www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 16(4), Sep-Dec 2025, 212-232. ISSN: 2007-1558.2
https://doi.org/10.61467/2007.1558.2025.v16i4.1021

Adaptive Traffic Congestion Prediction Using LSTM Networks and Physics-Based Models

Ernesto de la Cruz-Nicolás[1,2], Hugo Estrada-Esquivel[1], Alicia Martínez-Rebollar[1], Odette Pliego-Martínez[1,3], Eddie Clemente[1]
[1] Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico
[2] Tecnológico Nacional de México/Tecnológico de Cuautla
[3] Tecnológico Nacional de México/Tecnológico de Milpa Alta
{d21ce090, hugo.ee, alicia.mr, d21ce092, eddie.ct}@cenidet.tecnm.mx

Abstract. Accurately modelling and predicting traffic congestion is essential for effective traffic management in dynamic urban environments. This study introduces a hybrid model that combines a deep neural network architecture, designed to capture long-term temporal dependencies, with physics-based approaches to traffic modelling. The deep learning component identifies complex temporal patterns and non-linear behaviours in traffic evolution, while the physics-based component incorporates dynamic constraints that enhance the model's robustness and ability to generalise. This integration leverages the predictive power of machine learning while maintaining consistency with the fundamental principles governing vehicle flow. Experiments conducted using real-world traffic data demonstrate strong predictive performance, establishing a solid foundation for the development of intelligent transportation management systems and providing advanced tools to support adaptive mobility optimisation in congested urban settings.
Keywords: Traffic Prediction, LSTM Neural Networks, Hybrid Modeling, Traffic Congestion, Traffic Optimization

Article Info
Received March 14, 2025
Accepted June 14, 2025

1 Introduction

Traffic congestion is a major challenge in urban transportation, negatively impacting mobility, air quality, and economic productivity. Its complexity arises from the nonlinear and dynamic behavior of traffic systems, influenced by driver behavior, infrastructure, and unexpected events. Addressing this problem demands advanced models capable of accurate forecasting and adaptive control to support real-time decision-making.

In this context, deep learning, particularly Long Short-Term Memory (LSTM) networks, has shown promise due to its ability to capture temporal dependencies in traffic time-series data (Villarroya et al., 2022). Unlike traditional machine learning methods, which often struggle with generalization in highly variable conditions, LSTMs can effectively learn and anticipate recurring congestion patterns.

Simultaneously, physics-based traffic models such as METANET and Nagel-Schreckenberg have played a foundational role in traffic simulation. METANET, a macroscopic model based on differential equations, simulates speed, density, and flow evolution. It is widely used for highway control due to its computational efficiency and effectiveness in applications like dynamic speed limits and ramp metering. However, it struggles to account for sudden disruptions like accidents or lane closures (Wang et al., 2022). In contrast, the Nagel-Schreckenberg model, based on cellular automata, simulates vehicle interactions using simple rules for acceleration, braking, and safe distance. It reproduces emergent phenomena such as spontaneous jams and stop-and-go waves (Gupta & Santhanam, 2021), offering scalability for large simulations but lacking the ability to model continuous dynamics or external factors like weather or infrastructure quality.

To overcome these limitations, recent research explores hybrid models that integrate deep learning with physics-based approaches. Combinations of CNNs, RNNs, Restricted Boltzmann Machines (RBMs), and Stacked Autoencoders (SAEs) with traffic physics components have demonstrated improvements in feature extraction and traffic evolution modeling (Kashyap et al., 2022).
This article proposes a hybrid model that combines LSTM networks with physics-based models such as METANET and Nagel-Schreckenberg. LSTMs capture historical patterns and latent trends, while the physics-based components ensure structural consistency and adherence to traffic flow principles. This synergy enhances predictive performance and enables more effective adaptive control strategies in complex urban environments. Comparative experiments demonstrate that integrating deep learning with physics-based models significantly improves forecasting accuracy and responsiveness, supporting more efficient and sustainable urban traffic management.

2 Related work

Traffic prediction and adaptive control have been intensively studied in recent decades, driven by the increasing demands of urban mobility and the need for efficient traffic management (Bridge, 2024). Three main approaches dominate the literature: deep learning models, physics-based methods, and hybrid frameworks combining both paradigms. This section reviews representative studies in each category, outlining their contributions, limitations, and evolution toward integrative models such as the one proposed here.

The first line of research focuses on neural network-based prediction, especially LSTM architectures, due to their ability to model temporal patterns. Villarroya Sánchez (2021) applied LSTM to forecast traffic in Valencia, outperforming Multilayer Perceptrons, particularly on congested roads. Ascencio et al. (2022) developed a real-time LSTM system effective in handling long time sequences. Zhao et al. (2017) used LSTM with urban sensor data, achieving precise predictions by capturing temporal patterns. Nuli et al. (2022) extended this to short- (15 min) and medium-term (1 h) forecasts with real-time data. Similarly, Rivera-Picado and Meneses-Guzmán (2022) analyzed hourly vehicle flow on Route 27 (San José–Caldera) using intelligent transport system models. While these methods offer high accuracy, they face limitations in interpretability and causal inference, restricting their application in real-time traffic control.

The second approach involves physics-based models grounded in macroscopic dynamics and conservation principles. METANET (Pan et al., 2022) and the Nagel-Schreckenberg model (Harsono & Arai, 2024) are widely used to simulate traffic flow. Shehzad Khattak et al. (2021) explored flow–density–speed relationships to assess stability under congestion. Romanowska and Jamroz (2021) employed differential equations to analytically model traffic congestion. From a control theory perspective, Uppaluru et al. (2022) introduced a continuous-time congestion management framework, while Rastgoftar and Jeannin (2021) proposed a finite-state abstraction rooted in physical laws. Ziad et al. (2020) conducted an empirical analysis of congestion in Guayaquil, identifying its causes and proposing mitigation measures. Despite their theoretical rigor, these models often oversimplify real-world variability, limiting predictive accuracy.

A third and more recent trend focuses on hybrid models that combine data-driven learning with physics-based constraints. Khan et al. (2023) proposed a hybrid LSTM–simulation model that improved congestion estimation. Feroz Khan and Ivan (2023) emphasized long-term strategies for urban traffic control. Florentino and Gismondi (2024) implemented a traffic light control system using YOLOv4 and heuristics. Itu and Danescu (2024) used CNNs to estimate traffic speed, stressing socioeconomic relevance. More recent architectures integrate multiple components: Wang and Chen (2025) combined Transformers with TCNs for global and local dependency capture; Yang et al. (2024) proposed TARGCN, merging temporal attention, GRUs, and graph convolutions; Chen et al. (2025) introduced STFGCN, integrating weather and external factors; Wei et al. (2024) developed TSGDC with channel attention and sequence decomposition; and Wu et al. (2024) presented STNE, fusing GCNs with multidimensional LSTMs, outperforming conventional baselines.

Building on these advances, the model proposed in this study combines LSTM networks with physics-based modeling, aiming to overcome the limitations of previous approaches in terms of accuracy and interpretability. Unlike earlier studies that focused solely on either prediction or theory, this hybrid framework enables forecasts up to 24 hours in advance. It utilizes a high-resolution urban mobility dataset from the Tlalpan Borough, recording data every five minutes, 24/7. This rich dataset supports accurate trend detection, dynamic adaptation to traffic demand, and the development of real-time control strategies.

Furthermore, this work aligns with recent contributions: Chawla et al. (2024) and Ravichandran (2024) leverage big data and machine learning for real-time congestion prediction, exploiting spatiotemporal features and scalable streaming pipelines; Kaleem et al. (2024) propose energy-efficient models for smart transportation, targeting reduced power use and emissions in edge/cloud settings; Khatri et al. (2021) assess ML approaches in VANETs, benchmarking routing, prediction, and reliability under dynamic

topologies; León et al. (2022) review AI in network data environments, surveying methods for large-scale telemetry and noting challenges in heterogeneity, privacy, and bias; Mauricio and Torres (2023) detail Lima's traffic light control system, outlining system architecture, coordination strategies, and observed operational improvements; and Rehmat Ullah et al. (2021) compare vehicular traffic simulation platforms, contrasting microscopic vs. macroscopic tools, calibration workflows, and scalability trade-offs.

3 Methodology

This section outlines the methodology used to develop the traffic congestion prediction model, which integrates a deep neural network designed for sequence modeling with a physics-based representation of traffic dynamics. The process is divided into four main phases: data collection, preprocessing, model development, and prediction implementation. The first phase involves defining the case study and gathering relevant data, focusing on identifying sources that effectively characterize urban traffic behavior. Key variables such as vehicle speed, congestion levels, traffic density, and roadway incidents are collected to ensure the model accurately reflects real-world traffic conditions.

In the second phase, the data undergoes preprocessing and exploratory analysis to refine and prepare it for modeling. This step includes detecting and removing outliers to improve consistency, normalizing variables to facilitate comparisons across different scales, and identifying behavioral patterns that inform and optimize the modeling process. The third phase centers on developing the physics-based model, which describes the evolution of traffic speed through a differential equation. This equation captures essential elements of traffic dynamics, including recovery toward free-flow speed, the impact of congestion, and speed reductions caused by road incidents. To enable integration with the neural network architecture, the equation is discretized using Euler's method, ensuring compatibility with deep learning frameworks.

In the fourth phase, the predictive model is constructed by embedding the discretized physical formulation within the neural network structure. This integration allows the model to generate predictions based not only on historical traffic data but also on the underlying physical laws that govern traffic behavior. By combining data-driven learning with theoretical consistency, this hybrid approach improves predictive accuracy and ensures that model outputs remain grounded in realistic traffic dynamics. The following sections provide a detailed explanation of each phase within this methodology.

3.1 Definition of the case study and data collection

This study focuses on the Tlalpan Borough of Mexico City, a critical zone for metropolitan mobility due to its intricate road network and its connection to major traffic corridors such as Periférico Sur, Insurgentes Sur, and Calzada de Tlalpan. The importance of this area stems from the high traffic volumes it experiences during peak hours, which significantly impact vehicle flow and exacerbate urban mobility challenges, as illustrated in Fig. 1.

