lead to remarkable improvements, including up to 28% reductions in both queue lengths and CO_2 emissions, demonstrating the powerful role of intelligent systems in enhancing urban mobility and sustainability.

While previous research has explored AI-based traffic monitoring, few studies have incorporated tourism-related contextual factors into predictive models. This work addresses that gap by applying machine learning to real-world traffic conditions in a tourism-intensive urban mobility node. Through the analysis of video footage, this study offers practical insights into how cities with high tourist activity can better anticipate and manage traffic challenges, eventually improving mobility for both residents and visitors. From a tourism geography perspective, traffic congestion represents a spatial manifestation of visitor pressure on urban infrastructure. Integrating AI-based traffic forecasting within geospatial frameworks allows for a better understanding of tourist mobility flows and their temporal–spatial variability across destinations. This approach aligns with Stoyanova et al. (2025), who used GIS to assess transport accessibility to cultural sites in Sofia, demonstrating how spatial analysis can inform mobility planning and congestion mitigation in urban tourism contexts.

Tourist areas present unique challenges due to seasonal surges, diverse transport modes, and the need for seamless mobility experiences. By applying intelligent traffic systems in these contexts, we seek to enhance both operational efficiency and the overall tourist experience, contributing to more sustainable and adaptive urban tourism infrastructure.

This study aims to develop a real-time, context-aware traffic forecasting framework for tourism-heavy urban mobility nodes, using computer vision and ensemble learning to predict short-term congestion patterns. Specifically, it seeks to (1) extract high-frequency traffic data through YOLO–SORT detection and tracking, (2) integrate contextual features such as weather, time of day, and day type to enhance predictive accuracy, (3) evaluate model performance using CatBoost and

LightGBM for 15-minute-ahead forecasts, and (4) classify congestion regimes and quantify uncertainty through Gaussian Mixture Models GMM and conformal prediction, providing actionable insights for tourism mobility management. The case study focuses on Southampton, UK, a city whose geography and economy are dominated by its port. Southampton's waterfront functions as a port–city interface where cruise activity, local travel, and service traffic converge on a constrained street network. Town Quay sits between the Old Town and cruise facilities, operating as a tourism-heavy urban mobility node that concentrates arrivals, departures, and pedestrian/vehicle mixing. This setting intensifies peak-period pressure on access and circulation, making congestion a recurring management problem rather than an isolated traffic anomaly.

RESEARCH METHODOLOGY

The research methodology illustrated in Figure 1 follows a structured, data-driven framework for traffic analysis and prediction. The data acquisition and processing stage involves video data collected from the Town Quay webcam undergoing a vehicle detection and tracking pipeline using the YOLO algorithm and the Simple Online Realtime Tracking SORT tracking model, as detailed in Section 2.2. The derived vehicle counts are then combined with contextual data (such as weather conditions, time of day, and day type (business or leisure)) in the contextual data integration phase, as detailed in Section 2.3. Subsequently, this dataset feeds into the predictive modeling framework, where CatBoost and LightGBM are employed. The framework includes model training and evaluation using chronological data splits and assesses performance using metrics such as MAE, RMSE, Bias, and R^2.

Finally, uncertainty quantification is performed using conformal prediction and Gaussian Mixture Model (GMM) thresholds to ensure robustness and reliability in traffic forecasting, as detailed in Section 2.4.

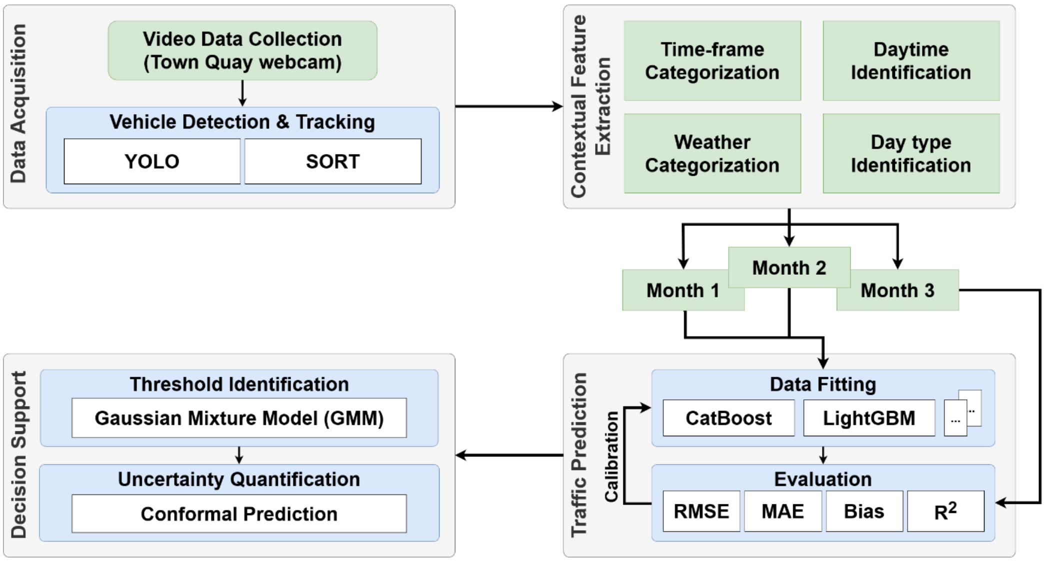

Figure 1. Research Methodology - Grey: Module, Blue: Model, Green: Input

Figure 2. Map illustration of the case study geospatial location, showing geospatial data source (surveillance camera) Town Quay in Southampton, UK.

