International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

STTBoost: A Hybrid Road Traffic Congestion Prediction Model Using Transformer and XGBoost

Deepti Soni, Shraddha Masih

Received: 14/01/2024 **Accepted**: 10/02/2024 **Published**:18/03/2024

Abstract: Accurate traffic congestion prediction is crucial for effectively managing urban transportation systems. Traditional models often focus on either speed or density factors independently, which limits their ability to capture the complex dynamics of real-world traffic conditions. This paper proposes a hybrid approach that integrates both speed and density factors to enhance the realism and accuracy of congestion prediction. The standard TCI is modified to combine density with speed, and a modified Traffic Congestion Index (M_TCI) is proposed. A new hybrid model called STTBoost (Spatio-temporal Transformer Boost) is also proposed, which uses the strengths of Transformer networks and XGBoost to predict traffic congestion. The Transformer model captures spatial-temporal dependencies, whereas XGBoost excels at handling nonlinear patterns. The integration of these two models is used for robust predictions, especially when contextual features, such as road characteristics, road incidents, and temporal patterns, are included. The proposed hybrid model, tested on real-world datasets, demonstrates significant improvements in prediction accuracy, offering a powerful tool for modern traffic management systems. This approach addresses the shortcomings of existing models by providing a more comprehensive and realistic method for congestion forecasting.

Keywords: *Traffic congestion prediction, Free-stream velocity, Number of vehicles, Geohash-temporal information, Transformer, XGBoost.*

1. Introduction

The rapid urbanization and increasing number of vehicles have made traffic congestion a critical issue in modern cities, leading to delays, economic losses, and environmental challenges. Effective traffic management and accurate congestion prediction are essential for mitigating these issues, optimizing road networks, and improving commuter experiences. Traditional congestion prediction models [1-5] have often relied on isolated factors such as speed or traffic volume, failing to capture real-world traffic patterns' complex dynamics. However, congestion is influenced by several interconnected factors, including vehicle density, road network characteristics, and external conditions, such as weather and time of day [6,7]. A more comprehensive approach is required to predict and manage congestion accurately.

A significant limitation of conventional approaches is that traffic volume alone does not always indicate congestion. High traffic volume on major roads might simply reflect busy but free-flowing conditions, while low traffic volume near residential areas or schools may still lead to congestion due to lower speed limits and complex road conditions. Therefore, integrating speed and density is crucial for a more realistic prediction. To address this, we introduce the concept of free-stream velocity, representing the ideal speed under no congestion, and combine it with vehicle density to reflect the actual capacity and congestion level of roads. This approach, captured in a modified Traffic Congestion Index (M_TCI), enhances the accuracy of congestion predictions by considering both factors in conjunction.

(Research Scholar, SCSIT, Davv, Indore, (M.P))
Gmail Id: dsdeepti989@gmail.com
(Professor, SCSIT, Davv, Indore, (M.P))

Geohash, a spatial encoding technique, also plays a key role in improving congestion prediction by allowing the model to capture spatial relationships between road segments. Geohashing divides geographical areas into uniform grids, making it possible to incorporate spatial proximity and interactions between regions into the predictive model [8,9]. This is particularly valuable in urban environments where traffic congestion often spreads spatially over time. When combined with temporal embeddings, geohash provides a structured way to model how congestion propagates across different parts of a city, offering more profound insights into traffic patterns.

Recent advancements in machine learning, particularly with Transformer networks, have enabled the modeling of complex spatio-temporal dependencies in traffic data. Transformers excel at capturing sequential patterns over time, making them well-suited for predicting congestion evolution. However, Transformers alone may struggle to capture non-linear relationships that are common in traffic flow. To overcome this, we propose a hybrid model that combines the temporal learning capabilities of Transformers with XGBoost, which excels at handling non-linear dependencies. This hybrid model integrates geohash-encoded spatial data, speed, density, and contextual features such as road type and weather conditions, events occurring, and landmarks, resulting in a more robust and accurate congestion prediction framework.

Transformer model was initially proposed by Vaswani et al. [10], and its key components—the attention mechanism, encoder, and decoder—form the black box that underpins the model's core structure. The complex nature of its parallelized computation provides significant advantages over RNN in terms of both accuracy and performance, which is why it has been widely applied in fields such as Natural Language Processing (NLP) [11] and Computer Vision (CV) [12]. Ahmed et al. extended its use to mine temporal features in time-series data. However, using the Transformer to extract spatiotemporal patterns is less common [13].

This paper presents a novel approach that merges these elements to create a more comprehensive congestion prediction model. The proposed model significantly outperforms traditional methods by better reflecting the realities of road traffic through a modified TCI and advanced machine learning techniques. This hybrid model offers a scalable and powerful tool for modern traffic management in urban environments. The main contributions of this paper are as follows:

Integration of Speed and Density Factors: The paper introduces a modified Traffic Congestion Index (M_TCI) that combines vehicle speed and density, providing a more comprehensive and realistic measure of congestion compared to traditional models that only consider one factor.

Geohash Embedding for Spatial Representation: Geohash encoding enables the model to capture spatial relationships between road segments, resulting in a more structured representation of geographical areas. This enhances the model's ability to understand and predict how congestion spreads across different regions in urban environments.

Temporal Embedding for Time-Based Dependencies: The model incorporates temporal embeddings to capture time-based patterns in traffic data, such as rush hour trends and daily variations. This helps in accurately modeling the evolution of congestion over time.

Hybrid Model Combining Transformer and XGBoost: The paper proposes a hybrid model that combines Transformer networks' spatiotemporal learning capabilities with XGBoost's nonlinear prediction strengths. This fusion improves the accuracy of congestion prediction by combining the best features of two algorithms.

Incorporation of Contextual Features: The model integrates contextual factors such as road characteristics, road incidents, landmark, weather conditions, and external events, which are essential for capturing the full complexity of traffic congestion and enhancing prediction accuracy.

Improved Prediction Accuracy: Through experiments on real-world datasets, the proposed model demonstrates significant improvements in congestion prediction accuracy over traditional methods, making it a more effective tool for urban traffic management.