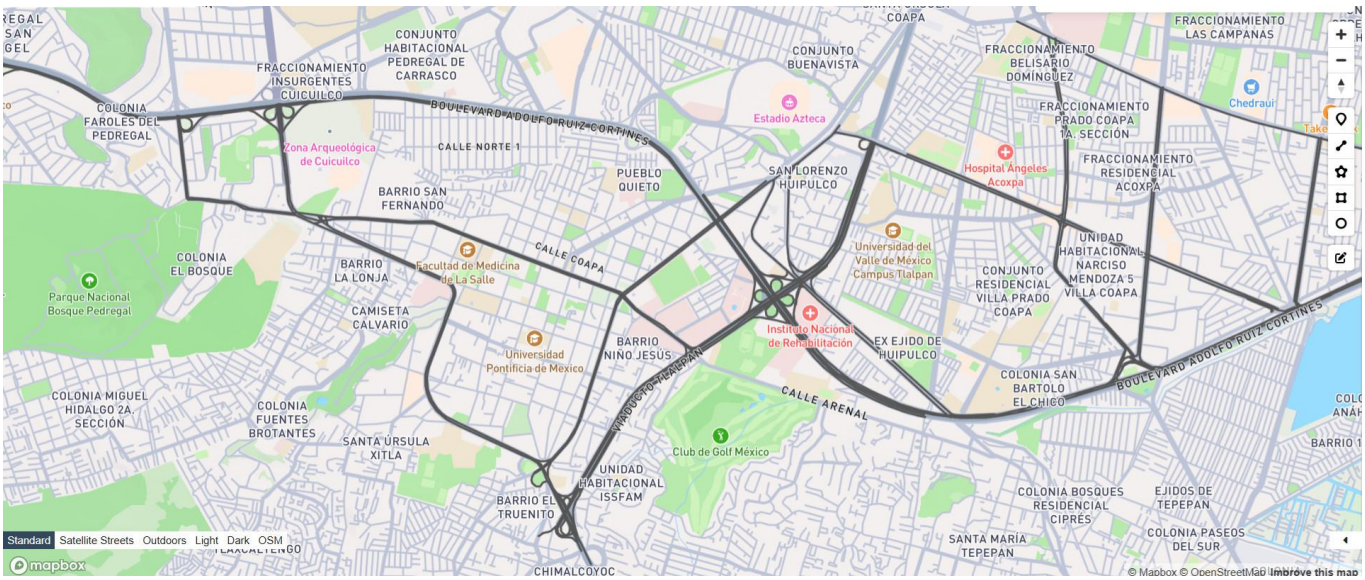

Fig. 1. Road map of the main highways in Tlalpan borough, Mexico City.

Data collection was carried out using the Here Maps API (HERE Technologies, 2025), which provides real-time information on traffic flow (measured in vehicles per minute) and road incidents, including accidents, closures, and severe congestion levels. The data was collected at 5-minute intervals, operating continuously 24 hours a day from February 1 to October 30, 2024. This process resulted in a dataset containing approximately 30,000,000 records, composed of 18 attributes that describe traffic conditions and reported incidents throughout the road network. This dataset, referred to as Traffic_Incidents, is described in Table 1, which outlines the recorded variables in detail.

Given the study's focus on understanding traffic dynamics in high-demand urban areas, the analysis centers on the primary avenues of Tlalpan Borough one of the most congested traffic zones in Mexico City. Table 2 presents a representative sample of the Traffic_Incidents dataset, offering an initial overview of the data structure and content. This dataset forms the foundation for analyzing traffic behavior and identifying congestion patterns, supporting the development of predictive and adaptive control strategies for vehicle flow in the study area.

Table 1. Description of the variables in the Traffic_Incidents dataset

Data	ID	Attribute	Description
Traffic	X1	description	Road name where traffic data is collected.
	X2	TrafficLength	Road segment length, measured in meters.
	X3	Speed	Current traffic speed on the road, in kilometers per hour (km/h).
	X4	speedUncapped	Unrestricted traffic speed.
	X5	FreeFlow	Average speed under free-flow conditions
	X6	jamFactor	Traffic flow impact index, rated on a scale from 0 to 10.
	X7	Number_Segments	Number of road segments into which the analyzed section is divided.
	X8	Daily_Traffic	Total traffic volume recorded on the road throughout a day.
	X9	Traffic_Hour	Time of day when the traffic measurement was recorded.
	X10	Traffic_Minute	Specific minute when the traffic data was captured.
Incidents	X11	IncidentLength	Length of the road segment affected by the incident, measured in meters.
	X12	startTime	Incident start time, recorded in ISO 8601 format (UTC).
	X13	endTime	Estimated incident end time, expressed in ISO 8601 format (UTC).
	X14	entryTime	Time when the incident was recorded in the system.
	X15	roadClosed	Indication of full road closure, represented as a boolean value (True/False).
	X16	criticality	Incident severity level, categorized as critical, major, or minor.
	X17	Type	Type of incident (e.g., road closure, accident, construction).
	X18	junctionTraversability	Accessibility status of the affected intersection.

Table 2. Excerpt from the Traffic_Incidents dataset showing traffic conditions on various avenues in Tlalpan (values rounded to two decimals due to space constraints). A full sample of 648,831 records is available in (De la Cruz Nicolás, 2025)

X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15
0.15	0.0	0.01	0.0	0.26	0.23	0.13	1.0	0.82	0.05	0.91	0.57	0.62	0.0
0.67	0.39	0.03	0.09	0.46	0.4	1.0	0.26	0.45	0.83	0.21	0.24	0.94	0.0
0.31	0.31	0.03	0.09	0.22	0.19	0.93	0.61	0.0	0.21	0.86	0.43	0.2	0.0
0.38	0.11	0.02	0.0	0.36	0.31	0.13	0.61	0.09	0.62	0.06	0.39	0.27	0.0
0.38	0.35	0.01	0.0	0.21	0.18	0.0	0.65	0.0	0.72	0.39	0.41	0.92	0.0
0.21	0.0	0.02	0.0	0.24	0.21	0.93	0.83	0.73	0.85	0.95	0.77	0.79	1.0
0.33	0.19	0.01	0.0	0.25	0.22	0.97	0.65	0.64	0.2	0.04	0.45	0.83	0.0
0.33	0.19	0.05	0.09	0.28	0.25	1.0	0.65	0.27	0.6	0.86	0.01	0.5	0.0
0.33	0.31	0.02	0.09	0.23	0.2	0.13	0.78	0.82	0.29	0.77	0.65	0.99	1.0
0.32	0.0	0.02	0.0	0.39	0.34	0.93	0.78	0.0	0.88	0.65	0.97	0.74	0.0

3.2 Data preprocessing and exploratory analysis

This section outlines the strategies used for cleaning, transforming, and conducting exploratory analysis on the Traffic_Incidents dataset, which contains 30,000,000 records collected from the Here Maps API for the Tlalpan Borough of Mexico City. These

steps are critical to ensuring data quality, consistency, and representativeness, which are essential for identifying mobility patterns and assessing the impact of road incidents on traffic flow.
The data preprocessing phase includes detecting and correcting outliers, imputing missing values, removing redundant entries, and standardizing variables to ensure that the dataset is accurate and suitable for predictive modeling. Following this, various transformation techniques are applied to optimize the dataset's structure such as encoding categorical variables, converting date formats, and normalizing numerical values.

Finally, the exploratory data analysis (EDA) process is carried out to uncover key insights into traffic dynamics within Tlalpan. This involves identifying relationships among critical variables like *vehicle speed, congestion levels, segment length, and incident severity*. Using a combination of visualization tools and statistical analysis, both temporal and spatial trends are revealed, offering valuable input for the development of traffic prediction models and adaptive traffic control strategies.

Data cleaning and validation. The first step in data preprocessing focused on identifying and correcting potential inconsistencies within the dataset. This began with verifying the completeness of each column to address missing or null values. For categorical variables such as *trafficability, incident type, and severity*, missing values were imputed using the most frequently occurring category. For numerical variables, the median was used to minimize bias and maintain data integrity. Duplicate records especially those with *identical date, time, and location* were removed to eliminate redundancy and prevent overestimation in subsequent analyses. Lastly, a validation step was performed to ensure that key variables, including *current speed, unrestricted speed, free-flow speed, and affected segment length*, fell within reasonable and expected ranges based on established findings from previous urban mobility studies.

Transformation and normalization. To enhance data interpretability and suitability for predictive modeling, several transformations were applied. First, the datetime fields (*startTime*, *endTime*, and *entryTime*) were converted to numerical values as seconds since the UNIX epoch (1970-01-01 00:00:00 UTC). This was done in three steps: parsing the original text into datetime objects using pd.to_datetime(), converting them to integers in nanoseconds via astype(int), and dividing by 10^9 to obtain seconds. This standardization allows consistent time comparisons, event duration calculations, and normalization for traffic forecasting.

Second, categorical variables such as criticality, Type, and junctionTraversability were encoded numerically using *LabelEncoder()* from sklearn.preprocessing. Each variable was transformed with *fit_transform(),* and the corresponding encoders were stored in a dictionary (label_encoders) for potential inverse decoding.

Finally, to mitigate differences in feature scales especially for *speed, traffic flow, and segment length Min-Max* normalization was applied across all numeric features. This ensured a uniform value range, reducing scale-induced bias in subsequent modeling and improving convergence in learning algorithms.

Exploratory data analysis. After completing the data preprocessing, a comprehensive exploratory analysis was performed to uncover significant mobility patterns within the study area.

Vehicle speed distribution. The analysis of actual vehicle *speed* relative to *free-flow speed* revealed notable temporal variability, particularly during peak traffic hours. In these high-demand periods, speeds ranged from complete stops (0 km/h) to an average of approximately 30 km/h, reflecting frequent stop-and-go conditions and episodic congestion. These fluctuations indicate unstable flow dynamics, often caused *by traffic signals, intersection delays, bottlenecks, or minor incidents* such as collisions and lane obstructions. Such disruptions break the continuity of movement, leading to vehicle clustering and longer travel times.

Conversely, during off-peak hours, vehicle speeds remained consistently close to the theoretical free-flow level, indicating lower traffic density and minimal resistance. This contrast highlights the cyclical nature of urban congestion and the need for dynamic management strategies that adapt to varying demand levels. The observed speed patterns support the use of real-time monitoring and adaptive control to mitigate congestion impacts during critical periods. This behavior is visualized in Fig. 2, which depicts the distribution of traffic speeds across different times of day and traffic conditions. In practice, the distribution compresses around the free-flow mode during nighttime, with reduced variance and fewer extreme low-speed outliers. This clear separation between off-peak and peak regimes enables threshold-based triggers (e.g., occupancy, queue length, or jam-factor) for time-of-day plans, variable cycle lengths, and priority/phasing adjustments. It also opens the door to short-horizon predictive control that preempts incipient breakdowns as demand ramps toward the morning and evening peaks, reinforcing the value of integrating streaming telemetry with adaptive signal policies and performance dashboards. Moreover, Fig. 2 typically shows shoulder periods with mild multimodality during transitions into and out of peak, which can be exploited via ramp-up/ramp-down plans and transit signal priority without degrading off-peak efficiency. Incorporating exogenous signals such as incident reports, weather, or planned

works—into these triggers further improves robustness. Finally, the stable off-peak free-flow provides a dependable baseline for KPIs (e.g., 85th-percentile speed, Travel Time Index, Buffer Index), aiding calibration and post-implementation evaluation.

