Contents lists available at www.goldenlightpublish.com/journals/jcemi/2-toc

Journal of Construction Engineering, Management & Innovation

Journal homepage: www.goldenlightpublish.com/journals/2/jcemi

Enhancing bike-sharing demand forecasting with spatio-temporal learning: A comparative study on a local dataset

Melike Aygün Çakıroğlu[1], Suat Özdemir[2]

[1] Distance Education, Research and Application Centre, Abdullah Gül University, Kayseri, Türkiye
[2] Computer Engineering Department, Hacettepe University, Ankara, Türkiye

* Corresponding author: M. Aygün Çakıroğlu (melikeaygun@gmail.com)

https://doi.org/10.31462/jcemi.2026.652
Received 15 December 2025; Revised 19 March 2026; Accepted 08 April 2026; Available online 22 June 2026

Keywords
Artificial intelligence
Bike-sharing systems
Sustainability
Spatio-temporal forecasting
Graph neural networks

Abstract

Bike-sharing systems rely on accurate short-term demand forecasts to prevent shortages, surpluses, and costly rebalancing operations. Accurate predictions are also essential for enhancing the environmental, operational, and social sustainability of these systems, as improved forecasting may help reduce truck-based redistribution, potentially lowering emissions and supporting more efficient resource usage and more reliable urban mobility services. In this study, we evaluate a wide spectrum of forecasting approaches—ranging from classical time-series models (ARIMA, Prophet) and ensemble learners (Random Forest, XGBoost) to spatio-temporal deep learning models (LSTM variants, ST-GCN, GraphWaveNet, and GNN)—using a local station-level dataset. We incorporate both temporal history and spatial dependencies among stations under a unified evaluation protocol (24-hour look-back, one-hour-ahead prediction). Our findings show that integrating spatial context consistently improves accuracy. The spatio-augmented Random Forest (RF-ST) achieves the best performance, reducing error rates by over 10% compared to its temporal-only counterpart and by more than 35% relative to ARIMA and Prophet. Graph-based neural models (e.g., GNN) deliver comparable accuracy, further confirming the benefits of explicit spatial modeling. These results highlight the potential sustainability implications of spatio-temporal forecasting, suggesting that more accurate station-level predictions may support more sustainable rebalancing strategies, potentially help reduce operational inefficiencies, and strengthen the long-term viability of bike-sharing systems as a sustainable transportation mode.

1. Introduction

Bike-sharing systems are a key component of smart and sustainable urban mobility, yet their effectiveness hinges on accurate short-term demand forecasts to avoid stock-outs, full docks, and costly truck rebalancing that degrade user experience and efficiency. Hence, demand forecasting is a central research area in intelligent transportation and service quality [1, 2]. Prior work spans deep learning [3, 4], machine-learning baselines [5-7], and hybrid temporal models that couple ARIMA with deep components to capture both long- and short-term dependencies [2, 8]. These approaches leverage historical usage and external drivers such as weather and calendar effects, but many primarily emphasize temporal signals.

A persistent gap in the literature is the limited treatment of spatial dependencies among stations. Nearby stations often exhibit correlated usage patterns, and spatial context can shift demand across neighborhoods and time. However, many existing approaches rely primarily on temporal signals, which restricts their ability to capture spatial heterogeneity in bike-sharing usage. The insufficient modeling of spatial interactions may also reduce operational efficiency, as operators depend on accurate local demand patterns to plan station rebalancing and resource allocation.

This limitation also has potential implications for operational efficiency and sustainability. When spatial demand variations are not adequately captured, operators may rely on less efficient rebalancing strategies, which can increase operational costs and resource usage. Prior studies (e.g., Subramanian et al. [2]) highlight that improved demand forecasting can support more efficient and sustainable urban mobility systems. This study is primarily a forecasting-focused analysis, and sustainability considerations are included as potential application outcomes rather than directly measured impacts.

This study's objective is to quantify how much explicit spatial modeling improves short-term, station-level forecasting over purely temporal counterparts under a unified evaluation protocol (24-hour look-back, one-hour-ahead prediction). By situating forecasting within a sustainability framework, we emphasize that reducing prediction errors can potentially contribute to: (i) lower rebalancing mileage and associated carbon emissions, (ii) reduced waste of operational resources, (iii) improved service reliability and equitable access across neighborhoods, and (iv) enhanced system resilience through data-driven decision-making.

This study contributes by (i) using a previously unstudied local city-scale dataset (Kayseri; 84 stations, ~1.53M rentals), (ii) injecting lightweight spatio-temporal (ST) features into classical and ensemble learners (RF-ST, XG-ST) while benchmarking graph-based and sequence deep models under identical splits/metrics, and (iii) reporting station-wise significance to provide reproducible, operation-oriented evidence. Unlike prior studies that focus mainly on accuracy improvements, our work discusses the potential sustainability implications of modeling choices, demonstrating how spatially enriched prediction methods can support greener, cost-efficient, and user-centered bike-sharing operations. Moreover, unlike sustainability-oriented studies such as Subramanian et al. [2], which conceptually link AI-driven demand forecasting to greener mobility outcomes, our work provides a comprehensive, model-rich, station-level empirical evaluation that empirically evaluates how ST modeling may support operational sustainability through improved forecasting accuracy.

To address the gap, we incorporate ST information across model families—augmenting classical time-series and tree ensembles with ST features and evaluating ST deep models (e.g., LSTM variants and GNN-based methods)—on the local dataset. Through this design, our study evaluates not only whether ST modeling improves numerical performance but also how these improvements can contribute to sustainable urban transport management by reducing operational burdens and enhancing long-term system viability. Our results show that encoding spatial relations alongside temporal history yields markedly better station-level forecasts and, by extension, more reliable operations. In our experiments, RF-ST achieves the best overall accuracy, with GNN and XG-ST close behind, highlighting a practical accuracy–efficiency trade-off for operators.

Our contributions are as follows:

• We unify and benchmark temporal, ST, and graph-based models under identical data splits and metrics.

• We provide station-wise statistical tests and operational insights for practical deployment in local bike-sharing systems.

• We explicitly link model performance improvements to their environmental, operational, and social sustainability implications.

Finally, the paper is organized as follows: Section 2 details the dataset, features, and models; Section 3 explains the experimental protocol and metrics; Section 4 reports comparative results; Section 5 discusses model behavior and practical implications; Section 6 concludes with key takeaways and limitations/future work.

