Asia Pacific Science Press

Journal of Advances in Engineering and Technology
ISSN(O):3005-7507
ISSN(P):3005-7493
Vol. 2 No.4 (2025)

MADENet: Explainable AI-Driven Bike-Sharing Demand Forecasting for Sustainable Urban Mobility

Jianbo Huang*, Yanbin Zheng, Lihua Lan, Jia Chen

School of Computer Application, Guilin University of Technology, Guilin, 541000, China

Corresponding author: Jianbo Huang, lucassivan@163.com

Abstract: With rapid urbanization and increasing motorization, bike-sharing systems have emerged as sustainable solutions for urban “last-mile” connectivity. However, existing demand forecasting approaches face a critical trade-off between predictive accuracy and operational interpretability, limiting their practical deployment in municipal deci-sion-making contexts where both reliable predictions and transparent insights are essen-tial. The study proposes MADENet, a novel neural architecture that systematically ad-dresses this accuracy-interpretability challenge. The framework integrates three key inno-vations: multi-head attention mechanisms to dynamically capture cross-regional demand dependencies and temporal periodicity patterns; adaptive dropout with early-stopping regularization to mitigate overfitting in high-dimensional spatio-temporal scenarios; and multilayer perceptron components to model complex nonlinear interactions between het-erogeneous external factors and urban mobility patterns.Experimental evaluation demon-strates MADENet’s superior performance, achieving 95.1% prediction accuracy (R^2=0.9515, MAE=0.2320) and outperforming 15 baseline algorithms with MAE improvements rang-ing from 7.7% to 70% across different algorithmic paradigms. Embedded SHAP and LIME explainable AI frameworks systematically identify hour-of-day, temperature, and humid-ity as dominant spatio-temporal drivers while quantifying their nonlinear interactions with demand patterns.These innovations provide transparent operational protocols for station layout optimization, dynamic fleet rebalancing, and evidence-based policy formu-lation, ultimately advancing data-driven governance of sustainable urban mobility sys-tems through actionable insights that bridge algorithmic predictions with practical urban planning requirements.

Keywords: Bike-Sharing; Explainable AI; Smart City; Sustainable Transportation

Published: Oct 15, 2025

DOI: https://doi.org/10.62177/jaet.v2i4.662

1.Introduction

The rapid urbanization observed in recent decades has intensified challenges such as traffic congestion, environmental degradation, and strained public transportation systems [1]. A critical contributor to these issues is the overreliance on private vehicles, which ac-count for approximately 11% of global carbon dioxide emissions [2] and impose severe economic burdens—for instance, traffic congestion alone costs China an estimated USD 35 billion annually [3]. In response to these dual crises of sustainability and urban effi-ciency, bike-sharing systems have emerged as a transformative mobility solution. Since their inception, these programs have experienced explosive global growth: active bike-sharing fleets surged from 700,000 to 2 million in the United States between 2013 and 2016, with over 2,000 systems now operating nearly 10 million bicycles worldwide [4]. By providing affordable, eco-friendly "last-mile" connectivity, bike-sharing networks reduce reliance on private cars while alleviating congestion and emissions.

Accurate demand forecasting and bicycle allocation remain critical bottlenecks for optimizing bike-sharing system efficiency. While traditional time-series models such as ARIMA and linear regression [5-10] often fail to capture intricate spatio-temporal de-pendencies and heterogeneous external factors, recent deep learning approaches have improved predictive performance through temporal dependency modeling [11-12]. How-ever, these solutions face a fundamental accuracy-interpretability trade-off: high-dimensional meteorological and urban infrastructure data exacerbate overfitting risks in sparse datasets, while the inherent opacity of neural networks undermines stake-holders' trust in critical urban planning decisions where interpretability is essen-tial—such as justifying infrastructure investments or optimizing fleet rebalancing strate-gies [13-15].