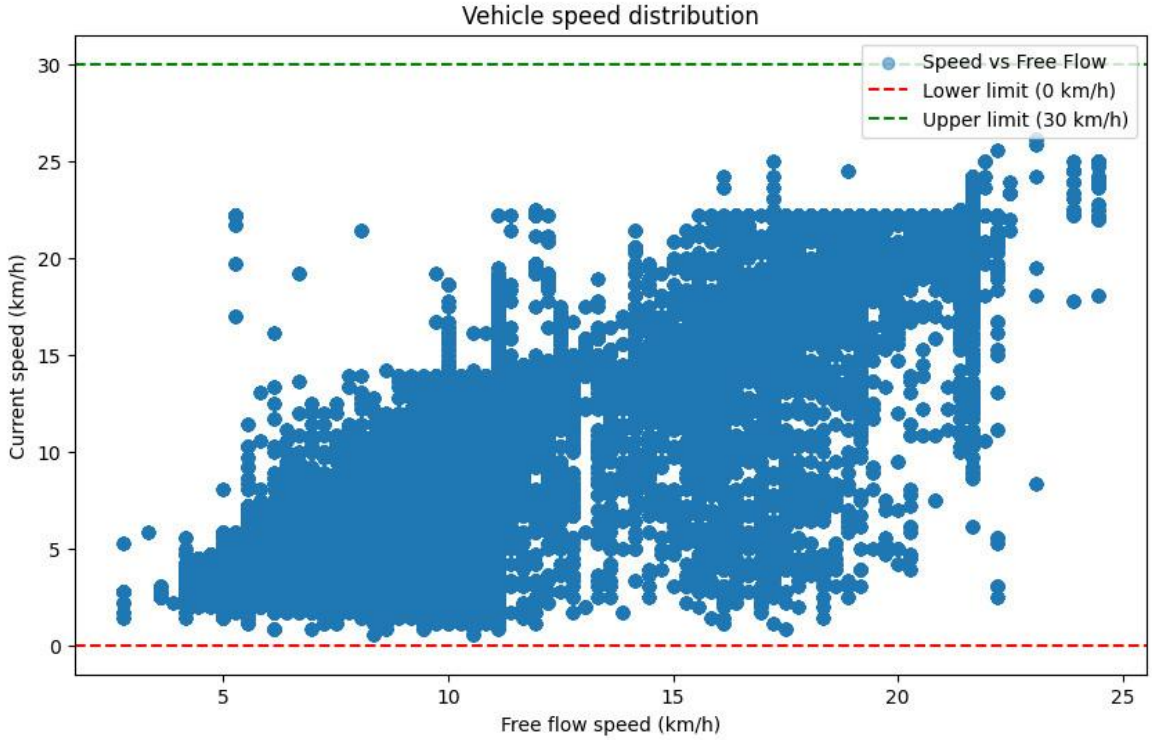

Fig. 2. Scatter plot illustrating the relationship between free-flow speed and current vehicle speed.

Congestion Factor and Affected Segment Length. A clear, near-linear relationship was identified between the *congestion factor* and the *length of road segments* affected by traffic incidents. Higher congestion factors indicative of more severe disruptions were consistently associated with longer impacted segments. This pattern reflects the nature of major *incidents such as collisions, full-lane blockages, multi-vehicle accidents, or weather-related obstructions*, which not only increase traffic density but also expand the spatial reach of their effects within the network.

This phenomenon is explained by traffic shockwave propagation: when an incident occurs, abrupt vehicle deceleration or stoppage generates a backward-moving delay wave through the traffic stream. As upstream vehicles encounter the disturbance, queuing intensifies, throughput decreases, and delays extend over increasingly longer road segments. The longer the disruption remains unresolved, the farther the shockwave propagates. From a kinematic-wave perspective, the propagation speed and amplitude depend on the local slope of the fundamental diagram (flow–density relation) and the severity of the capacity drop at the bottleneck. In dense regimes, even small perturbations can nucleate stop-and-go waves that coalesce and spill back into adjacent intersections or ramps, especially under rubbernecking and lane-changing frictions. Heterogeneous driver reactions further amplify the wave and increase the risk of secondary incidents. Dissipation occurs only once bottleneck capacity is restored and/or upstream demand falls below the effective capacity, allowing queues to discharge and the wavefront to recede hence the value of rapid clearance, coordinated signal/ramp metering, and speed harmonization to smooth inflows and dampen growth.

The correlation between incident severity and spatial spread was especially evident in dense urban areas, where high traffic volumes, complex intersections, and limited diversion capacity amplify the impact. In such contexts, the absence of real-time rerouting or adaptive signal control systems allows congestion to grow unchecked. Without targeted intervention, localized disturbances escalate into broader network-level delays.

These findings underscore the importance of early detection and localized incident response to contain the spatial propagation of congestion. Intelligent monitoring systems and predictive traffic management tools can significantly reduce both response time and the extent of disruption. As shown in Fig. 3, empirical results confirm this trend, displaying a strong and consistent correlation between congestion factor values and the length of affected segments. This reinforces

the need for proactive, adaptive control strategies in high-impact corridors to mitigate cascading delays and preserve urban mobility.

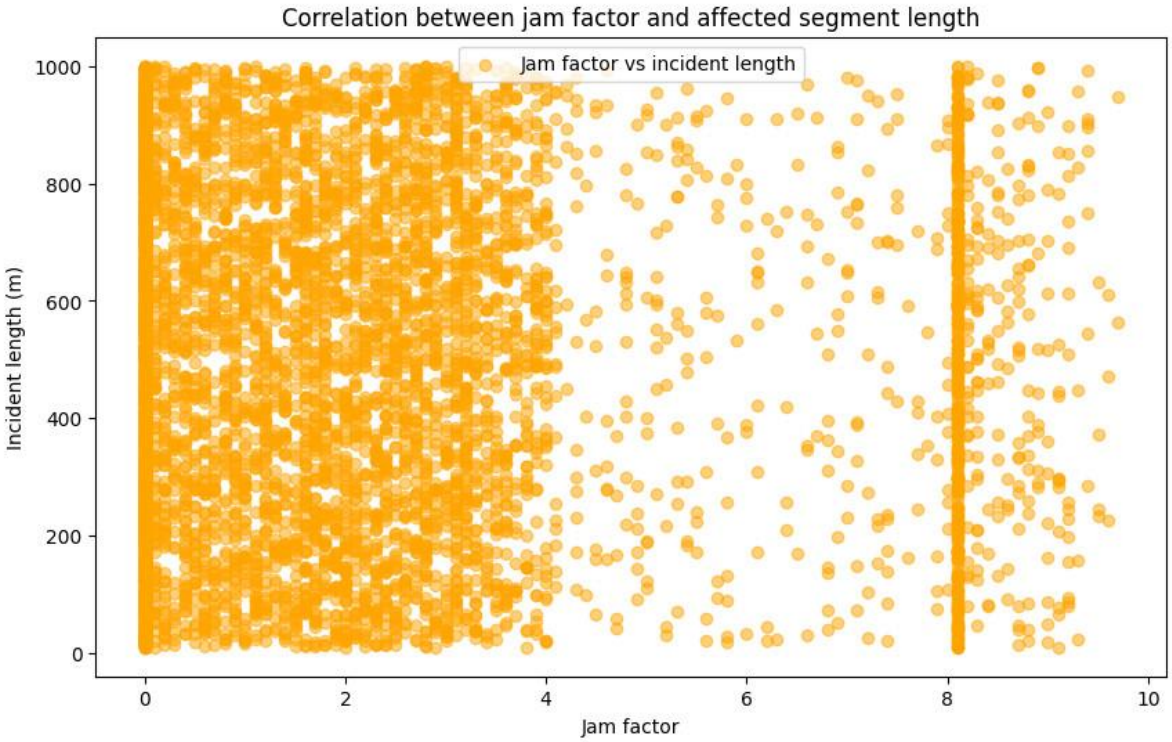

Fig. 3. Scatter plot illustrating the correlation between congestion factor and affected segment length.

Temporal distribution of incidents. A bimodal pattern was identified in the occurrence of road closures and accidents, with peak incidences observed between 7:00–9:00 AM and 6:00–8:00 PM, corresponding to periods of highest traffic flow (see Fig. 4).

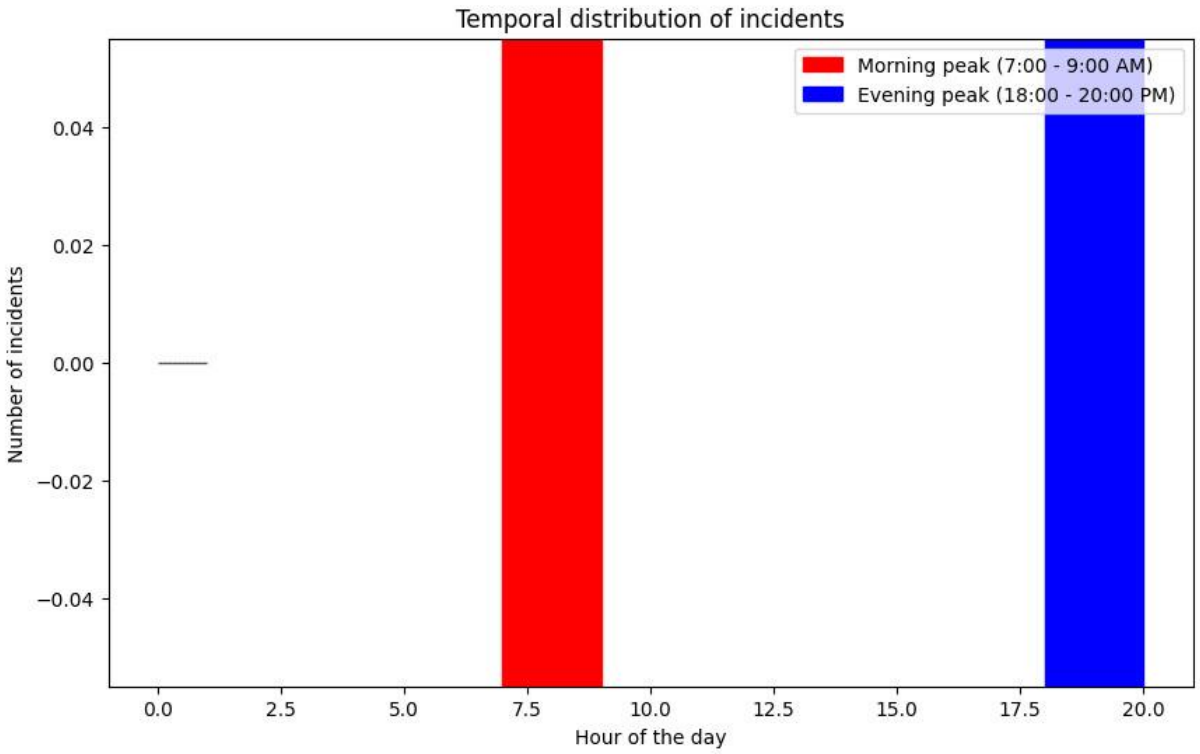

Fig. 4. Histogram illustrating the temporal distribution of road closures and accidents.

Impact of incidents on trafficability. Events were categorized by severity and type, revealing that 77% of incidents labeled as "critical" resulted in full road closures. In contrast, lower-severity incidents more commonly led to partial closures or intermediate traffic restrictions.

Data visualization. Data visualization plays a vital role in interpreting and analyzing traffic patterns. Clear and effective visual representations help uncover trends, detect anomalies, and support informed decision-making in traffic management. To facilitate this, several key visualizations were developed to examine traffic evolution, as described below:

To evaluate the distribution of vehicle speed and congestion levels, boxplots were used to visualize data dispersion, median values, and the presence of outliers. Three key variables were analyzed from the normalized dataset: current traffic speed (*Current Speed*), free-flow speed (*Free Flow Speed*), and the congestion factor (*Jam Factor*). The results show high variability in current *speed*, with a median notably lower than the free-flow speed, indicating recurring traffic slowdowns. The presence of outliers suggests that unusually high or low speeds occur at certain times of day, likely due to traffic incidents or fluctuating demand. In contrast, the congestion factor displays an asymmetric distribution, with elevated values concentrated during specific time intervals highlighting the occurrence of critical congestion episodes. The dispersion of this variable also aids in identifying the most impacted road segments and the frequency at which they reach saturation.