2. Methodology

In this section, we describe the methodology employed to improve demand forecasting in bike-sharing systems by integrating ST features into both traditional and modern forecasting models. Our methodological choices also aim to support sustainable bike-sharing operations by enabling more accurate and equitable decision-making, potentially reducing inefficient rebalancing, and ultimately lowering the environmental footprint associated with redistribution activities. We detail the dataset used, the models implemented, the experimental setup, and the evaluation metrics to assess the performance of each model. The overall methodological workflow is illustrated in Fig. 1, which summarizes the major stages of the study from data collection to model evaluation.

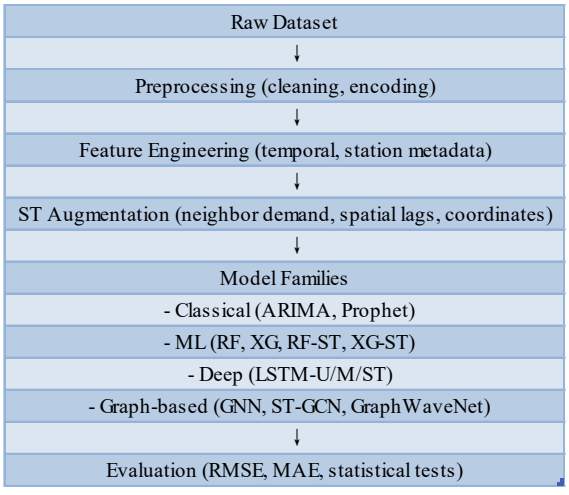

Fig. 1. Overall framework of the study, illustrating the full pipeline from raw data through preprocessing, feature engineering, ST augmentation, model development, and evaluation

2.1. Dataset and data collection

We use Kayseri's bike-sharing data (2023–2024; April–November operation), covering 84 stations, 1,838 total capacities, and ~1.53M rentals with start/end times and stations. We clean missing/outliers, encode categorical variables, and derive temporal (hour, weekday) and spatial (station coordinates, neighbor proximity) features for all models. It should also be noted that the Kayseri bike-sharing system operates seasonally. During winter months, when cycling demand is extremely low due to weather conditions, the system is temporarily suspended and bicycles undergo maintenance operations. Consequently, the dataset used in this study covers the active operational period between April and November, and therefore does not include full annual seasonal cycles.

In this study, hourly demand is defined as the number of bike rental trips originating from each station within a given hour. Thus, the forecasting task focuses on predicting trip departures rather than net station flow (inflow minus outflow).

To provide a clearer characterization of the dataset, we computed several descriptive statistics at the station–hour level. The hourly demand values range from 0 to 93 trips, with an average of 1.26 trips per station-hour. Across the complete station–hour matrix, approximately 70.45% of entries contain no recorded trips, reflecting the naturally sparse usage patterns observed in large-scale bike-sharing systems, where many stations experience prolonged periods of zero demand. Using an IQR-based rule, 7.19% of non-zero observations were flagged as statistical outliers, typically corresponding to peak-demand hours at major transit and activity hubs. At the station level, the number of non-missing hourly observations varies substantially (min = 722 hours, max = 7,641 hours), indicating heterogeneity in operational periods and usage intensity across the network. These statistics provide an empirical foundation for understanding spatial and temporal variability in demand prior to modeling. Although the dataset contains a high proportion of zero-demand observations, this sparsity reflects the natural usage patterns of bike-sharing systems, where many stations experience periods of inactivity. In this study, we did not apply explicit zero-inflation modeling techniques, since the forecasting models used (e.g., tree ensembles and neural networks) are capable of learning from sparse demand patterns without requiring a separate zero-generation process.

2.2. Data preprocessing and feature engineering

The raw data was cleaned and preprocessed by handling missing values, removing outliers, and transforming categorical variables into numerical features. In addition, temporal and spatial features were extracted, including time of day, day of the week, and geographical proximity between stations. These features were essential for incorporating ST dynamics into the forecasting models.

To ensure reliable forecasting performance, several preprocessing steps were applied before model development. First, raw trip logs were aggregated into station-level hourly demand series based on trip origin stations. Data cleaning procedures were then conducted to remove inconsistent records and correct timestamp irregularities. Missing hourly observations were handled based on the structure of the raw dataset. Since bike-sharing trip logs are recorded only when a rental transaction occurs, the absence of a record for a given station-hour interval indicates that no trips took place during that period rather than a missing observation. Therefore, such intervals were represented as zero demand in the constructed time series. This approach reflects the operational characteristics of bike-sharing systems, where stations frequently experience periods of inactivity.

Outlier detection was performed using the interquartile range (IQR) rule to identify unusually high demand values that could distort model training. These observations were retained when they corresponded to plausible peak demand periods, such as commuting hours or major activity hubs, since they represent meaningful operational patterns rather than measurement errors.

After cleaning, several temporal and spatial features were engineered to enrich the predictive signals. Temporal features included hour-of-day, day-of-week, and cyclic encodings of time variables to capture daily usage patterns. Spatial features were derived from station coordinates and neighboring station relationships using a geographic proximity graph. For spatio-temporal feature construction, neighbor demand signals were computed as a weighted average of nearby stations using a Gaussian kernel. The weights were derived from pairwise geographic distances and then row-normalized, ensuring that the contributions of neighboring stations sum to one. Thus, the spatial lag feature represents a normalized weighted aggregation of neighbor demand rather than a simple arithmetic average. In this study, the spatial adjacency matrix was constructed using geographic proximity based on station coordinates, implemented through a Gaussian kernel and k-nearest neighbor filtering. This distance-based formulation was preferred because it provides a simple, interpretable, and reproducible representation of spatial interactions between stations. More complex alternatives, such as correlation-based or dynamic graphs, were not considered in order to maintain a consistent and transparent experimental framework across all models. These engineered features enable the models to capture both temporal dynamics and spatial dependencies across stations, which are essential for accurate station-level demand forecasting in bike-sharing systems.

