



OPEN ACCESS

EDITED BY

Moustafa Nasralla,
Prince Sultan University, Saudi Arabia

REVIEWED BY

Sohaib Bin Altaf Khattak,
Prince Sultan University, Saudi Arabia
Mehre Munir,
Prince Sultan University, Saudi Arabia

*CORRESPONDENCE

Adnan Nadeem,
✉ adnan.nadeem@iu.edu.sa
Akbar Ali,
✉ aakbarali18@gmail.com
Noureen Zafar,
✉ noureen_zafar@uau.edu.pk
Muhammad Shiraz,
✉ muhammad.shiraz@fuuast.edu.pk

RECEIVED 15 July 2025

ACCEPTED 07 October 2025

PUBLISHED 30 October 2025

CITATION

Ali A, Nadeem A, Zafar N and Shiraz M (2025) F-GGRU: a sensor-driven deep learning framework for smart city weather-aware traffic congestion prediction.
Front. Commun. Netw. 6:1666487.
doi: 10.3389/frcmn.2025.1666487

COPYRIGHT

© 2025 Ali, Nadeem, Zafar and Shiraz. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

F-GGRU: a sensor-driven deep learning framework for smart city weather-aware traffic congestion prediction

Akbar Ali^{1*}, Adnan Nadeem^{2*}, Noureen Zafar^{3*} and Muhammad Shiraz^{1*}

¹Department of Computer Science, Federal Urdu University of Arts, Science and Technology, Islamabad, Pakistan, ²Faculty of Computer and Information Systems, Islamic University of Madinah, Medina, Saudi Arabia, ³Department of Computer Science, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan

The deployment of various sensors including inductive loops, radars, GPS devices, cameras and floating car data (FCD) in intelligent transportation systems generates a stream of heterogeneous data, further complicated by exogenous factors like weather conditions and temporal patterns (e.g., peak hours, weekends). For urban traffic development planning, the accurate prediction of congestion under the influence of these exogenous factors remains a major challenge. The proliferation of these diverse data sources creates a complex prediction environment, demanding advanced analytical frameworks. To address this issue, we propose a novel Fusion-based Generative Adversarial Network with Gated Recurrent Unit (F-GGRU) framework. The F-GGRU develops a generic data pipeline for integrating and preprocessing multi-source data, featuring advanced techniques for outlier removal, fuzzy logic-based automatic labeling, and Generative Adversarial Networks (GANs) for class balancing. Extensive experimentation was conducted on a novel real-time dataset from the Safe City Islamabad Pakistan (SCIP) project, integrating heterogeneous and exogenous features. The results demonstrate that our proposed F-GGRU framework achieves superior performance, with 98% accuracy, 0.99 precision, 0.98 recall, and a 0.98 F1-score. This significantly outperforms a suite of benchmark models, including Logistic Regression, Random Forest, XGBoost, and deep learning baselines like ANN, which achieved accuracies between 77% and 83% with correspondingly lower precision, recall, and F1-scores. Significantly, hyperparameter tuning and validation on a second independent dataset (CityPulse, Aarhus) confirmed the proposed framework robustness and generalizability, achieving even higher performance 99.42% accuracy and 0.99 AUC. These findings affirm that the F-GGRU framework is a robust and generalizable solution for real world traffic congestion prediction in smart cities.

KEYWORDS

internet of things, generative adversarial networks, GRU, smart cities, weather, conditions, intelligent transportation system, sensors

1 Introduction

Currently, the rapid expansion of sensor technologies, internet connectivity, and high-volume data generation has reshaped the landscape of intelligent systems, especially in the context of Internet of Things (IoT)-driven applications and smart cities. The continuous stream of real-time, heterogeneous data captured from diverse sensor sources has unlocked vast potential for real-world insights and decision-making. However, it also introduces new complexities related to the efficient integration, interpretation, and utilization of such multifaceted information (Ren et al., 2023). Extracting meaningful patterns from these rich datasets requires innovative approaches capable of merging varied data types and applying advanced learning techniques to uncover actionable intelligence. In this context, leveraging sensor-derived data from urban surveillance infrastructures, such as the smart city, presents a compelling opportunity to explore intelligent traffic solutions in dynamically evolving environments.

The Intelligent Transportation System (ITS) serve as a foundational element in the transformation of urban regions into smart cities by enabling adaptive and data-driven traffic solutions. The growing demand for accurate and timely traffic forecasting necessitates the integration of real-time inputs from diverse sensor modalities, including fixed and mobile surveillance units, environmental monitoring systems, and external contextual sources. Modern ITS frameworks capitalize on IoT to construct interconnected platforms that combine multi-source, multi-sensor, and multi-model data streams. In this context, our research leverages the integrated pipelined dataset obtained from heterogeneous features collected from smart city real time environment, combining traffic surveillance with exogenous features from weather sensing data to enable more precise traffic speed and congestion prediction through our proposed hybrid F-GGRU framework.

Traffic congestion arises from a range of exogenous contributing factors (Rehborn and Koller, 2014), including weather conditions, peak-hour load (Ali et al., 2021a), road maintenance, accidents, and adverse weather conditions (Romanowska and Budzyński, 2022a). Weather condition sensors are instrumental in monitoring atmospheric parameters such as rainfall, wind speed, temperature, humidity and visibility offering important inputs for both real-time and predictive traffic models (Du et al., 2022) (Vargas et al., 2021). Within this context, sensors form the technological backbone of the smart city, which goals to establish a smart, responsive and secure urban environment. The Safe City Islamabad, Pakistan (SCIP) network includes 2,758 cameras monitor high-definition surveillance radar-based cameras allowing near-total coverage of the SCIP (Khattak, 2025). Smart City authorities (Authors Anonymous, 2022) state that the initiative aims to ensure a digital record is maintained for every vehicle that enters the city. Additional by mobile patrol vehicles and surveillance drones, these sensor networks are linked to a centralized control hub, enabling continuous data collection. These extensive sensor deployments not only enhance urban smart city but also produce rich spatiotemporal datasets helping as a fundamental enabler of intelligent transportation research and congestion prediction.

There are two principal approaches to alleviating traffic congestion on urban road networks. The first involves expanding

infrastructure by increasing the number of freeway lanes or constructing new roads. However, this solution demands significant land acquisition and financial investment, which may not be practical or sustainable in densely populated urban areas. Alternatively, a more efficient and scalable strategy is to implement intelligent traffic control mechanisms that optimize the use of existing road infrastructure. These control strategies depend heavily on accurate congestion prediction, empowering authorities to proactively manage traffic flow and mitigate congestion before it increases.

According to the United Nations, it is estimated that by 2050 nearly 68% of the global population will reside in urban regions, reflecting a significant shift toward urbanization (United Nation, 2023).

Country wise rapid pace of urbanization, coupled with exponential growth in vehicle ownership, has introduced serious challenges to traffic management, road safety, environmental sustainability, and overall urban livability. Pakistan population has grown tremendously, with census data indicating an increase from 132.35 million in 1998 to 241.49 million in 2023. Projections suggest this number will reach approximately 255.22 million by 2025 (Authors Anonymous, 2021). Alongside population growth, the number of people residing in urban areas has risen from 32.5% in 1998 to 38.82% in 2023, highlighting a marked demographic shift toward cities. This surge in urban population and vehicular density underscores the urgent need for intelligent, sensor-driven traffic solutions to address congestion and ensure sustainable urban mobility.

Furthermore, in the United States 3.3 Billion Gallons of fuel are wasted in 2022 due to the traffic congestion (U.S Department of Energy, 2025). Exogenous features (Weather conditions, peak hours, week days and weekends) (Ali et al., 2021a) can significantly influence traffic flow and congestion levels (Romanowska and Budzyński, 2022b; Agarwal et al., 2005; Lin et al., 2015). Rain (Mashros et al., 2014), fog (Ali et al., 2024), or snow often reduce visibility and road friction, leading to slower driving speeds and increased travel time. Severe weather may also cause accidents, lane closures, or disruptions in traffic signals, further compounding delays. Even mild weather changes can alter driver behavior, contributing to fluctuations in congestion patterns across urban roads.

In previous research work, for the traffic congestion prediction several Mobile Crowded Sensing (MCS) (Ali et al., 2021b), ML (YixinLi and Zhang, 2024; Yu and Xie, 2024) and DL (Zafar et al., 2022a; Lartey et al., 2021) techniques have been presented, including KNN, Support Vector Machines, RF, and deep networks such as GRUs, Convolutional Neural Network (CNN) and Long Short-Term Memory. Whereas these techniques have achieved varying accuracy, some of the techniques deal to adapt to real-world complexities such as temporal variations, missing data and sudden disturbances caused by external factors like weather. Traditional models often treat traffic data in isolation, supervising the impact of exogenous variables and failing to integrate heterogeneous data sources effectively. Furthermore, imbalanced class distributions and the lack of dynamic feature fusion methods limit their generalizability. However, their performance often declines in real-world situations due to issues such as imbalanced class distributions, missing or noisy data,

temporal irregularities, and an inability to effectively integrate exogenous factors like weather conditions or peak hours and holidays. The existing techniques tend to deal with traffic data sources separately, lacking robust fusion strategies for heterogeneous and exogenous inputs. Our proposed F-GGRU framework overcomes these limitations through a novel hybrid pipelined combination of techniques: F-GGRU framework integrates fuzzy logic to fuse heterogeneous and exogenous features, use fuzzy logic-automatic labeling for classification, GANs for class balancing to balance the imbalanced class, and enhances temporal learning through a gated recurrent structure. This empowers the F-GGRU framework to capture more nuanced patterns in smart cities dynamic urban environments, contributing improved robustness, and predictive accuracy compared to existing approaches (Ali et al., 2021b; YixinLi and Zhang, 2024; Yu and Xie, 2024; Zafar et al., 2022a; Al-Qarafi et al., 2022; Yasir et al., 2022; Zafar et al., 2022b; Zhao et al., 2019; Zhong et al., 2024).

1.1 The following are scientific contributions to intelligent transportation system and smart cities

- The F-GGRU framework integrates underutilized smart city traffic observation sensor heterogeneous (FCD) data with real-time exogenous (weather conditions information, peak hour, week days and weekend) information transforming passive traffic records into actionable congestion prediction insights improving the predictive power of ITS and supporting real-time traffic management decisions.
- Enhanced data handling through applying automatic fuzzy logic for intelligent labeling and to balance class distribution use GANs.
- The usage of a Gated Recurrent Unit in F-GGRU framework empowers effective modeling of time-dependent traffic patterns, allowing the F-GGRU framework to learn long-term and short-term dependencies in sequential data while maintaining computational efficiency.

After the introduction presented in this Section 1, the literature review description presented in Section 2. The proposed Hybrid F-GGRU framework is presented in Section 3. Section 4 presents result analysis and discussion explaining the results. Conclusion and future research direction presented in Section 5.

2 Literature review

The integration of weather conditions into traffic congestion prediction models has garnered significant attention in recent years. Adverse weather events, such as fog, rain, snow, and humidity, have been empirically shown to influence traffic flow characteristics, including speed, capacity, and headways. For instance (Romanowska and Budzyński, 2022b) conducted a comprehensive study on a Polish expressway, revealing that average vehicle speeds could decrease by up to 19% and road capacity by 18% under adverse weather conditions, compared to normal conditions.