1. Research Area and Context

This research examines traffic dynamics at Town Quay in Southampton, UK. This tourism-heavy urban mobility node is a vibrant waterfront district where cruise ship operations and seasonal tourism create recurring congestion challenges. It represents a complex and dynamic traffic environment, shaped by the constant flow of visitors, port operations, and local commuters. Cruise ship arrivals bring surges of tourists, while nearby businesses and offices generate steady work-related traffic, creating short-term fluctuations that make congestion management particularly challenging. Figure 2 illustrates the geospatial location of the case study and indicates the geospatial data source (surveillance camera).

Crucially, the study area borders Southampton’s medieval defensive walls and the old town heritage zone. Consequently, traffic queuing at this junction creates a visual and atmospheric barrier between the maritime entry point (the port) and the cultural heritage site (the old town). This spatial proximity classifies Town Quay as a tourism-heavy urban mobility node where modern mobility friction directly degrades the historic landscape's interpretative value. To address these challenges, this research develops a real-time, data-driven framework designed to predict congestion patterns 15 minutes in advance. Continuous video feeds are used to monitor traffic flow, turning raw data into actionable insights.

2. Data Acquisition and Processing

2.1. Video Data Collection

The primary data source consists of real-time video footage obtained from a publicly accessible online webcam located at the Southampton Town Quay junction, as shown in Figure 3. The video footage provides continuous visual coverage of the intersection, capturing diverse traffic conditions under varying temporal and environmental contexts. We collected data for a three-month period between July–September 2025.

(a)  (b)

Figure 3. Screenshots from the Southampton Town Quay Junction live webcam. (a) nighttime, (b) daytime
(Source: Authors’ own processing, July-Septemper, 2025; https://www.skylinewebcams.com/en/webcam/united-kingdom/england/southampton/southampton-town-quay.html)

2.2. Vehicle Detection and Tracking

The machine learning models employed include YOLOv5 for object detection and SORT for vehicle tracking across consecutive video frames. YOLOv5 is selected for its high detection accuracy and computational efficiency, making it suitable for real-time applications (Jaiswal & Agrawal, 2024). SORT complements YOLO by maintaining consistent vehicle tracking across frames.

This methodology demonstrates how machine learning models can be leveraged to analyze real-world traffic dynamics in tourism-heavy areas, offering a scalable and adaptive solution for urban mobility management. In particular, the use of YOLO is highly advantageous compared to traditional manual counting or sensor-based methods. YOLO enables real-time, automated vehicle detection from video feeds, providing accurate, high-frequency data without requiring costly infrastructure or labor-intensive surveys (Adhithya, 2025). By employing machine learning models such as YOLO and SORT, urban planners and local authorities in tourism-heavy areas can obtain timely and accurate insights into mobility patterns. This data-driven approach supports more responsive traffic management, smarter resource allocation, and an overall improved experience for visitors. To ensure reliability, the detection and tracking results were validated against manual counting on randomly selected video samples, confirming the accuracy of the automated approach.

3. Contextual Feature Extraction

In addition to vehicle counts generated by the YOLO–SORT processing pipeline, the research integrates a comprehensive set of contextual variables to capture the multifaceted nature of urban traffic dynamics. By incorporating weather conditions, time-of-day segmentation, and day-type classification, the model achieves a more realistic representation of traffic behavior in real-world environments.

Weather conditions are set as categorical features distinguished between clear, foggy, and rainy periods. Weather plays a crucial role in shaping driving behavior and road capacity (Romanowska & Budzyński, 2022). Rain, for example, can reduce visibility and speed, causing localized slowdowns even in non-peak hours. Integrating weather data enables the model to account for such external perturbations, improving both the accuracy and interpretability of congestion prediction.

In coastal cities like Southampton, which attract large numbers of tourists, especially during warmer months, weather also directly affects touristic traffic patterns. Warmer, sunnier conditions tend to attract more visitors to outdoor attractions, increasing traffic volumes and congestion in key areas (Wilkins & Horne, 2024). Conversely, adverse weather such as heavy rain or fog can deter discretionary travel (Peng et al., 2018), leading to reduced tourist mobility and shifting traffic burdens to indoor venues or alternative transport modes. In this research, weather conditions were systematically extracted from video footage spanning the entire observation period.