These findings underscore the value of boxplot analysis for understanding urban mobility dynamics and improving traffic management decisions. This approach supports the development of evidence-based strategies to mitigate congestion, as illustrated in Fig. 5.

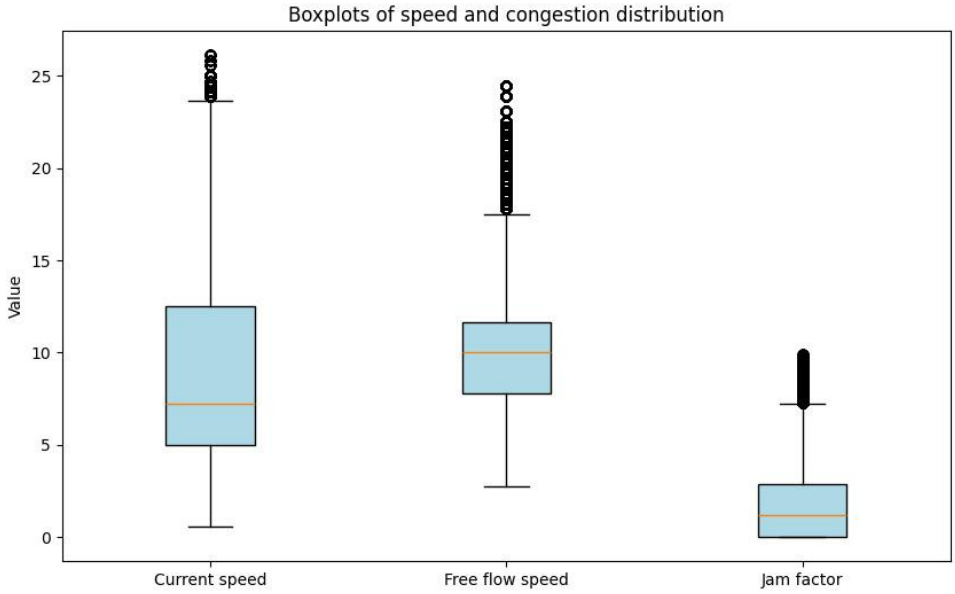

Fig. 5. Boxplots illustrating the distribution of vehicle speed and congestion factor.

The heat maps illustrate the georeferenced distribution of incidents along major avenues, highlighting high-incidence zones concentrated on Calzada de Tlalpan and Insurgentes Sur.

In Fig. 6, traffic congestion patterns within Tlalpan Borough are shown across different times of the day. On the left, section (a) represents congestion during the morning peak hours (7:00–9:00 AM), where elevated traffic levels are observed throughout the area, with congestion hotspots appearing at key strategic points. In the center, section (b) displays conditions during off-peak hours (10:00 AM–4:00 PM), reflecting an overall reduction in congestion, although certain areas continue to experience high vehicle density. On the right, section (c) illustrates congestion during the evening peak hours (6:00–8:00 PM), revealing patterns similar to those seen in the morning and emphasizing the influence of work-related schedules on daily traffic flow. Spatially, these hotspots tend to cluster along major corridors Periférico Sur, Insurgentes Sur, and Calzada de Tlalpan and at signalized intersections and freeway ramps, where spillbacks frequently affect feeder streets. A directional asymmetry is also evident: northbound flows are typically more constrained in the morning, while southbound movements bear the brunt during the evening commute. Even during off-peak, residual pockets persist near commercial frontages, school zones, and chronic bottlenecks, underscoring the need for targeted corridor coordination and adaptive signal timing.

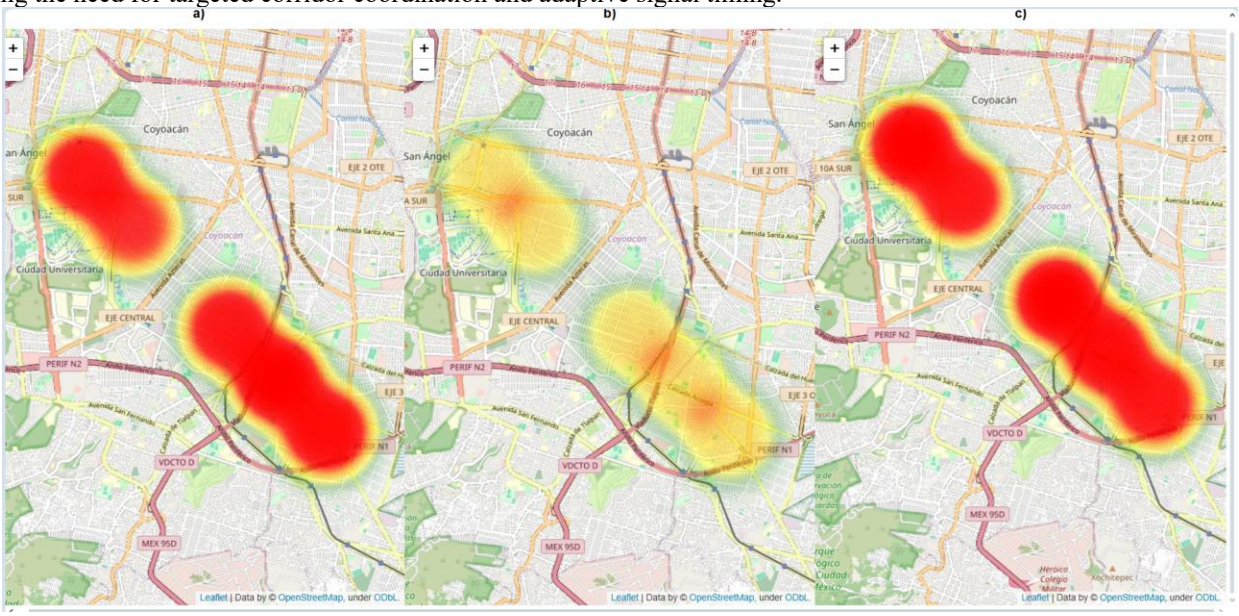

Fig. 6. Heat maps illustrating the spatial distribution of traffic congestion in Tlalpan Borough across different time periods.

Time series analysis: The time series illustrates the evolution of congestion throughout the day, highlighting the *hours* and *days* of the week with the greatest impact on mobility (see Fig. 7).

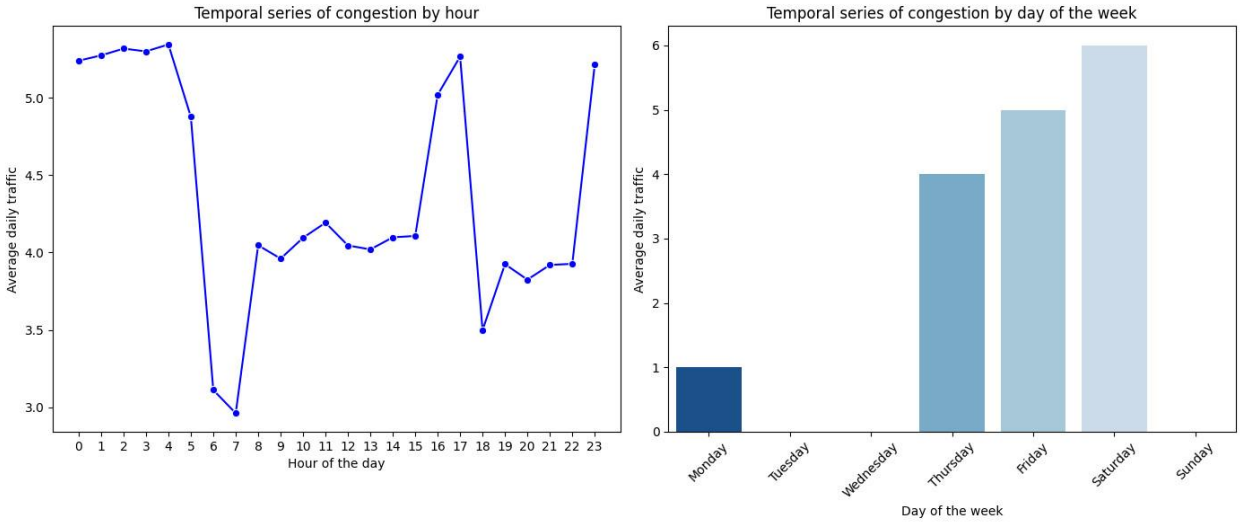

Fig. 7. The time series of traffic congestion reveal distinct patterns throughout the day and week.

3.3 Development of the physics-based model

The traffic speed evolution model is grounded in the macroscopic dynamics of vehicular flow, represented by a differential equation that captures the effects of congestion, road restrictions, and incidents on traffic speed (*X3*). This formulation describes how various factors interact to influence mobility across the road network. The construction of the model is outlined below, detailing the key components that define its structure.

The evolution of traffic speed is driven by three primary elements. First, *free-flow speed* and roadway restrictions play a central role. Under ideal conditions, the actual *traffic speed* (*X3*) tends to converge toward the *free-flow speed* (*X5*), which reflects the average velocity observed in the absence of major obstacles. However, this natural tendency is moderated by congestion and the physical constraints of the road infrastructure. Second, the impact of congestion is expressed through the *congestion factor* (*X6*), which measures the level of traffic disruption on a scale from 0 to 10. Congestion leads to reduced speeds, and its effect is proportional to the length of the *road segment* (*X2*); in shorter segments, congestion typically dissipates quickly, while in longer segments, it results in sustained deceleration.

Third, *traffic incidents* impose an additional slowdown. Events such as accidents, lane closures, or construction activities further reduce speed depending on their severity and duration. This impact is modeled using variables such as the length of the affected *segment* (*X11*), the severity of the *incident* (*X16*), and the type of event (*X17*).

Taking these factors into account, the following differential equation is proposed to describe the evolution of *traffic speed*:

$$\frac{dX3}{dt} = -\alpha \frac{X3}{X5} + \beta \frac{X6}{X2} + \gamma + F(X11, X15, X16, X17) \quad (1)$$

Where:
α, β and γ are adjustment coefficients that define the sensitivity of speed to each factor.
$F(X11, X15, X16, X17)$ is a function that models the impact of incidents on speed reduction.

The free-flow speed adaptation term, represented by $-\alpha \frac{X3}{X5}$ The f*ree-flow speed* adaptation term describes the tendency of traffic to regain its normal *speed* under *free-flow* conditions. However, this recovery is not instantaneous; rather, it depends on the relationship between current speed (*X3*) and free-flow speed (*X5*).

When *X3* is significantly lower than *X5*, traffic tends to accelerate to approach the free-flow speed. As *X3* nears *X5*, acceleration gradually decreases until a steady state is reached.

The congestion factor (*X6*) measures the impact of congestion on traffic on a scale from 0 to 10. The speed reduction caused by congestion is modeled through the term $\beta \frac{X6}{X2}$, Where *X2* represents the segment length. In short segments, congestion effects tend to dissipate more quickly, whereas in long segments, they can cause a more pronounced deceleration. The β coefficient regulates the sensitivity of speed to this effect, determining the degree of reduction induced by congestion.