The study (Ali et al., 2021b) proposed MCS based dynamic traffic efficiency framework for traffic congestion prediction and avoidance. The real time vehicular traffic data collected through GetApp mobile application. The recommended MCS based dynamic traffic efficiency framework allocate the fastest available route with specific time slot to the commuter to follow and reached to the destination on time, but it does not explore other potential influencing factors in depth, such as weather conditions, which could also significantly impact congestion levels.

In research study (Zafar et al., 2022a) the author presented LSTM-GRU model that combine heterogeneous data sources collected data from sensors sources, holiday data, tracking company, OSM road, Google, peak hour data and weather data. Weather data from weather APIs and Open Street map for mapping. The authors have provided an exploratory data analysis using GRU, LSTM, CNN and their hybrid integration. The combination of LSTM + GRU hybrid gave the finest output with 6.67% MAPE and 4.5% RMSE. As classification the LSTM-GRU model yields 95% accuracy.

This study (Pragalathan and Schramm, 2024) employs the Neural Prophet (NP) model to advance the prediction of urban traffic dynamics by incorporating exogenous variables such as meteorological conditions and public holidays. By integrating classical time-series analysis with neural network architectures, the NP model is capable of capturing non-linear and seasonally varying patterns in traffic flow. The model's responsiveness to external influences particularly rain and calendric events demonstrates its robustness in urban forecasting contexts. These findings emphasize the necessity of including environmental and temporal factors in traffic prediction models, particularly in metropolitan areas where such variables exert a significant influence on vehicular movement.

In research study (Sun et al., 2021) the authors developed an online traffic flow prediction framework that integrates a Bi-LSTM framework with CNN. Real time data collected from the IoT sensors situated at intersection of Hongzhe Road and Qingnian Road, in Suqian City, Jiangsu Province, China. The dataset, structured as a time series, is managed by a Bi-LSTM network helping as the generator, while a CNN operates as the discriminator within the GAN model. To evaluate predictive performance, the study applies metrics including MSE, MAE, and binary entropy. According to the reported results, this GAN-based approach demonstrates improved accuracy compared to separate Bi-LSTM and ARIMA models. The study does not touch weather impact on traffic flow.

The study (Solanki et al., 2023) incorporates weather data alongside traffic and Twitter messages to enhance traffic flow predictions. The deep learning model Bi-directional LSTM Stacked Auto Encoder architecture is used, it aims to improve accuracy in predicting traffic congestion under varying weather conditions, contributing to effective traffic management. Twitter messages data and traffic and weather datasets are used.

The paper (Valarmathi and Dhanalakshmi, 2024) discusses using genetic algorithms for optimization and IoT for real time data collection of weather-adaptive traffic monitoring, enabling dynamic adjustments in traffic management strategies. This integration helps predict and mitigate traffic congestion by analyzing real-time weather data and optimizing traffic flow accordingly.

The paper (Dong et al., 2010) develops traffic estimation and prediction models that account for traffic response to extreme weather, enabling real-time traffic management systems to predict congestion and implement advisory and control strategies effectively, thus mitigating weather impacts on traffic flow. Loop detectors and roadside sensors data. Vehicle probes data for traffic conditions. Deficiency in current weather-responsive traffic management practices. Need for improved traffic estimation models for inclement weather.

The TransGTR-MCA model (Cui, 2024) incorporates weather factors, specifically precipitation, to enhance traffic flow predictions. By considering these external conditions, the model improves adaptability and accuracy in predicting traffic congestion, particularly in data-scarce urban environments. Limitations of the proposed work are insufficient optimization for long-term prediction accuracy and need for better adaptation to external factors.

The proposed Multilevel-Gated Recurrent Unit (MGRU) model (Sravani et al., 2024) incorporates weather conditions and vehicle numbers to enhance traffic congestion prediction accuracy, achieving a notable accuracy of 0.887 and a Mean Absolute Error of 82.34, outperforming existing methods like Conv-Bi-LSTM.

A Weather interaction-aware spatio-temporal attention network (WST-ANet) model (Zhong et al., 2024) proposed effectively predicts traffic flow by integrating weather factors, enhancing adaptability to varying weather conditions. It utilizes a spatio-temporal weather collaboration insight module, improving accuracy in forecasting traffic congestion under different weather scenarios.

The work in (Yasir et al., 2022) highlights that weather significantly impacts traffic congestion levels. By utilizing machine learning models MLP Regressor, Stacking Regressor, SVR used for predictions. Historical traffic volume data utilized for training and assessment, the study effectively predicts congestion dynamics, taking into account various weather parameters alongside time of day and holiday indicators to enhance forecasting accuracy. Through this model, congestion of a road can be predicted 1 week in advance with an average RMSE of 1.12. Consequently, this model can be used to take preventive measure beforehand.

Recent advances in traffic prediction have demonstrated the strengths and limitations of different modeling approaches. Early studies such as (Wu et al., 2018) and (Polson and Sokolov, 2017) focused primarily on temporal patterns of traffic flow using deep neural networks and recurrent models. While these methods achieved promising accuracy for short-term prediction, they lacked the ability to incorporate exogenous factors, making them less reliable under unusual or disruptive conditions.

Subsequent works introduced weather and accident information to improve robustness. For example, (Zhong et al., 2024), highlighted that rainfall and visibility significantly affect congestion, while Gu et al. (2016) emphasized the role of accident reports in traffic disruption. Although these studies improved predictions (Wang et al., 2016) under specific scenarios, they were limited by narrow boundary considerations and often struggled with the data imbalance problem, where rare events such as severe accidents or extreme weather were underrepresented.

The study (Yixinli and Zhang, 2024) investigates the relationship between various traffic modes and congestion levels, establishing a high-fidelity prediction model that analyzes multi-modal traffic and congestion data across different time frames. It employs machine learning techniques, including decision tree, logistic regression, KNN and random forest (RF) models. The RF model achieved accuracy of 99.88% after optimization. The study primarily focuses on traffic volume as the most important predictor of congestion, but it does not explore other potential influencing factors in depth, such as weather conditions, road infrastructure, or driver behavior, which could also significantly impact congestion levels. The paper does not specifically address the impact of weather conditions on traffic congestion prediction. The paper establishes a robust prediction model using various machine learning techniques; it does not address the addition of real-time data sources or application of the model in dynamic traffic management systems, which could enhance the practical utility of the findings in real-world scenarios.

The study (Hazarika et al., 2024) introduced an edge machine learning framework for adaptive traffic signal control in intelligent transportation systems. Their method utilized lightweight object detection models deployed at the edge to monitor vehicle density in real time and dynamically adjust traffic light phases. While their approach significantly reduced delays at intersections and improved traffic coordination, the study did not explicitly consider exogenous influences such as weather variability or special events, which often disrupt flow patterns. This highlights a key research direction, as integrating environmental conditions into traffic prediction frameworks can complement edge-based management systems by providing more resilient and context-aware congestion forecasting.

Despite the fact that various of the above studies on vehicular traffic congestion prediction presents valuable insights and robust models, these also has certain limitation like it does not explore other potential influencing factors in depth, such as weather conditions, which could also significantly impact congestion levels that should be acknowledged. These limitations indicate areas for future research and improvement, highlighting the need for a broader framework to understanding and predicting traffic congestion effectively on the basis of heterogeneous and exogenous data. There are some limitations of widely used models such as LSTM, Bi-LSTM and CNN-based hybrids. The LSTMs often suffer from gradient vanishing, while Conv-bi-LSTM has high computational cost. These shortcomings motivated the development of our FG-GGRU model, which integrates fuzzy logic with gated recurrent units to address these issues.

3 Proposed hybrid fusion-based generative adversarial network applied gated recurrent unit framework

To effectively capture the sequential and temporal dynamics inherent in traffic flow data, this study employs a Gated Recurrent Unit (GRU)-based neural architecture at the core of the predictive framework. GRUs are a refined variant of traditional Recurrent Neural Networks (RNNs), specifically designed to preserve relevant information over long sequences while discarding irrelevant signals through an efficient gating mechanism.

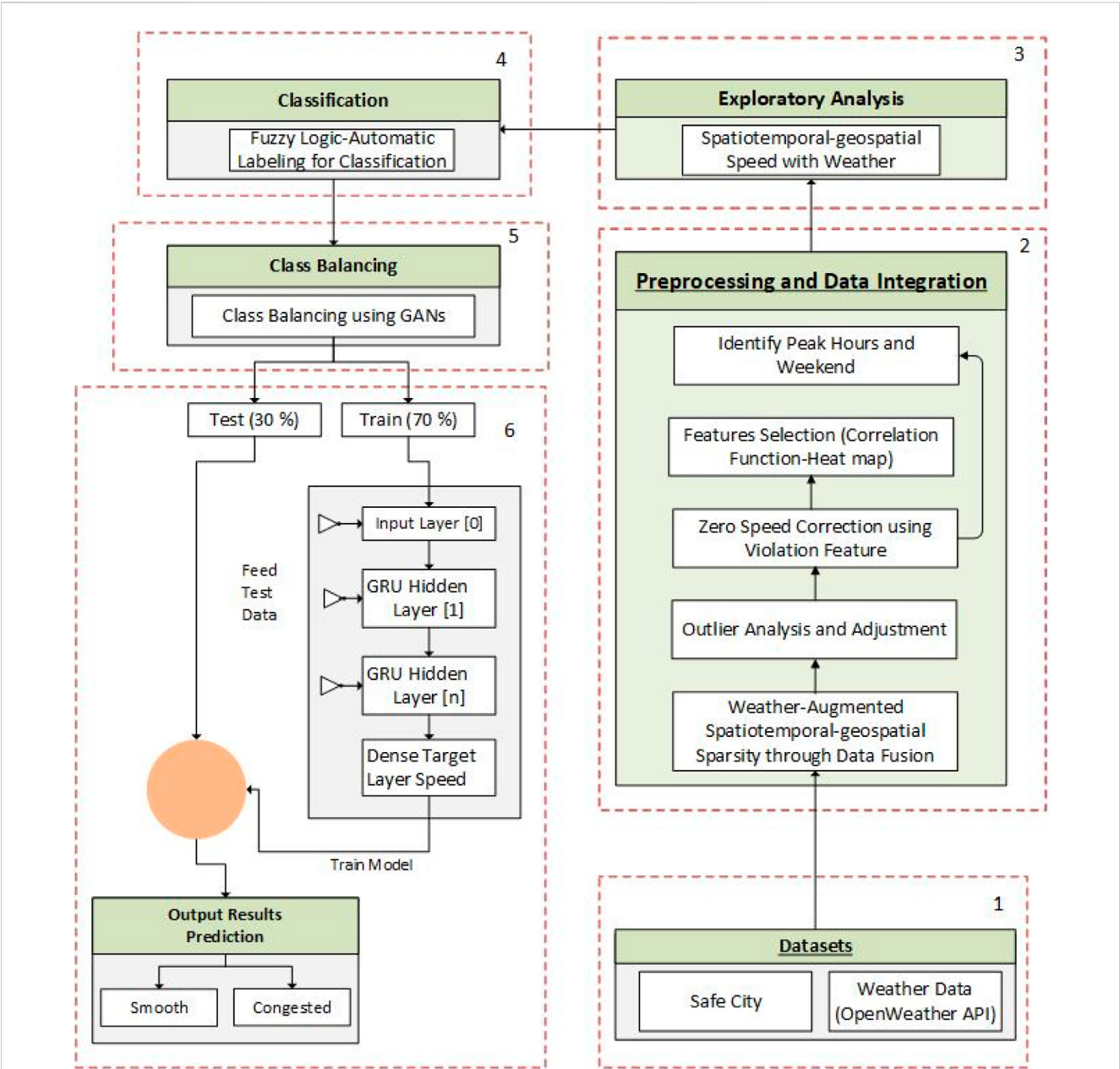


FIGURE 1
A novel hybrid features F-GGRU framework for traffic congestion prediction under weather conditions and temporal patterns.