The daytime versus nighttime distinction (06:00–18:00 and 18:00–06:00, respectively) captures the cyclical patterns associated with human activity and mobility demand. Daytime periods generally correspond to higher vehicle counts due to commuting, commercial deliveries, and tourism-related travel, whereas nighttime periods exhibit reduced traffic intensity and different vehicle composition (Goerguelue et al., 2025). By including this temporal feature, the model is able to learn context-dependent traffic behaviors and adapt its forecasts to diurnal variations. Moreover, traffic patterns at Town Quay strongly depend on day type, classified as business or leisure. Business days (weekdays) feature regular commuting and logistics, while leisure days (weekends and holidays) coincide with higher tourist movement. This distinction helps the model capture shifts in traffic driven not only by time but also by behavioral patterns.

By combining contextual variables with vehicle counts, the model gains a richer understanding of traffic behavior, enabling accurate short-term predictions and better generalization. This integration transforms it from a purely data-driven tool into a context-aware forecasting system that adapts to dynamic urban conditions. Although based on computer vision, the framework can be extended into GIS workflows to map predicted congestion zones and link them to tourism hotspots.

4. Predictive Modeling Framework

4.1. Model Selection

In this research, two ensemble learning algorithms, CatBoost and LightGBM were analyzed. These models are widely recognized for their ability to handle structured tabular data, capture complex nonlinear relationships, and deliver computational efficiency in large-scale predictive tasks (Bentéjac et al., 2021). CatBoost is particularly effective in managing categorical features and mitigating overfitting (Prokhorenkova et al., 2018), whereas LightGBM is optimized for speed and memory efficiency, making it suitable for high-dimensional datasets for the traffic domain (Cheng et al., 2022).

4.2. Model Data Fitting and Evaluation

The prediction framework produces short-term forecasts of traffic conditions 15 minutes in advance. The dataset includes time-stamped entries with the following explanatory variables: date x_1, time x_2, day type x_3, day/night classification x_4, and weather conditions x_5. The target variable y is the observed vehicle count in each 15-minute interval. The analysis follows a temporal split: Data from the first two months, $t_1$and t_2, are used for model fitting, while data from the third month, t_3, are reserved for evaluation. Let n denote the number of days in $t_1 \cup t_2$. Since each day contains 96 intervals of 15 minutes, the training feature matrix is $X = (\mathbf{x}_i)_{i=1}^{96n}$, where each feature vector is $\mathbf{x}_i = \{x_1^i, x_2^i, x_3^i, x_4^i, x_5^i\}$, and the corresponding target value is the vehicle count y_i. For validation, let m denote the number of days in t_3. The evaluation matrix is $\ddot{X} = (\ddot{\mathbf{x}}_i)_{i=1}^{96m}$, constructed analogously from the same five explanatory variables. This formulation follows the standard supervised-learning design-matrix representation for regression modelling (James et al., 2021).

Both models were evaluated using standard regression performance metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Bias, and the Coefficient of Determination (R^2). These metrics collectively evaluate prediction accuracy, quantify error magnitude, and measure the models' ability to generalize to unseen data.

4.3 Conformal Prediction and Gaussian Mixture Model (GMM) Integration

Following the selection of CatBoost as the optimal model, a conformal prediction framework is implemented to extend point forecasts into uncertainty-aware outputs. Traffic congestion levels are identified using a GMM applied to the predicted traffic volume distributions. The model estimated five components, each representing a distinct congestion regime, with transition thresholds at approximately 113.4, 240.6, 363.8, and 441.0 vehicle counts. Conformal prediction intervals are computed alongside point forecasts to quantify uncertainty, adapting dynamically to traffic intensity. Narrow intervals during low-flow periods indicate high confidence, whereas wider intervals during peak congestion reflect increased variability. This combined approach enables both probabilistic uncertainty assessment and multi-level classification of traffic states, supporting robust interpretation of congestion dynamics.

Model alignment with observed traffic volumes is assessed using three complementary statistical measures: Pearson's correlation coefficient (r), Spearman's rank correlation (ρ), and the Concordance Correlation Coefficient (CCC). Pearson and Spearman correlations are computed to assess relational strength and ranking preservation, whereas the CCC serves as an integrated measure of accuracy and precision. Comparative analysis across CatBoost and LightGBM models enables identification of the model with superior concordance for subsequent uncertainty quantification and risk classification.

RESULTS AND ANALYSIS

1. Traffic Volume Trends Derived from YOLO-SORT Processing

This section presents findings from the YOLO-SORT model applied to Town Quay traffic footage. The analysis examines how contextual factors (day type, weather, and time of day) affect congestion, measured by vehicle count. The dataset spans three months, with traffic aggregated at 15-minute intervals. Each record includes attributes such as date, day type (business or leisure), time of day (daytime or nighttime), and weather (clear, foggy, or rainy).

