Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 11 Issue 01
2026

Research Article
https://doi.org/10.56979/1101/2026/1436

CAST-FC: Context-Aware Spatio-Temporal Feature Selection for Classification in Heterogeneous Urban Environments

Muhammad Nabeel Asghar[1*]

[1]Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Al-Ahsa, Saudi Arabia.
*Corresponding Author: Muhammad Nabeel Asghar. Email: masghar@kfu.edu.sa

Received: February 11, 2026 **Accepted**: May 26, 2026

Abstract: The growth in complexity of traffic systems in urban settings has rendered crash severity prediction as an increasingly difficult task mainly due to high heterogeneity of data in space and time. Conventional feature selection algorithms often make an assumption that the relevance of certain features holds true in the whole data set without any consideration for possible heterogeneities that exist within different portions of it. To overcome this limitation, this paper proposes a novel Context-Aware Spatio-Temporal Feature Selection (CAST-FS) technique which leverages local contextual information when selecting features from heterogeneous urban traffic data sets. In particular, the data set is partitioned into homogeneous groups according to its geographic and temporal characteristics by applying K-means clustering algorithm, and then a LightGBM classifier is applied locally within each group to estimate local feature importances. The local feature importance scores obtained in the process are integrated using cluster size-based weights to produce a global ranking of features. After that, selected feature subset which consists of the highest-ranking features is used to build a final classifier. The experiment in this study is conducted on the publicly available New York City Motor Vehicle Collisions dataset. It uses a binary classification approach for the task of crash severity prediction. In order to evaluate the effectiveness of the proposed approach, both threshold-independent and threshold-dependent evaluation metrics, such as AUC-ROC, Average Precision, accuracy, precision, recall, and F1-score, are employed. Results show that compared to global LightGBM-based and mutual information approaches CAST-FS demonstrates higher performance in AUC-ROC and Average Precision while shrinking the number of used features from 61 to 12. At the same time, imbalanced data classes lead to poor recall results when applying a fixed threshold.

Keywords: Classification; Motor Vehicle Collision; Feature Selection; LightGBM.

---

1. Introduction

The city has become a highly instrumented ecosystem where sensors and other data collection devices continuously monitor transport operations, public safety incidents, environmental indicators, and mobility flows. These data sources provide rich information about diverse phenomena of urban dynamics with detailed geographic and temporal resolution, and have played a key role in the development of applications like crash severity prediction, demand forecasting, incident detection, and traffic state classification. With the advent of open data policies, many of these datasets have been released to the public domain, and have been extensively used in the evaluation of machine learning techniques within real-world scenarios. Yet analyzing such data is extremely challenging given issues of spatial heterogeneity, temporal variation, and high-dimensional feature spaces.

One core issue concerns the uneven relevance of the features within an urban region, or through time. For instance, road safety analysis often considers various environmental factors, such as lighting, roadway design, traffic volume, or weather conditions. While these features can affect the likelihood and severity of a crash, the effect size may differ across residential and commercial regions, as well as between nighttime and peak traffic periods. It has been demonstrated that traffic incidents show significant spatiotemporal dependencies and the same predictor may have varying effects across space and time [1], [2]. Likewise, pedestrian-related features, urban morphology, and intersection density have all shown varying degrees of impact under localized urban form characteristics [3], [4].

Regardless of these challenges, the vast majority of popular feature selection techniques rely on the premise that there is a common structure to all the samples. While the filter methods (chi-square scoring, mutual information, correlation-based selection) compute relevance scores by aggregating statistics globally over the entire dataset [5], wrapper and embedded approaches base their evaluation of feature importance on model performance, which usually implies the assumption that the relevancy does not change between the samples [6]–[8]. This technique can prove useful in most cases, though, it fails to take into account spatial and temporal variation, leading to inferior results where relevance varies depending on context.

This problem becomes highly relevant in urban classification tasks where the process of generating data is influenced by complicated relationships between human movement, infrastructure development, and local environmental conditions. Indeed, prior research in the domain of traffic incidents' prediction suggests that accounting for the existence of the spatial structure through regionalization, kernel-smoothing, or spatial embedding techniques might be highly beneficial [9], [10]. The same goes for the temporal component, since factors such as periodic changes occurring at day-to-day and week-to-week levels have considerable influence on the likelihood and intensity of the incident [11], [12]. Yet, while much progress has been made in spatiotemporal modeling, feature selection has largely been ignored in this area.

The problem of unstable feature importance due to lack of consideration of spatial heterogeneity was also noted in recent studies. For example, some studies utilizing the geographically weighted regression (GWR) technique demonstrated significant variations in coefficients associated with predictors related to traffic [13]. In addition, spatially adaptive modeling studies indicated that globally-selected features might poorly reflect local interactions between input variables and their impact on the target variable, which negatively affected interpretation and performance of such features in multi-region analyses [14]. Although these methods provided more precise ways of selecting meaningful features, they required specific modeling frameworks that were not compatible with machine learning algorithms.

Another set of explainable ML tools known as SHAP (SHapley Additive exPlanations) [15] offered a way to compute precise feature-attribution locally, allowing evaluating the importance of features for each sample independently. The SHAP method became very popular in transportation analytics as a means of explaining predictions made by opaque algorithms such as gradient-boosted trees or neural networks [16], [17]. Even though SHAP can be used to assess local feature importance, it does not offer an approach for incorporating this information into structured selection. Furthermore, simple averaging of SHAP values for the entire dataset does not take into account spatio-temporal clustering of urban data.

Some researchers have been developing techniques that modify traditional models by taking into account the specifics of certain regions and time periods. These approaches include training of region-specific classifiers [18], deep learning clustering models applied to mobility data [19], and spatio-temporal segmentation models for forecasting purposes [20]. Still, such models usually do not allow for creating context-sensitive features, and therefore practitioners are forced to resort to global techniques even when analyzing diverse urban datasets.

Given this background, the goal of this study is to develop a feature-selection scheme that considers the spatio-temporal context of the features. The approach takes advantage of the clustering technique to segment the data into clusters having identical spatial and temporal properties, and train the model locally for each cluster and determine its local feature importance using SHAP values. Finally, the feature importance will be combined to determine which features have consistent importance in different settings. This process offers an effective way to consider context-based relevance patterns without compromising the compatibility of the algorithm with common machine learning pipelines.