2.3. Models and techniques

We compare four families under a common setup (one-hour-ahead target; 24-hour look-back; identical split/metrics; hyperparameters in Table 1):

Table 1. Model-specific configurations summarizing the input features, architectural components, and key hyperparameters used across all forecasting models evaluated in this study

Model	Inputs (delta vs. common)	Key architecture / hyperparams	Approx. Trainable Parameters	Notes
ARIMA	—	Candidate orders: (1,1,1)/(0,1,1)/(1,1,0)/(2,1,2)	N/A	Rolling 1-step ahead on test
Prophet	—	Weekly seasonality on; daily/yearly off; no holidays	N/A	Robust to gaps/outliers
LSTM-U	24 self-lags	LSTM(64, tanh) → Dense(1); Adam; up to 20 epochs; EarlyStopping(pat=3, batch=64)	16,961	Scaling: MinMax (train)
LSTM-M	Demand + hour, weekday + (lat, lon)	LSTM(64, ReLU) → Dense(1); Adam; up to 20 epochs; EarlyStopping(pat=3, batch=64)	17,985	MinMax (train); window=24
LSTM-ST	Self lags + kNN neighbor aggregate (σ=1.5 km, k=5, row-norm) + cyclic time	LSTM(64, tanh) → Dense(1); Adam; up to 20 epochs; EarlyStopping(pat=3, batch=64)	18,241	StandardScaler (train)
GNN	Last-24h demand (z-scored) + sin/cos(hour), weekday	Linear(24→64) → GCN(64) → GCN(64) → Linear(64+4→1); ReLU; dropout=0.1; residual; Adam (lr=1e-3, wd=1e-4); up to 40 epochs; EarlyStopping(pat=10)	9,989	Graph: geo-proximity (σ=1.5 km), kNN=5, row-norm, self-loops, symmetric norm
RF	Last-24 self-lags only	RFRegressor(n_estimators=500, max_features="sqrt", min_samples_leaf=5, random_state=42, n_jobs=-1)	N/A	No explicit calendar vars
RF-ST	Self-lags + kNN neighbor lags (σ=1.5 km, k=5, row-norm) + cyclic time	Same RF as above	N/A	—
XG	Last-24 self-lags only	XGRegressor(objective="reg:squarederror", n_estimators=100, learning_rate=0.1, random_state=42, n_jobs=-1)	N/A	Very short series skipped (≥ 10 post-lag samples)
XG-ST	Self-lags + hour, weekday + (lat, lon)	XGRegressor(n_estimators=200, learning_rate=0.1, max_depth=6, subsample=0.9, colsample_bytree=0.9, random_state=42, n_jobs=-1)	N/A	MinMax (global); window=24
ST-GCN	Seq_len=24 (+ optional cyclic time)	Linear(24→16, ReLU) → A·X → Linear(16→1); Adam (lr=0.01); up to 40 epochs; EarlyStopping(pat=5)	417	Graph: Haversine + Gaussian(σ=1.5 km), row-norm
GraphWave Net	W = 24; X standardized over lag windows; Y standardized per station.	1×1 Conv (1→32) → 4 dilated temporal convolution + graph convolution blocks with residual/skip connections → 1×1 Conv (64→128→1); Adam (lr = 1e-3, wd = 1e-4); up to 40 epochs; EarlyStopping (patience = 5)	54,465	Haversine + Gaussian (σ=1.5 km), k = 5, row-normalized, bidirectional supports with self-loops

(i) Classical baselines. The Autoregressive Integrated Moving Average (ARIMA) models trends/seasonality from historical demand, serving as a temporal-only reference [9-14]. Prophet offers an additive trend/seasonality/holiday formulation that is robust to missing/outliers and provides quick, interpretable baselines [15-18].

(ii) Sequence deep learning. Long Short-Term Memory (LSTM) (univariate (LSTM-U) and multivariate (LSTM-M)) captures nonlinear temporal dependencies; a spatio-temporal LSTM (LSTM-ST) augments demand sequences with calendar features as well as neighboring station demand (Fig. 2).

(iii) Tree/boosting. Random Forest (RF) and eXtreme Gradient Boosting (XG) are strong nonlinear tabular learners; their spatio-augmented variants add coordinates, spatial lags, and neighbor signals to encode spatial dependence—RF-ST [19, 20, 32] and XGBoost–Spatio [21] (XG-ST)—building on RF [22-24].

(iv) Graph-based models. Spatio-Temporal Graph Convolutional Network (ST-GCN) learns joint space–time patterns on a station graph (edges from proximity/correlation) with temporal modules (and, in some variants, attention) [25]. GraphWaveNet combines adaptive graph convolutions with dilated causal temporal convolutions for long-range dependencies [26-28].

The selection of model families in this study was designed to provide a balanced comparison across different methodological paradigms used in demand forecasting. Classical time-series models (ARIMA and Prophet) were included as widely adopted baseline approaches that capture temporal dynamics using statistical formulations.

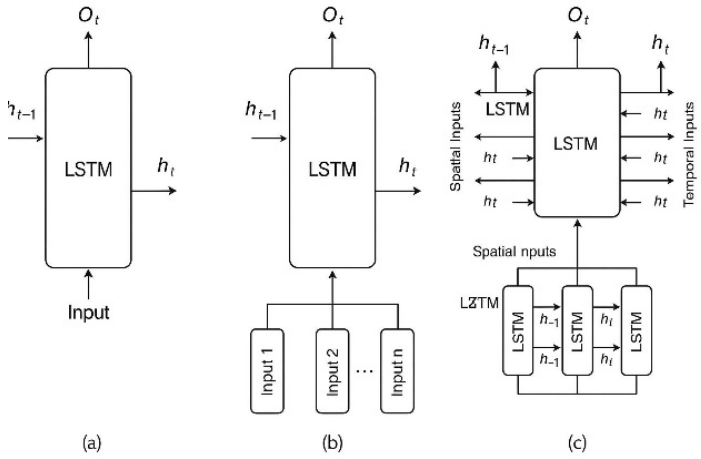

Fig. 2. Architectures of LSTM, Multi-variate LSTM, and LSTM-ST Models

Tree-based machine learning models (Random Forest and XGBoost) were selected due to their strong performance on tabular data and their ability to model nonlinear feature interactions. LSTM-based architectures represent sequence deep learning models capable of capturing complex temporal dependencies in time-series data. Finally, graph-based neural networks were included to explicitly model spatial relationships among stations through graph structures. By evaluating these model families under a unified experimental framework, the study aims to assess how progressively richer representations of temporal and spatial dependencies influence forecasting performance.