As shown in Table 1, on business days, daytime traffic under clear weather conditions exhibits the highest total volume (572,913 vehicles), with an average of 265.23 vehicles per 15-minute interval, reflecting the intensity of commuter flows and commercial transport activity in this port-adjacent urban corridor. Nighttime traffic on business days declines sharply, consistent with reduced professional mobility outside working hours. Leisure days show elevated traffic volumes during daytime hours, particularly under clear and foggy conditions, suggesting strong touristic engagement with the surrounding attractions. The relatively high average vehicle counts during foggy leisure periods (233.28 vehicles per interval) indicate that recreational travel persists even in suboptimal weather, likely driven by planned visits and seasonal tourism demand. The high traffic on 'foggy' leisure days suggests tourists are less deterred by weather than commuters, a key behavioral insight for tourism and transport managers. Nighttime traffic on leisure days, while lower, remains notable and may be attributed to evening leisure activities, hospitality services, and ferry-related movements. These findings underscore the dual function of Town Quay as both a commuter hub and a tourism gateway.

The observed traffic distribution suggests that leisure periods, typically associated with increased tourist activity, exert a notable influence on vehicular flow. Daytime traffic volumes on leisure days under clear weather conditions (average 220.95 vehicles per 15-minute interval) approach those recorded on business days (265.23 vehicles), despite the absence of routine commuting. This pattern indicates that tourism-driven mobility substantially contributes to daytime congestion, particularly in favorable weather scenarios conducive to recreational travel. Conversely, nocturnal traffic remains consistently low across all conditions, reflecting limited tourist movement during late hours. Although adverse weather conditions, such as rain and fog, reduce overall traffic volumes, the persistence of leisure-related travel under these circumstances suggests that tourism demand is less sensitive to meteorological variability than business-related commuting.

Table 1. Traffic Volume Summary by Day Type, Time of Day, and Weather Conditions

Day Type	Time of the Day	Weather Condition	Total Vehicle Count	Average Vehicle Count per 15 Minutes
Business	Day	Clear	572913	265.23
		Foggy	27212	188.97
		Rainy	190443	220.42
	Night	Clear	143521	66.44
		Foggy	6733	46.75
		Rainy	48746	56.41
Leisure	Day	Clear	194439	220.95
		Foggy	7465	233.28
		Rainy	62625	186.38
	Night	Clear	46538	52.64
		Foggy	1389	49.60
		Rainy	14935	44.44

Table 2. Comparative Performance Metrics of CatBoost and LightGBM Models

Model	RMSE	MAE	Bias	R^2
CatBoost	43.62	28.06	3.59	0.902
LightGBM	46.12	28.40	0.89	0.891

2. Predictive Performance of Ensemble Learning

The predictive results confirm the effectiveness of ensemble learning in capturing traffic congestion dynamics in tourism-intensive urban settings. As shown in Table 2, the CatBoost model outperformed LightGBM achieving lower RMSE and MAE values alongside a higher R^2, indicating stronger explanatory power. Although LightGBM exhibited slightly lower bias compared to CatBoost, both models maintained minimal error levels. As illustrated in Figure 4, the one-to-one plots confirm that both models capture the overall relationship between observed and predicted vehicle counts.

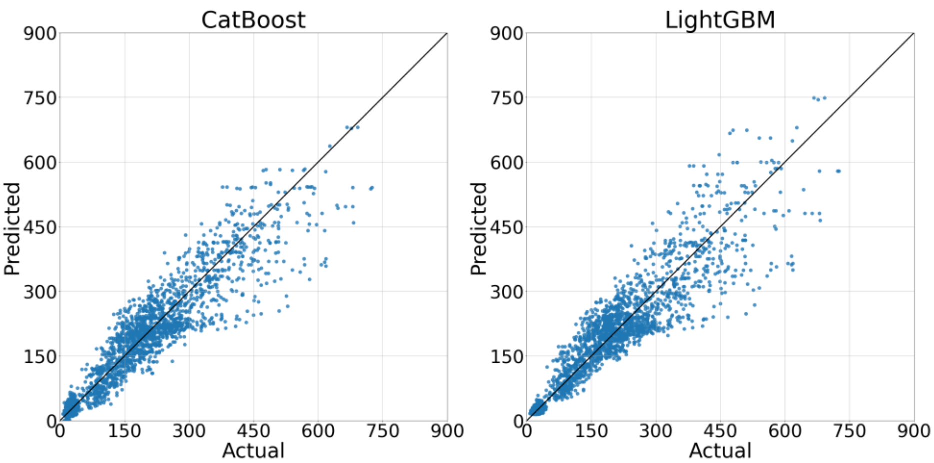

Figure 4. One-to-One Comparison of Predicted Traffic Volumes: CatBoost vs. LightGBM (Source: Authors' own processing, 2025)

The majority of points align closely with the identity line, indicating limited deviations across most of the range. The most noticeable discrepancies appear at higher traffic volumes, where both models tend to slightly misestimate actual counts. CatBoost provides slightly closer adherence to the identity line in this upper range, consistent with its stronger fit, whereas LightGBM produces a smoother trend but with marginally larger deviations from the observed values.