Traffic incidents affect *traffic speed* based on their severity and extent. To model this impact, the following function is defined:

$$F(X11, X15, X16, X17) = X11 + X15 + \mathrm{I}(X16 = \mathrm{critical}) + \mathrm{I}(X17 = roadClosure) \quad (2)$$

Where:
X11 (*IncidentLength*): measures the extent of the affected segment.
X15 (*roadClosed*): indicates whether the road is completely closed (1: Yes, 0: No).
X16 (*criticality*): classifies the severity of the incident (critical, major, minor).
X17 (*Type*): specifies the type of incident (roadClosure, accident, construction).
I (·): is an indicator function that takes the value 1 if the condition is true and 0 otherwise.

The γ parameter controls the magnitude of the impact of incidents on speed reduction. Therefore, the final model is expressed as follows:

$$\frac{dX3}{dt} = -\alpha \frac{X3}{X5} + \beta \frac{X6}{X2} + \gamma\big(X11 + X15 + \mathrm{I}(X16 = \mathrm{critical}) + \mathrm{I}(X17 = roadClosure)\big) \quad (3)$$

Where:
The first term describes the natural tendency of traffic to recover its *free-flow speed*.
The second term models the speed reduction caused by congestion.
The third term accounts for the impact of traffic incidents, considering their severity and type.

3.4 Development of the prediction model using LSTM networks

The prediction model, based on Long Short-Term Memory (LSTM) networks, is designed to capture the temporal evolution of *traffic speed* (*X3*) by leveraging the mathematical formulation of the physical model introduced in Section 3.3. This differential equation outlines how traffic speed evolves as a function of congestion and traffic incidents, providing the foundation for a hybrid modeling approach. In this framework, the mathematical structure of traffic dynamics is integrated with a deep learning model to improve prediction accuracy.

The LSTM network is trained on time series data from the Traffic_Incidents dataset, enabling it to learn the relationships between traffic speed (*X3*), congestion (*X6*), and incident-related variables (*X11, X15, X16, X17*). Predictions are generated at discrete time intervals, with the differential equation solved numerically using methods adapted to the memory mechanisms of the LSTM architecture. This approach allows the model to account for both short-term fluctuations and long-term trends in traffic behavior.

Discrete formulation of the differential equation. To integrate the differential equation into the LSTM-based model, a time discretization process was applied, converting the continuous formulation into a sequence of discrete states. This transformation was performed using the Euler numerical method.

$$X_3^{t+288} = X_3^t + \Delta t \cdot f(X^t) \quad (4)$$

Where:
$f(X^t)$ Is the expression that describes the rate of change of *X3*. It is a first-order method, easy to implement, but it can become unstable if Δt is not sufficiently small.

The differential equation was discretized using the Euler method, as Long Short-Term Memory (LSTM) networks operate on sequences of discrete data. As a result, the continuous formulation was transformed into a discrete representation, as shown in

Equation 4. This discretization allows the model to capture the temporal evolution of traffic speed without requiring an analytical solution to the differential equation. It aligns naturally with the structure of LSTM networks, which rely on prior states to predict future outcomes.

Moreover, the Euler method is particularly well-suited for this application, given that traffic flow typically changes gradually and the data is collected at consistent time intervals. This approach not only simplifies the implementation of the neural network but also preserves numerical stability while mirroring how traffic data is observed and processed. Once discretized, the data is structured into time sequences tailored for training the LSTM network, where each time step reflects the dynamic evolution of the system. This enables the network to learn underlying patterns in variable behavior, providing an effective and flexible means of approximating the differential equation, as illustrated in Equation 5.

$$X_3^{t+288} = X_3^t + \Delta t \left(-\alpha \frac{X_3^t}{X_5^t} + \beta \frac{X_6^t}{X_2^t} + \gamma(X_{11}^t + X_{15}^t + I(X_{16}^t = critical) + I(X_{17}^t = roadClosure)) \right) \quad (5)$$

Where Δt represents the data sampling interval, which in this case is 5 minutes.

The differential equation used to model the evolution of *traffic speed* (*X3*), represented by Equation 5, describes how this parameter changes over time based on the conditions affecting traffic dynamics. The equation states that the *future speed* X_3^{t+288} It is the result of the *current speed* (X_3^t) plus a series of adjustments determined by key factors, such as the tendency of traffic to recover its free-flow speed, the impact of congestion, and the influence of traffic incidents.

In the graph in Fig. 8, two time series are represented one in dotted blue, which shows the *current traffic speed* (X_3^t), and another in red with square markers, which illustrates the *projected speed* (X_3^{t+288}), obtained from the equation. The comparison between these two series allows for the analysis of how the estimated speed varies relative to the observed speed, thereby validating the model's behavior.

The first term of the equation $-\alpha \frac{X_3^t}{X_5^t}$, represents the tendency of traffic to return to the *free-flow speed* (*X5*). In the graph in Fig. 8, this effect is reflected in periods where the projected speed stabilizes or shows an increasing trend after being reduced by an external factor. In other words, in the absence of significant disturbances, traffic tends to recover a higher speed.
The second term $\beta \frac{X_6^t}{X_2^t}$, models the influence of traffic congestion on speed reduction. The "*jam factor*" (*X6*) serves as an indicator of the degree to which *traffic flow* is affected, with its impact varying according to the length of the *road segment* (*X2*).

In the graph, this effect is evident in the declines in projected *speed*, highlighting moments when congestion significantly restricts *traffic flow* (X_3^{t+288}) compared to the *current speed* (X_3^t), especially during periods of higher congestion.
The third term, $\gamma(X_{11}^t + X_{15}^t + I(X_{16}^t = critical) + I(X_{17}^t = roadClosure))$, incorporates the impact of *traffic incidents*, considering factors such as *incident length* (*X11*), total *road closure* (*X15*), and *incident severity* (*X16*), where a critical value has a greater impact. Additionally, it includes the *type of incident* (*X17*), distinguishing between accidents, construction, and total closures.

In the graph, this effect is observed as abrupt drops in projected speed, indicating moments when traffic incidents significantly reduce *traffic flow* (X_3^{t+288})), showing events where an incident has drastically decreased vehicular flow.

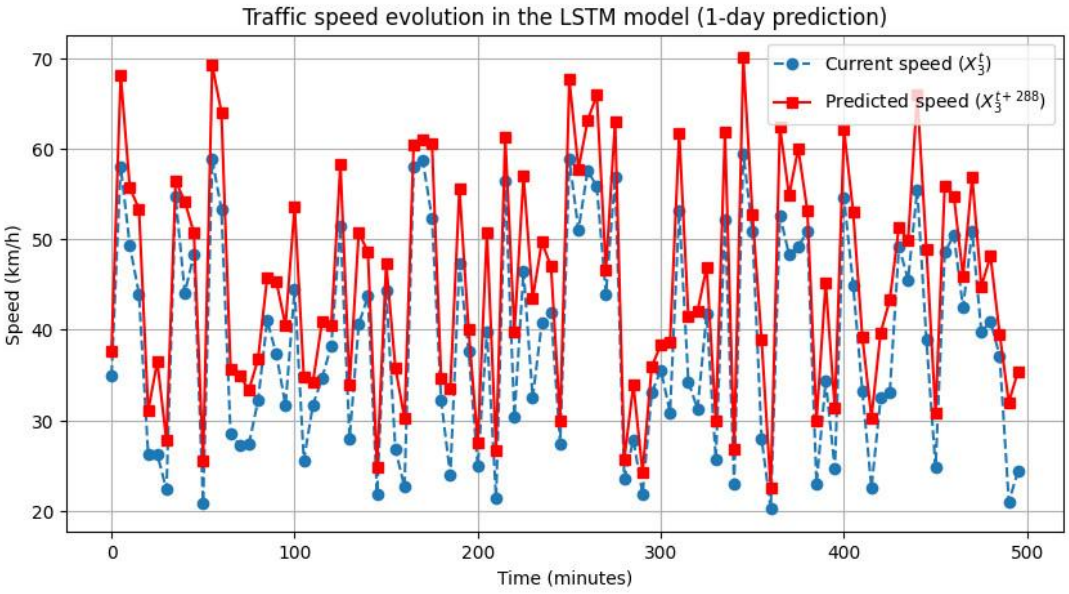

Fig. 8. Evolution of traffic speed in the LSTM model.

The integration of the physical model with the Long Short-Term Memory (LSTM) network is accomplished through a hybrid formulation that combines the mathematical representation of traffic dynamics with deep learning applied to time series data. In this approach, the differential equation governing the evolution of *traffic speed* is discretized using the Euler method and embedded as a constraint within the LSTM's prediction process.

During training, the LSTM learns temporal patterns in traffic speed from historical data while adjusting its predictions based on the physical components of the equation capturing speed recovery under free-flow conditions, the effects of congestion, and the impact of traffic incidents. This hybrid model not only captures complex dependencies within the data but also maintains alignment with the fundamental principles of traffic flow, enhancing both the interpretability and robustness of its predictions.

As a result, the system delivers more accurate forecasts and adapts effectively to varying traffic scenarios, making it well-suited for real-time traffic control and decision-making applications.

LSTM Model Architecture. The LSTM model architecture is designed to capture the temporal evolution of traffic speed, enabling medium-term predictions with a 1-day horizon. To achieve this, the input window is extended to 288-time steps (*Equivalent to 24 hours of data at 5-minute intervals*), allowing the model to capture daily trends and recurring congestion patterns.

The model consists of two hierarchical LSTM layers: The first layer, with 256 units, captures long-term temporal dependencies and the second layer, with 128 units, reduces dimensionality while retaining memory of learned sequences.
Next, two dense layers with 64 and 32 neurons and *ReLU* activation transform the learned representation before the final prediction. The output layer, with a single neuron and linear activation, generates the future speed estimate X_3^{t+288}, predicting *traffic speed* one day in advance.

To prevent overfitting and stabilize training, regularization techniques are applied, including 30% Dropout in the LSTM layers and Batch Normalization after each recurrent layer. The model is optimized using the Adam optimizer with an initial learning rate of 0.0005, ensuring stable and efficient convergence for long-term predictions. Additionally, the RMSE (Root Mean Square Error) loss function is used to minimize prediction error and improve model accuracy. The detailed model structure is presented in Table 3, while its graphical configuration is illustrated in Fig. 9.

Table 3. LSTM Model Architecture for Traffic Speed Prediction

Network Input	Model Layers	Regularization Techniques	Optimization
*Time sequences of the attributes: *X3, X5, X6, X2, X9, X11, X13, X14, X15, X16, X17*. *Sequence length: 288-time steps (equivalent to 1 day of data at 5-minute intervals).	*LSTM Layer 1: 256 units with tanh activation to capture long-term temporal dependencies. *LSTM Layer 2: 128 units with tanh activation to reduce dimensionality and retain long-term memory. *Dense Layer 1: 64 neurons with ReLU activation to transform the learned representation. *Dense Layer 2: 32 neurons with ReLU activation to refine the representation before the output. *Output Layer: 1 neuron with linear activation to predict X_3^{t+288} (speed in 24 hours).	*Dropout (30%) in LSTM layers to prevent overfitting. *Batch Normalization after each LSTM layer to enhance training stability.	*Loss function: RMSE (Root Mean Square Error) to minimize prediction error. *Optimizer: Adam, with an initial learning rate of 0.0005, to improve convergence in long-term predictions.