2. Related Work

The prediction of road traffic congestion has evolved significantly, transitioning from traditional statistical models to advanced machine learning (ML) and deep learning (DL) techniques. This shift is driven by the need to manage the growing complexity of urban traffic data, including non-linear relationships, spatial-temporal dependencies, and external factors such as weather and public events. This literature review covers key contributions in traffic congestion prediction from 2020 to 2023, focusing on applying statistical models, ML, and DL approaches, emphasizing their predictive capabilities and limitations.

2.1 Statistical Models

Statistical models have been foundational in traffic prediction, primarily relying on time-series data to forecast traffic congestion. Among these, Autoregressive Integrated Moving Average (ARIMA) models remain widely used for short-term traffic forecasting due to their simplicity and effectiveness in capturing linear relationships in traffic data. However, these models struggle with non-linear traffic patterns and complex spatial-temporal interactions. A study by Alghamdi et al. (2021) highlighted the limitations of ARIMA models in dealing with real-world traffic conditions, especially during peak congestion hours. The authors proposed a hybrid ARIMA-LSTM model to address the ARIMA model's shortcomings in capturing non-linear dependencies in traffic flow. Other extensions, such as Seasonal ARIMA (SARIMA), have also been applied to address periodic variations in traffic patterns, but they still face challenges with high-dimensional and highly volatile data [14,15].

2.1 Machine Learning Models

Machine learning models have gained traction in traffic congestion prediction because they can handle non-linear data patterns more effectively than statistical models. Various approaches, such as Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Forests (RF), have been applied to predict congestion based on historical traffic data, road conditions, and external factors like weather and accidents. Support Vector Machines (SVM) have been used in congestion prediction due to their robustness in handling non-linear traffic data. Kamaruddin et al. (2022) applied SVMs to classify traffic conditions and predict future congestion based on historical traffic and weather data. Their model outperformed traditional statistical methods in accuracy but struggled with large datasets due to computational limitations [16].

Similarly, Random Forest models have been applied to congestion prediction by training on features such as traffic speed, volume, and density. Xu et al. (2021) proposed an ensemble-based Random Forest model that integrates multiple traffic-related datasets to improve accuracy in congestion forecasting [17]. Another popular machine learning model in recent years is XGBoost, which is known for its efficiency in handling large datasets and its ability to capture complex relationships. In 2023, Zhou et al. demonstrated the efficacy of XGBoost in traffic congestion prediction by incorporating real-time traffic data and external factors, such as weather and road incidents. Their model was highly accurate in predicting short-term congestion but faced challenges in long-term forecasting [18].

2.2. Deep Learning Models

Deep learning models, especially those designed to capture spatial and temporal dependencies, have become state-of-the-art traffic congestion prediction approaches. Long Short-Term Memory (LSTM) networks, which are particularly adept at capturing temporal relationships, have been extensively used in recent studies. For example, Yang et al. (2022) developed an LSTM-based model to predict congestion by learning temporal patterns in traffic data. Their model achieved higher accuracy than traditional models, particularly for long-term predictions. However, LSTM models often struggle with spatial dependencies, which are crucial in traffic data [5]. Convolutional Neural Networks (CNNs) are used to capture spatial patterns by treating traffic data as grid-like structures to address this. Zhu et al. (2021) introduced a CNN-LSTM hybrid model that combines CNN's spatial learning capabilities with LSTM's temporal features to enhance prediction accuracy. This hybrid model demonstrated significant improvements in forecasting traffic congestion across different regions of a city [20]. In addition to LSTM and CNN models, Graph Neural Networks (GNNs) have gained attention for their ability to represent traffic networks as graphs, capturing the irregular nature of road networks and spatial dependencies between road segments. Li et al. (2022) proposed a GNN-based model to predict traffic congestion by learning from the graph structure of a city's road network. The model outperformed traditional CNN and LSTM models by effectively capturing spatial dependencies and dynamic traffic patterns [21]. More recently, Transformer-based models have shown exceptional performance in traffic congestion prediction due to their ability to capture long-range dependencies in both spatial and temporal dimensions. Originally developed for natural language processing, transformer models are well-suited to processing traffic sequences because of their parallel processing capabilities. Wang et al. (2023) introduced a Spatio-Temporal Transformer (STTF) model for traffic congestion prediction, which integrates road network structures and spatiotemporal dependencies. The STTF model demonstrated superior performance in congestion prediction compared to other deep learning models, particularly in handling large-scale traffic datasets. However, it does not deal with non-linear features for prediction [22].

From 2020 to 2023, significant advancements have been made in traffic congestion prediction, from traditional statistical models to more sophisticated machine learning and deep learning techniques. While applicable, statistical models like ARIMA and SARIMA struggle with the complexities of modern traffic data. Machine learning models such as SVM, Random Forests, and XGBoost have improved prediction accuracy by capturing non-linear dependencies, but they are often limited in handling large-scale spatio-temporal data. Deep learning models, particularly LSTM, CNN, GNN, and Transformer-based models, have emerged as state-of-the-art, offering robust performance in capturing both temporal and spatial dependencies. Hybrid models that combine these approaches and integrate contextual features and geohash embeddings represent the most promising direction for future research in traffic congestion prediction.

3. Notations

In this section, we introduce the key notations used in the proposed STTBoost hybrid model for predicting traffic congestion using hourly traffic data

3.1 Traffic Congestion Index (M_TCI)

A new Traffic Congestion Index using speed and density is proposed called as M_TCI. We use $\tilde{v}_t$ to denote the average speed of vehicles in km/h during the hour, v_{fs} is the ideal speed under no congestion conditions and $v \leq v_{fs}$, else M_TCI= 0, β is the adjustment constant to weigh density's impact, N_{ref} is the ideal operational capacity per kilometer before severe congestion or gridlock happens, D_t is the traffic density/km, $and\ D = N_t/0.7$, where N_t is the number of vehicles passing through the area in 1 hour considering heterogeneous traffic.

Then TCI is calculated as follows,

$$M\_TCI = \left(1 - \frac{\tilde{v}_t}{v_{fs}}\right) \times \left(1 + \beta \times \frac{D_t}{N_{ref}}\right) \quad (1)$$

After calculating the TCI values, it is classified into six classes showing the congestion status. The classes used here are taken from the Highway Capacity Manual (HCM,2010) [23].