Further insights are provided in Figure 5, which compares the kernel density estimates of the observed and predicted distributions. The bandwidth for these estimates is determined using Silverman's rule of thumb (Silverman, 1986) to achieve an optimal balance between smoothness and detail. The close alignment between the actual and predicted curves indicates that both algorithms accurately reproduce the underlying traffic volume patterns. However, CatBoost demonstrates marginally superior performance. CatBoost's predicted distribution aligns closely with actual data, especially in the lower range (0–150), where peaks nearly overlap, indicating better capture of fine traffic variations. LightGBM produces a smoother curve but slightly underestimates the first peak and over-smooths around the second mode (~250 vehicles). Overall, both models generalize well, but CatBoost offers marginally higher accuracy and sensitivity to local variations.

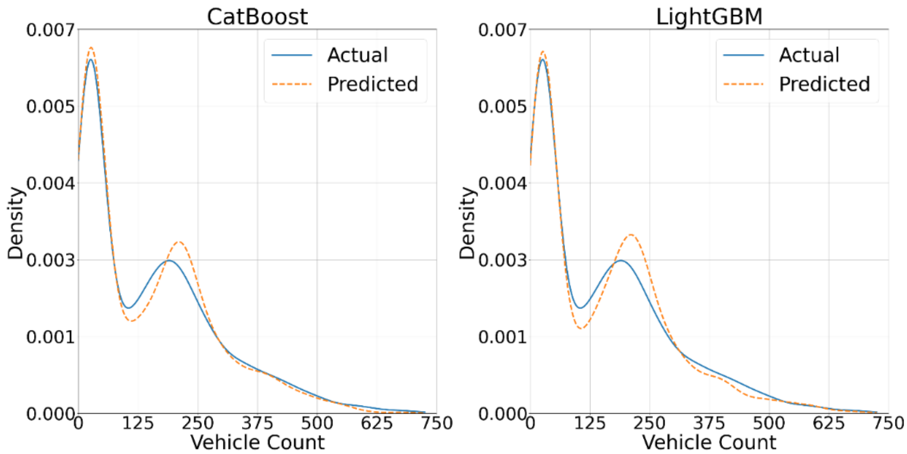

Figure 5. Kernel Density Estimates of Traffic Volume Distributions: Observed vs. Predicted (CatBoost and LightGBM)
(Source: Authors' own processing, 2025)

These outcomes suggest that the framework effectively incorporates predictive models to capture congestion dynamics under conditions of variability and peak demand. The ability of both methods to approximate the underlying distributions while keeping systematic bias low ensures that the predictions provide a realistic basis for subsequent analysis. Importantly, the models are not the primary focus of evaluation but serve to demonstrate that the framework operates reliably with advanced machine learning tools. This establishes the foundation for extending the approach toward higher-level tasks, such as translating continuous predictions into categorical assessments of congestion risk.

3. Threshold Identification and Uncertainty Quantification

The GMM identifies four transition thresholds at approximately 113.4, 240.6, 363.8, and 441.0 vehicle counts, corresponding to five distinct congestion levels. The fitted component statistics are summarized in Table 3. The component means (35.9, 178.5, 286.6, 425.0, and 607.8) increase monotonically with traffic intensity, whereas the mixture weights show that most observations belong to lower and mid-level regimes. The relatively uniform variance across components indicates consistent dispersion within the identified traffic states.

Table 3. Statistical properties of the fitted Gaussian components used to define congestion thresholds

Component	Mean Count	Variance	Weight	Congestion Level
0	35.9	1,394.4	0.503	Level 1 (Very Low)
1	178.5	1,394.4	0.265	Level 2 (Low)
2	286.6	1,394.4	0.142	Level 3 (Moderate)
3	425.0	1,394.4	0.064	Level 4 (High)
4	607.8	1,394.4	0.026	Level 5 (Severe)

The conformal prediction intervals show that the majority of observed traffic counts fall within the estimated uncertainty bounds, indicating consistent calibration across all congestion regimes. The intervals adjust adaptively to variations in traffic intensity: they remain narrow during stable, low-traffic periods, reflecting lower predictive uncertainty, and expand considerably during high-density periods to capture the increased variability typical of peak congestion.

The multi-level classification derived from the GMM thresholds provides a finer segmentation of traffic conditions compared to the conventional three-class (Low–Medium–High) scheme. The first two levels correspond primarily to nighttime and off-peak hours, where traffic flow is smooth and model confidence remains high. Intermediate levels represent transitional states, often associated with the buildup and dissipation phases of morning and evening peaks. The highest level captures sustained congestion, where traffic counts are high and prediction intervals widen substantially, reflecting the increased uncertainty inherent in unstable flow conditions.