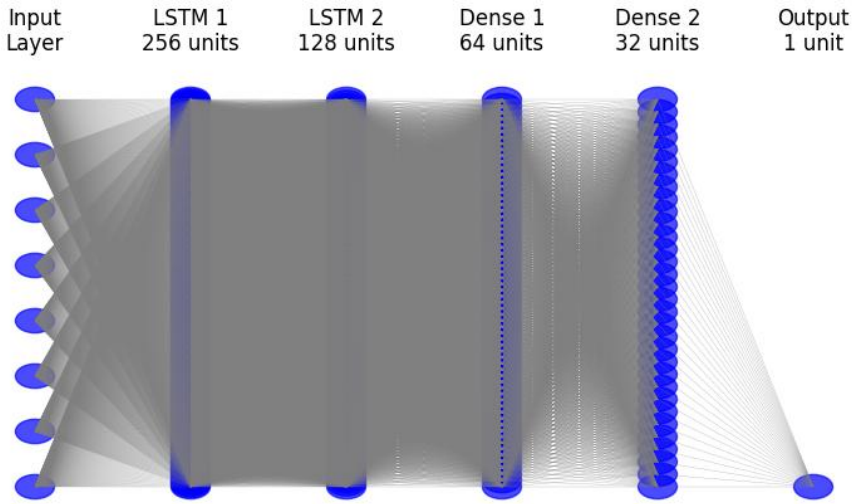

Fig. 9. The representation of the LSTM neural network architecture for traffic speed prediction consists of two levels of visualization.

In addition to the variables described in Table 3, the proposed LSTM network incorporates an additional input: the rate of change in speed, estimated using the discretized differential equation presented in Equation 5. This enables the model to not only learn from historical traffic data but also adapt its predictions according to physical traffic flow constraints. To ensure consistency with traffic laws, a regularization term is added to the loss function, which penalizes deviations from the expected dynamic behavior. In practice, if the LSTM predicts a speed that does not align with the evolution defined by the differential equation, the model incurs a higher training error, encouraging convergence toward physically plausible solutions.

The LSTM architecture is designed to predict the future velocity X_3^{t+288} , equivalent to a 24-hour forecast horizon, with a focus on modeling capacity, numerical stability, and computational efficiency. It consists of two hierarchically arranged LSTM layers followed by two fully connected dense layers and a final linear output layer. The first LSTM layer, composed of 256 units, captures long-term temporal dependencies such as traffic speed trends and congestion propagation. The tanh activation function is employed to normalize neuron outputs within the range [-1, 1], mitigating the risk of gradient explosion and improving training stability. The second LSTM layer, with 128 units, compresses the learned representation into a more compact form, reducing

dimensionality while preserving essential features. This output is then processed by a dense layer with 64 neurons and *ReLU* activation, enhancing the alignment of the representation with the data distribution and improving generalization. A second dense layer with 32 neurons refines the representation further, contributing to more accurate predictions. The final layer is a linear output neuron that produces continuous speed values, allowing the network to reflect fine-grained variations in traffic dynamics.

To prevent overfitting and enhance model robustness, a dropout rate of 30% is applied to both LSTM layers. This regularization mechanism reduces overreliance on specific features seen during training. Additionally, batch normalization is applied after each LSTM layer to maintain stable neuron activation distributions, reduce internal covariate shift, and support faster convergence. The network is trained using the Adam optimizer, selected for its adaptive learning rate adjustment and superior convergence behavior compared to traditional optimizers such as stochastic gradient descent. An initial learning rate of 0.0005 is used to ensure training stability during long-sequence forecasting. The loss function is Root Mean Square Error (RMSE), chosen for its sensitivity to large deviations a key requirement in traffic forecasting, where abrupt changes often indicate critical shifts in urban mobility.

The model processes sequences of 288 time steps, representing one full day of traffic data collected at 5-minute intervals. This temporal window allows the network to capture intra-day traffic patterns, including peak-hour fluctuations and daily trends, while balancing prediction accuracy and computational cost. The sequence length is strategically selected to provide sufficient context for learning traffic dynamics without introducing unnecessary noise or storage overhead. To justify the architectural and training choices, a sensitivity analysis was performed by systematically varying key hyperparameters such as the number of LSTM units, dropout rate, sequence length, and learning rate while holding the others constant. As summarized in Table 4, the configuration using 256 units in the first LSTM layer, 128 in the second, a dropout rate of 0.3, and a sequence length of 288 yielded the best trade-off between predictive performance and training stability. These results support the design decisions shown in Table 3 and illustrated in Fig. 9.

Table 4. Hyperparameter Sensitivity Analysis

Hyperparameter	Tested Values	Optimal Value	Effect on RMSE	Remarks
Window Size	144, 288, 432	288	Higher RMSE for 144; plateau after 288	Captures full daily pattern
LSTM Units Layer 1	128, 256, 384	256	Lower RMSE with 256; marginal gains with 384	Captures long-term dependencies best
LSTM Units Layer 2	64, 128, 256	128	Best at 128; drop in performance with 256	Balances dimensionality and memory
Dropout Rate	0.1, 0.3, 0.5	0.3	0.3 minimizes overfitting and RMSE	Best regularization trade-off
Learning Rate	0.0001, 0.0005, 0.001	0.0005	0.0005 gives most stable and low RMSE	Ensures convergence stability

Model Training. The Long Short-Term Memory (LSTM) model was trained on a dataset of 30,000,000 records, split into 70% for training, 15% for validation, and 15% for testing to ensure coverage of diverse traffic conditions. All numerical features were scaled to the [0, 1] range using Min-Max normalization, selected for preserving data distribution, improving interpretability in time series, and optimizing the tanh activation function in LSTM layers by preventing saturation and enhancing gradient flow. Compared to Z-score normalization, Min-Max also provided greater robustness to outliers, ensuring stable learning. The model was trained over 100 epochs with a batch size of 64, balancing convergence and computational efficiency.

To prevent overfitting, Early Stopping was applied with a patience of 10 epochs without improvement in validation loss. A learning rate scheduler adjusted the initial rate of 0.0005, reducing it by half if the validation loss plateaued for 5 epochs, allowing finer updates in later training phases.

The Adam optimizer was used for its adaptive gradient adjustment and faster convergence than traditional methods like stochastic gradient descent. RMSE was the primary loss function due to its sensitivity to large deviations critical for detecting severe congestion while MAPE was tracked as a complementary metric for assessing relative accuracy under varying conditions. The model was implemented in TensorFlow (Abadi et al., 2016) and Keras (Chollet, 2015) with GPU acceleration, processing sequences of 288 time steps (equivalent to one day at 5-minute intervals). Metrics were tracked using TensorBoard, enabling real-time monitoring and iterative refinement. This configuration yielded a stable, generalizable model capable of forecasting traffic

speed up to 24 hours in advance, resistant to overfitting and adaptable to diverse urban conditions, making it suitable for real-time traffic management systems.

For transparency and reproducibility, the full implementation and source code are publicly available in (De la Cruz Nicolás, 2025). As shown in Fig. 10: section (a) displays the RMSE loss curve over training and validation, indicating convergence and stable generalization; section (b) illustrates normalized traffic speed distributions across the day, highlighting congestion peaks; section (c) shows the inverse correlation between *traffic speed* and the jamFactor index, validating its use in traffic modeling; and section (d) compares predicted versus actual speeds, confirming high accuracy and alignment with observed traffic trends.

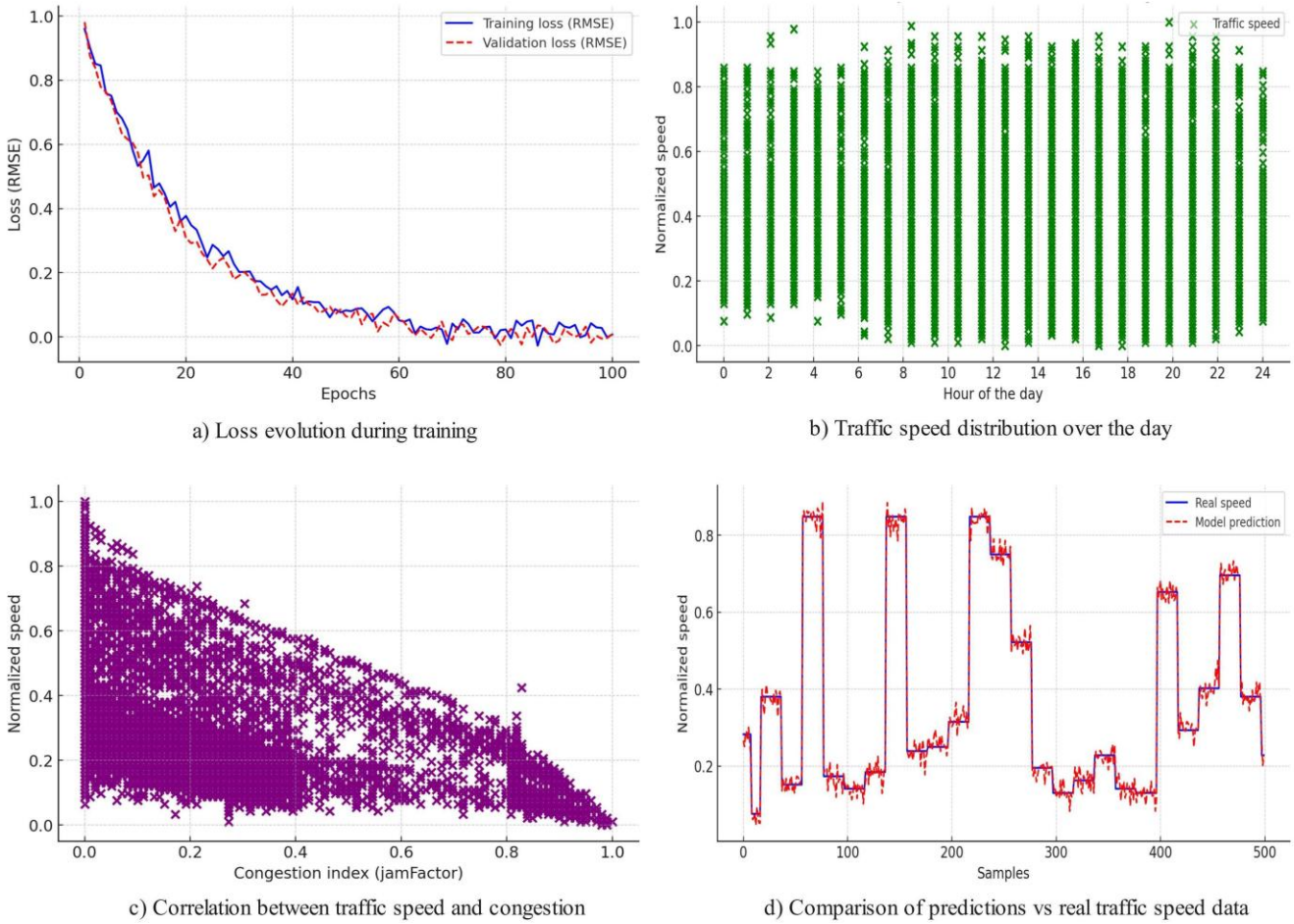

Fig. 10. Analysis of LSTM Model Training for Traffic Prediction.