Compared to Long Short-Term Memory (LSTM) networks, which utilize three gates (input, forget, and output) and a memory cell, GRUs reduce architectural complexity by combining the forget and input mechanisms into a single update gate and introducing a reset gate. This simplification results in fewer parameters, reduced computational overhead, and faster training making GRUs well-suited for time-sensitive and resource-limited applications such as traffic congestion prediction.

We propose hybrid F-GGRU framework shown in Figure 1, which builds on the fundamental GRU design and improves the model's learning potential in two significant ways. In order to ensure that the framework learns representative patterns from both congested and smooth traffic situations, we first implement a GAN-based class

balancing strategy to address label imbalance in the binary congestion classification problem mathematically algorithmic representation of F-GGRU framework presented in algorithm 1 to algorithm 6. We develop a multisource data fusion approach that combines exogenous (like weather, peak hours, and holidays) and heterogeneous (like Smart City Automatic Number Plate Recognition Cameras, patrol, and drone sensor) inputs into a single time-series sequence. Through encoding and passing these enhanced inputs via the GRU layers, the F-GGRU framework is able to learn contextual impacts and temporal dependencies simultaneously. The fusion approach make the model adaptable to abrupt changes and noise in the data, enhancing its capacity to represent the complexity of real world urban transportation network.

3.1 Mathematical model explanation of the F-GGRU framework

The core of the proposed F-GGRU framework is based on the standard Gated Recurrent Unit (GRU) architecture, which is enhanced through data fusion and synthetic class balancing. The GRU cell is designed to capture temporal dependencies by controlling the flow of information across time steps using gating mechanisms. Here $\tilde{y}_t \in \mathbb{R}^n$ denote the fused input vector at time step t , constructed by integrating heterogeneous features and exogenous attributes. This enriched input forms the sequential input to the F-GGRU framework.

The cell computations are defined as follows:

i. Update Gate:

$$a. z_t = \sigma(W_z \cdot \tilde{y}_t + U_z \cdot h_{t-1} + b_z) \quad (1)$$

ii. Reset Gate:

$$a. r_t = \sigma(W_r \cdot \tilde{y}_t + U_r \cdot h_{t-1} + b_r) \quad (2)$$

iii. Candidate Activation:

$$a. \tilde{h}_t = \tanh(W_h \cdot \tilde{y}_t + U_h \cdot (r_t \odot h_{t-1}) + b_h) \quad (3)$$

iv. Final Hidden State:

$$a. h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (4)$$

$$\therefore \tilde{y}_t = [y_t^{FCD}, y_t^{weather}, y_t^{peak\ hours}, y_t^{weekday}, y_t^{weekend}]$$

Where symbols stands for:

- $\tilde{y}_t$ is the input at time step t ,
- h_{t-1} is the previous hidden state,
- h_t Final hidden state at time t ,
- z_t and r_t are the update and reset gates respectively,
- $\tilde{h}_t$ is the candidate activation,
- $\odot$ denotes element-wise multiplication,
- σ Sigmoid activation function (outputs between 0 and 1),
- W_z, W_r, W_h Weight matrices for input $\tilde{y}_t$ in the update, reset, and candidate gates,
- U_z, U_r, U_h Weight matrices for hidden state h_{t-1} ,
- b_z, b_r, b_h Bias terms for the respective gates
- $\tanh$ Hyperbolic tangent activation function (outputs between -1 and 1).

These Equations 1–4 collectively describe the temporal learning mechanism of the GRU cell within the F-GGRU framework. The update gate z_t regulates how much of the previous hidden state is carried forward, while the reset gate r_t determines how much past information to forget. By feeding the GRU with a multi-source fused input $\tilde{y}_t$ the F-GGRU framework effectively captures contextual dependencies and temporal dynamics for robust traffic congestion prediction.

3.1.1 Loss function and GAN-based class balancing

To enhance the reliability of the model in handling imbalanced congestion labels, a GAN-based synthetic oversampling mechanism

is applied during preprocessing. Specifically, a Vanilla Generative Adversarial Network (GAN) is utilized to generate artificial feature vectors that mimic the statistical distribution of the underrepresented class (e.g., congested or smooth traffic states). The GAN contains two neural components: A generator $G(z; \theta_g)$, which maps input noise $z \sim N(0, 1)$ to synthetic traffic feature vectors $\hat{x}$. A discriminator $D(x; \theta_d)$, which attempts to distinguish between real input samples x and generated samples $\hat{x}$.

The generator and discriminator are trained via the following minimax loss function in Equation 5:

$$\min_G \max_D L_{GAN}(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))] \quad (5)$$

Where:

- p_{data} is the distribution of real traffic data samples.
- p_z is the prior distribution over latent noise vectors.
- $D(x)$ represents the probability that sample x is real.

When trained, the generator G is used to synthesize new samples for the minority class, thereby producing a balanced training dataset that ensures fair learning during the model's optimization phase.

3.1.2 F-GGRU framework loss function

For the prediction task, the F-GGRU framework is trained as a binary classifier using the Binary Cross-Entropy (BCE) loss function, defined by Equation 6:

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=1}^N [x_i \log(\hat{x}_i) + (1 - x_i) \log(1 - \hat{x}_i)] \quad (6)$$

Where:

- $x_i \in \text{Error! Bookmark not defined.}$ is the true label for the i -th sample (0 = Congested, 1 = Smooth).
- $\hat{x}_i \in$ is the predicted probability from the F-GGRU framework.
- N is the number of training samples.

The loss function encourages the model to produce output probabilities that align closely with the ground truth, whereas the GAN-balanced dataset ensures that both classes contribute equally to learning, preventing bias toward the dominant class.

3.2 Algorithmic representation of F-GGRU framework

The proposed hybrid features F-GGRU framework pipeline can be expressed as a sequence of algorithmic steps applied to the raw dataset D_{raw} . The pipeline includes data preprocessing, feature engineering, class balancing, and final model prediction.

1. Data Preprocessing and Feature Engineering

1.1. Dataset Preparation:

Input: Raw Smart City traffic and Weather conditions data dataset D_{raw}

Output: Cleaned and integrated dataset D_{clean}

- 1.2. Load Dataset:
 - 1.2.1. s_i : Speed
 - 1.2.2. $t_i = (h_i, m_i, \text{sec}_i)$: Time (Hour, Minute, Second)
 - 1.2.3. $d_i = (\text{day}_i, \text{month}_i, \text{hour}_i)$: Date
 - 1.2.4. g_i : Geographic Sector
- 1.3. Perform data cleansing:
 - 1.3.1. Remove or impute missing values
 - 1.3.2. Handle outliers
- 1.4. Generate additional features:
 - 1.4.1. **Peak**
Hour: $\text{Is_Peak_Hour}_i = 1$ if $t_i \in \{7 - 9, 16 - 19\}$, else 0
 - 1.4.2. **Weekend**: $\text{Is_Weekend}_i = 1$ if day_of week (d_i) $\in \{6, 7\}$, else 0
- 1.5. Merge with weather dataset W:
 - 1.5.1. Join D_{raw} and W on (g_i, d_i)
 - 1.5.2. Aggregate and return enriched dataset D_{clean} .
2. Spatiotemporal and Weather-Speed Correlation:

Input: D_{clean}

Output: Correlation statistics

 - 2.1. Group data by (g_i, t_i, d_i)
 - 2.2. Compute correlation between:
 - 2.2.1. Speed vs. Weather attributes (temp, humidity, rainfall, etc.)
 - 2.2.2. Speed vs. Time attributes (peak hours, weekends)
 - 2.3. Store correlation matrix for model insight.
3. Fuzzy Logic-Based Traffic Speed Labeling:

Input: Clean dataset with speeds

Output: Fuzzy traffic speed labels

 - 3.1. Define fuzzy sets for Speed: {Low, Medium, High}
 - 3.2. Apply membership functions:
 - 3.2.1. Low: Congested
 - 3.2.2. Medium/High: Smooth
 - 3.3. Generate fuzzy rules and assign traffic label $y_i \in \{\text{Congested, Smooth}\}$.
4. GAN-Based Class Balancing:

Input: Labeled dataset D_{label} with imbalance

Output: Balanced dataset D_{balanced}

 - 4.1. Encode categorical features into numeric format
 - 4.2. Train a Conditional GAN (cGAN) using minority class labels
 - 4.3. Generate synthetic samples for minority class
 - 4.4. Merge synthetic and real data
 - 4.5. Shuffle to obtain D_{balanced} .
5. Train-Test Preparation:

Input: D_{balanced}

Output: Normalized and reshaped training/testing datasets

 - 5.1. Split D_{balanced} into X (features) and Y (labels)
 - 5.2. Normalize features using Min-Max scaling

Reshape input for F-GGRU: (n, timesteps, features)

: Where n is the number of samples, *timesteps* denotes the sequence length (look-back window), and *features* represents the number of variables at each time step.

 - 5.3. Prepare training and testing sets with equal class distribution.
6. FG-GGRU Framework Training and Prediction

Input: Training dataset (X train, Y train)

Output: Predicted traffic congestion labels

 - 6.1. Define FG-GGRU architecture:
 - 6.1.1. Input layer (1,d) (1, d) (1,d)

- 6.1.2. Fusion mechanism (traffic + weather + temporal features)
 - 6.1.3. GRU layers
 - 6.1.4. Dense layer with sigmoid activation
 - 6.2. Fusion function combines heterogeneous and exogenous features with weighted rules
 - 6.3. Compile model with:
 - 6.3.1. Loss = Binary Cross-Entropy
 - 6.3.2. Optimizer = Adam
 - 6.4. Train on (X_{train}, Y_{train})
 - 6.5. Evaluate on test data (X_{test}, Y_{test})
 - 6.6. Output binary predictions: {Congested, Smooth}.
- Objective is to train the F-GGRU framework to predict binary traffic congestion labels using fused inputs.

3.3 Data sources

For the F-GGRU framework evaluation used the real time heterogeneous and homogenous features data.

3.3.1 Heterogeneous features

We acquired heterogeneous features FCD data from the real time SCIP (Khattak, 2025) dataset. SCIP uses a network of advanced sensors of surveillance tools for urban monitoring and manually traffic management. We collect data of key camera types include high-definition sixty four CCTV, facial recognition systems linked with NADRA, Automatic Number Plate Recognition (ANPR), at strategic points across the city. These cameras are supported by additional sensors like RFID readers, radar systems for vehicle speed tracking, and flashlight systems for night visibility. The system integrates with a centralized command and control center and supports technologies like e-challan issuance and real-time traffic surveillance, obtaining valuable data for further traffic analysis and traffic congestion prediction. Traffic data of SCIP for the year 2023 have been collected from SCIP through proper channel. The key features of SCIP FCD include Date_Time, Date, Hour, District, Reg_No, Violation, Speed, Geographic_Sector, Police_Station, Camera_Name, latitude and longitude initially.

3.3.2 Exogenous features

The weather conditions data is collected from OpenWeather API (API, 2025) on the basis of latitude, longitude, time and date. The homogenous raw dataset have 28 features dt, dt_iso, timezone, city_name, lat, lon, temp, visibility, dew_point, feels_like, temp_min, temp_max, pressure, sea_level, nd_level, humidity, wind_speed, wind_deg, wind_gust, rain_1h, rain_3h, snow_1h, snow_3h, clouds_all, weather_id, weather_main, weather_description, weather_icon.