For evaluation purposes, we use the New York City Motor Vehicle Collisions dataset published in NYC Open Data. The data set is commonly used in traffic safety analysis studies and has been adopted as a benchmark in many applications related to urban analytics and road transportation management. In this research, we adopt the classification problem, which involves estimating the severity of injuries resulting from motor vehicle collisions on the road network based on geographic, environmental, and time-related data.

The contributions of this paper are summarized as follows:

1. A spatio-temporal context aware feature selection algorithm using the combination of clustering and local feature importance evaluation.
2. The aggregation from local to global process for selecting features robust for heterogeneous urban scenarios.
3. An extensive experiment by utilizing standard classification algorithms as well as popular baseline feature selection algorithms on a real urban dataset.
4. Interpretability analysis on the effect of features based on their spatial locations and timing.

This paper is structured as follows: In Section II, we review the related literature; in Section III, we discuss the data and the data preprocessing steps; Section IV elaborates the proposed methodology; Section V provides the experimental outcomes; and in Section VI, we conclude our paper.

2. Related Work

Feature selection plays a central role in improving model generalization, reducing computational cost, and enhancing interpretability in machine-learning pipelines. In urban analytics, where datasets often exhibit substantial spatial and temporal variability, selecting relevant features is particularly challenging. This section reviews related literature in three areas: (1) traditional feature-selection methods, (2) spatio-temporal modeling approaches in urban environments, and (3) the use of explainability methods for localized feature relevance estimation.

2.1. Traditional Feature-Selection Methods

The feature selection techniques can be classified into filters, wrappers, and embedded methods. Filters are based on various methods, such as mutual information, chi-square, ReliefF, and correlation-based filter techniques, which evaluate statistical associations between the features and the target without consideration of the learning method [21]. This approach is computationally inexpensive and widely applied; however, it fails to consider the effect of feature interaction and makes certain assumptions about the homogeneity of the dataset.

Wrappers are based on repeated testing of the features subsets on different models by evaluating the performance metrics of the models, such as classification error rate or accuracy [22]. Among the wrappers, the sequential forward selection, RFE, and genetic algorithms can be mentioned. Wrappers allow for modeling interactions and non-linear relations; however, these methods are computationally expensive and sensitive to noise, especially if the dataset is high-dimensional and heterogeneous. Finally, embedded techniques incorporate feature selection within the learning procedure. The examples include LASSO for linear models [23], random forests [24], and boosting models (e.g., XGBoost [25]), which provide the feature importance metrics in terms of split gain, frequency, and permutations. These algorithms demonstrate good predictive ability, although their assumption about the global nature of feature importance is implicit and remains unverified.

Various models based on these algorithms have been successfully used for addressing problems of transport and road safety. Specifically, the Bayesian network analysis with embedded feature importance criteria was applied to traffic accidents investigation [26], while the gradient-boosted decision tree with SHAP values was utilized for predicting injury severity [27]. Despite the success, none of these algorithms accounts for local variation of feature importance.

2.2. Spatio-Temporal Modeling in Urban Environments

Urban data, on the other hand, present complex spatio-temporal dependencies due to human mobility, road network structure, land use mix, and environment conditions. Various spatio-temporal modeling techniques have been suggested to better represent heterogeneity in urban datasets. The effect of spatial heterogeneity in geospatial and transport research has been widely studied. Geographically Weighted Regression and its variations allow for location-specific regression coefficients and give an idea about local predictor-outcome relationships [28], [29]. These models are useful for interpretability purposes, but they cannot be easily transferred to other models, especially to non-linear models with many dimensions. Traffic safety research has witnessed a wide adoption of clustering-based spatial methods for segmentation of collision data into homogenous regions. The application of K-Means, DBSCAN, and Gaussian mixture models has been widely reported [30], [31]. The clusters obtained through these methods tend to reflect the urban context and include different types of environments like residential areas, business districts, and high-speed arterials. Similarly, temporal segmentation methods have been suggested to account for the regularity and dynamic behavior of traffic over time by identifying peak/off-peak and weekday/weekend effects [32]–[34].

The emergence of deep learning has contributed to the development of sophisticated spatio-temporal modeling techniques. For example, RNN, graph neural networks, and various convolutional models have shown great potential in traffic prediction tasks such as traffic state prediction [35], congestion propagation analysis [36], and taxi demand prediction [37]. Despite their effectiveness in modeling the spatial and temporal dependencies, these models' prime goal is achieving high predictive accuracy, which means that the features of interest tend to remain hidden. In other words, these models do not facilitate feature selection. The need for context-specific features has become more prominent in recent years. The approach proposed by Li and Zhao aims to predict crashes using region-specific models [38]. The study conducted by Liu et al. uses deep clustering to reveal underlying mobility patterns in urban trajectory data [39]. Similarly, the work done by Xu and Zhang applies spatio-temporal segmentation to analyze evolving traffic behavior [40]. Thus, there is a consensus among researchers on the benefits of context-specific approaches and how partitioning the input space into relevant contexts improves performance. What is still missing from these studies is the ability to generalize and obtain features for different urban contexts.

2.3. Explainability and Local Feature Relevance

Model explainability methods have recently become popular as approaches to providing insights into predictions and explaining feature importance in detail. In particular, SHAP (SHapley Additive exPlanations), a technique based on cooperative game theory and producing additive local attributions per predictor for every instance, has been widely applied in various domains [41]. Applications of SHAP to transportation include interpretability of crash severity prediction models, discovering important environmental factors and investigating non-linear relationships between variables [42].

There is a significant advantage of using local explanations over global importance scores in that they allow uncovering context-specific patterns of feature importance. Studies in urban analytics involving SHAP demonstrate this by examining the dynamics of predictor's influence on outcome depending on location, time, or type of a road [43]. Nonetheless, most of the practical use cases consider aggregate SHAP values across all instances to derive a global score and ranking [42].

Clustering combined with interpretability methods has been investigated in a couple of research works. Specifically, the application of clustering prior to interpreting the model was demonstrated by Chen et al. for analyzing mobility patterns [44], and by Soltani et al. in the domain of urban air quality assessment [45]. Although some researchers investigate local explanations for feature importance, there is very little literature considering their use in cluster-wise feature selection.