3. Experimental Setup and Results

3.1. Model evaluation and performance metrics

We evaluate per-station hourly series with a chronological 80/20 split and a unified 24-h look-back, rolling one-step-ahead protocol. Because the forecasting models were trained separately for each station, the chronological 80/20 split was applied independently to each station's hourly demand series. Therefore, the exact calendar boundaries of the training and testing sets vary across stations depending on their data availability and operational duration. In all cases, the training set corresponds to the earliest 80% of observations and the test set to the most recent 20% of observations within the April 2023–November 2024 operating period. Metrics are per-station RMSE/MAE, and are reported as mean ± std across stations. Scale-independent metrics such as MAPE were not included due to the high proportion of zero-demand observations in the dataset. Since MAPE involves division by the true value, it becomes undefined or unstable when actual demand values are zero or close to zero. Given the zero-inflated nature of bike-sharing demand at the station level, RMSE and MAE were preferred as more stable and interpretable evaluation metrics. Unless specified otherwise: negative predictions are clipped to 0; hours missing in the raw logs are imputed as 0; series with ≤ 25 post-lag samples are excluded; optimization uses mean squared error; statistical significance is assessed via Friedman + Nemenyi ($\alpha = 0.05$). For the Friedman test, models were ranked separately for each station based on their RMSE values, where lower error corresponds to a better rank (rank 1 = best). These ranks were then averaged across all stations to obtain the mean rank of each model. The Friedman test was applied to these per-station rankings, followed by a Nemenyi post-hoc test to assess pairwise differences.

To ensure a fair and efficient comparison across model families, hyperparameters were tuned on a small validation subset using short grid searches around commonly adopted defaults. For tree-based models (RF, XGBoost), depth, learning rate, and number of estimators were varied to balance bias–variance trade-offs; for LSTM variants, hidden size, learning rate, and patience were adjusted to stabilize convergence; and for graph-based models (GNN, ST-GCN, GraphWaveNet), hidden dimension and weight-decay parameters were selected to prevent overfitting on the relatively small graph. The final configurations shown in Table 1 correspond to the best validation settings under this unified tuning protocol ('pat' denotes the patience parameter in early stopping). Table 1 also reports the approximate number of trainable parameters for the neural and graph-based models to improve transparency regarding model complexity. The validation subset corresponded to the last 10% of the training data, and grid searches were performed over narrow ranges (e.g., max_depth ∈ {4, 6, 8}, n_estimators ∈ {200, 500, 800}). Approximate trainable parameter counts are reported only for neural and graph-based models. Classical statistical models and tree-based ensemble methods are marked as N/A because their complexity is not typically expressed in trainable-parameter form. The search spaces were intentionally kept compact in order to maintain a fair and computationally tractable comparison across all model families and 84 station-level forecasting tasks. In particular, for graph-based and deep learning models, we prioritized commonly used parameter ranges from prior forecasting studies and adjusted them conservatively to reduce the risk of overfitting on a relatively small station graph. Therefore, the tuning strategy was designed to identify stable and

comparable configurations rather than to exhaustively optimize each model individually.

A single chronological split was adopted to preserve the temporal structure of the data and avoid information leakage that would arise from randomized resampling. To prevent information leakage, all spatio-temporal features, including neighbor-based lag variables, were constructed using only past observations within each input window. For each prediction step, both temporal lags and spatial features were derived exclusively from historical data (t−24 to t−1), ensuring that no future information from the test period was used during feature construction. Following common practice in station-level demand forecasting, all performance metrics were computed independently for each of the 84 stations and summarized using station-wise means, standard deviations, and non-parametric statistical tests. While repeated holdout or blocked cross-validation could offer additional robustness, such extensions are considered outside the scope of this study due to the substantial computational cost of re-training all models.

Since the forecasting task is defined at the station level, the models generate hourly demand predictions independently for each station in the network. This design enables operators to anticipate demand shortages or surpluses at specific stations and supports fine-grained operational decisions such as localized rebalancing. In addition, station-level forecasts can be aggregated to estimate demand patterns for broader urban regions. For example, stations located within the same district or mobility corridor can be grouped to derive regional demand estimates. Such aggregation could support higher-level planning tasks such as district-level resource allocation or infrastructure planning.

Although rolling-origin evaluation can provide a more robust assessment for non-stationary time series, it was not adopted in the present study due to the substantial computational burden of repeatedly re-training all model families across 84 station-level series, particularly for deep and graph-based architectures. Instead, a single chronological holdout split was used to preserve temporal order, prevent information leakage, and ensure a fair and consistent comparison across all models under the same experimental protocol. In addition, because the Kayseri bike-sharing system operates only during the active April–November period, the dataset does not include a full annual cycle; therefore, the dominant recurring patterns in the data are primarily daily and weekly rather than full-year seasonality. Nevertheless, rolling-origin or repeated temporal validation constitutes an important direction for future work.

3.2. Results

Table 2 summarizes the predictive performance of all models. Classical time-series baselines (ARIMA, Prophet) yielded the weakest results, while both neural and ensemble approaches showed substantial improvements. Incorporating spatial dependencies consistently enhanced accuracy: RF-ST achieved the best overall performance (2.67 RMSE, 1.67 MAE), followed by GNN and XG-ST.

The Friedman test confirmed significant performance differences among models ($\chi^2\_F = 418.205$, $p = 8.34\times10^{-83}$). The corresponding effect size, measured using Kendall's coefficient of concordance (W), was approximately 0.45, indicating a moderate to strong level of agreement among model rankings and suggesting that the observed performance differences are not only statistically significant but also practically meaningful. Post-hoc Nemenyi tests (see Fig. 3) showed that ST models significantly outperformed most temporal baselines, underscoring the value of spatial information in bike-sharing demand forecasting.

Table 2. Comparative model performance (RMSE/MAE mean ± std across stations)

Model	RMSE	MAE
ARIMA	4.2341 ± 3.7493	3.2484 ± 3.2945
Prophet	5.5153 ± 15.4679	3.9516 ± 11.8305
LSTM-U	3.0914 ± 1.8371	2.0313 ± 1.3653
LSTM-M	2.9128 ± 1.4797	1.9498 ± 1.0707
LSTM-ST	2.9422 ± 1.4663	1.8169 ± 0.9925
GNN	2.7627 ± 1.3934	1.7281 ± 0.9893
RF	2.9685 ± 1.6494	1.9119 ± 1.1612
RF-ST	2.6677 ± 1.2351	1.6707 ± 0.8664
XG	3.0321 ± 1.7356	1.9144 ± 1.1729
XG-ST	2.8248 ± 1.3853	1.7909 ± 0.9976
ST-GCN	3.6468 ± 2.1099	2.4420 ± 1.5910
GraphWaveNet	3.0251 ± 1.7875	2.0174 ± 1.3315