3.4 Fusion of heterogeneous and exogenous features based on correlation analysis

To enhance the representational strength of F-GGRU traffic prediction framework, this study introduces a data integration pipeline step 1 to step 3 in Figure 1, that fuses spatiotemporal traffic records (heterogeneous) with contextual environmental (exogenous) features that integrate the final dataset for F-GGRU

TABLE 1 Hybrid features space of heterogeneous and exogenous features for integrated dataset.

Heterogeneous (FCD) feature	Data type	Exogenous (weather, peak hours, weekend) feature	Data type
Speed	int64	feels like	float64
Violation	int64	Humidity	int64
Geographic Sector	int64	wind_speed	float64
Year	int64	clouds_all	int64
Month	int64	weather_main	int64
Day	int64	weather_description	int64
Hour	int64	Date	object
Minute	int64	Label	int64
Second	int64	Is Peak Hour	int64
		Is_Weekend	int64

framework training, both heterogeneous features (traffic-related attributes) and exogenous features (weather and temporal variables) were integrated into a single feature pool (SCIP). Since including all available attributes may introduce redundancy and noise, a correlation analysis was applied to identify the most relevant predictors. Using a heat map of pairwise correlations, weakly associated or redundant variables were removed, while strongly correlated features were retained.

The resulting integrated feature set provides a balanced representation of traffic dynamics and external influencing factors. Core traffic variables such as speed and geographic sector were preserved alongside exogenous indicators including temperature, humidity, wind speed, cloud cover, and categorical weather conditions. Temporal and contextual variables such as peak hour and weekend flag, indicators were also included, as their correlations with congestion patterns were significant. By fusing both heterogeneous and exogenous features under a correlation-driven selection strategy, the final dataset ensures that the FG-GGRU framework learns congestion patterns that are both data-rich and context-aware, thereby enhancing robustness under boundary conditions.

By integration spatial location, time of observation, and ambient environmental conditions into a unified structure, the pipeline takes a more comprehensive dataset features presented in Table 1 of revealing patterns otherwise obscured in isolated data streams. This multi-dimensional fusion not only improves the data density in sparse geospatial segments but also forms the basis for more accurate and context-aware congestion prediction.

3.5 Outlier analysis and adjustment

The identified outliers are fixed through OSM maximum speed limit for the respective route. The vehicle speed where exceeding the maximum limit of speed on a segment of road is replaced with the standard maximum speed limit value respectively. This step is critical for minimizing the influence of noise and ensuring that the fused heterogeneous and selected homogenous features data

accurately represent typical urban mobility patterns observed in the smart city network in real time environment.

3.5.1 Zero speed correction using violation feature

This step addresses records in the selected dataset where vehicles consistently report zero speed, which may not always indicate congestion but could result from long stops, sensor errors, or stationary conditions. To improve the dataset, the violations feature from FCD is selected that specifically identifying cases where zero speed persists beyond a realistic time threshold. These flagged records are corrected by replacing the zero values with the minimum observed non-zero speed. This approach enhances the accuracy of congestion labeling while preserving the natural flow patterns in the smart city traffic data. The speed value s_i at time step i is corrected using the following Equation 7:

$$s'_i = \begin{cases} \min(S > 0) & \text{if } s_i = 0 \\ s_i & \text{otherwise} \end{cases} \quad (7)$$

Where:

- s_i : Original speed value at time i .
- s'_i : Corrected speed value.
- $\min(S > 0)$: The minimum non-zero speed in the dataset.

3.5.2 Identify peak hours and weekend

This step involves deriving temporal features that capture recurring patterns in traffic behavior. Based on historical analysis of the Smart City data, specific time intervals such as morning, afternoon and evening commute hours are flagged as peak hours, while we use a calendar data source to identify the effect of holidays (weekend) on traffic. The behaviors of commuter and traffic and patterns of traffic are highly dependent on holiday data. The calendar data source features contain Name, DateTime, and Type and using the day-of-week attribute. These derived indicators help differentiate routine traffic flow from irregular patterns, enabling the model to account for variations caused by daily schedules and weekend travel dynamics. Incorporating these temporal features improves the model's ability to predict congestion with greater

TABLE 2 Presents number of vehicles observed in peak vs. non-peak hour and weekday vs. weekend.

Total records	Peak hour records	Non-peak hour records	Weekday records	Weekend records
3149881	1316877	1833004	2266293	883588
-----	(41.81%)	(58.19%)	(71.95%)	(28.05%)

contextual awareness. Further statistical analysis presented in 3.3 Exploratory Analysis Section and presented in Table 2.

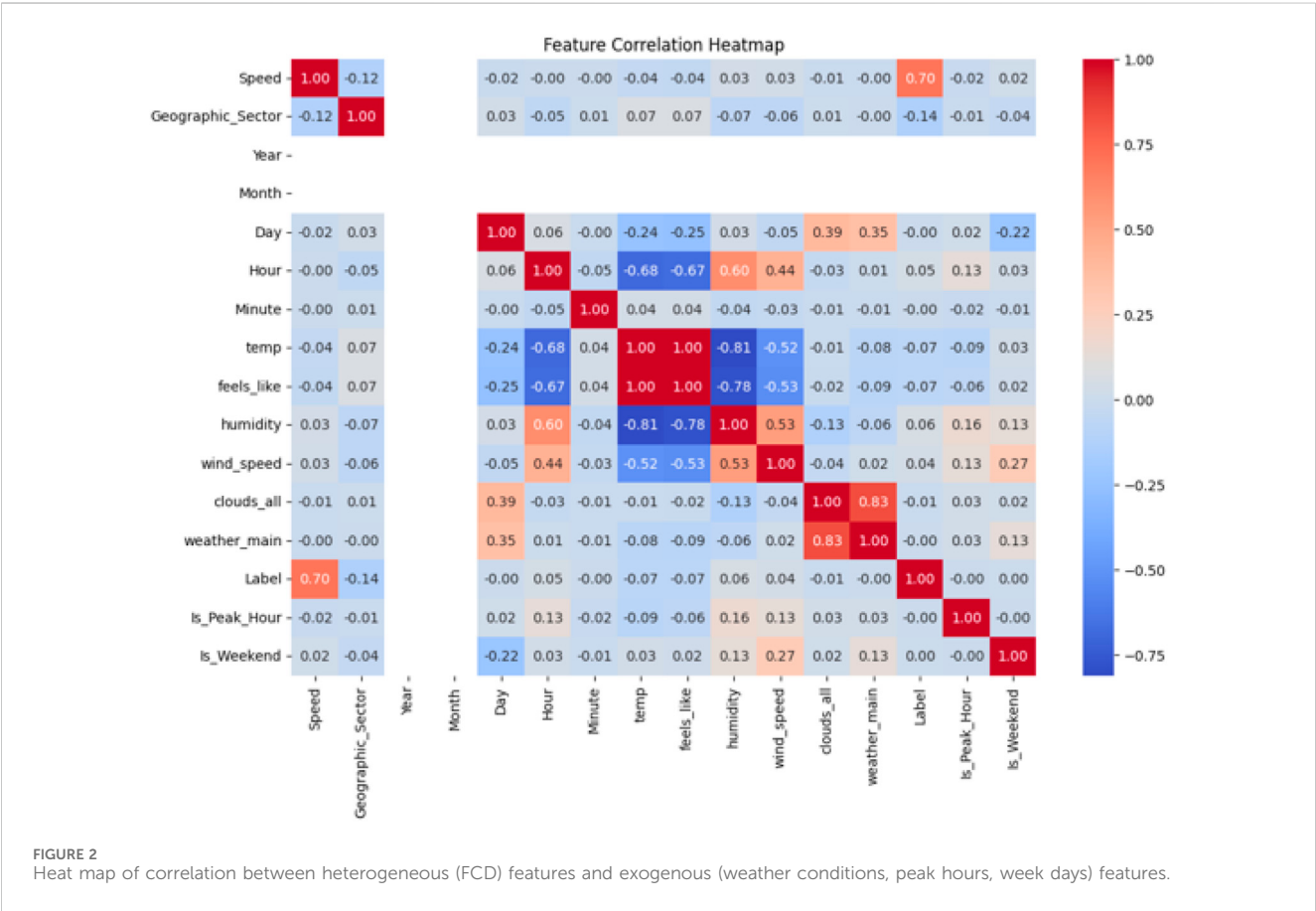
3.6 Exploratory data analysis

This step involves examining how traffic speed varies across different locations, time intervals, and weather conditions. By analyzing the merged dataset from heterogeneous traffic data of Smart City and exogenous data of weather sources, patterns are uncovered showing how weather condition like rainfall, humidity, cloudy, and wind speed influence vehicle movement in specific geographic sectors and time frames. Heatmaps, trend plots, and correlation matrices are used to visualize these relationships, providing critical insights into how congestion behavior shifts under varying spatiotemporal and weather contexts. The analysis shown in Figure 2 forms a foundation for building weather-aware traffic prediction models.

The analysis of the Average Speed Performance Index (SPI) across weekdays and weekends reveals clear distinctions in traffic behavior within smart cities urban road network. During weekdays (Monday to Friday), SPI values remain relatively high (above 0.68) during the early morning hours (midnight to 4 a.m.), indicating free-flowing traffic shown in Figure 3. However, a sharp decline occurs between 5 a.m. and 9 a.m., marking the onset of peak congestion driven by office commutes and school activity. The lowest SPI values are consistently observed between 7 a.m. and 9 a.m., with Friday exhibiting a slightly more pronounced drop likely due to early closures and pre-prayer movement. A mild recovery follows during midday (10 a.m.–2 p.m.), though SPI remains below off-peak levels, with a second dip observed between 3 p.m. and 6 p.m. representing evening rush hour. After 7 p.m., the SPI gradually rises, stabilizing (above 0.64) post-9 p.m. as traffic dissipates shown in Figure 3.

In contrast, weekend traffic patterns (Saturday and Sunday) display a smoother flow with less pronounced rush-hour fluctuations shown in Figure 3. SPI values stay elevated through the early morning (midnight to 5 a.m.) and gradually decline from 6 a.m., reaching their lowest between 9 a.m. and 5 p.m. Sunday shows slightly reduced performance during afternoon hours, reflecting increased recreational or commercial activity. Unlike weekdays, the absence of sharp troughs indicates more dispersed travel behavior. After 6 p.m., SPI steadily recovers toward the late evening on both days.