Current literature reveals significant methodological gaps in bike-sharing demand forecasting, where existing research typically addresses predictive accuracy and model interpretability as separate processes, resulting in suboptimal frameworks for urban mo-bility management. Contemporary deep learning approaches predominantly focus on maximizing prediction performance through increasingly complex architectures, while traditional interpretable models sacrifice accuracy for transparency, creating a funda-mental dichotomy that fails to meet the dual requirements of municipal decision-making contexts. The unique challenges in bike-sharing demand prediction—dynamic spa-tio-temporal dependencies, complex me-teorological-mobility interactions, and stringent requirements for both operational accuracy and transparent decision-making sup-port—demand specialized methodological considerations that contemporary approaches inadequately address.

To address these limitations, this investigation introduces MADENet, a novel neural architecture synergistically combining multi-head attention mechanisms with adaptive dropout regularization and early-stopping protocols for enhanced bike-sharing demand forecasting. The framework systematically addresses the accuracy-interpretability trade-off by embedding explain-able AI techniques directly into the prediction pipeline, enabling efficient capture of cross-regional demand correlations while mitigating overfit-ting without compromising computational efficiency.

The principal contributions encompass:

• Novel Architecture Design: Introduction of an innovative neural framework integrat-ing multi-head attention mechanisms that dynamically capture cross-regional de-mand dependencies and temporal periodicity patterns, coupled with multilayer per-ceptron components and adaptive regularization strategies specifically tailored for spatio-temporal bike-sharing data;

• Superior Predictive Performance: Achieved exceptional forecasting accuracy (95.1% accuracy, R^2=0.9515, MAE=0.2320) outperforming 15 baseline algorithms with MAE improvements ranging from 7.7% to 70% across different algorithmic paradigms in-cluding deep learning, ensemble methods, and traditional statistical approaches;

• Comprehensive Explainability Integration: Embedded SHAP and LIME interpretabil-ity frameworks ensuring operational transparency for municipal decision-making applications, systematically identifying key spatio-temporal drivers including hour-of-day, temperature, and humidity while quantifying their nonlinear interac-tions with urban mobility patterns.

The remainder of this paper is organized as follows: Section 2 reviews existing methodologies in bike-sharing demand prediction and highlights unresolved challenges. Section 3 presents preliminary work, including data preprocessing and exploratory visu-alization analysis of temporal and environmental factors influencing bike-sharing de-mand. Section 4 details the MADENet architecture, covering its attention mechanisms and regularization strategies. Section 5 presents experimental results and comparative anal-yses, exploring interpretability outcomes via SHAP and LIME to connect algorithmic be-havior with practical urban mobility strategies. Finally, Section 6 concludes by discussing implications for sustainable smart city governance and outlines future research directions.

2.Related Works

The rapid growth of bike-sharing systems as sustainable urban mobility solutions has intensified the need for accurate demand forecasting. Data-driven approaches, partic-ularly machine learning, have propelled this field forward, yet existing frameworks still struggle to capture dynamic spatio-temporal dependencies while maintaining operational interpretability for practical deployment [5,13,30].

Early research predominantly relied on statistical methods such as time series analy-sis and linear regression [6–10]. Although these approaches provided foundational in-sights, they often failed to capture nonlinear interactions arising from weather fluctua-tions, special events, and complex urban dynamics. Studies that integrated seasonal and weather factors [5–6,16] established valuable baselines but exhibited limited adaptability to real-world volatility. Sathishkumar et al. [7] incorporated weather and usage data into a Gradient Boosting Machine to predict hourly bike-sharing demand, yet its robustness to rare events remained unclear.

Machine learning models marked a paradigm shift by addressing nonlinear effects and temporal interactions. Random Forest and Gradient Boosting Machines (GBMs) pro-duced promising results [28-29], while Schnieder [18] demonstrated that temperature, distance, wind, and elevation accounted for 21–27% of potential e-bike usage. Hu et al. [19] employed a grid search-optimized XGBoost model for Washington rental data, but it did not adequately handle abrupt real-time demand shifts. Lee et al. [14] illustrated that in-corporating air pollution, traffic, and socio-economic variables can enhance predictive performance, reinforcing the potential of diverse data sources.

Deep learning approaches have attempted to bridge gaps by jointly modeling spatial and temporal complexities. LSTMs outperformed GBMs for weekend demand [10], and CNN-based models introduced spatial awareness through demand heatmaps [11]. How-ever, rigid grid structures often clashed with organic urban layouts. Li et al. [17] intro-duced STG2Vec, an attention-based graph embedding model, to learn heterogeneous spa-tio-temporal patterns for improved demand prediction. Li et al. [20] developed a Spa-tial-Temporal Memory Network (STMN) to capture short-term spatio-temporal patterns more effectively.