4. Alignment with Ground Truth

To assess model fidelity, we evaluated concordance between predicted and observed traffic volumes using Pearson's correlation (r), Spearman's rank (ρ), and the Concordance Correlation Coefficient (CCC). CatBoost achieved $r = 0.951$, $\rho = 0.921$, and $CCC = 0.949$, indicating excellent agreement and strong preservation of rank and scale. LightGBM also performed well ($r = 0.945$, $\rho = 0.929$, $CCC = 0.944$), with slightly better rank ordering but marginally lower overall concordance. Both models align closely with ground truth, confirming the effectiveness of gradient boosting for traffic prediction. CatBoost's higher CCC and Pearson values support its selection for uncertainty quantification and conformal prediction.

DISCUSSION

The proposed framework effectively captures how contextual and temporal factors shape traffic dynamics in tourism-heavy urban environments. Variations in vehicle counts across day types, weather, and time periods reveal the multidimensional nature of congestion at Town Quay, where tourist surges and commuter peaks overlap. Increased daytime traffic on business days reflects commuter flows linked to the port and city center, while sustained volumes on leisure days (even under foggy conditions) underscore tourism's resilience and socio-economic significance.

A key behavioral insight is the persistence of high vehicle counts during foggy leisure days. Despite adverse weather conditions, leisure-driven mobility remains strong (average vehicle count = 233.28 per 15-minute interval). This indicates hedonic resilience among tourists: even when travel is discretionary, fixed schedules such as cruise or ferry departures drive mobility. For tourism and transport managers, this underscores that traffic mitigation cannot rely on unfavorable weather to naturally suppress demand; proactive interventions are required to maintain site accessibility and carrying capacity.

From a predictive perspective, CatBoost outperformed LightGBM, demonstrating the value of categorical feature handling and contextual variables for accurate short-term forecasting. Its lower RMSE and higher R^2 indicate strong sensitivity to local traffic fluctuations, though slight underestimation at peak volumes reflects the stochastic nature of congestion during cruise arrivals or weather disruptions.

The integration of conformal prediction and GMM-based classification adds significant interpretive value. Narrow intervals during off-peak periods and wider intervals during volatile conditions demonstrate adaptive confidence estimation, crucial for real-time decision-making in tourism-heavy areas. the GMM thresholds translate continuous predictions into actionable congestion levels, supporting interpretable alerts for urban traffic management and tourist routing.

High concordance metrics (Pearson, Spearman, $CCC > 0.92$) validate predictive reliability, confirming that the model captures both magnitude and structural relationships in traffic data. By differentiating leisure-day (tourist-driven) traffic from business-day (commuter-driven) traffic, the framework enables planners to implement dynamic routing strategies during cruise arrivals, reducing resident frustration and environmental impacts.

These tourism-sensitive insights have theoretical and practical implications: for urban tourism geography, congestion is conceptualized as a spatiotemporal manifestation of visitor pressure, revealing patterns that inform sustainable destination planning; for destination management, short-term predictive insights allow adaptive routing, smart signal control, and proactive mobility management; and for policymakers, the evidence supports event-based traffic interventions that maintain carrying capacity of a mobility node without reducing tourist satisfaction. In practical terms, city planners could use the 15-minute forecasts to trigger dynamic traffic routing on cruise ship arrival days and integrate predictive congestion analytics into urban tourism management systems to support adaptive visitor routing, reduce environmental stress on tourism corridors, and enhance the destination's sustainability profile.

However, this study has two key limitations. First, classifying days as simply "business" or "leisure" ignores hybrid scenarios (such as cruise arrivals on weekdays) where tourist flows spike despite being business days.

Second, reliance on video-based vehicle counts excludes pedestrian and multimodal traffic, which are critical in tourism-heavy urban mobility nodes. Future research should adopt event-based categorization and integrate multimodal mobility data to better capture complex demand patterns.

CONCLUSION

Conceptually, this study reframes congestion at a tourism-intensive urban mobility node as a spatiotemporal manifestation of visitor pressure on maritime built heritage. By integrating computer vision, contextual features and uncertainty-aware forecasting, the framework extends previous tourism mobility studies that relied on static accessibility indices or survey data. Instead of treating traffic as an exogenous constraint, congestion becomes a measurable, short-term indicator of node carrying capacity that can be monitored and anticipated in real time.

This research demonstrates the effectiveness of a geospatial machine learning framework for modeling and forecasting traffic dynamics at a tourism-heavy urban mobility node. Using computer vision-based detection (YOLO–SORT), contextual enrichment, and ensemble learning, the framework captures distinct mobility patterns: commuter-driven peaks on business days and sustained tourism flows on leisure days. A notable insight, the persistence of high traffic during foggy leisure periods, reveals tourists' lower sensitivity to adverse weather, underscoring the need for proactive traffic management even under suboptimal conditions.