Together, these patterns highlight distinct temporal traffic characteristics between weekdays and weekends shown in



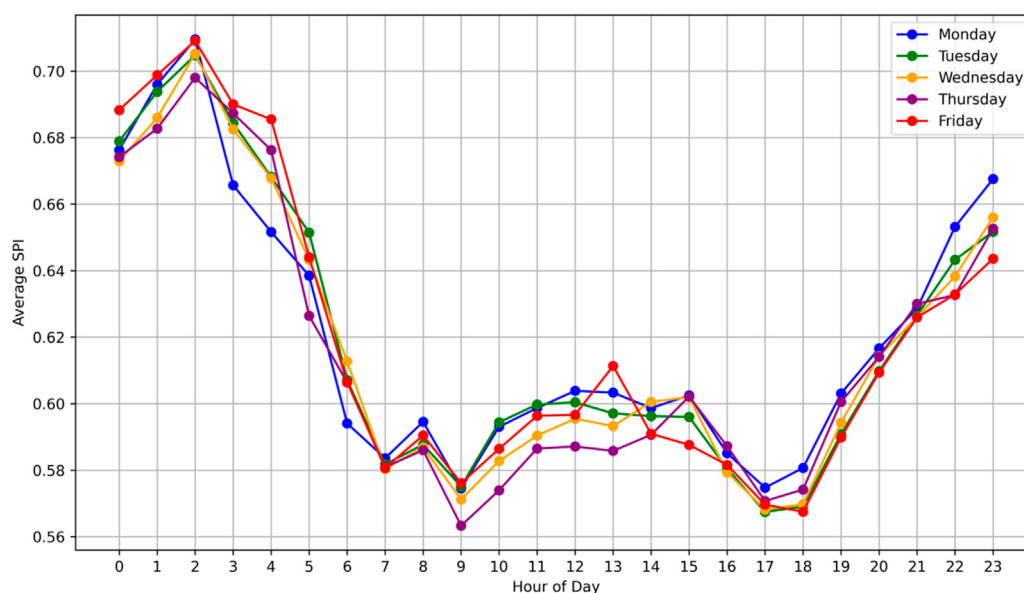


FIGURE 3
Speed performance index variations on weekdays.

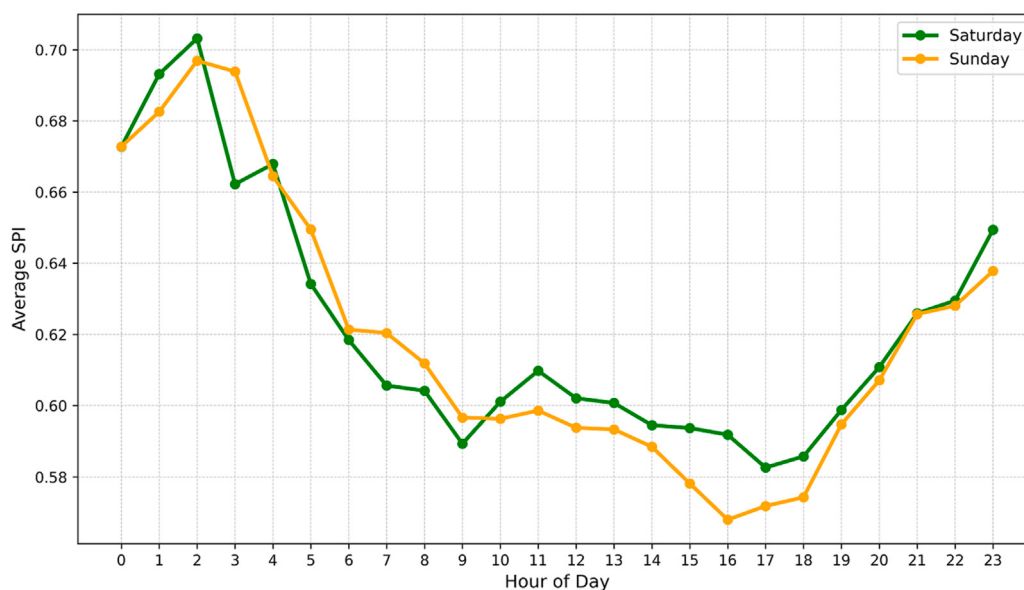


FIGURE 4
Speed performance index variation on weekends.

Figure 4. They highlight the importance of incorporating time-of-day and day-of-week variations in predictive traffic models. The insights derived from Smart City real time sensor data support the development of intelligent, context-aware urban traffic management systems. Figure 5 shown total number vehicles per hour.

The comparison of vehicle record volumes across defined time categories tells a diverse distribution pattern between peak and non-peak hours are presented in Table 2. As per the implemented time segmentation logic, seven out of 24 h (approximately 29%) are

considered peak hours, specifically 8–9 a.m., 2–3 p.m., 6–8 p.m. The remaining 17 h (about 71%) are classified as non-peak hours. Despite this uneven distribution of hours, the recorded traffic data show that 1833004 vehicle records occurred during non-peak hours, while 1316877 vehicles record were observed in the peak-hour window. This outcome highlights that although peak hours span fewer clock hours, they still account for a substantial portion of the total traffic volume approximately 41.81% of all vehicle records for December 2023. This indicates significant

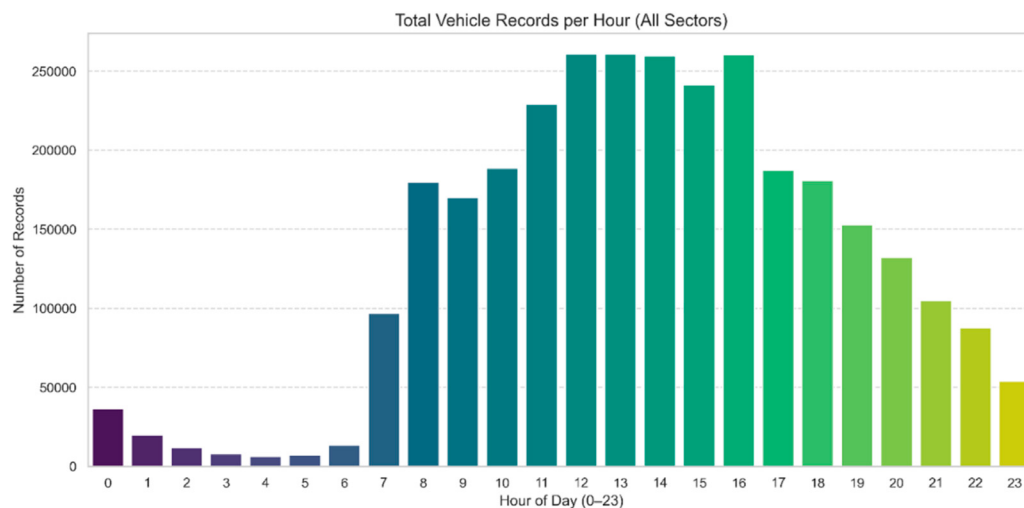


FIGURE 5
Vehicles record per hour.

congestion pressure during concentrated periods of urban activity, such as morning and evening commutes or mid-day institutional movement. On the other hand, the larger share of vehicle records during non-peak hours reflects steady urban mobility outside conventional congestion windows. These findings support the relevance of temporal features in traffic modeling and underscore the importance of capturing both regular and irregular traffic behavior in predictive systems.

3.6.1 Feature importance analysis

To enhance the interpretability of the proposed F-GGRU framework, a feature importance analysis was performed. We applied statistical features selection techniques over the integrated dataset obtained from FCD SCIP and weather sources. The results shown that traffic speed is the most influential predictor of congestion, confirming its strong and direct association with traffic flow conditions. Among exogenous variables, humidity, cloud cover, and wind speed emerged as significant contributors, underscoring the impact of weather factors on congestion dynamics. Temporal indicators such as peak hour and weekend also demonstrated strong influence, reflecting the critical role of time-of-day and day-of-week variations in shaping traffic behavior. In contrast, features such as feels_like showed limited predictive value and was excluded during correlation-based selection. Overall, this analysis provides transparency into how the model utilizes heterogeneous and exogenous inputs, reinforcing that the F-GGRU framework not only delivers higher predictive accuracy but also offers interpretable insights into congestion patterns under boundary conditions.

3.7 Fuzzy logic-automatic labeling for classification

Fuzzy logic automatic labeling is used for classification. The mathematical formulation for our traffic congestion classification

problem, particularly a binary classification task (Congested vs. Smooth) based on traffic data features like Speed, Hour, Geographic_Sector, and weather condition.

We define a binary classification problem to predict:

$y \in (0, 1)$

Where:

- $y = 1$: Congested
- $y = 0$: Smooth

Input Features:

Let the feature vector be: $x = [x_1, x_2, \dots, x_n] \in \mathbb{R}_n$

We have selected total twenty features, here $\mathbb{R}_n$ is set of features. Our selected features are shown in Table 2. We use a logistic regression-based model as the core classifier for Prediction Function by Equation 8:

$$\hat{y} = f(x; \theta) = \sigma(w^T x + b) \quad (8)$$

Where.

- $\hat{y}$: Predicted probability congestion
- $(z) = \frac{1}{1+e^{-z}}$: Sigmoid activation function
- $w \in \mathbb{R}^n$: Weight vector
- $b \in \mathbb{R}$: Bias term

Classification Rule: We may adjust the threshold $t \in (0, 1)$ depending on class imbalance.

$$\hat{y}_{\text{label}} = \begin{cases} 1 & \text{if } \hat{y} \geq 0.5 \quad (\text{Congested}) \\ 0 & \text{if } \hat{y} < 0.5 \quad (\text{Smooth}) \end{cases} \quad (9)$$

We use binary cross-entropy (log loss) as a Loss Function:

$$\mathcal{L}(\hat{y}, y) = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] \quad (10)$$

The optimization objective over m training samples is:

$$\mathcal{I}(\theta) = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(\hat{y}^{(i)}, y^{(i)}) \quad (11)$$

Evaluation Metrics are:

$$\text{Accuracy} = \frac{TC + TS}{TC + TS + FC + FS} \quad (12)$$

$$\text{Precision} = \frac{TC}{TC + FS} \quad (13)$$

$$\text{Recall} = \frac{TC}{TC + FS} \quad (14)$$

$$F1 - \text{Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

- *TC* = True Congested
- *FC* = False Congested
- *TS* = True Smooth
- *FS* = False Smooth

Evaluation metrics of the classification model presents mathematically by Equations 12–15 presents. In the proposed scenario, accuracy measures the proportion of correctly predicted instances among the total number of samples. It provides a general effectiveness measure by computing how often the model's predictions match the actual class labels. Whereas informative, it may be less reliable when class imbalance is present.

In our scenario Precision evaluates the ratio of true positive predictions to all in-stances predicted as positive. This metric emphasizes the model accuracy by identifying how many of the predicted congestion alerts (positives) are truly congested. It is crucial in applications wherever false positives carry a significant cost. Recall calculates the proportion of actual positive cases that the model correctly identifies. It measures the model's ability to detect all relevant traffic congestion events. High recall ensures that the system minimizes the chances of overlooking actual congested conditions, which is critical for real-time urban traffic management. The F1-score is the harmonic mean of precision and recall. It delivers a balanced evaluation metric that accounts for both false negatives and false positives. This is especially useful when seeking a trade-off between capturing all congestion events and ensuring the accuracy of predictions.

3.8 Class balancing using generative adversarial network (GAN)

In real-world traffic datasets such as those derived from Smart City, class imbalance is a common issue particularly when congestion events are underrepresented compared to normal traffic conditions. To address this skewed distribution, GAN are employed to balance the imbalance class without disturbing the sequence of data. This approach mitigates bias in the training process, ensuring that the learning algorithm does not disproportionately favor the majority class and keeps predictive fairness across all classes.

A standard GAN architecture comprises two neural networks: a Generator $G(z; \theta_g)$ and a Discriminator $D(x; \theta_d)$, where random noise vector sampled represented by z , from a prior distribution

$p_z(z)$, and x is a real data sample drawn from the true distribution $p_{data}(x)$. The generator learns to produce synthetic data $G(z)$ that mimics the true distribution, whereas the discriminator attempts to distinguish between real and generated samples.

The minimax optimization objective is formally defined as:

$$\min_G \max_D L_{GAN}(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))] \quad (16)$$

Through iterative adversarial training, the generator gradually improves its ability to produce high-fidelity minority class data points that the discriminator cannot distinguish from real samples. Once the GAN reaches equilibrium, synthetic minority data is combined with the original dataset to form a balanced training set. This enhances model generalizability, reduces classification bias, and improves performance on previously underrepresented classes during evaluation.