Table 1: Traffic Congestion Index

TCI Values	Congestion Status
0.0−2.0	*Unhindered flow conditions*
2.0−4.0	*Nominal flow conditions*
4.0−6.0	*Emerging congestion*
6.0−8.0	*Escalated congestion*
8.0−10.0	*Critical gridlock*

3.2 Geohash(G)

Each location is represented by a geohash code, which encodes geographic coordinates into a string for spatial representation. We consider five geohash areas in Istanbul, each covering approximately 0.7 km. Let G_i represent the geohash location, where I =1,2,3,4.

3.3 Road Network Structure Graph for STTBoost Model

The road network structure for our STTBoost hybrid model is based on a geospatial representation of five geohash locations in the Istanbul City, each covering approximately 0.7 km [24]. These geohash areas act as the primary units for spatial representation, and their connectivity and interactions are fundamental to predicting traffic congestion. The road network is divided into geohash grids, each representing a specific spatial unit on the map. This geohash structure simplifies the complexity of the urban road network, making it easier to model spatial dependencies and variations in traffic patterns across different locations. Each geohash area is represented as a node in the graph, denoted by G_i , where i refers to the unique geohash location. The interaction between these geohash areas is critical to understanding how congestion propagates across regions. For instance, congestion in one geohash area can have ripple effects on neighboring areas due to traffic spillover.

The spatial attention mechanism in the STTF model is intended to capture these relationships. It learns how traffic congestion in one geohash influences nearby geohashes, helping the model predict more accurately for spatially connected locations. The road network structure graph also incorporates temporal dynamics. Traffic congestion is not only spatially dependent but also varies by hour, day of the week, and month. These temporal variations are captured using the temporal attention layer in the STTF model, which integrates time-based features like hourly congestion patterns. The model provides a comprehensive representation of traffic dynamics across the city, capturing both spatial and temporal dependencies. The graph structure can be visualized as a network of nodes called a geohash area connected by edges representing the road infrastructure between them. Each node contains features related to the current traffic state, such as the number of vehicles and speed, and temporal factors, such as the hour of day. The spatio-temporal attention framework dynamically captures how congestion levels evolve over time and space, allowing for more accurate predictions of future congestion levels.

The road network structure graph in our model leverages the geohash-based spatial framework integrated with the temporal dynamics of traffic. By modeling both spatial and temporal relationships within the road network, the STTBoost hybrid model can more accurately predict congestion levels in Istanbul.

4. STTBoost hybrid model

In this section, we introduce the structure of the proposed STTBoost model and the functions of each part.

The STTBoost model is built upon the STTF Transformer model and XGBoost. The whole framework incorporates the Transformer framework for the purpose of extracting spatio-temporal feature extraction and leverages XGBoost for prediction of traffic congestion levels. The dataset is preprocessed and divided into training, validation and test set and after model's training on training data, performance of the model is evaluated.

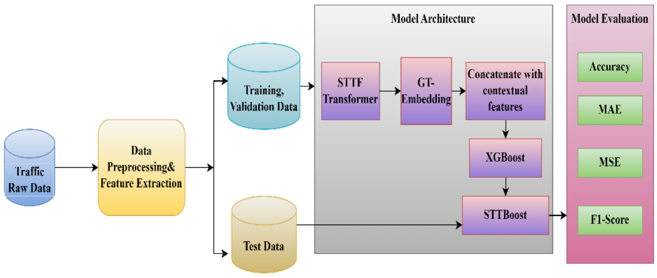

Fig.1: Proposed Framework of STTBoost

The STTBoost hybrid model is specifically tailored to predict urban traffic congestion. The Transformer framework consists of the spatiotemporal attention layers (spatial and temporal) and geohash embedding layers, which enable it to handle spatial information from different locations in Istanbul and temporal data related to hourly, daily, and weekly traffic variations. Here, the input data X includes spatio-temporal data at different time steps. Each X_{ti} consists of features describing traffic conditions excluding contextual information for a specific time t_i, features such as Traffic Congestion Index (M_TCI), number of vehicles and average speed from five geohash locations. This data is processed through the spatial and temporal attention layers and STTF model outputs a D-dimensional vector, which encapsulates the spatio-temporal relationships of the traffic data. This vector, along with additional structured data (like lag features, weather, road incident, event etc.), is passed into XGBoost, which handles the final classification step to predict the Congestion Level Class. The integration of XGBoost allows for efficient handling of structured data, improving the model's robustness compared to using STTF alone. The model predicts the Congestion Level Class across different time horizons (e.g., hour-by-hour predictions or weekly trends), optimizing performance for traffic congestion prediction in Istanbul.

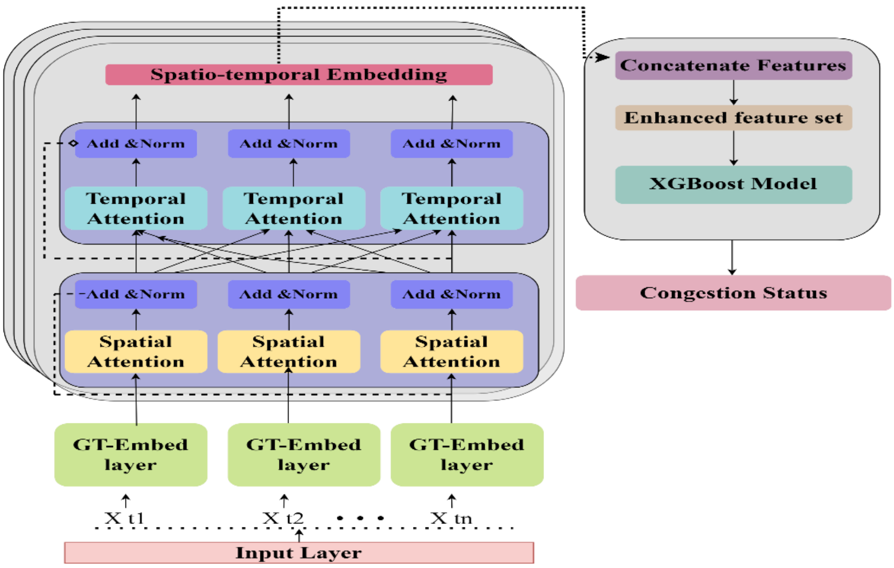

Fig. 2: STTBoost Model Structure

This hybrid model not only captures spatial and temporal dependencies but also processes structured features more effectively, leading to improved accuracy in traffic congestion forecasting over traditional models. The output of the model is the predicted congestion levels for future time steps based on the input historical data. Each module in the framework is set to output a D-dimensional vector to facilitate smooth connections between layers and achieve the highest performance.