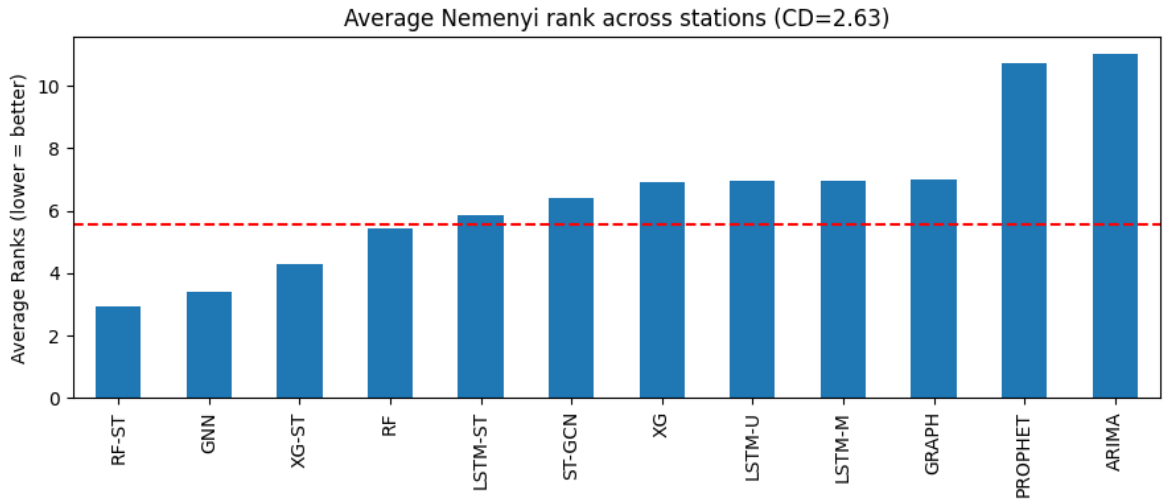

Fig. 3. Average Nemenyi rank across stations (lower = better)

4. Discussion

The results of this study demonstrate that ST modeling substantially improves forecasting accuracy for bike-sharing demand, consistent with recent literature. Compared to purely temporal baselines, models that explicitly integrate spatial information consistently achieved lower errors. In particular, the RF-ST emerged as the best-performing model (RMSE = 2.67, MAE = 1.67), closely followed by the GNN and XG-ST. Friedman and Nemenyi tests confirmed that these ST models significantly outperformed classical time-series approaches (ARIMA, Prophet) as well as standard machine learning and LSTM variants. The unusually high standard deviation observed for the Prophet model indicates substantial variability across stations. This behavior can be attributed to the sparse and highly heterogeneous nature of the dataset, where many stations exhibit prolonged zero-demand periods and occasional sharp peaks. Prophet's additive structure is less robust under such conditions, leading to unstable predictions for certain stations. Although a detailed outlier station analysis was not conducted, this variability highlights a limitation of Prophet in handling zero-inflated and highly imbalanced demand patterns.

These findings align with several contemporary studies:

- Zhao et al. [29] propose a Robust Spatio-Temporal Demand Prediction (RST) model which integrates POIs, weather, road networks, and a dynamic hierarchical structure, showing strong improvements in NY and Beijing datasets. Their work stresses the importance of spatial heterogeneity and differentiating station behavior.
- Kim et al. [30] demonstrate how region generation via soft clustering allows spatial units to share characteristics and improves predictive performance.
- Sardinha, Finamore & Henriques [31] argue that contextual features such as weather, calendar effects, and spatial awareness contribute significantly; their models show improvements, though not always statistically significant, when such context is included.
- Subramanian et al. [2] explicitly frame bike-sharing demand forecasting as a tool for enhancing sustainable transportation, framing AI-driven prediction models as enablers of more energy-efficient and environmentally responsible operations, and more efficient operations in smart cities.
- Recent studies published in the Journal of Construction Engineering, Management & Innovation (JCEMI) have also emphasized the role of data-driven and intelligent systems in improving operational efficiency and decision-making in infrastructure and facility management contexts. For instance, Suliman et al. [32] highlight how IoT-based approaches can support real-time data utilization and system optimization, which aligns with the present study's focus on leveraging data-driven forecasting for improved bike-sharing operations.

Compared with these studies, our results reinforce several key insights:

- *Spatial dependencies are crucial for sustainable operations.*

The superiority of RF-ST and GNN confirms that modeling inter-station relationships (e.g., via proximity or graph structures) is critical. Ignoring spatial correlations may result in less efficient operations, may potentially increase redistribution effort and associated costs, since operators compensate for poor predictions with more frequent truck-based rebalancing.

- *Combining features improves accuracy and resource efficiency.*

Multivariate LSTM, RF-ST, and XG-ST all outperform their temporal-only counterparts, confirming that temporal + spatial + demand history jointly enhances prediction. This improved performance is operationally meaningful: fewer forecast errors can facilitate more efficient scheduling, may help reduce wasted labor hours, and more balanced bike availability across neighborhoods—supporting equitable and socially sustainable mobility.

- *Graph-based learning is promising but architecture-dependent.*

While GNN performed competitively, ST-GCN and GraphWaveNet showed lower accuracy in our dataset. This can be attributed not to a lack of data volume, but to the relatively moderate size and sparsity of the station graph (84 nodes), which limited the expressive power of deep graph convolutions. The static proximity-based graph construction may have failed to capture dynamic correlations such as commuter flows or event-driven demand changes. Moreover, the high parameterization and sensitivity to hyperparameters of these models made them prone to underfitting under this moderate graph density. Although the dataset itself is large, the network topology was not complex enough for deep graph models to fully exploit ST dependencies.