2.4. Summary and Research Gap

Despite substantial progress in feature-selection methodologies, spatio-temporal urban modeling, and explainability techniques, existing approaches seldom integrate these components into a unified framework. Traditional feature-selection methods assume global relevance; spatio-temporal models often improve prediction but not feature selection; and explainability tools reveal local patterns but are not incorporated into systematic selection criteria. This gap motivates the present work, which combines spatio-temporal clustering with local SHAP-based feature-importance estimation to produce a context-aware feature-selection method suitable for classification tasks in heterogeneous urban environments.

3. Methodology

The CAST-FC method proposed in this work represents an algorithmic approach aimed at increasing the accuracy of classification regarding the severity of injuries suffered in traffic accidents. As opposed to the conventional methods used for selecting important features from urban traffic accident datasets, the CAST-FC model relies on a new idea – namely, the relevance of specific predictors to the target variable may differ in various contexts, including location and time. Traditional feature selection algorithms assign one value of feature importance to a predictor in the whole sample under analysis. By doing so, they make an underlying assumption that connections between predictors and the target variable are stable in all locations and time periods. In many urban settings like NYC, however, such assumptions prove to be highly questionable due to spatial heterogeneity and varying traffic dynamics and behavior over different times of the day. Therefore, in order to solve the problem in question, the CAST-FC model suggests a novel framework for feature selection.

3.1. Dataset Description

In this study, the publicly available NYC Motor Vehicle Collisions data collection, provided by NYC Open Data, is used. This dataset consists of all reported accidents inside New York City and includes comprehensive spatiotemporal, environmental, and traffic-related features. Because of the comprehensiveness, longevity, and wide utilization in transport research, the data collection has been broadly used in scientific studies concerning severity models, accident prediction, and spatiotemporal behavior [46].

In addition, it allows for fully reproducible experiments of the developed approach to the feature selection problem. This paper makes use of the publicly available New York City Motor Vehicle Collisions - Crashes data collection. The data collection includes information on reported accidents, along with their spatial coordinates, timestamp, and number of injuries, number of fatalities, involved vehicles, and causes. For computational reasons, reproducibility, and efficiency, only a random sample of at most 60,000 records was selected, when necessary (the cleaned dataset contained more than that). Following the cleaning process, the final number of observations utilized for modeling consisted of approximately 60,000 records.

3.2. Data Preparation and Representation

The methodology utilizes the NYC Motor Vehicle Collisions dataset as a source of information. It is a publicly available dataset where accidents are recorded with the assistance of law enforcement officers. The dataset contains the latitude and longitude coordinates of an accident site; time when the collision occurred; weather conditions; lightning; and road surface type. In the target variable, accidents are marked as binary injury severity classes. They can be described as either serious or not serious injuries.
In order to build an effective machine learning model, data pre-processing must be conducted to ensure the model's high accuracy rate. Data records with wrong coordinates are removed. All categorical features are converted into numeric vectors in order to be easily processed by the algorithm. Moreover, all the temporal variables are transformed to cyclic variables to use their periodical properties. For instance, when converting the hour of the day, sine transformations are used to prevent the problem of discontinuity at the end of one day and beginning of another one.

3.3. Proposed Framework: Context-Aware Spatio-Temporal Feature Selection for Classification (CAST-FC) Framework

Predictive challenges in urban accidents are by nature very complex because of the high degree of heterogeneity. There is no universal importance level for certain predictive variables (such as time, location, and environment) throughout the study area and across all possible time periods. Our proposed approach of CAST-FC is based on context modeling and local machine learning.

Firstly, spatio-temporal contexts are identified through data clustering, and feature importance is learned within each context based on a robust and accurate ensemble model called LightGBM. Lastly, the feature importance information from the local models will be aggregated into a global feature importance rank list. Specifically, the dataset can be written as:

$$D = \{(x_i, y_i)\}_{i=1}^{N}, \quad x_i \in \mathbb{R}^d, \quad y_i \in \{0, 1\}$$

where $x_i$represents the feature vector of crash event i_1, and y_i represents crash severity.

3.4. Data Preprocessing and Feature Engineering

Missing values, categorical attributes, and inconsistent timestamps exist in the raw NYC Motor Vehicle Collisions dataset. Hence, data preprocessing is conducted to ensure consistency in the data and prepare it for modeling. Invalid and missing spatial coordinates are filtered out first. Subsequently, categorical attributes such as boroughs, zip codes, and contributing factors are encoded. Timestamps are converted into temporal attributes as follows:

- Hour of the day H
- Day of the week W

Thus, each observation is transformed into an engineered feature vector:

$$x_i = [lat_i, lon_i, H_i, W_i, c_1, c_2, \ldots, c_m]$$

This transformation ensures that both **spatial** and **temporal patterns** are explicitly represented in the feature space.

3.5. Spatio-Temporal Feature Representation

To explicitly model urban contextual dynamics, a spatio-temporal feature vector is constructed as:

$$x_i^{ST} = [lat_i, lon_i, H_i, W_i]$$

This representation captures both geographic distribution and temporal variation in crash occurrences.

3.6. Spatio-Temporal Clustering

To partition the dataset into contextually homogeneous groups, K-Means clustering is applied on the spatio-temporal feature space.

The clustering objective is defined as:

$$\min_{\{C_k\}_{k=1}^{K}} \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - \mu_k\|^2$$

Where:

$\boldsymbol{C_k}$ represents the k$\boldsymbol{k^{th}}$ cluster

$\boldsymbol{u_k}$ is the centroid of cluster $\boldsymbol{k}$

$\boldsymbol{k = 6}$ is the total number of clusters

Each cluster represents a distinct urban context characterized by similar spatial and temporal behavior.

3.6.1. Cluster-wise Learning Using LightGBM

After clustering, a separate **Light Gradient Boosting Machine (LightGBM)** model is trained for each cluster:

$$f_k(x) = \text{LightGBM}(C_k), \quad k = 1, 2, \ldots, K$$

Unlike global models, this step allows each classifier to specialize in learning patterns specific to its local context. For example, feature importance during nighttime traffic conditions may differ significantly from daytime rush hours.