4.1 GT-Embedding Layer

The GT Embedding Layer in the proposed STTBoost model combines spatial and temporal embeddings to capture traffic dynamics effectively. In this modified version, we adopt cyclical encoding (sin/cos transformation) for temporal features to represent the periodic nature of traffic patterns better. The resulting embeddings from both spatial and temporal dimensions are then integrated using a novel concatenation technique.

4.1.1 Geohash Embedding Layer (Spatial Embedding)

We embed the geohash locations to capture spatial dependencies in the road network. The geohash encoding inherently captures proximity between locations, making it well-suited for embedding into a spatial feature space. Each geohash area is encoded as a D-dimensional vector. This vector representation captures the spatial relationships between locations by encoding each geohash region's traffic characteristics (e.g., vehicle density, road capacity) into a feature vector. The geohash codes are embedded into a dense vector space using an embedding layer in the model. These embeddings represent the spatial information, allowing the model to learn spatial dependencies between traffic zones.

The final spatial embedding for each geohash location G_i is denoted as

$$se_{vi} \in R^D \qquad (2)$$

where vi represents the geohash location, and D is the dimensionality of the embedding.

4.1.2 Temporal Embedding Layer

We use cyclical encoding with sine and cosine transformations to capture the cyclical nature of time features (such as hours of the day and days of the week). This encoding ensures that time steps like 11 PM and 12 AM are treated as close in the temporal space, which one-hot encoding fails to capture. For each time t, the hour of the day is encoded using sine and cosine transformations, and these transformations project the hour of the day into a cyclical feature space. Similarly, the day of the week is encoded using equations (5) and (6) and captures the cyclical nature of the week (Monday to Sunday).

$$\sin\_hour(t) = \sin\left(\frac{2\pi.hour}{24}\right) \qquad (3)$$

$$\cos\_hour(t) = \cos\left(\frac{2\pi.hour}{24}\right) \qquad (4)$$

$$\sin\_day(t) = \sin\left(\frac{2\pi.day}{7}\right) \qquad (5)$$

$$\cos\_day(t) = \cos\left(\frac{2\pi.day}{7}\right) \qquad (6)$$

These cyclical encoding provides a continuous representation of time, ensuring that the temporal embeddings reflect the periodic nature of traffic data. These temporal features are combined into a D-dimensional vector using an embedding layer $te_{tj} \in R^D$, where tj represents the time step.

4.1.3 Method for Concatenating Spatial and Temporal Embeddings

To integrate the spatial and temporal embeddings, we propose a new concatenation method that maintains the structure of both embeddings while allowing the model to learn from both spatial and temporal dependencies. Rather than using simple concatenation, we propose using the Hadamard product (element-wise multiplication) [25] to combine the spatial and temporal embeddings. This allows each element of the spatial embedding to interact with the corresponding element of the temporal embedding, capturing both the geographic and temporal variations simultaneously. The combined spatiotemporal embedding for geohash location vi at time step tj is computed as

$$ste_{vi,tj} = se_{vi} \odot te_{tj}, \qquad (7)$$

where $\odot$ denotes the Hadamard product. This operation effectively integrates spatial and temporal information into a unified representation that can capture complex interactions between location and time. The resulting embedding $ste_{vi,tj}$ is then passed through subsequent layers of the STTF model, ultimately being combined with structured features (such as TCI, vehicle count, weather, etc.) before being passed to the XGBoost model for congestion prediction.

4.2 Attention Layers

In the STTBoost model, we utilize two primary attention blocks: Spatial Attention and Temporal Attention. Both blocks are designed to capture dependencies and relationships in the traffic data across space and time.

4.2.1 Spatial Attention Block

The Spatial Attention block focuses on capturing the interactions between different geohash locations. Traffic conditions in one location often influence neighboring areas, and this block helps the model learn these spatial dependencies. It takes the spatial embeddings (generated from geohash embeddings) and calculates attention scores between locations to determine how much influence one location has on another. The block outputs a refined spatial embedding that incorporates the spatial relationships between locations, helping the model to understand spatial spillover effects in traffic.

4.2.2 Temporal Attention Block

The Temporal Attention block captures the relationships between traffic patterns at different time steps. Traffic evolves over time, and this block helps the model learn how past traffic data influences future predictions. This block processes the temporal embeddings (created using cyclical encoding for time features) and computes attention scores between different time steps to highlight which past time periods are most relevant for predicting current or future congestion, and the temporal embedding is refined with attention to time dependencies, ensuring that the model can recognize time-based patterns such as daily rush hours or weekly trends.

The spatial and temporal attention blocks operate in parallel to capture both spatial and temporal dynamics in traffic data. The attention scores are calculated automatically using the Scaled Dot-Product Attention mechanism and the model computes these scores using the Query, Key, and Value matrices, and the attention weights are learned through training and enhance model's ability to focus on the most important locations and time periods, making it better equipped to predict traffic congestion patterns in a complex urban environment like Istanbul.

4.3 Spatio-Temporal Embedding

The Spatio-Temporal Embedding (shown in pink) is the final representation that captures both the spatial and temporal dependencies in the model. This block represents the combined output of both the Spatial Attention and Temporal Attention blocks. It synthesizes the learned spatial relationships between different geohash locations and the temporal dependencies between time steps, providing a rich representation for each geohash at a given time. Together, this embedding represents the spatio-temporal dynamics of the traffic data, allowing the model to predict how congestion evolves over both space and time.

4.4 Concatenation with Contextual Features

After generating the Spatio-Temporal Embedding, it is concatenated with additional contextual features- road incident, weather, holiday and landmark proximity. The spatio-temporal embedding $ste_{vi,tj}$ is concatenated with these structured features into a combined feature set. This combined feature set serves as the final input to the XGBoost model, which makes the final predictions for congestion level based on this enriched set of features. The concatenation ensures that both learned embeddings (spatio-temporal dynamics) and external structured data are integrated for more accurate predictions. This Spatio-Temporal Embedding provides a comprehensive representation of the traffic situation at each geohash and time step, and by concatenating it with contextual features, we ensure the model incorporates both deeply learned representations and external factors for robust traffic prediction.