3.9 Impact of hybrid F-GGRU framework on society

The hybrid F-GGRU framework acts a valuable role in administrative urban development by highlighting places that frequently involvement traffic congestion. This awareness enables city planners to avoid further constructions of public commercial centers, educational institutions and patrol pumps in congestion proven locations in city infrastructure they are already overburdened areas. The hybrid F-GGRU framework not only improves daily mobility for citizens but also supports long-term urban sustainability and improves the livability of cities.

4 Results and discussion

This study evaluates the effectiveness of Gated Recurrent Unit (GRU) models across four experimental configurations designed to enhance traffic congestion prediction accuracy. The experiments were systematically structured as follows: (i) using raw FCD alone, (ii) GAN balanced classification FCD (iii) Combine FCD with weather conditions (FCD–weather) features from weather APIs, and (iv) Applying GANs for class balancing over the merged FCD–weather dataset. These configurations reflect a progressive enhancement in data richness and preprocessing sophistication, enabling the assessment of each augmentation's contribution to the model's predictive performance.

The dataset used in this analysis comprises minute-level traffic records for the month of December 2023, collected from the Smart City real time surveillance system, covering seven major urban sectors. Traffic behavior was analyzed at fine temporal granularity, where each segment was annotated using fuzzy logic based labeling strategies derived from vehicle speed and time-based thresholds. Our analysis described in below four scenarios.

4.1 Comparative analysis of F-GGRU framework and benchmark models

In our study, a range of machine learning, classical deep learning, and advanced deep learning models were implemented

TABLE 3 Benchmark models accuracy results on heterogeneous FCD features.

Model setup	Accuracy (%)	Precision	Recall	F1-score
Logistic Regression	74	0.56	0.74	0.6425
Decision Tree	65	0.67	0.65	0.66349
Random Forest	70	0.67	0.70	0.6872
Gradient Boosting	75	0.71	0.75	0.66187
XGBoost	75	0.70	0.75	0.670464
AdaBoost	75	0.71	0.75	0.644089
KNN	71	0.67	0.713	0.6877
Artificial Neural Network	75	0.70	0.752	0.6625
GRU	76	0.81	0.77	0.79

on below four types of features i. FCD features dataset, ii. GAN based FCD features dataset, iii. Integrated heterogeneous (FCD) and exogenous features dataset, iv. GAN based balanced Integrated heterogeneous (FCD) and exogenous features dataset. The Benchmark Models include widely used approaches such as AdaBoost (Wang et al., 2016), XGBoost (Yu and Xie, 2024), Decision Trees (Lartey et al., 2021) as base learners alongside Linear Regression (Hazarika et al., 2024), and K-Nearest Neighbors KNN (API, 2025). Additionally, Artificial Neural Networks (ANNs) and CNN were applied. Deep learning model employed in the analysis include Gated Recurrent Unit (GRU). Furthermore, hybrid architectures were explored through F-GGRU framework. Given the spatiotemporal characteristics of the dataset, F-GGRU framework demonstrated relatively stronger performance. Among these, the F-GGRU framework yielded the most favorable outcomes. The subsequent sections provide a concise overview of the F-GGRU framework, followed by an explanation of the hybrid approach integrating this model.

4.1.1 Models training and evaluation

The entire dataset is split into 70% training and 30% testing sets represented in algorithm 5. During the training phase, the model learns to minimize the prediction error between actual and predicted speeds. Once trained, the model is evaluated using standard performance metrics, such as accuracy, precision, recall, and F1-score, especially focusing on the congested class to validate its reliability under real-world conditions. The results are then presented in terms of the predicted traffic state, facilitating decision-making for traffic control authorities.

4.1.2 Scenario 1: benchmark models and GRU analysis on unbalanced floating car data features dataset

In the initial evaluation using the selected FCD features presented in Table 1 without class balancing, performance variation was observed across classical ML and DL models and the models results shown in Table 3. Logistic Regression, although simple, showed relatively moderate accuracy (74%) and a notably low precision (0.56), indicating high false-positive rates. Decision Tree and Random Forest showed balanced precision and recall, but

accuracy remained in the 65%–70% range. Notably, ensemble models like Gradient Boosting, XGBoost, and AdaBoost reached slightly better accuracy (~75%), though their F1-scores hovered around 0.66. Among all models, GRU with linear activation stood out with an F1-score of 0.79 and the highest precision (0.81), highlighting its superior capability in capturing sequential dependencies in temporal traffic data even without class balancing.

4.1.3 Scenario 2: benchmark models and F-GGRU framework analysis on generative adversarial networks based balanced floating car data features dataset

Upon applying GANs to balance class distributions of FCD features presented in Table 1, a significant performance boost was evident across nearly all models and results shown in Table 4. Logistic Regression, which previously had weak precision, achieved a perfect recall (1.0) and improved F1-score (0.85). Similarly, ensemble models like Gradient Boosting, AdaBoost, and XGBoost consistently achieved an accuracy of 83% with strong F1-scores around 0.85, indicating balanced predictive strength. The F-GGRU framework further excelled in this scenario, reaching an accuracy of 84% and maintaining high precision (0.78) and recall (0.98), underscoring the effectiveness of combining generative resampling with temporal deep learning in imbalanced classification tasks.

4.1.4 Scenario 3: benchmark models and F-GRU framework analysis on unbalanced integrated floating car data and exogenous features dataset

Integrating external exogenous features presented in Table 1 with the FCD features presented in Table 1 provided additional context for congestion prediction, and models responded well to the enriched feature space. Integrating exogenous features presented in Table 1 with FCD features dataset provided additional context for congestion prediction, and models responded well to the enriched feature space. While benchmark models Logistic Regression and Decision Trees showed modest improvements, reaching up to 74% accuracy and F1-scores of 0.85 and 0.76 respectively, ensemble techniques retained their strength, maintaining accuracy levels

TABLE 4 Benchmark models and F-GGRU framework accuracy on GAN applied on FCD features.

Model setup	Accuracy (%)	Precision	Recall	F1-score
Logistic Regression	83	0.74	1.0	0.85
Decision Tree	77	0.78	0.75	0.76
Random Forest	80	0.77	0.85	0.81
Gradient Boosting	83	0.75	0.99	0.85
XGBoost	83	0.75	0.98	0.85
AdaBoost	83	0.75	0.99	0.85
KNN	80	0.77	0.87	0.82
ANN	83	0.75	0.98	0.85
F-GGRU Framework	84	0.78	0.98	0.85

TABLE 5 Benchmark models and F-GRU accuracy results on unbalanced integrated features space dataset.

Model setup	Accuracy (%)	Precision	Recall	F1-score
Logistic Regression	74	0.74	1	85
Decision Tree	65	0.78	0.75	76
Random Forest	69	0.77	0.83	0.80
Gradient Boosting	75	0.75	0.98	0.85
XGBoost	75	0.75	0.98	0.85
AdaBoost	75	0.75	0.99	0.85
KNN	71	0.77	0.87	0.82
ANN	75	0.75	0.98	0.85
F-GRU	83	0.89	0.87	0.88

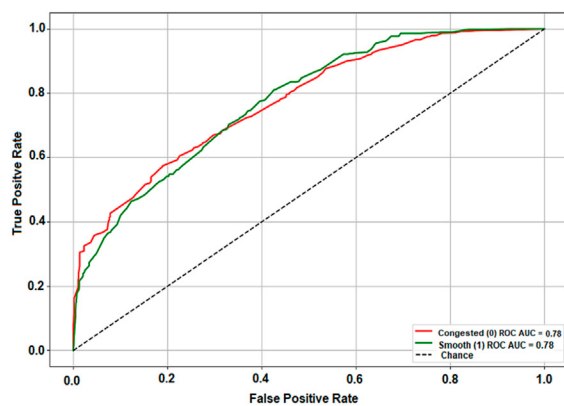
TABLE 6 Accuracy table of benchmark models and F-GGRU framework on GANs applied integrated data.

Model setup	Accuracy (%)	Precision	Recall	F1-score
Decision Tree	77	0.77	0.77	0.77
Random Forest	79	0.80	0.80	0.80
KNN	81	0.82	0.81	0.81
Logistic Regression	83	0.87	0.83	0.83
Gradient Boosting	83	0.87	0.84	0.83
XGBoost	83	0.87	0.84	0.83
AdaBoost	83	0.87	0.83	0.83
ANN	83	0.86	0.83	0.83
F-GGRU Framework	0.98	0.99	0.98	0.98

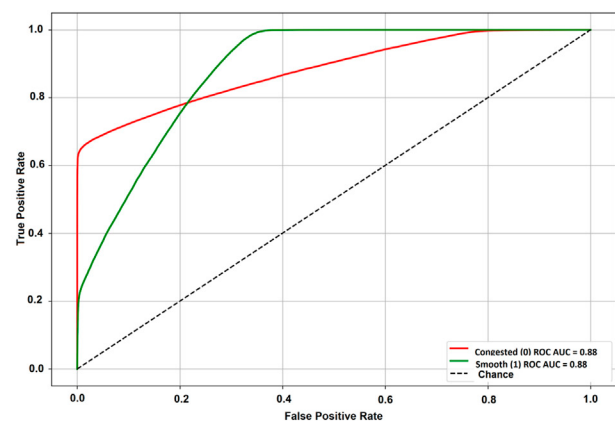
around 75% with improved recall and F1-scores (~ 0.85). Notably, the fuzzy logic based labeling GRU (F-GRU) framework achieved 83% accuracy and a superior F1-score of 0.88, suggesting that weather context synergizes well with temporal dependencies in traffic flow modeling. Comparative results of models presented in Table 5.

4.1.5 Scenario 4: benchmark models and F-GGRU framework analysis on generative adversarial networks based balanced integrated features space

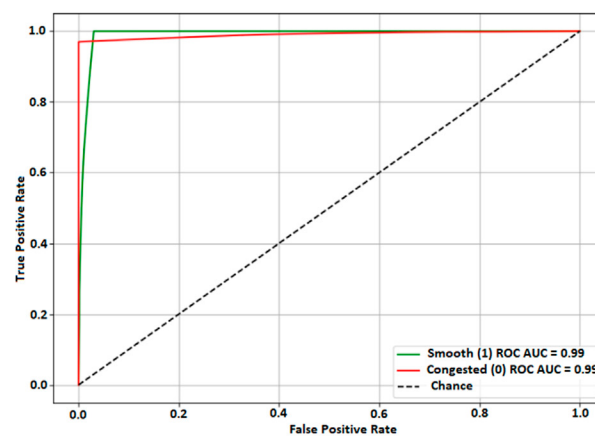
This scenario produced the most pronounced results presented in Table 6, where both class balancing (via GAN) and multi-source



(a). ROC Curve for both classes Congested and Smooth using AdaBoost Model



(b). ROC Curve for both classes Congested and Smooth using ANN Model



(c). ROC Curve for both classes Congested and Smooth using F-GGRU Framework

FIGURE 6

(A) ROC curve for both classes congested and smooth using AdaBoost model (B) ROC Curve for both classes congested and smooth using ANN model (C). ROC curve for both classes congested and smooth using F-GGRU framework.

data fusion (FCD + weather) were employed. All models benefited from this approach, with ensemble methods like Gradient Boosting, Ada-Boost, and XG-Boost reaching consistent scores accuracy at 83%, precision at 0.87, and F1-scores around 0.83. KNN also showed strong performance (81% accuracy and 0.81 F1-score). The F-GGRU framework, however, dominated this setting with a remarkable 98% accuracy and F1-score of 0.98, authenticating its robustness in sequence modeling and its adaptability to fused and balanced datasets.