- *Trade-offs matter.*

While GNNs achieve high accuracy, they are computationally demanding and highly dependent on graph design. One possible explanation for the strong performance of RF-ST is that tree-based ensemble methods tend to be relatively robust to moderate hyperparameter variation, whereas deep learning and graph-based architectures are often more sensitive to design and optimization choices. Therefore, the observed superiority of RF-ST in this study should be interpreted within the scope of the present experimental setup rather than as a universal claim of model superiority under exhaustive tuning. By contrast, tree-based ST extensions (RF-ST, XG-ST) offer competitive accuracy with lower complexity, making them attractive for operational deployment. The superior performance of RF-ST can be attributed to both data characteristics and model properties. The dataset is highly sparse and zero-inflated, conditions under which tree-based ensemble methods are known to perform robustly. In contrast, graph neural networks rely on learning smooth spatial representations, which may be less effective for irregular and discontinuous demand patterns. Furthermore, RF-ST incorporates spatial information through explicitly engineered features, while GNN models must learn these relationships implicitly, which can be challenging in relatively small and static graph structures. These factors collectively help explain the observed performance difference. The performance differences between LSTM-ST and graph-based models should be interpreted in light of how spatial information is incorporated into each modeling approach. While both model families include spatio-temporal components, they differ significantly in representation strategy. LSTM-ST integrates spatial information indirectly through feature augmentation, whereas graph-based models attempt to learn spatial dependencies explicitly through graph convolutions. The results suggest that spatial information is indeed valuable for improving forecasting accuracy, as evidenced by the strong performance of RF-ST and LSTM-ST compared to purely temporal models. However, the effectiveness of graph-based approaches appears to depend on the quality and expressiveness of the underlying graph structure. In the present dataset, where the network is relatively small and spatial interactions may be irregular, explicitly learned graph representations may not provide a substantial advantage over feature-based spatial integration. This indicates that, while spatial context plays a critical role in bike-sharing demand prediction, simpler and more direct methods of incorporating spatial information can sometimes be more effective than more complex graph-based architectures under certain data conditions.

- *Environmental and societal implications.*

While this study does not directly quantify environmental impacts such as emissions or fuel consumption, improved forecasting accuracy can help support more efficient rebalancing operations, which may contribute to sustainability objectives. Although operational logs such as rebalancing routes or fuel consumption were not available in our dataset, the sustainability implications of improved predictive accuracy can be contextualized using evidence from prior research. Fan et al. [23] showed that a 10–18% improvement in demand forecasting accuracy resulted in a 9–17% reduction in rebalancing route cost. Similarly, Ashqar et al. [22] demonstrated that reducing station-level prediction error by approximately 10% led to an 8–12% decrease in rebalancing movements in a large-scale system. Subramanian et al. [2] further frame accurate demand forecasting as a key enabler of sustainable bike-sharing operations by supporting more efficient resource allocation and system planning. Given that our spatio-temporal models achieve 10–35% lower error compared to temporal baselines, it is reasonable to expect that such improvements may be associated with proportional reductions in rebalancing mileage and associated emissions, consistent with these earlier findings. While these estimates are illustrative rather than empirical, they provide a literature-grounded indication of the potential sustainability benefits achievable through more accurate station-level forecasting. This perspective is also consistent with broader sustainability frameworks, where global policy initiatives such as the Paris Agreement emphasize reducing emissions through more efficient and data-driven system-level decision-making [33].

The improvements demonstrated by ST models have broader sustainability implications:

- Potential reduction in emissions due to fewer unnecessary truck trips.
- Potentially lower energy consumption in daily operations.
- Potential improvements in service reliability, supporting mode shift from cars to bicycles.
- Potentially more equitable distribution of bikes across districts, improving access for different user groups.

To further assess the robustness of graph-based models, we conducted a sensitivity analysis by varying the Gaussian kernel width (σ = 1.0, 1.5, 2.0 km) and neighborhood size (k = 3, 5, 7). The results indicate that GNN and GraphWaveNet models exhibit stable performance across different graph specifications, with only minor variations in error metrics. In contrast, the ST-GCN model showed higher sensitivity to these parameters, with more noticeable fluctuations in RMSE

and MAE values. As summarized in Table 3, the variation ranges remain relatively narrow for GNN and GraphWaveNet, confirming their robustness to adjacency parameter choices, whereas ST-GCN demonstrates a wider performance spread across configurations. It should be noted that the best-performing configurations identified in the sensitivity analysis were not used in the main experiments, as all models were evaluated under a unified setup to ensure fair comparison across model families. Importantly, these findings suggest that the relative underperformance of certain graph-based models cannot be attributed solely to the choice of adjacency parameters. Instead, the results point to dataset-related factors, such as sparse demand patterns and limited spatial dependencies, as more influential in determining model performance. This analysis was conducted for graph-based models only, as the Gaussian kernel width and neighborhood size are specific to the construction of spatial adjacency matrices and do not apply to non-graph models. Overall, the limited variation across configurations further supports the stability and robustness of the reported model comparisons.

From a practical perspective, more accurate demand forecasts can help support more efficient rebalancing decisions and may contribute to lower operational costs, improved user satisfaction, and environmental benefits through lower operational emissions. Thus, the superiority of ST models, particularly RF-ST and GNN, holds not only academic but also operational significance for bike-sharing operators.

Beyond methodological comparisons, the results of this study also have important practical implications for bike-sharing system operations. Accurate station-level demand forecasts can support more efficient rebalancing strategies by allowing operators to anticipate shortages or surpluses of bicycles at specific stations. Instead of relying on reactive redistribution, operators can schedule proactive rebalancing operations based on predicted demand patterns.

In addition, demand forecasting can assist in workforce and fleet planning for daily operations. By identifying stations that are likely to experience high demand during certain hours, operators can allocate redistribution vehicles and staff more efficiently.

Forecasting models can also support long-term planning decisions. For example, station-level demand predictions can inform infrastructure expansion, optimal station placement, and capacity planning in growing urban areas. In medium-sized systems such as Kayseri's network, these insights are particularly valuable for improving accessibility across different districts and ensuring balanced service coverage.

Finally, integrating predictive models into real-time decision-support systems could further enhance operational efficiency. Machine learning models such as RF-ST provide a favorable balance between accuracy and computational efficiency, making them suitable candidates for deployment in practical bike-sharing management platforms.

From an operational perspective, implementing demand forecasting models in real bike-sharing systems requires an automated data processing pipeline. Trip logs collected from docking stations can be aggregated into hourly demand series and periodically fed into forecasting models. In practice, models such as RF-ST or XG-ST can be retrained at regular intervals (e.g., weekly or monthly) as new data become available, ensuring that the models adapt to evolving usage patterns.

Once trained, the models can generate short-term demand forecasts in near real-time using the most recent observations and temporal features. These predictions can be integrated into operator dashboards or decision-support systems that assist in scheduling rebalancing operations, allocating redistribution vehicles, and monitoring station-level demand conditions. Due to their relatively low computational requirements compared to deep graph architectures, tree-based models such as RF-ST may be particularly suitable for operational deployment in medium-sized bike-sharing systems.