5. Experiments

To evaluate the practical performance of our model, we conduct experiments using large-scale datasets from Istanbul, which provide hourly traffic density data across various locations in the city.

5.1 Datasets

We have gathered dataset from the city of Istanbul [29]. The 13-month period runs from April 2022 until May 2023. The dataset is separated into test, validation, and training sets. We project the level of congestion in May 2023 for the following five geohash point of interest (POI) regions: sxkb6p, sxk3xe, sxkb97, sxk8yv, and sxk92w as shown in Fig:10. The dataset consists of 29,715 rows. It is trained using thirteen different factors that depict the actual traffic. For one geohash(sxk3xe), we have predicted and compared the real congestion levels with the result of the model. Fig.3 shows the density of vehicles over the data collected for one year, where the red color shows the high-density area in Istanbul.

Fig.3: Map reflecting traffic density across Istanbul City

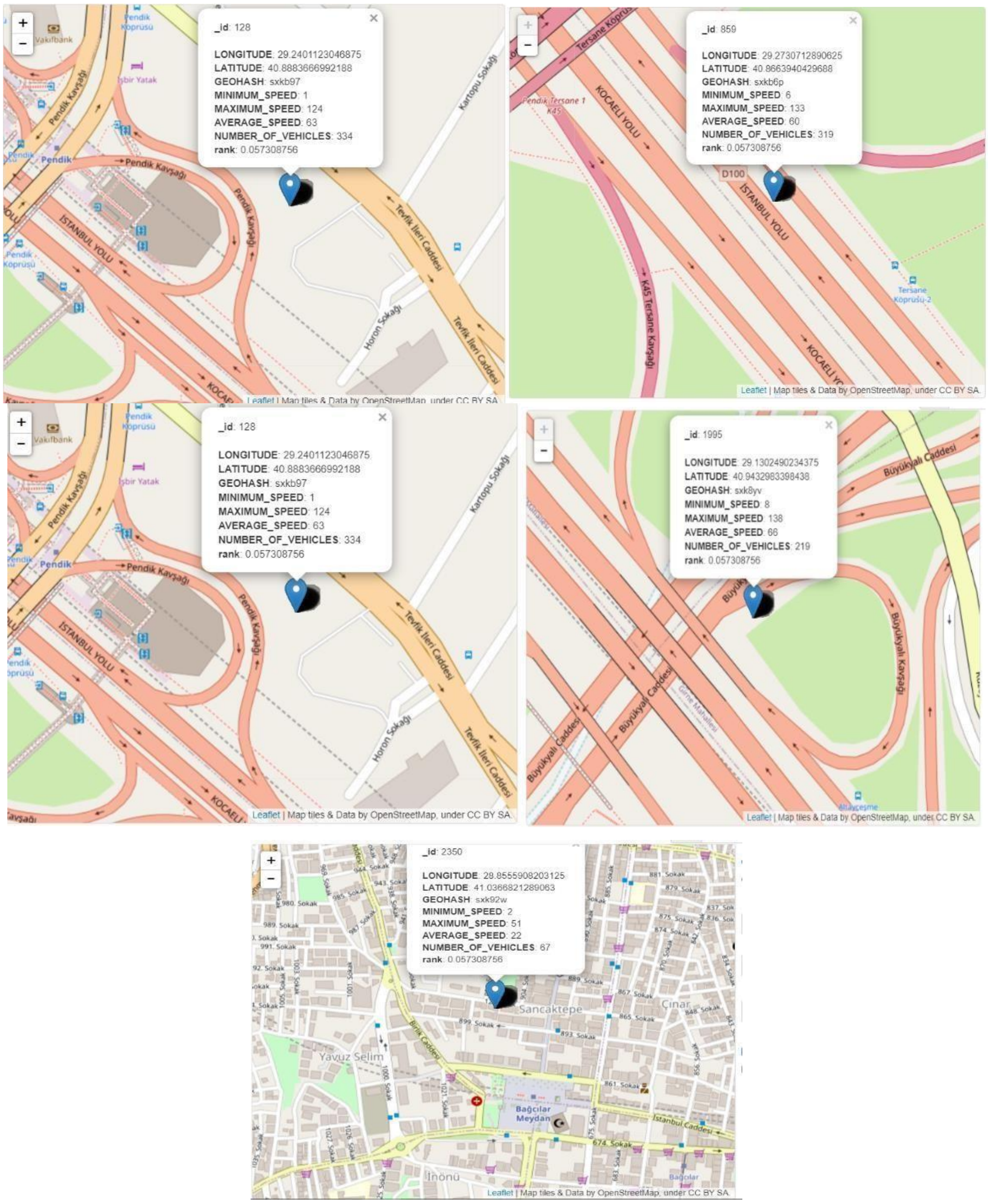

Fig.4: Map reflecting POIs

5.2 Experimental Configuration

The STTBoost model represents traffic congestion on an hourly, weekly, and monthly basis, ensuring it can capture both short-term and long–term traffic variations over time. We use the Adam Warm-up optimizer [27], initializing the learning rate at 0.001, with a warm-up step size of 4000 and batch size of 20, ensuring stable training.

The STTF part of the model has three key hyperparameters: Number of encoder layers (L), L = 4, ensuring the model has sufficient depth to capture spatio-temporal dependencies. Number of attention heads (Q), Q = 8, the multi-head attention mechanism captures different aspects of the spatial and temporal relationships in traffic data. Output vector dimension (D), D = 128, allowing enough capacity to model the complexity of the data. A dropout rate of 0.3 is set to prevent overfitting, and Xavier initialization [33] is used to initialize the weights in the network. For the XGBoost component, the following configuration is used to optimize the final prediction of the Congestion Level. The learning rate is set at 0.05 to balance convergence speed and performance. Max Depth of 6 to ensure the model captures sufficient interactions between the features without overfitting. A subsample ratio of 0.8 ensures that XGBoost leverages different portions of the data to generalize better. The number of boosting rounds is set to 300, allowing enough iterations for the model to converge [28]. This configuration provides a robust hybrid approach that leverages both the STTF model for feature extraction and XGBoost for final congestion classification.