The inclusion of weather variables such as humidity, wind speed, temperature, and overall weather conditions demonstrated a measurable impact on prediction accuracy, reflecting the influence of environmental dynamics on congestion trends. Furthermore, the application of GANs to balance class representation effectively mitigated the skew typically observed in real-world traffic datasets, especially between congested and smooth instances. The results highlight that the final F-GGRU framework, trained on the GAN-augmented and exogenous features (weather conditions, peak hours, weekend)-enhanced dataset, outperformed other configurations in terms of classification metrics including F1-score, precision, and Recall. This confirms the benefit of both data balancing and

multimodal data integration in smart traffic prediction systems. This benchmarking not only establishes the superiority of the proposed F-GGRU framework but also fulfills the requirement of comparing its performance against state-of-the-art deep learning ANN and in Table 6 mentioned machine learning algorithms.

To evaluate the classification performance of each model, the Receiver Operating Characteristic (ROC) curves for both traffic congestion classes—Congested (0) and Smooth (1)—were plotted for AdaBoost shown in Figure 6a, ANN shown in Figure 6b, and the proposed hybrid F-GGRU framework shown in Figure 6c.

The ROC curve for the AdaBoost model indicates moderate discriminative power, with AUC values of 0.78 for both classes, suggesting limited sensitivity in differentiating between congested and smooth traffic under the current feature representation. The performance improved considerably with the ANN model, where both classes achieved an AUC of 0.88, reflecting enhanced predictive capacity and more reliable generalization over unseen data.

However, the most significant improvement was observed in the F-GGRU framework with 98% accuracy, which delivered near-perfect separation with an AUC of 0.99 for both Congested and

TABLE 7 Hyperparameter tuning results for the F-GGRU framework on SCIP dataset

Trial	Learning rate	Batch size	Hidden units (Layers)	Dropout	Activation	Accuracy	Precision	Recall	F1-score	AUC
1	0.001	64	(64, 32, 16)	0.2	ReLU	0.9847	0.9702	1.0000	0.9849	0.9912
2	0.001	128	(128, 64, 32)	0.3	Tanh	0.9848	0.9704	1.0000	0.9850	0.9914
3	0.001	256	(128, 64, 32)	0.4	ReLU	0.9846	0.9701	1.0000	0.9848	0.9911
4	0.0005	64	(64, 32, 16)	0.2	Tanh	0.9848	0.9703	1.0000	0.9849	0.9913
5	0.0005	128	(128, 64, 32)	0.3	Tanh	0.9848	0.9704	1.0000	0.9850	0.9914
6	0.0005	256	(256, 128, 64)	0.3	ReLU	0.9847	0.9703	1.0000	0.9849	0.9912
7	0.0001	64	(64, 32, 16)	0.2	ReLU	0.9847	0.9702	1.0000	0.9849	0.9912
8	0.0001	128	(128, 64, 32)	0.3	Tanh	0.9848	0.9704	1.0000	0.9850	0.9913
9	0.0001	256	(128, 64, 32)	0.4	ReLU	0.9846	0.9701	1.0000	0.9848	0.9911
10	0.001	64	(128, 64, 32)	0.3	Tanh	0.9847	0.9702	1.0000	0.9849	0.9912
11	0.001	128	(256, 128, 64)	0.3	ReLU	0.9847	0.9703	1.0000	0.9849	0.9912
12	0.0005	128	(128, 64, 32)	0.3	Tanh	0.9848*	0.9704*	1.0000*	0.9850*	0.9914*
13	0.0005	64	(256, 128, 64)	0.2	ReLU	0.9847	0.9702	1.0000	0.9849	0.9912
14	0.0005	256	(128, 64, 32)	0.4	Tanh	0.9846	0.9701	1.0000	0.9848	0.9911
15	0.0001	128	(128, 64, 32)	0.3	ReLU	0.9847	0.9703	1.0000	0.9849	0.9912
16		128	(128, 64, 32)	0.3	Tanh	0.9848	0.9704	1.0000	0.9850	0.9913

Smooth classes. The ROC curves of F-GGRU framework closely follow the top-left corner, indicating a very high true positive rate with a minimal false positive rate. This level of precision can be attributed to the model's ability to incorporate temporal patterns and weather-related variations more effectively than traditional ensemble or feed forward architectures. In summary, the progression from AdaBoost to F-GGRU framework highlights the benefits of deep temporal modeling and fusion strategies in enhancing traffic congestion prediction accuracy.

Across all scenarios, the proposed F-GGRU framework with linear activation function consistently outperformed from traditional ML models, Multilevel-Gated Recurrent Unit (MGRU) model (Sravani et al., 2024) and Conv-Bi-LSTM and GRU-LSTM (Zafar et al., 2022b), especially when supported by balanced data and contextual weather features. GAN-based augmentation and data fusion significantly enhanced classification robustness, reducing bias and improving generalizability. These findings validate the effectiveness of integrating generative resampling with temporal deep learning frameworks for smart city traffic analytics.

4.2 Hyperparameter tuning and optimal F-GGRU framework configuration

4.2.1 Hyperparameter tuning of F-GGRU on SCIP dataset

A thorough hyperparameter tuning process was carried out to optimize the proposed F-GGRU framework. Sixteen experimental trials were conducted by systematically adjusting the learning rate,

batch size, model depth, hidden units, dropout rate, and activation functions. The results Table 7 demonstrated that the model achieved consistently strong performance across all configurations, with validation accuracy of approximately 98.5%, precision of 97.0%, recall of 100%, F1-score of 98.5%, and AUC values exceeding 0.991. Among the tested configurations, the combination of a learning rate of 0.0005, batch size of 128, hidden units (128, 64, 32), dropout rate of 0.3, and tanh activation function achieved the best trade-off between accuracy and generalization, and was therefore selected as the final model.

4.2.2 Hyperparameter tuning of F-GGRU on CityPulse dataset

The hyperparameter tuning of the F-GGRU framework analysis presented in Table 8 on the CityPulse weather conditions data integrated with FCD dataset (Aarhus, 2025), conducted across 16 trials, successfully identified an optimal configuration (Trial 5: LR = 0.001, Batch = 256, Units = (64,64,32), Dropout = 0.2, Activation = Tanh) that achieved exceptional performance (Accuracy = 0.9942, AUC = 0.9976) while consistently maintaining perfect recall (1.000) across all trials. The results demonstrated a pronounced sensitivity to batch size, with larger batches (256) significantly outperforming smaller ones (128), and revealed the superiority of a wider, symmetric network architecture paired with moderate regularization. This rigorous tuning process not only optimized the model for the CityPulse dataset but also, when contrasted with the robust performance on the SCIP dataset, provided strong evidence for the generalizability and robustness of the F-GGRU framework across diverse real-world sensing data environments.

TABLE 8 Hyperparameter tuning results for the F-GGRU framework on CityPulse dataset.

Trial	Learning rate	Batch size	Hidden units (Layers)	Dropout	Activation	Val_Acc	Precision	Recall	F1-score	AUC
1	0.001	128	(64, 64, 32)	0.2	tanh	0.9845	0.9701	1.0000	0.9850	0.9910
2	0.001	128	(64, 64, 32)	0.3	tanh	0.9852	0.9710	1.0000	0.9854	0.9915
3	0.001	128	(128, 64, 32)	0.2	tanh	0.9868	0.9725	1.0000	0.9861	0.9920
4	0.001	128	(128, 64, 32)	0.3	tanh	0.9860	0.9718	1.0000	0.9859	0.9917
5	0.001	256	(64, 64, 32)	0.2	tanh	0.9942	0.9870	1.0000	0.9942	0.9976
6	0.001	256	(64, 64, 32)	0.3	tanh	0.9935	0.9865	1.0000	0.9932	0.9973
7	0.001	256	(128, 64, 32)	0.2	tanh	0.9928	0.9855	1.0000	0.9927	0.9970
8	0.001	256	(128, 64, 32)	0.3	tanh	0.9918	0.9842	1.0000	0.9920	0.9965
9	0.0005	128	(64, 64, 32)	0.2	tanh	0.9850	0.9705	1.0000	0.9850	0.9912
10	0.0005	128	(64, 64, 32)	0.3	tanh	0.9848	0.9704	1.0000	0.9850	0.9914
11	0.0005	128	(128, 64, 32)	0.2	tanh	0.9856	0.9715	1.0000	0.9856	0.9918
12	0.0005	128	(128, 64, 32)	0.3	tanh	0.9848	0.9704	1.0000	0.9850	0.9914
13	0.0005	256	(64, 64, 32)	0.2	tanh	0.9920	0.9850	1.0000	0.9925	0.9960
14	0.0005	256	(64, 64, 32)	0.3	tanh	0.9915	0.9846	1.0000	0.9920	0.9962
15	0.0005	256	(128, 64, 32)	0.2	tanh	0.9922	0.9852	1.0000	0.9926	0.9967
16	0.0005	256	(128, 64, 32)	0.3	tanh	0.9918	0.9845	1.0000	0.9920	0.9963

4.3 F-GGRU framework validation on additional real-world dataset and robustness analysis

To further establish the robustness and generalizability of the proposed F-GGRU framework, validation was conducted on an independent real-world dataset, the CityPulse dataset (Aarhus, 2025). This dataset integrates urban Floating Car Data (FCD) with weather conditions collected from Open Data Aarhus in Denmark, providing diverse traffic dynamics and environmental conditions that differ from the SCIP dataset. The evaluation confirmed that the framework consistently sustained high predictive performance, achieving validation accuracies above 98.4%, precision around 97%, perfect recall, and AUC values exceeding 0.991 across both datasets. Notably, the SCIP dataset favored a deeper hidden-layer configuration with moderate dropout, while the CityPulse dataset achieved optimal results with a more symmetric architecture and larger batch size. Despite these structural variations, the framework demonstrated stable generalization across heterogeneous data sources. These results validate the adaptability of the F-GGRU model and reinforce its potential for deployment in practical intelligent traffic management systems.

5 Conclusion

This research study developed the F-GGRU framework, a comprehensive methodology for fusing heterogeneous traffic data FCD with exogenous factors like weather conditions and

temporal patterns into a unified hybrid feature space. A key innovation lies in its integrated pipeline, which combines fuzzy logic-based automatic labeling, GAN-driven class balancing, and advanced temporal modeling through a gated architecture. Validated against a suite of classical machine learning (Random Forest, XGBoost, etc.) and deep learning (GRU, ANN) benchmarks which achieved accuracies of only 77%–83% with poor performance on the critical “congested” class the proposed F-GGRU framework demonstrated superior performance, achieving over 98% accuracy, near-perfect recall (1.00), and a high ROC-AUC value (0.99). Significantly, rigorous hyperparameter tuning and validation on a second, independent real-world dataset (CityPulse) confirmed the framework’s robustness and generalizability, where it achieved even higher performance (99.42% accuracy, 0.9976 AUC). This proves the model is not a singular solution but an adaptable and robust tool for diverse urban environments. These findings underscore that the framework achievement is rooted in its synergistic combination of intelligent preprocessing (addressing class imbalance with GANs), feature fusion (incorporating impactful weather data), and sequential deep learning. The F-GGRU is therefore not only a highly accurate predictive tool but also a scalable and practical solution for real-time traffic monitoring systems. Its ability to reliably differentiate between traffic conditions enables proactive measures such as dynamic route guidance and adaptive traffic signal control. By providing traffic management authorities with a state-of-the-art, generalizable solution, this work contributes directly to reducing travel delays, optimizing road infrastructure, and enhancing decision-making for smarter, more responsive urban mobility.