CatBoost emerged as the most accurate predictive model, complemented by conformal prediction and GMM-based classification for uncertainty quantification and congestion categorization. These components collectively provide a scalable, interpretable solution for short-term traffic forecasting in complex urban environments. However, limitations include the binary day-type classification, which overlooks hybrid scenarios such as cruise arrivals on weekdays, and the exclusion of pedestrian and multimodal flows. Future research should incorporate event-based categorization, multimodal mobility indicators, and real-time data streams to enhance predictive accuracy.

Integration with GIS dashboards and advanced geo-visualization tools could further support adaptive traffic management, eventually contributing to a 'Digital Twin' of the tourism-intensive urban mobility node. This would allow destination managers to simulate the impact of cruise schedules on carrying capacity before the vessels even arrive.

Author Contributions: Conceptualization, E.E. and E.F.; methodology, E.E. and A.K.; software, E.E.; validation, E.E. and E.F.; formal analysis, E.E. and A.K.; investigation, E.E. and E.F.; data curation, E.E. and E.F. and A.K.; writing - original draft preparation, E.E. and E.F.; writing - review and editing, V.V. and A.K. and B.F.; visualization, E.E. and B.F.; supervision, E.F. and V.V.; project administration, B.F. All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study may be obtained on request from the corresponding author.

Acknowledgements: The research undertaken was made possible by the equal scientific involvement of all the authors concerned.

Conflicts of Interest: The authors declare no conflict of interest.

Use of Generative AI: Not applicable.

REFERENCES

Adhithya, N. (2025). Real-time vehicle detection and counting for traffic management system. In S. J. Nanda, R. P. Yadav, A. H. Gandomi, & M. Saraswat (Eds.), *Data Science and Applications. ICDSA 2024. Lecture Notes in Networks and Systems, vol. 1266*. Springer, Singapore. https://doi.org/10.1007/978-981-96-2647-2_23

Artemyev, A., Abdreyeva, S., Baiburiev, R., Beisakhmet, A., Kaliyeva, A., Abzhapparova, A., Savanchiyeva, A., & Dávid, L. D. (2025). Study of the tourist and recreational load in the “Kolsai Lakes” State National Nature Park, Kazakhstan. *Geojournal of Tourism and Geosites, 58*(1), 9–17. https://doi.org/10.30892/gtg.58101-1386

Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review, 54*(3), 1937–1967. https://doi.org/10.1007/s10462-020-09896-5

Chelabi, O., Ziane, M., & Chadli, M. (2025). The issue of mobility and the road network within cities and their surrounding areas: Case study of the city of Sour El Ghozlane, Bouira Province, Algeria. *Geojournal of Tourism and Geosites, 59*(2), 714–724. https://doi.org/10.30892/gtg.59218-1450

Cheng, W., Li, J. L., Xiao, H. C., & Ji L.–(2022). Combination predicting model of traffic congestion index in weekdays based on LightGBM-GRU. *Scientific Reports, 12*(1), 2912. https://doi.org/10.1038/s41598-022-06975-1

Chong, K. L. (2016). The impact of traffic congestion on tourist behaviour: Case study of Chiang Mai. *Proceedings of the 14th APacChrie Conference,* 290–300, Bangkok, Thailand.

Cui, R., & Zhang, J. (2025). Exploring the relationship between the built environment and spatiotemporal heterogeneity of urban traffic congestion during tourism peaks: A case study of Harbin, China. *ISPRS International Journal of Geo-Information, 14*(12), 470. https://doi.org/10.3390/ijgi14120470

El Mokhi, C., Erguig, H., Hmina, N., & Hachimi, H. (2025). Intelligent traffic management systems: A literature review on AI-based traffic light control. In *The Future of Urban Living: Smart Cities and Sustainable Infrastructure Technologies,* 154–171, Springer, Cham. https://doi.org/10.1007/978-3-031-98334-4_15

Fedorov, A., Nikolskaia, K., Ivanov, S., Shepelev, V., & Minbaleev, A. (2019). Traffic flow estimation with data from a video surveillance camera. *Journal of Big Data, 6*(73), 1–14. https://doi.org/10.1186/s40537-019-0234-z

Gao, Y., Liao, Y., Wang, D., & Zou, Y. (2021). Relationship between urban tourism traffic and tourism land use: A case study of Xiamen Island. *Journal of Transport and Land Use, 14*(1), 761–776. https://doi.org/10.5198/jtlu.2021.1799

Ghazali, W. N. W. B. W., Zulkifli, C. N. A. B., & Ponrahono, Z. (2019). The effect of traffic congestion on quality of community life. In *Carving the Future Built Environment: Environmental, Economic and Social Resilience. European Proceedings of Multidisciplinary Sciences, 2,* 759–766. Future Academy. https://doi.org/10.15405/epms.2019.12.77

Goerguelue, M. E., Kostorz-Weiss, N., Voss, A. S., Kagerbauer, M., & Vortisch, P. (2025). Effects of daytime vs. nighttime on travel mode choice and use patterns: Insights from a ride-pooling survey in Germany. *Applied Sciences, 15*(14), 7774. https://doi.org/10.3390/app15147774