5. Conclusions

This study is primarily a forecasting-focused analysis, and sustainability considerations are included as potential application outcomes rather than directly measured impacts.

In this study, we evaluated a wide range of models for short-term bike-sharing demand forecasting. RF-ST achieved the lowest average error (RMSE = 2.67, MAE = 1.67), while GNN and XG-ST also performed competitively. Incorporating spatial information consistently improved accuracy across model families.

From a construction and infrastructure management perspective, improved station-level demand forecasting supports more efficient allocation of physical assets, workforce planning for rebalancing, and lifecycle optimization of shared mobility infrastructure.

Beyond predictive performance, these improvements may support more efficient and sustainable bike-sharing operations, as discussed in Section 4.

Table 3. Sensitivity analysis summary for graph-based models under varying σ and k

Model	RMSE Range	MAE Range	Best Configuration (σ, k)
ST-GCN	3.35 – 3.73	2.27 – 2.68	(1.0, 5)
GNN	2.75 – 2.78	1.72 – 1.74	(2.0, 7)
GraphWaveNet	3.02 – 3.05	2.01 – 2.07	(2.0, 7)

Overall, the study demonstrates that improving demand forecasting accuracy is not merely a technical task but a strategic pathway toward sustainable transportation management. ST models—particularly RF-ST and GNN—offer a practical and effective means for operators to optimize fleet usage, reduce operational burdens, and strengthen the long-term viability of bike-sharing systems within smart city contexts. These findings reinforce the role of data-driven forecasting as a core component of sustainable urban mobility planning.

5.1. Limitations & future work

This study has several limitations. First, the analysis relies on a single chronological train–test split applied to each station-level time series. While this approach preserves the temporal structure of the data, alternative evaluation strategies such as rolling-origin validation could provide additional robustness. Second, external contextual factors such as weather conditions, public events, or holiday effects were not incorporated into the current modeling framework. Additionally, the Kayseri bike-sharing system operates seasonally, and the dataset therefore covers only the active operational period between April and November, excluding winter months when the system is temporarily suspended. Furthermore, the graph structure used in graph-based models was static and based solely on geographic proximity, which may not fully capture dynamic inter-station relationships.

The sustainability implications discussed in the paper were inferred rather than measured directly, due to the lack of operational data such as fuel consumption logs or truck routing traces. As a consequence, the quantified environmental and operational impacts of improved forecasting accuracy remain approximate rather than empirical.

Another limitation concerns the generalizability of the findings. The analysis is based on a single-city dataset (Kayseri), which represents a medium-sized bike-sharing system with 84 stations. While the results provide valuable insights, they may not directly generalize to larger metropolitan systems with denser networks, more complex mobility patterns, and higher demand variability. In such settings, the relative performance of different model families—particularly graph-based approaches—may differ. Therefore, future studies should evaluate the proposed framework on larger and multi-city datasets to assess its robustness and scalability under diverse urban conditions.

Another limitation of the study is the absence of external contextual variables such as weather conditions, public events, or holiday indicators. These factors are known to influence bike-sharing demand, particularly in urban mobility systems where weather variability can significantly affect cycling behavior. Due to the limited availability of consistent contextual records aligned with the station-level demand logs, such variables were not incorporated into the current modeling framework.

Future research will incorporate repeated holdout evaluation, richer contextual variables, and dynamic graph structures to capture evolving station interactions. Additionally, integrating real operational data—such as rebalancing routes, vehicle emissions, or user accessibility metrics—will enable a more rigorous assessment of the sustainability benefits of ST forecasting. Future research could also integrate meteorological data, event schedules, or holiday calendars to further enhance forecasting accuracy and better capture demand fluctuations driven by external environmental or social factors.

Declarations

Conflict of Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This research received no external funding.

Author Contributions

M. Aygün Çakıroğlu: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization. S. Özdemir: Conceptualization, Methodology, Writing-Review & Editing, Supervision.

Acknowledgments

We gratefully acknowledge Kayseri Transportation Company for providing access to the dataset used in this research.

Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Ethics Committee Permission

Not applicable.

Use of generative AI and AI-assisted technologies

The authors used ChatGPT (OpenAI) only for minor language editing and grammar checking. The authors have reviewed all content and take full responsibility for the final manuscript.

References

[1] Lim H, Chung K, Lee S (2022) Probabilistic forecasting for demand of a bike-sharing service using a deep-learning approach. Sustainability 14(23): 1-18. https://doi.org/10.3390/su142315889.

[2] Subramanian M, Cho J, Sathishkumar V-E, Murugesan A, Chinnasamy R (2023) Enhancing sustainable transportation: AI-driven bike demand forecasting in smart cities. Sustainability 15(18): 1-19. https://doi.org/10.3390/su151813840.

[3] Yi S, Zhang L, Lu S, Liu Q (2023) Short-term demand prediction of shared bikes based on LSTM network. Electronics 12(6). https://doi.org/10.3390/electronics12061381.

[4] Li X, Xu Y, Zhang X, Shi W, Yang Y, Li Q (2022) Improving Short-term bike sharing demand forecast through an irregular convolutional neural network. Transportation Research Part C: Emerging Technologies 147: 1-16. https://doi.org/10.1016/j.trc.2022.103984.

[5] Gao C, Yong C (2022) Using machine learning methods to predict demand for bike sharing. Information and Communication Technologies in Tourism 2022, ENTER 2022. Springer: 282-296. https://doi.org/10.1007/978-3-030-94751-4_25.

[6] Butt M-A, Danjuma S, Bin Ilyas M-S, Butt U-M, Shahid M, Tariq I (2023) Demand prediction on bike sharing data using regression analysis approach. Journal of Innovative Computing and Emerging Technologies 3(1). https://doi.org/10.56536/jicet.v3i1.52.

[7] Kim T-Y, Park M-J, Shin J, Oh S (2021) Prediction of bike share demand by machine learning. International Journal of Business Analytics 9(1): 1-16. https://doi.org/10.4018/ijban.288513.

[8] Dastjerdi A-M, Morency C (2022) Bike-Sharing demand prediction at community level under covid-19 using deep learning. Sensors 22(3). https://doi.org/10.3390/s22031060.

[9] Mariati N-P-A-M, Setiawati L-P-E, Dewi N-L-P-S (2023) Inflation value forecasting post covid-19 in denpasar using ARIMA. International Journal of Application on Economics and Business 1(3). https://doi.org/10.24912/ijaeb.v1i3.1165-1169.