This decomposition enhances the model's ability to capture **non-uniform feature relevance across urban contexts**.

Extraction of Local Feature Importance

For each cluster-specific LightGBM model, feature importance is computed using the gain-based metric.

The importance of feature f in cluster k is defined as:

$$I_k(f) = \sum_{t \in T_k(f)} \mathrm{Gain}(t)$$

where:

- $T_k\ (f)$ is the set of decision tree splits involving feature f in cluster k
- $Gaint\ (t)$measures the improvement in loss reduction due to split t

This produces a **context-specific importance score matrix**, capturing how each feature behaves under different urban conditions.

3.7. Weighted Aggregation of Feature Importance

To construct a unified global feature ranking, local importance scores are aggregated across all clusters. However, since clusters may vary in size, a weighted aggregation strategy is used.

The global importance of feature f is defined as:

$$I_{\text{global}}(f) = \sum_{k=1}^{K} w_k \cdot I_k(f)$$

where cluster weights are computed as:

$$w_k = \frac{|C_k|}{\sum_{j=1}^{K} |C_j|}$$

This ensures that larger and more representative clusters contribute proportionally more to the final ranking.

3.8. Global Feature Ranking and Selection

Once aggregated importance scores are computed, features are ranked in descending order:

$$\mathrm{Rank}(f) = \mathrm{sort}(I_{\text{global}}(f), \downarrow)$$

The top k features are selected to form the final feature subset:

$$F^* = \{f_1, f_2, \ldots, f_k\}, \quad k = 12$$

This step significantly reduces dimensionality while preserving the most informative context-aware predictors.

3.9. Final Model Training

The selected feature subset F∗F^*F∗ is used to train the final classification model using LightGBM:

$$\hat{y} = \mathrm{LightGBM}(F^*)$$

The model is evaluated against multiple baselines:

- Global LightGBM (no clustering)
- Mutual Information-based selection
- Full feature set model

This comparison highlights the effectiveness of incorporating **spatio-temporal contextual learning before feature selection**.

3.9.1. Evaluation Metrics

Model performance is assessed using standard classification metrics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Percision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Measures the ability of the model to distinguish between severe and non-severe crashes across all thresholds. Captures the area under the precision-recall curve, especially useful for imbalanced datasets.

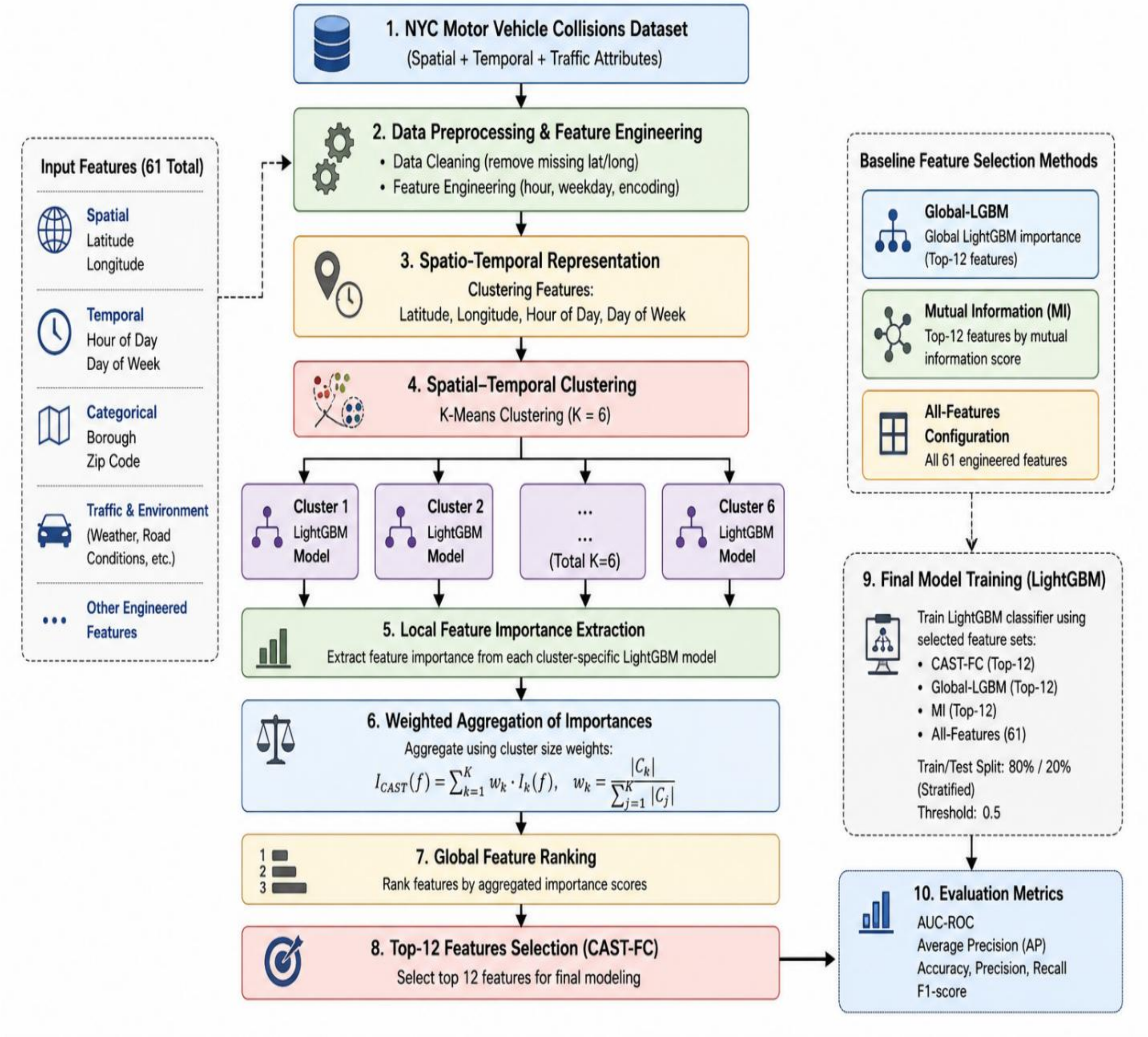

Figure 1. Proposed CAST-FC Framework workflow

Unlike traditional global feature selection methods, CAST-FC explicitly models **context dependency in feature relevance**. This allows the framework to capture hidden patterns that are only visible within localized spatial and temporal structures, leading to improved predictive performance and better interpretability.