Jaiswal, S. K., & Agrawal, R. (2024). A comprehensive review of YOLOv5: Advances in real-time object detection. *International Journal of Innovative Research in Computer Science and Technology, 12*(3), 75–80. https://doi.org/10.55524/ijircst.2024.12.3.12

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R (2nd ed.).* Springer, New York, USA. https://doi.org/10.1007/978-1-0716-1418-1

Jo, M. W., Kim, H. & Shin, H. J. (2016). Understanding traffic congestion to improve tourist satisfaction in local tourism. *International Journal of Tourism and Hospitality Research, 30*(4), 85–95. https://doi.org/10.21298/IJTHR.2016.04.30.4.85

Khule, P., Mendhkar, V., Shinde, S., Shinde, S., & Chikane, Y. R. (2025). Real-time traffic monitoring and management using video surveillance. *International Journal of Advanced Research in Science, Communication and Technology, 5*(11), 322–327. https://doi.org/10.48175/IJARSCT-25851

Kibiro, E., & Mwangi, M. (2025). Towards sustainable urban tourism: Addressing environmental degradation in Nairobi, Kenya. *International Journal of Tourism & Hospitality Reviews, 12*(2), 22–29. https://doi.org/10.18510/ijthr.2025.1222

Krabokoukis, T., & Polyzos, S. (2023). A bibliometric analysis of integrating tourism development into urban planning. *Sustainability, 15*(20), 14886. https://doi.org/10.3390/su152014886

Kulkarni, R. R. (2024). Evaluating the impact of intelligent traffic management systems on urban congestion and emission levels. *International Journal of Transportation Engineering and Traffic System, 10*(2), 15–19.

Lohmann, G., & Pearce, D. G. (2010). Conceptualizing and operationalizing nodal tourism functions. *Journal of Transport Geography, 18*(2), 266–275. https://doi.org/10.1016/j.jtrangeo.2009.05.003

Nyong, A. E., Ofem, N. O., Umo, U. A., Ebagu, T. E., Benjamin, I. J., Ben, A. O., Arrey, M. E. V., Ebong, E., Ebiala, P. O., Okoro, L. A., Nwobu, E. N., Eja, I. E., James, D. D., Utitofon, I. V., & Ndubuisi, I. G. (2024). Exploring intra-city transportation system as a catalyst for sustainable tourism development in Calabar, Nigeria. *Geojournal of Tourism and Geosites*, *56*(4), 1856–1866. https://doi.org/10.30892/gtg.56439-1353

Peng, Y., Jiang, Y., Lu, J., & Zou, Y. (2018). Examining the effect of adverse weather on road transportation using weather and traffic sensors. *PLOS ONE, 13*(10), e0205409. https://doi.org/10.1371/journal.pone.0205409

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems 31*, 6639–6649. Curran Associates Inc. https://doi.org/10.48550/arXiv.1706.09516

Putra, K. W., Suidarma, I. M., & Maitri, W. S. (2025). Understanding the influence of traffic congestion on tourist consumer behaviour: The role of Stoicism and Sad Ripu control in maintaining sustainable tourism. *Space and Culture, India, 13*(1), 146–162. https://doi.org/10.20896/af6s8z54

Romanowska, A., & Budzyński, M. (2022). Investigating the impact of weather conditions and time of day on traffic flow characteristics. *Weather, Climate, and Society, 14*(3), 823–833. https://doi.org/10.1175/WCAS-D-22-0012.1

Shahi, G., Gupta, M., Sharma, M. K, & Khare, M. D. (2024). Smart traffic management system using networked CCTV smart cameras. In *Futuristic Trends in IoT 3*, 218–231, IIP Series. https://doi.org/10.58532/v3bdio3p2ch7

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman and Hall, London, UK.

Stoyanova, V., Kazakov, B., Ravnachka, A., Patarchanova, E., Roukova, P., & Gartsiyanova, K. (2025). GIS-based assessment of the transport accessibility of the material anthropogenic tourist resources in Sofia, Bulgaria. *Geojournal of Tourism and Geosites, 61*(3), 1804-1813. https://doi.org/10.30892/gtg.61339-1549

Talaat, F. M., El-Balka, R. M., Sweidan, S., Gamel, S. A., & Al-Zoghby, A. M. (2025). Smart traffic management system using YOLOv11 for real-time vehicle detection and dynamic flow optimization in smart cities. *Neural Computing and Applications, 37*(24), 19957–19974. https://doi.org/10.1007/s00521-025-11434-9

Wilkins, E. J., & Horne, L. (2024). Effects and perceptions of weather, climate, and climate change on outdoor recreation and nature-based tourism in the United States: A systematic review. *PLOS Climate, 3*(4), e0000266. https://doi.org/10.1371/journal.pclm.0000266

Article history: Received: 16.12.2025 Revised: 05.04.2026 Accepted: 30.06.2026 Available online: 24.07.2026