5.3 Baselines and Measures

We selected five benchmark models for comparative experiments, covering traditional and state-of-the-art congestion prediction approaches. These models include ARIMA [29], a classic linear model for time series forecasting, LSTM [24], DCRNN (Diffusion Convolutional Recurrent Neural Network) [30]; and STTF [31], which are state-of-the-art deep learning models that focus on spatial-temporal patterns in traffic data, GMAN [32] and XGBoost [34].

All models' codebases are publicly available, enabling us to run them on our traffic dataset. We use three metrics to measure the models' predictive accuracy; these are F1-Score used for multi-class classification performance. Mean Absolute Error (MAE) which measures the average magnitude of errors between predicted and actual congestion levels. Root Mean Squared Error (RMSE) provides insight into the model's ability to minimize large errors in predictions. These metrics provide a comprehensive evaluation of the models, ensuring we capture both the precision and robustness of the traffic congestion predictions.

5.4 Experimental Results and Discussion

In this section, we evaluate the performance of the proposed STTBoost model by predicting the congestion levels for each hour during the weekday period of May 8, 2023 (00:00 to 23:00). The predictions are made for every 1-hour interval and compared against the actual congestion level as shown in Fig.5. The result of weekly congestion level is shown in Fig.6, hourly weekday congestion level is shown in Fig.7 and the impact of accident in congestion level is shown in Fig.8.

The TCI was calculated for each geohash location for every 1-hour interval, providing a continuous measure of traffic congestion. The congestion levels were then categorized into five classes (Unhindered flow, Nominal flow, Emerging congestion, Escalated congestion and Critical gridlock), which served as the target variable for all models. The STTBoost model utilized both spatial and temporal embeddings in conjunction with structured features like TCI, vehicle count, and road incident to predict the congestion level. The data of one geohash – sxk3xe is applied to the other models mentioned above for congestion prediction and to evaluate performance of STTBoost Model using key metrics and the result is shown in Fig.9.

5.4.1 Results Analysis

The results from the comparative experiments reveal that our proposed STTBoost model outperforms the other models across all key metrics, particularly in its ability to capture spatial-temporal dependencies in the data and incorporate contextual features (like weather, road incidents, etc.). LSTM and ARIMA models performed relatively well in modeling temporal dependencies but struggled to account for the spatial relationships between geohash locations. Both models exhibited higher errors when predicting congestion levels for locations with high traffic flow variability. GMAN, which handle spatio-temporal data with graph-based approaches, performed better than LSTM and ARIMA by capturing local spatial correlations. However, their performance was somewhat limited in cases where external structured features (like TCI) played a critical role in traffic flow. DCRNN and STTF performed better overall in capturing both the spatial and temporal aspects of the data, particularly with STTF integrating spatial and temporal embeddings through attention mechanisms. These models provided more accurate predictions during high-traffic periods, such as morning (7:00-9:00) and evening rush hours (17:00-19:00). The XGBoost model on its own produced strong predictions by leveraging structured data, but when combined with STTF in the STTBoost model, the predictions were further refined. This hybrid approach resulted in better handling of non-linear relationships and short-term fluctuations in congestion levels.

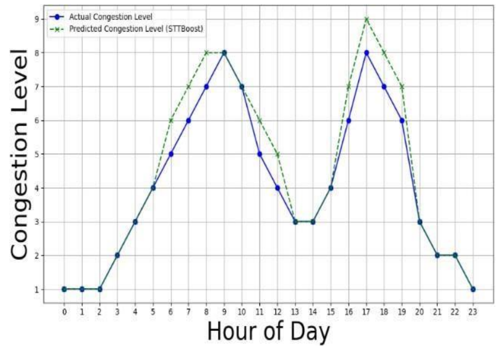

Fig.5: Actual vs predicted congestion levels on May 8,2023(Hourly)

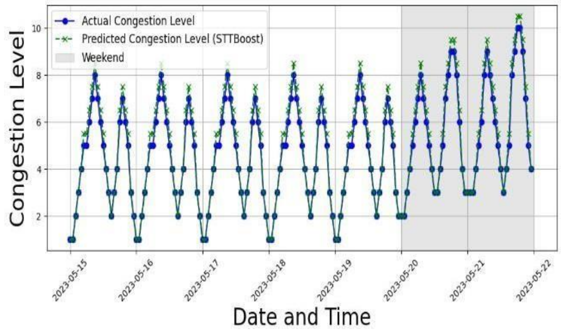

Fig.6: Actual vs predicted congestion level(weekly)

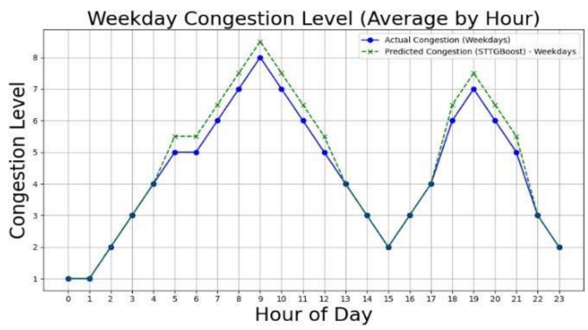

Fig.7: Weekday congestion level(hourly)

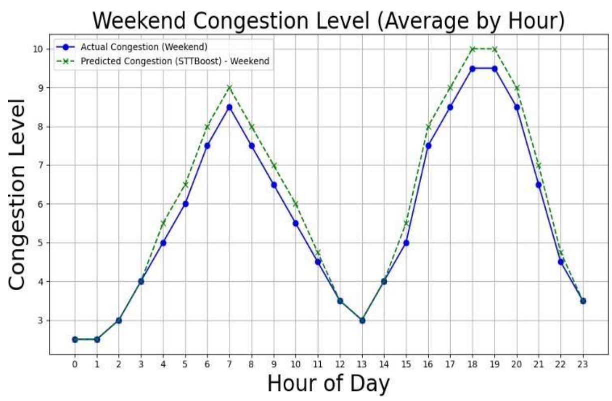

Fig.8: Weekend congestion level(hourly)