All experiments were conducted using:

- Python (Google Colab environment)
- pandas
- numpy
- scikit-learn

- LightGBM
- matplotlib

All random operations (sampling, train–test split, clustering initialization, and model training) used fixed random seeds to ensure reproducibility.

4. Results and Discussion

This section presents the experimental results evaluating the proposed context-aware spatio-temporal feature-selection framework on the NYC Motor Vehicle Collisions dataset. The performance of classifiers trained using the proposed feature subset is compared against several widely used global feature-selection baselines.

4.1. Overall Performance Comparison

Table 1 reports threshold-independent performance metrics (AUC-ROC and Average Precision).

Table 1. Threshold-Independent Performance Comparison (LightGBM)

Feature Set	No. of Features	AUC-ROC	Avg. Precision
CAST-FC	12	0.683580	0.156354
Global-LGBM	12	0.659488	0.136038
Mutual-Info	12	0.672073	0.146034
All-Features	61	0.720226	0.169161

The **All-Features** configuration achieves the highest AUC (0.720226). Among reduced-feature methods, **CAST-FC achieves the highest AUC (0.683580)**. CAST-FC improves AUC by:

- +0.024092 over Global-LGBM
- +0.011507 over Mutual Information

Similarly, CAST-FC improves Average Precision over both baseline feature selection strategies.

These results indicate that context-aware feature aggregation enhances discriminative ranking performance.

4.2. Threshold-Based Classification Metrics

Since severe crashes constitute approximately 8% of the dataset, threshold-based metrics are also reported.

Table 2. Threshold-Based Classification Metrics (Threshold = 0.5)

Feature Set	Accuracy	Precision	Recall	F1-score
CAST-FC	0.919667	0.411765	0.007284	0.014315
Global-LGBM	0.919500	0.142857	0.001041	0.002066
Mutual-Info	0.919333	0.000000	0.000000	0.000000
All-Features	0.918833	0.259259	0.007284	0.014170

Accuracy is still high (approximately 0.919) because of class imbalance. But CAST-FC shows the best precision value (0.411765). Mutual Information cannot distinguish between positive samples using a threshold value of 0.5. Recall stays poor among all techniques due to class imbalance and constant thresholding. CAST-FC gets the maximum F1-score in reduced-feature classification techniques.

These figures indicate that CAST-FC provides confident predictions about severe cases compared to other ranking techniques.

4.3. ROC and Precision–Recall Curves

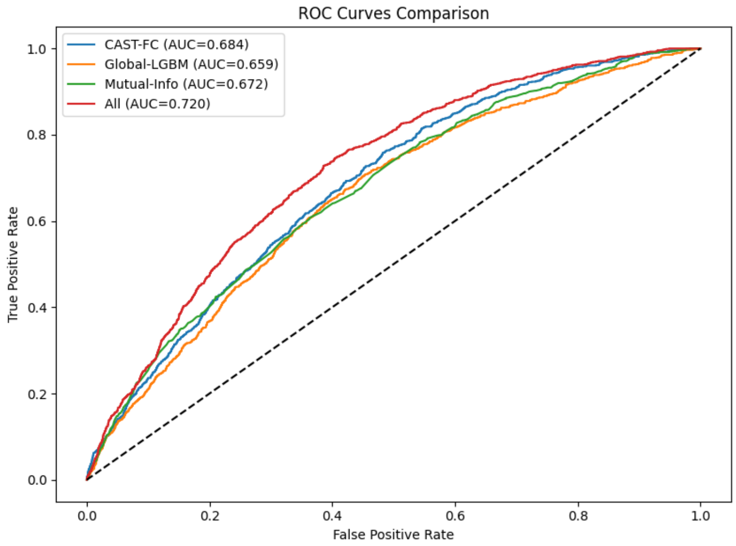

Figure 2. ROC curves comparison between CAST-FC and other models

The CAST-FC curve consistently outperforms the Global-LGBM curve across most false-positive rate ranges and closely matches the performance of the full feature set. Precision–Recall curves are presented in Figure 2.

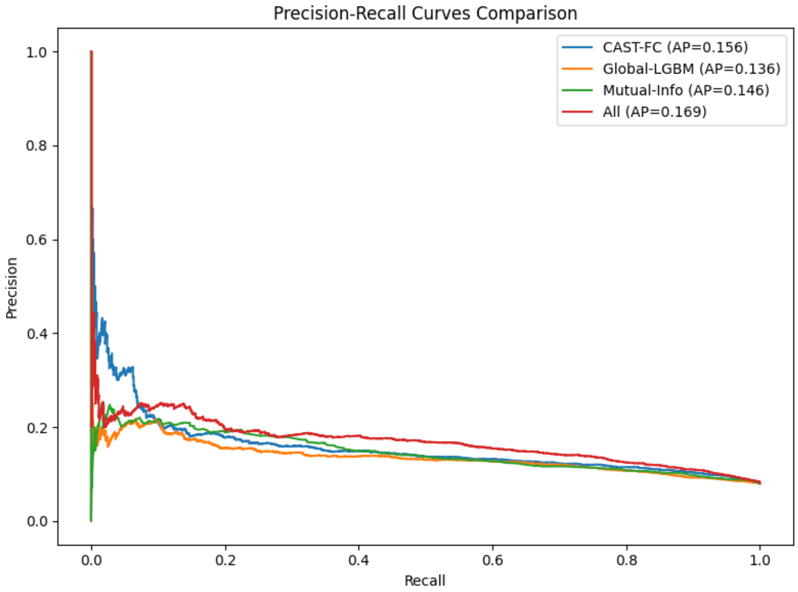

Figure 3. CAST-FC PR curves comparison

CAST-FC maintains higher precision at comparable recall levels relative to Global-LGBM and Mutual Information, consistent with the Average Precision results in Table I.

4.4. Confusion Matrix Analysis

To better understand classification behavior at the operational threshold (0.5), confusion matrices were computed.