[10] Kaur J, Parmar K-S, Singh S (2023) Autoregressive models in environmental forecasting time series: a theoretical and application review. Environ Sci Pollut Res 30(8): 19617–19641. https://doi.org/10.1007/s11356-023-25148-9.

[11] Yan Y (2024) Explore the impact of initial data coherence on ARIMA model prediction based on python. AEMPS 71(1): 34–41. https://doi.org/10.54254/2754-1169/71/20241386.

[12] Mondal P, Shit L, Goswami A (2014) Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices. IJCSEA 4(2): 13–29. https://doi.org/10.5121/ijcsea.2014.4202.

[13] Cheng H, Li M, Zhang H (2024) Research on the urban bike-sharing usage based on ARIMA model. Transactions on Computer Science and Intelligent Systems Research, 5: 166–172. https://doi.org/10.62051/v10qqh77.

[14] Chen P, Hsieh H, Su K, Sigalingging X-K, Chen Y, Leu J (2019) Predicting station level demand in a bike-sharing system using recurrent neural networks. Iet Intelligent Transport Systems 14(6): 554–561. https://doi.org/10.1049/iet-its.2019.0007.

[15] Kolari J-W, Sanz I-P (2022) Forecasting bank capital ratios using the prophet model by facebook. Journal of Finance Issues, 20(3). https://doi.org/10.58886/jfi.v20i3.4941.

[16] Navratil M, Kolkova A (2019) Decomposition and forecasting time series in the business economy using prophet forecasting model. Central European Business Review 8(4): 26–39. https://doi.org/10.18267/j.cebr.221.

[17] Xie C, Wen H, Yang W, Cai J, Zhang P, Wu R, Li M, Huang S (2021) Trend analysis and forecast of daily reported incidence of hand, foot and mouth disease in Hubei, China by Prophet model. Sci Rep, 11(1): 1445. https://doi.org/10.1038/s41598-021-81100-2.

[18] Belikov D, Arshinov M, Belan B, Davydov D, Fofonov A, Sasakawa M, Machida T (2019) Analysis of the diurnal, weekly, and seasonal cycles and annual trends in atmospheric CO2 and CH4 at tower network in Siberia from 2005 to 2016. Atmosphere 10(11). https://doi.org/10.3390/atmos10110689.

[19] Santoso H, Hidayatullah S (2024) Random forest-based assessment of mangrove degradation utilizing NDVI feature extraction in spatio-temporal analysis. Jurnal Nasional Pendidikan Teknik Informatika. JANAPATI 13(1): 58–65. https://doi.org/10.23887/janapati.v13i1.71173.

[20] Dapogny A, Bailly K, Dubuisson S (2019) Dynamic pose-robust facial expression recognition by multi-view pairwise conditional random forests. IEEE Transactions on Affective Computing 10(2): 167–181. https://doi.org/10.1109/TAFFC.2017.2708106.

[21] Schimohr K, Doebler P, Scheiner J (2023) Prediction of bike-sharing trip counts: comparing parametric spatial regression models to a geographically weighted XGBoost algorithm. Geographical Analysis, 55(4): 651–684. https://doi.org/10.1111/gean.12354.

[22] Ashqar H-I, Elhenawy M, Rakha H-A, Almannaa M, House L (2022) Network and station-level bike-sharing system prediction: a San Francisco bay area case study. Journal of Intelligent Transportation Systems 26(5): 602–612. https://doi.org/10.1080/15472450.2021.1948412.

[23] Fan Y, Wang G, Lu X, Wang G (2019) Distributed forecasting and ant colony optimization for the bike-sharing rebalancing problem with unserved demands. PLOS ONE 14(12): e0226204. https://doi.org/10.1371/journal.pone.0226204.

[24] Ngo T-T-T, Pham H-T, Acosta J-G, Derrible S (2022) Predicting bike-sharing demand using random forest. Journal of Science and Transport Technology 2(2): 13–21. https://doi.org/10.58845/jstt.utt.2022.en.2.2.13-21.

[25] Chen Z, Wu H, O'Connor N-E, Liu M (2021) A comparative study of using spatial-temporal graph convolutional networks for predicting availability in bike sharing schemes. IEEE International Intelligent Transportation Systems Conference (ITSC), 1299–1305. https://doi.org/10.1109/ITSC48978.2021.9564831.

[26] Chen R, Yao H (2023) Hybrid graph models for traffic prediction. Applied Sciences 13(15): 8673. https://doi.org/10.3390/app13158673.

[27] Cao D, Wang Y, Duan J, Zhang C, Zhu X, Huang C, Tong Y, Xu B, Bai J, Tong J, Zhang Q (2021) Spectral temporal graph neural network for multivariate time-series forecasting. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), https://doi.org/10.48550/arXiv.2103.07719.

[28] Lin L, He Z, Peeta S (2018) Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach. Transportation Research Part C: Emerging Technologies 97: 258–276. https://doi.org/10.1016/j.trc.2018.10.011.

[29] Zhao Y, Du B, Luo M, Wen H (2025) Robust spatio-temporal demand prediction for bike-sharing systems with dynamic hierarchical structure. CCF Trans. Pervasive Comp. Interact. 7(2): 229–245. https://doi.org/10.1007/s42486-024-00167-8.

[30] Kim K, Zhang P (2025) Enhancing spatiotemporal demand prediction in transportation systems through region generation using soft clustering. Transportation Research Part C: Emerging Technologies 179: 105258. https://doi.org/10.1016/j.trc.2025.105258.

[31] Sardinha C, Finamore A-C, Henriques R (2020) Context-aware demand prediction in bike sharing systems: incorporating spatial, meteorological and calendrical context. In Proceedings of BuildSys. ACM: 10. https://doi.org/10.48550/arXiv.2105.01125.

[32] Suliman A, Hanson T, Wachowicz M (2023) A conceptual use-cases mapping framework for IoT-based smart building systems. Journal of Construction Engineering, Management & Innovation. 6(4):239-265. https://doi.org/10.31462/jcemi.2023.04239265.

[33] Fidan F-S, Aydogan S, Akay D (2024) Investigating the carbon border adjustment mechanism transition process with linguistic summarization method: A situational analysis of exporting countries. Advanced Engineering Informatics. 61: 102528. https://doi.org/10.1016/j.aei.2024.102528.