Evaluation Metrics. The model's predictive performance was evaluated using three key metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). RMSE measures overall prediction error, emphasizing larger deviations; MAPE expresses error as a percentage, capturing relative performance under varying congestion levels; and R^2 quantifies the proportion of variance explained, indicating the model's generalization capacity. To assess stability, the model was trained in ten independent runs with different random seeds, varying weight initialization and data shuffling.

This allowed evaluation of consistency under training variability. Table 5 summarizes representative results, highlighting the best-performing configurations. Run 6 was selected as the reference due to its optimal trade-off between accuracy and robustness. It achieved an RMSE of 0.0201, MAPE of 6.36%, and R^2 of 0.9935, ranking among the top three across all metrics. Its reliability was further validated through sensitivity analysis and cross-validation. This methodology ensures that reported results reflect the model's typical behavior in real-world traffic conditions, avoiding reliance on a single training instance.

Table 5. Evaluation Metrics for the Hybrid LSTM-Physics Model in Traffic Speed Prediction

Run	RMSE	MAPE	R^2
1	0.02075	6.15	0.99497
2	0.01992	6.15	0.99327
3	0.02094	6.47	0.99357
4	0.02208	5.5	0.99208
5	0.0198	5.58	0.99296
6	0.0201	6.36	0.9935
7	0.02215	5.9	0.99235
8	0.0211	6.5	0.99388
9	0.01949	5.95	0.9929
10	0.02081	5.72	0.99321

To assess the model's robustness against variability in initialization, 30 independent training runs were performed and organized into three groups, each using a different random seed. A one-way ANOVA test was then conducted on the resulting RMSE values, producing an F-value of 11.989 and a p-value of 0.0002. These results indicate statistically significant differences among the groups ($p \leq 0.05$), confirming that the model's performance is sensitive to initial training conditions. This sensitivity underscores the importance of incorporating additional stabilization techniques, such as ensemble learning or more rigorous regularization strategies, to ensure consistent and reliable predictions (see Table 6).

Table 6. RMSE values obtained from 30 independent model runs, divided into three groups

Run	RMSE	Group	Run	RMSE	Group	Run	RMSE	Group
1	0.0209	Group A	11	0.0198	Group B	21	0.022	Group C
2	0.0204	Group A	12	0.0198	Group B	22	0.0205	Group C
3	0.021	Group A	13	0.0203	Group B	23	0.0208	Group C
4	0.0217	Group A	14	0.0188	Group B	24	0.0194	Group C
5	0.0203	Group A	15	0.0189	Group B	25	0.0202	Group C
6	0.0203	Group A	16	0.0197	Group B	26	0.0208	Group C
7	0.0218	Group A	17	0.0194	Group B	27	0.0197	Group C
8	0.0211	Group A	18	0.0203	Group B	28	0.021	Group C
9	0.0201	Group A	19	0.0195	Group B	29	0.0202	Group C
10	0.0209	Group A	20	0.0191	Group B	30	0.0204	Group C

Comparison with Baseline Models. To evaluate the performance of the proposed hybrid model (LSTM + Physics-based Equations), a direct comparison was made against three baseline models using three key metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). The results demonstrate that the hybrid model consistently outperforms all baseline models across all evaluated metrics. Specifically, it achieved the lowest RMSE (0.0201), indicating minimal average deviation between predicted and actual values. It also recorded the lowest MAPE (6.36%), highlighting strong relative accuracy under various congestion scenarios. Additionally, it achieved an exceptional R^2 value of 0.9935, showcasing its superior ability to capture the variability inherent in real-world traffic data. In comparison, the ARIMA model though appropriate for time series analysis struggled to adapt to the nonlinear characteristics of traffic, resulting in an RMSE of 0.0453 and a MAPE of 12.42%.

The linear regression model showed the weakest performance due to its overly simplistic structure, with an RMSE of 0.0678 and significantly higher MAPE. The conventional LSTM model, without the integration of physics-based constraints, outperformed traditional statistical models (RMSE = 0.0256, MAPE = 8.91%), but still fell short of the hybrid model. This highlights the importance of embedding domain-specific knowledge into predictive systems. Furthermore, the Temporal Convolutional Network (TCN) model, which uses dilated causal convolutions to capture long-range dependencies, yielded an RMSE of 0.0274, a MAPE of 9.12%, and an R^2 of 0.9761. Although it performed better than ARIMA and linear regression, it was slightly outperformed by the standalone LSTM model and clearly surpassed by the proposed hybrid model. These findings reinforce the value of integrating expert knowledge and physical constraints into deep learning frameworks for urban traffic forecasting. Overall, the results validate the effectiveness of the hybrid approach, underscoring its capacity to model complex traffic dynamics and significantly improve both prediction accuracy and robustness compared to traditional methods (see Table 7).

Table 7. Performance comparison of different models for 24-hour speed prediction

Model	RMSE	MAPE	R^2
ARIMA	0.0453	12.42	0.9321
Linear Regression	0.0678	18.75	0.8756
LSTM without Physics-based Equations	0.0256	8.91	0.9823
Hybrid Model (LSTM + Physics)	0.0201	6.36	0.9935
TCN	0.0274	9.12	0.9761

Simulation of the hybrid LSTM-physics model for traffic prediction in SUMO. The model's effectiveness in proactive traffic control was validated using simulations in the SUMO (Simulation of Urban Mobility) environment (Behrisch et al., 2011). A scenario grounded in real-world data from Tlalpan, Mexico City, was developed, modeling the primary road arteries and replicating traffic conditions across various levels of congestion. To integrate the hybrid model with the simulation, a direct connection was established through SUMO's Traffic Control Interface (TraCI), using Python scripts and the TraCI library to enable real-time data exchange. Three approaches were evaluated: a standalone LSTM model without physics integration, a traditional baseline model, and the proposed hybrid LSTM–physics model. The results revealed that the hybrid model significantly reduced average travel time by 26.3% compared to the alternatives, improved traffic flow stability by 35%, and effectively predicted congestion events. These outcomes underscore the advantages of combining neural network-based learning with physics-based equations, offering a robust and accurate forecasting framework essential for advanced traffic control and urban mobility optimization. To ensure full reproducibility, all SUMO/TraCI scenario files, including the .net.xml and .rou.xml files used for the integration, are publicly available and can be downloaded upon request (De la Cruz Nicolás, 2025).

Hybrid Integration and Ablation Study. As described in Sections 3.3 and 3.4, the hybrid model integrates a macroscopic differential equation with an LSTM neural network through two main mechanisms. First, the model includes the discrete time derivative of traffic speed (dX_3/dt), calculated via a first-order Euler approximation, as an additional input feature. This allows the network to account for physical dynamics such as return to free-flow speed and deceleration under congestion or incidents. Second, a penalty term is added to the loss function to regularize predictions that deviate from the behavior defined by the differential equation (see Equation 5), guiding the model toward physically consistent outputs.

These mechanisms enhance the model's stability and interpretability by reinforcing data-driven learning with structured domain knowledge, particularly under nonlinear or irregular traffic conditions. To quantify the impact of these physical integrations, an ablation study was conducted comparing four model configurations: full model, without the physical penalty, without the temporal derivative input, and without both. Results showed that removing the penalty term increased RMSE by 9%, while excluding the derivative input led to a 10.6% increase. The highest error was observed when both were removed, confirming their combined importance. These findings (summarized in Table 8) validate that embedding physical constraints in both input and loss functions significantly improves prediction accuracy, generalization, and consistency with real-world traffic behavior.

Table 8. Results of the ablation study on the proposed hybrid model

Model Configuration	Derivative (Input)	Physical Penalty	RMSE	MAPE
Full (Hybrid) Model	YES	YES	2.45	4.12%
Without Physical Penalty	YES	NO	2.67	4.38%
Without Derivative	NO	YES	2.71	4.49%
Without Physical Constraints (Pure LSTM Model)	NO	NO	2.78	4.63%

4 Results

The implementation and validation of the proposed hybrid model combining Long Short-Term Memory (LSTM) neural networks with physics-based traffic models demonstrated substantial improvements in predictive accuracy, early congestion detection, and adaptive control in the urban context of Tlalpan, Mexico City. The model achieved an RMSE of 0.0201, a MAPE of 6.36%, and an R^2 of 0.9935, indicating high precision and strong alignment with real-world traffic behavior. Compared to a standard LSTM model without physical integration, the hybrid approach reduced RMSE by 21.5% and MAPE by 28.6%. Against traditional statistical models such as ARIMA and Linear Regression, RMSE was reduced by 55.6% and 70.3%, respectively. These results confirm that embedding physical equations into deep learning architectures significantly enhances traffic modeling performance. The model also demonstrated the ability to anticipate sharp speed drops up to 15 minutes in advance, supporting real-time management applications. Prediction errors remained below 5% for critical events like accidents and closures, and between 6%

and 7% under normal traffic conditions. Simulation tests in the SUMO environment confirmed the model's utility for proactive traffic control. Adaptive strategies including dynamic signal timing, rerouting, and flow restrictions reduced average travel time by 26.3% and speed variability by 35%, improving overall traffic flow stability. Spatial-temporal analysis revealed consistent peak congestion between 7:00–9:00 AM and 6:00–8:00 PM. In these periods and locations, the model's predictions correlated strongly ($r = 0.92$) with real-time traffic data, confirming its robustness for operational deployment.

5 Limitations of the Proposed Model

Despite the strong performance of the proposed hybrid model that integrates LSTM neural networks with physics-based traffic modeling, certain limitations must be acknowledged to contextualize its scope and generalization. This section addresses key constraints related to data dependency, geographic bias, generalizability, and implementation feasibility. While the hybrid approach shows high accuracy and robustness, its results must be interpreted within the framework's boundaries. The model's dependence on real-time data from the HERE API is both a strength and a vulnerability. The data's granularity and timeliness enhance training and prediction, but changes in data availability, access policies, or schema formats could disrupt its operation. This highlights the need for standardized, reliable data sources in traffic prediction systems.

Regarding geographic bias, the dataset was collected exclusively from the Tlalpan Borough of Mexico City, an area with complex congestion patterns and heterogeneous traffic conditions. While this setting provides a strong foundation for model validation, applying the model to other urban regions would require retraining and adaptation to local dynamics. The integration of physical constraints improves stability within Tlalpan, but these same constraints may limit flexibility in cities with different infrastructures or mobility behaviors. In terms of generalization, the hybrid framework balances predictive accuracy with interpretability, but its performance may be affected by unmodeled variables such as weather conditions, large-scale events, or variations in public transport usage, which were not explicitly considered. Nevertheless, the current design serves as a foundation for future extensions that could incorporate these additional factors to improve explanatory power.

Finally, the training process based on 30 million records and extended temporal sequences demands significant computational resources. However, this data scale has enabled the construction of a representative and scalable system for urban mobility analysis, demonstrating that with a carefully designed architecture, high predictive performance is achievable in technologically equipped environments.

6 Conclusions

This study presents a hybrid model for adaptive traffic congestion prediction and control, integrating Long Short-Term Memory (LSTM) neural networks with differential equations from physical traffic models. The proposed approach addresses the need for systems that are accurate, robust, interpretable, and dynamically coherent, capable of modeling the complexity of urban traffic. A key strength lies in embedding physical constraints into the learning process, which enhances generalization, reduces reliance on historical data, and aligns predictions with real-world dynamics. The model predicts congestion events up to 24 hours in advance, enabling proactive strategies and real-time responses to incidents such as accidents or road closures. Operationally, it can be deployed within intelligent traffic control systems to adjust signal timings, optimize routing, and mitigate congestion across critical corridors.