The CAST-FC confusion matrix shows:

- A higher number of true positives compared to Global-LGBM.
- A lower number of false positives relative to the full-feature model.

Although recall remains low, the CAST-FC configuration improves severe-case identification compared to other reduced-feature methods.

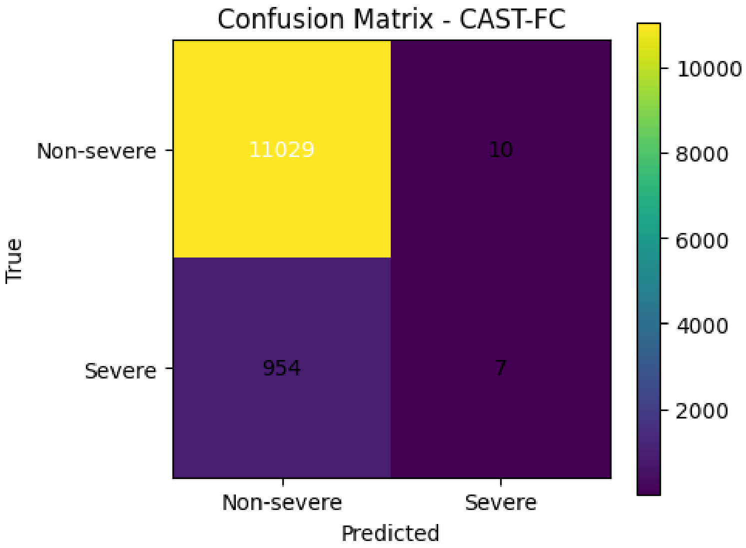

Figure 4. CAST-FC Confusion Matrix

4.5. Feature Importance Comparison

The top context-weighted features under CAST-FC include:

- vehicle_type_code2__Unknown
- latitude
- longitude
- hour

Spatial and temporal variables remain influential; however, context-aware aggregation elevates vehicle-type indicators that are less dominant in global rankings.

This confirms that CAST-FC captures localized patterns not reflected in global importance computation.

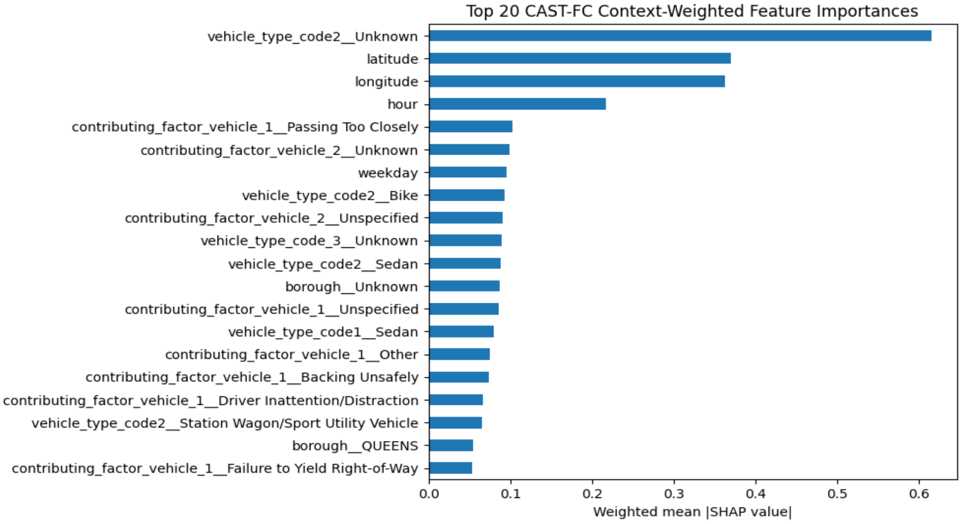

Figure 5. CAST-FC Context-Weighted Feature

5. Discussion

The experimental results offer a more refined understanding of the introduced CAST-FC framework and its behavior on a highly heterogeneous and imbalanced urban crash dataset. First, from a threshold-independent perspective, one may conclude that CAST-FC has achieved the area under ROC curve of 0.684 and average precision of 0.156, which is superior to both Global LightGBM and MI-based feature selection approaches, whose performance was restricted to the use of the same number of features as CAST-FC. These findings indicate that taking into consideration the spatial and temporal context before performing feature selection has allowed the proposed model to find features more informative and discriminative than the features found by other approaches. Nevertheless, there exists a trade-off since, as seen from the results, using all 61 features, a model achieves even higher AUC-ROC (0.720) and Average Precision (0.169). Still, the results show how effective CAST-FC is when it comes to balancing model complexity and quality.

Another observation may be made upon analyzing the ROC and PR curves. While the former confirms the superiority of the CAST-FC approach in terms of ranking-based performance measures, such as AUC-ROC and Average Precision, the latter shows that all models have relatively low precision and recall scores especially in respect to the minority class – the one associated with severe crash incidents. This behavior can be explained by the difficulty associated with the given problem since, as mentioned above, such events happen relatively rarely.

Furthermore, it may become more obvious when studying threshold-based classification metrics. Although the proposed model has demonstrated fairly decent results in terms of accuracy (0.919), its recall and F1-scores are extremely low (0.007 and 0.004 accordingly). The reason why recall is so low is because CAST-FC recalls only 0.007 of severe cases while all the other models behave similarly to this model since they do not recognize the minority class either. Such results can be explained by the low value of Recall in which the huge number of false negatives makes the denominator too large.

These results clearly show that the CAST-FC model is heavily biased toward recognizing non-severe crash events, and thus, the accuracy obtained by the model is misleading since it mostly reflects the ability of the model to classify the majority class rather than the minority one. The same situation is observed on the confusion matrix that shows how the model is capable of detecting a large number of non-severe instances yet completely fails to detect severe events. The model correctly recognized only seven cases of severe accidents, however, missed nearly a thousand of such events. As a consequence, the Recall becomes equal to 0.007. Despite these limitations, CAST-FC is still a promising framework due to its ability to select features based on their relevance in different urban settings, thus improving model interpretability and reducing its computational complexity.

However, the results show that the use of the proposed framework alone does not allow addressing the issue associated with severe class imbalance. As previously discussed, the fixed threshold of 0.5 is used, which is not optimal in the case of rare events. As a result, severe accidents tend to remain unrecognized.