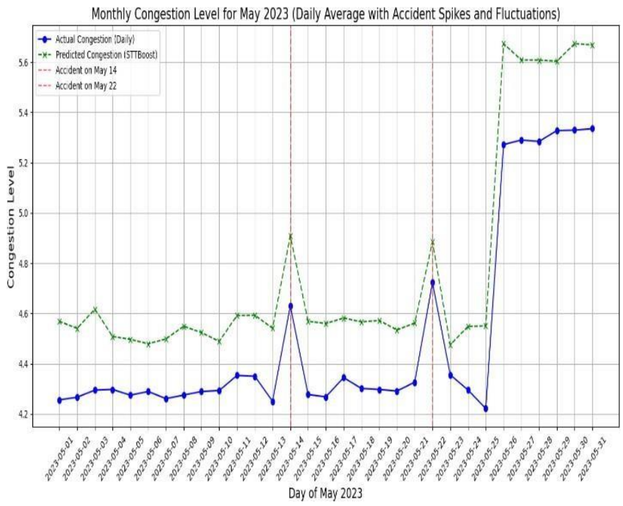

Fig.9: Monthly Congestion Level for May 2023 (Daily Average with Accident Spikes)

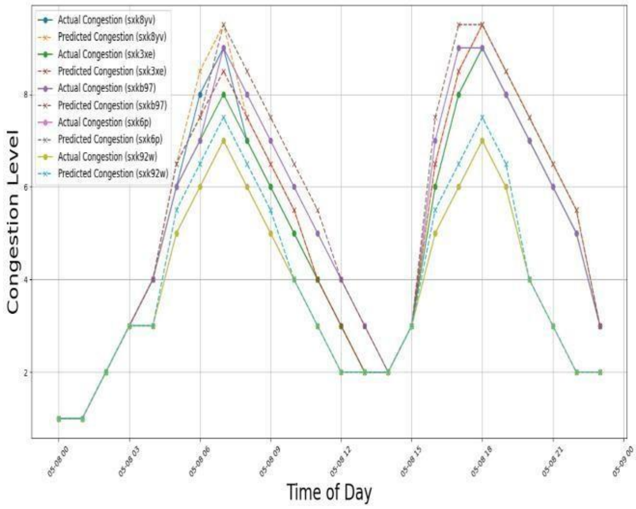

Fig.10: Hourly congestion level for five geohash

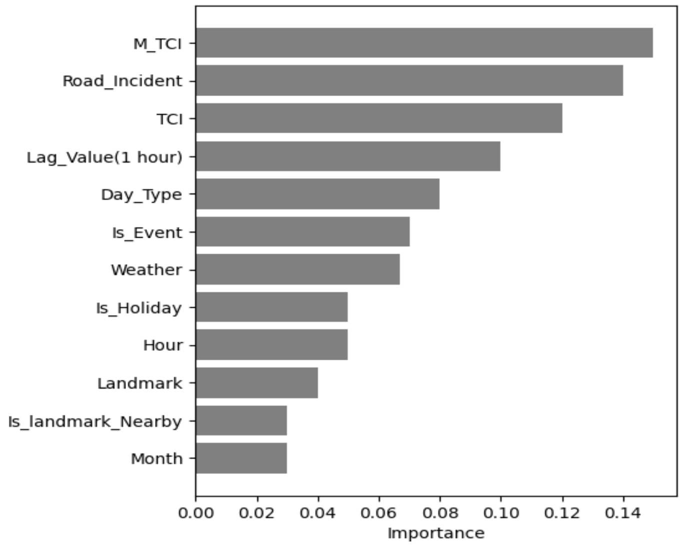

Fig.11: Feature importance in STTBoost Model

All the above plots, from Fig. 5 to Fig. 10 depict the STTBoost model consistently tracked the actual congestion patterns with minimal deviations, especially during peak hours and Fig.11 shows the importance of all factors taken by STTBoost to predict congestion.

5.4.2 Model Performance Evaluation

The STTBoost model achieved the highest F1-Score for predicting each congestion level class, indicating its superior ability to classify traffic congestion correctly. The RMSE for the STTBoost model was the lowest among all models, reflecting its ability to minimize large errors in congestion prediction. The model's average RMSE was 5.6% lower than that of the best-performing baseline model (DCRNN). The Mean Absolute Error for the STTBoost model was consistently lower across all hourly intervals, especially during peak congestion times. This indicates that our model was able to predict the actual congestion levels closely.

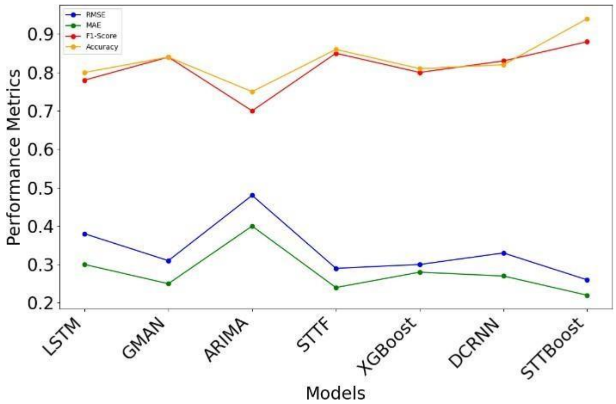

Fig.11: Comparison of models' vs Performance Metrics

Discussion

The results demonstrate that the STTBoost model significantly outperforms traditional models like ARIMA and LSTM, as well as graph-based models like GMAN, in predicting hourly traffic congestion levels. The attention mechanism in STTF, combined with the structured data processing power of XGBoost, allows the model to capture both spatial-temporal dependencies and contextual factors such as weather and traffic incidents.

In conclusion, the combination of spatio-temporal embeddings with XGBoost in the proposed model offers a robust solution for traffic congestion prediction. This hybrid approach demonstrates the capacity to model the complex relationships in urban traffic patterns more accurately than the state-of-the-art models evaluated. Future work can focus on expanding the model to predict congestion over more extended periods and in different urban settings.

6. Conclusion

In this work, we formulated a new traffic congestion index that integrates density with speed factors to better capture real-world traffic conditions. We also developed a novel hybrid model called STTBoost, designed to handle both short-term and long-term traffic predictions with enhanced accuracy. The STTBoost model leverages a spatio-temporal framework that includes an innovative information embedding module, which transforms road network structure and temporal data into learnable feature vectors. These vectors are processed through spatial and temporal attention modules, allowing the model to learn complex patterns in different directions. Our approach demonstrated superior prediction accuracy and efficiency compared to state-of-the-art algorithms, making it a robust solution for real-world traffic prediction tasks.