Validation was conducted using SUMO microscopic simulations with real traffic data from Tlalpan, Mexico City, where the model accurately captured peak-hour congestion patterns. All scenario files, source code, and implementation guides are publicly available to ensure full reproducibility. Experimentally, the hybrid model outperformed traditional approaches (ARIMA, linear regression, standalone LSTM) and advanced architectures like Temporal Convolutional Networks (TCNs), achieving lower RMSE and MAPE and a higher R^2. Sensitivity and ablation studies demonstrated that incorporating physical constraints—as input features and as penalty terms in the loss function led to significant performance gains. Robustness was tested via 30 independent training runs and confirmed through a one-way ANOVA test, showing statistically consistent results and reinforcing the framework's stability. The limitations section outlines key concerns: reliance on HERE API data, geographic specificity to Tlalpan, and the need for retraining in other urban settings. Despite these, the model provides a scalable base for future improvements, including the integration of external variables such as weather, events, and public transport variations. Future work should explore advanced spatiotemporal architectures (e.g., Temporal Transformers, Graph Neural Networks), multimodal data sources (e.g., IoT, cameras, meteorology), and reinforcement learning for dynamic control optimization. Overall, this hybrid integration of deep learning and physical modeling offers a significant step forward in intelligent traffic systems, delivering more accurate, explainable, and adaptive solutions for modern urban mobility challenges.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (pp. 265–283). USENIX Association. https://www.tensorflow.org/

Ascencio, J. A., Carlos, L., Martner, D., Agustín, P., & Rosales, B. (2022). Implementation of a traffic prediction model using deep learning (707). Instituto Mexicano del Transporte.

Behrisch, M., Bieker-Walz, L., Erdmann, J., & Krajzewicz, D. (2011). *SUMO – Simulation of Urban MObility: An overview*. In *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation* (pp. 63–68). ThinkMind. https://sumo.dlr.de

Bridge, K. I. I., Village, K., Zulkarnaen, D., Mayuni, S., & Mukti, E. T. (2024). Geometric Design of Coordinated Traffic Signal Intersections : Case Study of Four-Way Intersections at Kapuas II Bridge and Kapur Village. *Jurnal Teknik Sipil*, *24*(3), 1183–1200.

Chawla, P., Hasurkar, R., Bogadi, C. R., Korlapati, N. S., Rajendran, R., Ravichandran, S., Tolem, S. C., & Gao, J. Z. (2024). Real-time traffic congestion prediction using big data and machine learning techniques. *World Journal of Engineering*, *21*(1), 140–155. https://doi.org/10.1108/WJE-07-2021-0428

Chollet, F. (2015). *Keras* [Computer software]. https://keras.io/

De la Cruz Nicolás, E. (2025). *Traffic Incidents Dataset: Traffic conditions on avenues in Tlalpan* [Dataset]. Google Drive. https://goo.su/5bzzO

De la Cruz Nicolás, E. (2025). *Traffic modeling implementation and source code* [Source code and dataset]. https://goo.su/hYjud

De la Cruz Nicolás, E. (2025). *SUMO/TraCI simulation files for traffic control and mobility optimization* [Dataset]. https://goo.su/ZbSK

Feroz Khan, A. B., & 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. https://doi.org/10.56578/judm020404

Florentino, Y. I. A., & Gismondi, H. E. C. (2024). Intelligent Traffic Light Management System in Real Time. *Revista Iberica de Sistemas e Tecnologias de Información*, *2024*, 69–87. https://doi.org/10.17013/risti.54.69-87

Gupta, K., & Santhanam, M. S. (2021). Extreme events in Nagel–Schreckenberg model of traffic flow on complex networks. *European Physical Journal: Special Topics*, *230*(16–17), 3201–3209. https://doi.org/10.1140/epjs/s11734-021-00016-0

Harsono, T., & Arai, K. (2024). Modeling Micro Traffic Flow Phenomena Based on Vehicle Types and Driver Characteristics Using Cellular Automata and Monte Carlo. *International Journal of Advanced Computer Science and Applications*, *15*(8), 973–985. https://doi.org/10.14569/IJACSA.2024.0150896

HERE Technologies. (2025). *HERE Traffic API* [Web service]. https://developer.here.com/

Itu, R., & Danescu, R. (2024). Fully Convolutional Neural Network for Vehicle Speed and Emergency-Brake Prediction. *Sensors*, *24*(1). https://doi.org/10.3390/s24010212

Kaleem, S., Sohail, A., Babar, M., Ahmad, A., & Tariq, M. U. (2024). A hybrid model for energy-efficient Green Internet of Things enabled intelligent transportation systems using federated learning. *Internet of Things (Netherlands)*, *25*(101038). https://doi.org/10.1016/j.iot.2023.101038

Kashyap, A. A., Raviraj, S., Devarakonda, A., Nayak K, S. R., Santhosh, K. V., & Bhat, S. J. (2022). Traffic flow prediction models–A review of deep learning techniques. *Cogent Engineering*, *9*(1). https://doi.org/10.1080/23311916.2021.2010510

Khan, A., Fouda, M. M., Do, D. T., Almaleh, A., & Rahman, A. U. (2023). Short-Term Traffic Prediction Using Deep Learning Long Short-Term Memory: Taxonomy, Applications, Challenges, and Future Trends. *IEEE Access*, *11*, 94371–94391. https://doi.org/10.1109/ACCESS.2023.3309601

Khatri, S., Vachhani, H., Shah, S., Bhatia, J., Chaturvedi, M., Tanwar, S., & Kumar, N. (2021). Machine learning models and techniques for VANET based traffic management: Implementation issues and challenges. *Peer-to-Peer Networking and Applications*, *14*(3), 1778–1805. https://doi.org/10.1007/s12083-020-00993-4

León, D. A., Martínez Cuenca, J. G., Ardila Sánchez, I. A., & Mosquera Palacios, D. J. (2022). Artificial Intelligence for Traffic Control in Data Networks: A Review. *Entre Ciencia e Ingeniería.*, *16*(31), 17–24. https://doi.org/10.31908/19098367.2655

Mauricio, F., & Torres, T. (2023). The Lima Traffic Light Network Operations and Management System. *Industrial Data*, *26*(2), 93–118. https://doi.org/10.15381/idata.v26i2.24038.g20428

Nuli, S., Vikranth, N., & Gupta, K. A. (2022). Real-Time Traffic Prediction Using Neural Networks. *IOP Conference Series: Earth and Environmental Science*, *1086*(1). https://doi.org/10.1088/1755-1315/1086/1/012029

Pan, Z., Zhang, W., Liang, Y., Zhang, W., Yu, Y., Zhang, J., & Zheng, Y. (2022). Spatio-Temporal Meta Learning for Urban Traffic Prediction. *IEEE Transactions on Knowledge and Data Engineering*, *34*(3), 1462–1476. https://doi.org/10.1109/TKDE.2020.2995855

Rastgoftar, H., & Jeannin, J. B. (2021). A Physics-Based Finite-State Abstraction for Traffic Congestion Control. In *2021 American Control Conference* (pp. 237–242). IEEE

Ravichandran, N., Losanno, D., Pecce, M. R., & Parisi, F. (2024). Site-specific traffic modelling and simulation for a major Italian highway based on weigh-in-motion systems accounting for gross vehicle weight limitations. *Journal of Civil*

Structural Health Monitoring, *14*, 1739–1763. https://doi.org/10.1007/s13349-024-00809-6
Rivera-Picado, C., & Meneses-Guzmán, M. (2022). Vehicular Traffic Flow Prediction on Route 27 in Costa Rica. *Revista Tecnología En Marcha*, *35*, 138–148. https://doi.org/10.18845/tm.v35i4.5892
Romanowska, A., & Jamroz, K. (2021). Comparison of traffic flow models with real traffic data based on a quantitative assessment. *Applied Sciences*, *11*(21), 9914. https://doi.org/10.3390/app11219914
M. Rehmat Ullah, Khurram S. Khattak, Zawar H. Khan, Mushtaq A. Khan, Nasru Minallah and Akhtar N. Khan. (2021). Vehicular Traffic Simulation Software: A Systematic Comparative Analysis. *Pakistan Journal of Engineering and Technology*, *4*(1),66-78.
Uppaluru, H., Liu, X., Emadi, H., & Rastgoftar, H. (2022). A Continuous-Time Optimal Control Approach to Congestion Control. In *European Control Conference* (pp. 1572–1577). IEEE
Villarroya, C., Calafate, C. T., Onaindia, E., Cano, J. C., & Martinez, F. J. (2022). Neural Network-based Model for Traffic Prediction in the City of Valencia. *Procedia Computer Science*, *207*, 552–562. https://doi.org/10.1016/j.procs.2022.09.110
Villarroya Sánchez, C. (2021). *Neural Network-Based Model for Traffic Prediction in the City of Valencia*. Polytechnic University of Valencia
Wang, Y., Yu, X., Guo, J., Papamichail, I., Papageorgiou, M., Zhang, L., Hu, S., Li, Y., & Sun, J. (2022). Macroscopic traffic flow modelling of large-scale freeway networks with field data verification: State-of-the-art review, benchmarking framework, and case studies using METANET. *Transportation Research Part C: Emerging Technologies*, *145*(103904), 103904. https://doi.org/10.1016/j.trc.2022.103904
Ziad, T., Verdezoto, A., Felix, F., Montes, C., Beatriz, O., & Medina, R. (2020). Analysis of Traffic Congestion for the Improvement of a Main Road in Guayaquil, Ecuador. *Universidad Centroccidental Lisandro Alvarado, 21*(2), 4-23. http://dx.doi.org/10.13140/RG.2.2.21905.04960
Chen, Y. T., Liu, A., Li, C., Li, S., & Yang, X. (2025). Traffic flow prediction based on spatial-temporal multi factor fusion graph convolutional networks. *Scientific Reports, 15*(1), 1–17. https://doi.org/10.1038/s41598-025-96801-1
Wang, Y., & Chen, P. (2025). Network traffic prediction based on transformer and temporal convolutional network. *PLoS ONE, 20*(4), 1–22. https://doi.org/10.1371/journal.pone.0320368
Wei, S., Yang, Y., Liu, D., Deng, K., & Wang, C. (2024). Transformer-Based Spatiotemporal Graph Diffusion Convolution Network for Traffic Flow Forecasting. *Electronics, 13*(16). 3151. https://doi.org/10.3390/electronics13163151
Wu, C., Ding, H., Fu, Z., & Sun, N. (2024). Multi-Scale Spatio-Temporal Attention Networks for Network-Scale Traffic Learning and Forecasting. *Sensors, 24*(17), 1–16. https://doi.org/10.3390/s24175543
Yang, H., Jiang, C., Song, Y., Fan, W., Deng, Z., & Bai, X. (2024). TARGCN: temporal attention recurrent graph convolutional neural network for traffic prediction. *Complex and Intelligent Systems, 10*(6), 8179–8196. https://doi.org/10.1007/s40747-024-01601-1
Zhao, Z., Chen, W., Wu, X., Chen, P. C. Y., & Liu, J. (2017). LSTM network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems, 11*(2), 68–75. https://doi.org/10.1049/iet-its.2016.0208