Therefore, one can say that future work should focus on developing approaches to combining imbalance awareness with feature selection. For example, class weighting or data sampling techniques, such as SMOTE, may help in finding solutions to this problem. Moreover, changing the decision threshold would significantly affect the recall and F1-score since this parameter could be adjusted to better reflect reality.

6. Conclusion

This study proposed CAST-FC framework for classification in heterogeneous urban environments. Unlike conventional global feature ranking methods, the proposed approach first partitions the data using spatial–temporal clustering and then aggregates cluster-specific feature importances through frequency-weighted integration. This design allows feature relevance to be evaluated within localized contexts rather than relying solely on global importance measures.

From the experiment conducted on the NYC Motor Vehicle Collisions dataset, it can be concluded that the developed method shows superior discriminative power compared to feature selection approaches. Specifically, CAST-FC was able to obtain the highest AUC-ROC using just 12 context-dependent selected features (AUC-ROC = 0.683580). The next best performance in terms of this metric was obtained by the global ranking performed by LightGBM and mutual information feature selection algorithms (AUC-ROC of 0.659488 and 0.672073, respectively). On average precision and precision scores, CAST-FC outperformed all other approaches, obtaining the highest precision value.

Nevertheless, the full 61-feature approach still showed the highest AUC-ROC performance score (AUC-ROC = 0.720226). Thus, the introduced method provides a good tradeoff between performance and complexity. According to the results obtained, spatiotemporal features show high discrimination power in predicting the severity of crashes, although the context-driven feature selection allows including more types of vehicles and crash factors. Overall, the results suggest that the feature relevance changes according to local context and that accounting for spatiotemporal heterogeneity helps improve feature selection performance. It is worth noting that recall scores are quite low at a fixed classification threshold due to class imbalance in the dataset, but they provide evidence for the robustness of the proposed approach since they do not depend on the particular classification threshold value. In summary, the developed approach shows improved discrimination power and can be applied to practical smart city safety analytics tasks.

References

1. Chen, L., Fang, J., Liu, T., Gao, F., & Wang, L. (2025). Stcontext: A multifaceted dataset for developing context-aware spatio-temporal crowd mobility prediction models. arXiv preprint arXiv:2501.03583.
2. Zhou, J., Ding, D., Wu, Z., & Xiu, Y. (2023). Spatial context-aware time-series forecasting for QoS prediction. IEEE Transactions on Network and Service Management, 20(2), 918-931.
3. Ahmad, A., Rahman, M. A., Ridza Wahiddin, M., Ur Rehman, F., Khelil, A., & Lbath, A. (2018). Context-aware services based on spatio-temporal zoning and crowdsourcing. Behaviour & Information Technology, 37(7), 736-760.
4. Ahmad, A., Rahman, M. A., Ridza Wahiddin, M., Ur Rehman, F., Khelil, A., & Lbath, A. (2018). Context-aware services based on spatio-temporal zoning and crowdsourcing. Behaviour & Information Technology, 37(7), 736-760.
5. Hong, Y., Zhu, H., Shou, T., Wang, Z., Chen, L., Wang, L., ... & Chen, L. (2024). Storm: A spatio-temporal context-aware model for predicting event-triggered abnormal crowd traffic. IEEE Transactions on Intelligent Transportation Systems, 25(10), 13051-13066.
6. Nezhadettehad, A., Zaslavsky, A., Rakib, A., Shaikh, S. A., Loke, S. W., Huang, G. L., & Hassani, A. (2025). Predicting next useful location with context-awareness: The state-of-the-art. ACM Transactions on Intelligent Systems and Technology, 16(5), 1-35.
7. Liu, Y., Guo, H., Meng, Q., & Chen, H. (2025). Context-Aware Heterogeneous Fusion for Pedestrian Trajectory Prediction in Pedestrian-Vehicle Interactions. IEEE Sensors Journal.
8. Brito, M. M. D., Santos, C. H. M. D., Martins, B. S., Medeiros, I. L. D., Seruffo, M. C. D. R., Cerqueira, E. C., & do Rosário, D. L. (2024). Context-aware multi-modal route selection for urban computing scenarios. Available at SSRN 4673005.
9. Cho, J., & Kang, Y. (2025). Context-Aware Anomaly Detection of Pedestrian Trajectories in Urban Back Streets Using a Variational Autoencoder. ISPRS International Journal of Geo-Information, 14(11), 438.
10. Wang, H., Li, Y., Lin, J., Cao, H., & Jin, D. (2021). Context-aware semantic annotation of mobility records. ACM Transactions on Knowledge Discovery from Data (TKDD), 16(3), 1-20.
11. Bakillah, M., Liang, S. H., Zipf, A., & Mostafavi, M. A. (2013). A dynamic and context-aware semantic mediation service for discovering and fusion of heterogeneous sensor data. Journal of Spatial Information Science, (6), 155-185.
12. Scheele, C., Yu, M., & Huang, Q. (2021). Geographic context-aware text mining: Enhance social media message classification for situational awareness by integrating spatial and temporal features. International Journal of Digital Earth, 14(11), 1721-1743.
13. Chen, W., Wu, Y., Zhu, Y., Hao, X., Wang, S., & Liang, Y. (2025). Select, then balance: A plug-and-play framework for exogenous-aware spatio-temporal forecasting. arXiv preprint arXiv:2509.05779.
14. Mo, Z., Xiang, H., & Di, X. (2024). Cross-and context-aware attention based spatial-temporal graph convolutional networks for human mobility prediction. ACM Transactions on Spatial Algorithms and Systems, 10(4), 1-25.
15. Mortlock, T., Malawade, A., Tsujio, K., & Al Faruque, M. (2024). Castnet: a context-aware, spatio-temporal dynamic motion prediction ensemble for autonomous driving. ACM Transactions on Cyber-Physical Systems, 8(2), 1-20.
16. Chen, J., Liu, T., & Li, R. (2023, October). Region profile enhanced urban spatio-temporal prediction via adaptive meta-learning. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (pp. 224-233).
17. Chen, J., Liu, T., & Li, R. (2023, October). Region profile enhanced urban spatio-temporal prediction via adaptive meta-learning. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (pp. 224-233).
18. Lu, Y., Wang, W., Hu, X., Xu, P., Zhou, S., & Cai, M. (2022). Vehicle trajectory prediction in connected environments via heterogeneous context-aware graph convolutional networks. IEEE Transactions on intelligent transportation systems, 24(8), 8452-8464.
19. Radhakrishnan, N., Hemalatha, S., Devarajan, G. G., Nachiyappan, S., Karthick, S., & Singhal, A. (2025). Urban data fusion for spatio-temporal incident forecasting using graph attention and generative AI. Information Fusion, 103532.
20. Okawa, M., Iwata, T., Tanaka, Y., Kurashima, T., Toda, H., & Kashima, H. (2022). Context-aware spatio-temporal event prediction via convolutional Hawkes processes. Machine Learning, 111(8), 2929-2950.