Currently we integrated data sources of SCIP FCD from heterogeneous data sources and Weather, peak hours, week days, weekend from exogenous data sources in future we will integrate multiple data sources for further work. We have novel data set of SCIP, having various radar based cameras sensors that generate digital data. Incorporate events data such as accidents, road closures, public protests, planned events or major events from APIs like Google Traffic or Twitter, etc., to improve model context-awareness. Integrate the trained model into an edge computing environment or smart traffic control platform for real-time congestion prediction and response.

Another promising direction is the application of transfer learning to the smart city traffic domain. Pre-trained models trained on large-scale traffic or mobility datasets can be fine-tuned with city-specific data, enabling the framework to quickly adapt to new environments with limited labeled data. This strategy could enhance generalization, reduce training costs, and accelerate deployment in different urban contexts.

Data availability statement

Data available on request from the authors.

Author contributions

AA: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation, Writing – original draft, Writing – review and editing. AN: Funding acquisition, Validation, Writing – review and editing. NZ: Conceptualization, Data curation, Investigation, Software, Visualization, Writing – review and editing. MS: Project administration, Conceptualization, Investigation, Methodology, Supervision, Validation, Visualization, Writing – review and editing.

References

- Aarhus, D. O. D. A. (2025). Aarhus, Denmark (open data Aarhus) CityPulse Dataset Collection. Dataset Collection. Available online at: <http://iot.ee.surrey.ac.uk:8080/datasets.html>
- Agarwal, M., Maze, T. H., and Souleyrette, R. (2005). "Impacts of weather on urban freeway traffic flow characteristics and facility capacity," in *Proceedings of the 2005 mid-continent transportation research symposium*.
- Al-Qarafi, A., Alrowais, F., S. Alotaibi, S., Nemri, N., Al-Wesabi, F. N., Al Duhayyim, M., et al. (2022). Optimal machine learning based privacy preserving blockchain assisted internet of things with smart cities environment. *Appl. Sci.* 12 (12), 5893. doi:10.3390/app12125893
- Ali, A., Ayub, N., Shiraz, M., Ullah, N., Gani, A., and Qureshi, M. A. (2021a). Traffic efficiency models for urban traffic management using mobile crowd sensing: a survey. *Sustainability* 13 (23), 13068. doi:10.3390/su132313068
- Ali, A., Qureshi, M. A., Shiraz, M., and Shamim, A. (2021b). Mobile crowd sensing based dynamic traffic efficiency framework for urban traffic congestion control. *Sustain. Comput. Inf. Syst.* 32, 100608. doi:10.1016/j.suscom.2021.100608
- Ali, F., Khan, Z. H., Khattak, K. S., and Gulliver, T. A. (2024). The effect of visibility on road traffic during foggy weather conditions. *IET Intell. Transp. Syst.* 18 (1), 47–57. doi:10.1049/itr2.12432
- Api, O. (2025). OpenWeather API 13 September.
- Cui, S. (2024). A cross-city traffic flow prediction framework incorporating holidays and weather factors. *Appl. Comput. Eng.* 111, 159–164. doi:10.54254/2755-2721/111/2024ch0106
- Dong, J., Mahmassani, H. S., and Alfelori, R. (2010). Weather responsive traffic management: deployment of real-time traffic estimation and prediction systems. Transportation Research Board 89th Annual Meeting Transportation Research Board, 10–4094.
- Du, Y., Chen, Y., Li, X., Schönborn, A., and Sun, Z. (2022). Data fusion and machine learning for ship fuel efficiency modeling: part III—Sensor data and meteorological data. *Commun. Transp. Res.* 2, 100072. doi:10.1016/j.commtr.2022.100072
- Hazarika, A., Choudhury, N., Nasralla, M. M., Khattak, S. B. A., and Rehman, I. U. (2024). Edge ML technique for smart traffic management in intelligent transportation systems. *IEEE Access* 12, 25443–25458. doi:10.1109/access.2024.3365930
- Islamabad Scene (2021). Pakistan's population reaches 241 million Available online at: <https://www.islamabadscene.com/pakistans-population-reaches-241-million/>
- Khattak, A. (2025). Islamabad is getting thousands of AI cameras for better security. Available online at: https://propakistani.pk/2025/05/27/islamabad-is-getting-thousands-of-ai-cameras-for-better-security/?utm_source=chatgpt.com
- Lartey, B., Homaifar, A., Girma, A., Karimodini, A., and Opoku, D. (2021). "XGBoost: a tree-based approach for traffic volume prediction," in *2021 IEEE international conference on systems, man, and cybernetics (SMC)* (IEEE).
- Lin, L., Ni, M., He, Q., Gao, J., and Sadek, A. W. (2015). Modeling the impacts of inclement weather on freeway traffic speed: exploratory study with social media data. *Transp. Res. Rec.* 2482 (1), 82–89. doi:10.3141/2482-11
- Mashros, N., Edigbe, J. B., Hassan, S. A., Abdul Hassan, N., and Mohd Yunus, N. Z. (2014). Impact of rainfall condition on traffic flow and speed: a case study in johor and terengganu. *Jurnal Teknologi Sci. and Eng.* 70 (4). doi:10.11113/jt.v70.3490
- Polson, N. G., and Sokolov, V. O. (2017). Deep learning for short-term traffic flow prediction. *Transp. Res. Part C Emerg. Technol.* 79, 1–17. doi:10.1016/j.trc.2017.02.024

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research is funded by the Deanship of Scientific Research, Islamic University of Madinah, Madinah, Saudi Arabia.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Pragalathan, J., and Schramm, D. (2024). Urban traffic flow predictions with impacts of weather and holidays. *Emerg. Cutting-Edge Dev. Intelligent Traffic Transp.* doi:10.3233/ATDE240042
- Rehborn, H., and Koller, M. (2014). A study of the influence of severe environmental conditions on common traffic congestion features. *J. Adv. Transp.* 48 (8), 1107–1120. doi:10.1002/atr.154
- Ren, L., Jia, Z., Laili, Y., and Huang, D. (2023). Deep learning for time-series prediction in IIoT: progress, challenges, and prospects. *IEEE Trans. neural Netw. Learn. Syst.* 35, 15072–15091. doi:10.1109/tnnls.2023.3291371
- Romanowska, A., and Budzyński, M. (2022a). Investigating the impact of weather conditions and time of day on traffic flow characteristics. *Weather, Clim. Soc.* 14 (3), 823–833. doi:10.1175/wcas-d-22-00121
- Romanowska, A., and Budzyński, M. (2022b). Investigating the impact of weather conditions and time of day on traffic flow characteristics. *Weather, Clim. Soc.* 14 (3), 823–833. doi:10.1175/wcas-d-22-00121
- Safe City (2022). S.C.A. Available online at: <https://islamabadpolice.gov.pk/safecity.php>.
- SCIP (2022). Integrated_FCD_Weather_Dataset. Available online at: <https://docs.google.com/spreadsheets/d/1EiQFzKJnXSxxMDnrUk8CRg7mSgwI12L1/edit?usp=sharing&ouid=114614328391821557537&rtopof=true&sd=true2025>.
- Solanki, M., Rawat, V., Pandey, M., Singh, N., and Aswal, S. (2023). “Traffic Prediction Analysis (TPA) using machine learning methodologies,” in *2023 6th international conference on contemporary computing and informatics (IC3I)* (IEEE).
- Stravani, B., Shreyas, A. V., Abbas, H. M., Chanti, Y., and Punitha, S. (2024). “Traffic congestion prediction in smart cities using multilevel-gated recurrent unit,” in *2024 international conference on intelligent algorithms for computational intelligence systems (IACIS)* (IEEE).
- Sun, T., Sun, B., Jiang, Z., Hao, R., and Xie, J. (2021). Traffic flow online prediction based on a generative adversarial network with multi-source data. *Sustainability* 13 (21), 12188. doi:10.3390/su132112188
- United Nation (2023). Department of economic and social affairs United nation Available online at: <https://www.un.org/uk/desa/68-world-population-projected-live-urban-areas-2050-says-un-2025>.
- U.S Department of Energy (2025). Available online at: <https://www.energy.gov/eere/vehicles/articles/fotw-1359-sept-9-2024-traffic-congestion-united-states-wasted-33-billion#:~:text=Breadcrumb,of%20wasted%20fuel%20in%202022>.
- Valarmathi, V., and Dhanalakshmi, S. (2024). Internet of traffic surveillance System (IoTSS) with genetic Algorithm for optimized weather-adaptive traffic monitoring. *Bio-Inspired Intell. Smart Decision-Making*, 1–25. doi:10.4018/979-8-3693-5276-2.ch001
- Vargas, J., Alsweiss, S., Toker, O., Razdan, R., and Santos, J. (2021). An overview of autonomous vehicles sensors and their vulnerability to weather conditions. *Sensors* 21 (16), 5397. doi:10.3390/s21165397
- Wang, J., Gu, Q., Wu, J., Liu, G., and Xiong, Z. (2016). “Traffic speed prediction and congestion source exploration: a deep learning method,” in *2016 IEEE 16th international conference on data mining (ICDM)* (IEEE).
- Wu, Y., Tan, H., Qin, L., Ran, B., and Jiang, Z. (2018). A hybrid deep learning based traffic flow prediction method and its understanding. *Transp. Res. Part C Emerg. Technol.* 90, 166–180. doi:10.1016/j.trc.2018.03.001
- Yasir, R. M., Nower, N., and Shoyaib, M. (2022). Traffic congestion prediction using machine learning techniques. *arXiv Prepr. arXiv:2206.10983*. doi:10.48550/arXiv.2206.10983
- YixinLi, W., and Zhang, W. (2024). Research on traffic congestion prediction based on analyzable machine learning. *Eng. Technol.* 118, 102–110. doi:10.54097/sbj3av77
- Yu, W., and Xie, F. (2024). Research on traffic congestion prediction based on XGBoost. *Front. Traffic Transp. Eng.* 4 (1), 1–8. doi:10.23977/ftte.2024.040101
- Zafar, N., Haq, I. U., Chughtai, J. u. R., and Shafiq, O. (2022a). Applying hybrid LSTM-GRU model based on heterogeneous data sources for traffic speed prediction in urban areas. *Sensors* 22 (9), 3348. doi:10.3390/s22093348
- Zafar, N., Haq, I. U., Sohail, H., Chughtai, J. U. R., and Muneeb, M. (2022b). Traffic prediction in smart cities based on hybrid feature space. *IEEE Access* 10, 134333–134348. doi:10.1109/access.2022.3231448
- Zhao, J., Gao, Y., Bai, Z., Wang, H., and Lu, S. (2019). Traffic speed prediction under non-recurrent congestion: based on LSTM method and BeiDou navigation satellite system data. *IEEE Intell. Transp. Syst. Mag.* 11 (2), 70–81. doi:10.1109/itsmag.2019.2903431
- Zhong, H., Wang, J., Chen, C., Wang, J., and Guo, K. (2024). Weather interaction-aware spatio-temporal attention networks for urban traffic flow prediction. *Buildings* 14 (3), 647. doi:10.3390/buildings14030647