7. References

[1] Alghamdi, T., Elgazzar, K., Bayoumi, M., Sharaf, T., Shah, S.: Forecasting traffic congestion using ARIMA modeling. In 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC). IEEE, pp. 1227–1232 (2019)

[2] Zaki, J.F., Ali-Eldin, A.M., Hussein, S.E., Saraya, S.F., Areed, F.F.: Time aware hybrid hidden Markov models for traffic congestion prediction. Int. J. Elect. Eng. Inform. 11(1), 1–17 (2019)

[3] Wen, F., Zhang, G., Sun, L., Wang, X., Xu, X.: A hybrid temporal association rules mining method for traffic congestion prediction. Comput. Ind. Eng. 130, 779–787 (2019)

[4] Chen, F., Chen, Z., Biswas, S., Lei, S., Ramakrishnan, N., & Lu, C. T. (2020, November). Graph convolutional networks with kalman filtering for traffic prediction. In *Proceedings of the 28th international conference on advances in geographic information systems* (pp. 135-138).

[5] Momin, Khondhaker Al, et al. "Short duration traffic flow prediction using kalman filtering." *AIP Conference Proceedings*. Vol. 2713. No. 1. AIP Publishing, 2023.

[6] Majumdar, Sharmila, et al. "Congestion prediction for smart sustainable cities using IoT and machine learning approaches." *Sustainable Cities and Society* 64 (2021): 102500.

[7] Akhtar, M., Moridpour, S.: A review of traffic congestion prediction using artificial intelligence. J. Adv. Transport. 2021, 8878011 (2021)

[8] Al Jawarneh, Isam Mashhour, Luca Foschini, and Paolo Bellavista. "Polygon simplification for the efficient approximate analytics of georeferenced big data." *Sensors* 23.19 (2023): 8178.

[9] Yin, F.; Chen, H. Geohash encoding of k-anonymous location privacy protection scheme. J. Wuhan Univ. Sci. Ed. 2022, 68, 73–82.

[10] Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N; Kaiser, Łukasz; Polosukhin, Illia (2017). "Attention is All you Need" (PDF). Advances in Neural Information Processing Systems. 30. Curran Associates, Inc.

[11] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo,B.: Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)

[12] Lim, B., Arık, S.Ö., Loeff, N., Pfister, T.: Temporal fusion transformers for interpretable multi-horizon time series forecasting.Int. J. Forecast. 37(4), 1748–1764 (2021)

[13] Ahmed, Shams Forruque, et al. "Deep learning modelling techniques: current progress, applications, advantages, and challenges." *Artificial Intelligence Review* 56.11 (2023): 13521-13617.

[14] Alghamdi, A.S., et al. "Short-term traffic prediction using ARIMA-LSTM hybrid model." *Transportation Research Record* (2021).

[15] Giraka, Omkar, and Vasantha Kumar Selvaraj. "Short-term prediction of intersection turning volume using seasonal ARIMA model." *Transportation letters* 12.7 (2020): 483-490.

[16] Kamaruddin, S. et al. "Using support vector machines for traffic congestion prediction with weather data." *Journal of Advanced Transportation* (2022).

[17] Xu, M., et al. "Random Forest ensemble learning for accurate traffic congestion prediction." *IEEE Transactions on Intelligent Transportation Systems* (2021).

[18] Zhou, L., et al. "XGBoost-based congestion prediction for intelligent transportation systems." *Applied Sciences* (2023).

[19] Yang, H., et al. "LSTM-based traffic congestion prediction using temporal patterns." *Journal of Big Data Analytics in Transportation* (2022).

[20] Zhu, Y., et al. "A hybrid CNN-LSTM model for traffic congestion prediction in urban areas." *IEEE Access* (2021).

[21] Li, Z., et al. "GNN-based traffic congestion prediction using road network structure." *Journal of Transportation Engineering* (2022).

[22] Wang, X., et al. "STTF: A spatio-temporal transformer model for traffic congestion prediction." *International Journal of Computational Intelligence Systems* (2023).

[23] HCM 2010: Highway Capacity Manual. Washington, D.C.: Transportation Research Board, 2010.

[24] Tanberk, Senem, Mustafa Can, and Selahattin Serdar Helli. "Smart Journey in Istanbul: A Mobile Application in Smart Cities for Traffic Estimation by Harnessing Time Series." *2023 Innovations in Intelligent Systems and Applications Conference (ASYU)*. IEEE, 2023.

[25] Omran, Saleh, Ibtisam Masmali, and Ghaliah Alhamzi. "Banach fixed point theorems in generalized metric space endowed with the Hadamard product." *Symmetry* 15.7 (2023): 1325.

[26] Arifoğulları, Ömer, and Gülfem Işıklar Alptekin. "Public Transportation Data Analysis to Estimate Road Status in Metropolitan Areas: The Case of İstanbul." Procedia Computer Science 210 (2022): 12-18.

[27] Ma, J., Yarats, D.: On the adequacy of untuned warmup for adaptive optimization. arXiv preprint 7. arXiv: 1910. 04209 (2019)

[28] Liu, Liyuan, et al. "Understanding the difficulty of training transformers." *arXiv preprint arXiv:2004.08249* (2020).

[29] Alghamdi, T., Elgazzar, K., Bayoumi, M., Sharaf, T., Shah, S.:Forecasting traffic congestion using ARIMA modeling. In 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC). IEEE, pp. 1227–1232 (2019)

[30] Li, Y., Yu, R., Shahabi, C., Liu, Y.: Diffusion convolutional recurrent neural network: data-driven traffic forecasting. arXiv preprint arXiv: 1707. 01926 (2017)
[31] Shin, Dong-Hoon, Kyungyong Chung, and Roy C. Park. "Prediction of traffic congestion based on LSTM through correction of missing temporal and spatial data." *IEEE Access* 8 (2020): 150784-150796.
[32] Zheng, Chuanpan, et al. "Gman: A graph multi-attention network for traffic prediction." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 34. No. 01. 2020.
[33] Liang, Jiaqi. "A method for the detection of electricity theft behavior based on Xavier weight initialization." *Second International Conference on Electronic Information Technology (EIT 2023)*. Vol. 12719. SPIE, 2023.
[34] Zhan, Q. L., and Y. Li. "An evaluation method for economic loss of urban road traffic: XGboost algorithm." *Advances in transportation studies* (2023).