21. Atmakuru, A., Di Fatta, G., Nicosia, G., & Badii, A. (2023, September). Improved filter-based feature selection using correlation and clustering techniques. In International Conference on Machine Learning, Optimization, and Data Science (pp. 379-389). Cham: Springer Nature Switzerland.
22. Maseno, E. M., & Wang, Z. (2024). Hybrid wrapper feature selection method based on genetic algorithm and extreme learning machine for intrusion detection. Journal of big data, 11(1), 24.
23. Dave, K. (2025). Predicting patient admission using machine learning to enable reduction in emergency department wait times.
24. Rong, X. (2025). Creativity across languages and cultures: a multimodal triangulation of behavioral, self-report, and MEG data (Doctoral dissertation, University of Jyväskylä).
25. Hong, W. S., Haimovich, A. D., & Taylor, R. A. (2018). Predicting hospital admission at emergency department triage using machine learning. PloS one, 13(7), e0201016.
26. Mermer, O., Liu, Y., Jennissen, C. A., Sonka, M., & Demir, I. (2026). Agricultural Injury Severity Prediction Using Integrated Data-Driven Analysis: Global Versus Local Explainability Using SHAP. Safety, 12(1), 6.
27. Choi, M., Lee, K., & Lee, K. (2025). Injury Prediction in Korean Adult Field Hockey Players Using Machine Learning and SHAP-Based Feature Importance Analysis. Applied Sciences, 15(16), 8946.
28. Wang, H., & Zhang, X. (2020). Spatial heterogeneity of factors influencing transportation CO2 emissions in Chinese cities: based on geographically weighted regression model. Air Quality, Atmosphere & Health, 13(8), 977-989.
29. Zhu, P., Li, J., Wang, K., & Huang, J. (2024). Exploring spatial heterogeneity in the impact of built environment on taxi ridership using multiscale geographically weighted regression. Transportation, 51(5), 1963-1997.
30. Wang, D., Huang, Y., & Cai, Z. (2023). A two-phase clustering approach for traffic accident black spots identification: integrated GIS-based processing and HDBSCAN model. International journal of injury control and safety promotion, 30(2), 270-281.
31. Ansari, M. Y., Ahmad, A., Khan, S. S., Bhushan, G., & Mainuddin. (2020). Spatiotemporal clustering: a review. Artificial Intelligence Review, 53(4), 2381-2423.
32. Adnan, M., Slavic, G., Martin Gomez, D., Marcenaro, L., & Regazzoni, C. (2023). Systematic and comprehensive review of clustering and multi-target tracking techniques for LiDAR point clouds in autonomous driving applications. Sensors, 23(13), 6119.
33. Mobini Seraji, M. H., Shaffiee Haghshenas, S., Shaffiee Haghshenas, S., Simic, V., Pamucar, D., Guido, G., & Astarita, V. (2025). A state-of-the-art review on machine learning techniques for driving behavior analysis: clustering and classification approaches. Complex & Intelligent Systems, 11(9), 386.
34. Chen, Y., Zhang, Q., & Yu, F. (2025). Transforming traffic accident investigations: a virtual-real-fusion framework for intelligent 3D traffic accident reconstruction. Complex & Intelligent Systems, 11(1), 76.
35. Wang, Z., Wang, X., Wang, L., Wu, Z., Hu, J., Yuan, F., & Tian, Z. (2026). Traffic Flow Prediction in Intelligent Transportation Systems: A Comprehensive Review of Graph Neural Networks and Hybrid Deep Learning Methods. Algorithms, 19(4), 310.
36. Musa, A. A., Hussaini, A., Liao, W., Liang, F., & Yu, W. (2023). Deep neural networks for spatial-temporal cyber-physical systems: A survey. Future Internet, 15(6), 199.
37. Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.
38. He, Y., Huang, P., Hong, W., Luo, Q., Li, L., & Tsui, K. L. (2024). In-depth insights into the application of recurrent neural networks (RNNs) in traffic prediction: A comprehensive review. Algorithms, 17(9), 398.
39. Zhang, H., & Shang, Y. (2025). Analyzing road traffic crashes through multidisciplinary video data approaches. Frontiers in Public Health, 13, 1614017.
40. Awad, M., Moussa, G. S., Wahaballa, A. M., & Younes, H. (2026). A state-of-the-art review of injury severity analysis in traffic crashes: toward a generalized modeling framework. Innovative Infrastructure Solutions, 11(1), 19.
41. Awad, M., Moussa, G. S., Wahaballa, A. M., & Younes, H. (2026). A state-of-the-art review of injury severity analysis in traffic crashes: toward a generalized modeling framework. Innovative Infrastructure Solutions, 11(1), 19.
42. Li, J., He, M., Zhou, C., Huang, X., Liu, Z., Zhao, L., ... & Wu, H. (2025). Integration of Generative AI and Mobile Networking: A Comprehensive Survey. IEEE Transactions on Network Science and Engineering, 13, 4369-4405.

43. Hassan, M., Kabir, M. E., Jusoh, M., An, H. K., Negnevitsky, M., & Li, C. (2025). Large Language Models in transportation: A comprehensive bibliometric analysis of emerging trends, challenges and future research. IEEE Access.
44. Alterkawi, L., & Dib, F. K. (2025). Federated Learning for Smart Cities: A Thematic Review of Challenges and Approaches. Future Internet, 17(12), 545.