Recent hybrid approaches have shown promising results in addressing specific op-erational challenges. Yu et al. [30] integrated SARIMA with LSTM to predict bicycle flows around metro stations, while Wang et al. [38] proposed a model-data dual-driven ap-proach combining SARIMA and extended Long Short-Term Memory (xLSTM) networks, achieving high R-squared values (0.9928-0.9535) with 8% improvement over conventional LSTM. However, dual-component fusion introduces computational complexity and po-tential calibration challenges when adapting to different metropolitan contexts.

Graph Neural Networks (GNNs) emerged as a state-of-the-art paradigm by repre-senting stations as graph nodes [12]. However, their static graph architectures struggled to adapt to rapid demand fluctuations due to weather changes. Liang et al. [22] proposed a Domain-Adversarial Multi-Relational Graph Neural Network (DA-MRGNN) that lever-ages multimodal transport data, mitigating negative transfer between modes.

Recent advances have focused on improving GNN architectures for bike-sharing ap-plications. Behroozi et al. [36] proposed a gate graph convolutional neural network inte-grating trajectory, weather, and access data, though the framework may struggle with computational scalability due to dynamic graph topology changes. Qian et al. [37] devel-oped CGA-STNet for dockless bike-sharing demand prediction, integrating mul-ti-dimensional spatial features and time periodicity through Fourier transforms, achieving 16.3% MSE reduction over benchmark models but with limitations in handling long-er-term seasonal variations. Xiang et al. [41] combined dynamic time warping with spa-tio-temporal graph attention networks, using data-driven adjacency matrices and mul-ti-scale temporal features, though computational complexity may limit scalability.

Recent research has explored sophisticated frameworks addressing both prediction accuracy and operational challenges. Guo et al. [39] developed an XGBoost-based three-stage prediction approach that addresses the gap between observed bike pickup/drop-off records and true user demand by incorporating unsatisfied demand from empty or full stations. While demonstrating superior performance using Citi Bike data from New York, the framework's computational complexity and dependency on historical patterns may limit real-time implementation and transferability across different systems.

Beyond predictive modeling, recent research has explored optimization-driven ap-proaches. Shi et al. [40] introduced a generative-model-informed reinforcement learning approach for long-term inventory management in hybrid bike-sharing systems, utilizing a recurrent-attentive neural process (RANP) for demand prediction and a cooperative two-agent MDP framework for bike-ebike allocation optimization. While demonstrating superior performance, the framework's complexity may pose challenges for real-time de-ployment. Giner Fabregat et al. [42] developed an intelligent optimization framework for Barcelona's Bicing system, combining clustering analysis and machine learning-based demand prediction with optimization algorithms for efficient rebalancing strategies.

Despite noteworthy gains in accuracy, hybrid and deep learning models often lack explainability, limiting their usefulness in municipal decision-making contexts [25–26]. The "black-box" nature of these complex models raised concerns regarding transparency and stakeholder trust. Explainable AI (XAI) methodologies have thus emerged to balance predictive power with interpretability [27], offering mechanisms to unveil the decision logic of otherwise opaque models [25,26]. However, existing XAI applications in bike-sharing prediction rarely undergo thorough validation against empirical urban mo-bility patterns, underscoring the need for interpretable frameworks that align with re-al-world operational constraints [35].

Despite significant advances in bike-sharing demand forecasting, critical gaps re-main in achieving the optimal balance between predictive accuracy and operational in-terpretability. Current deep learning approaches are predominantly adapted from general time-series prediction without adequate consideration of urban mobility-specific require-ments, such as real-time deployment constraints, cross-regional transferability, and mu-nicipal decision-making transparency. The relationship between model architectural complexity and explanation reliability represents a fundamental but understudied aspect, potentially leading to urban planning decisions based on explanations derived from models that may not adequately capture the nuanced spatio-temporal dynamics of bike-sharing systems. Most studies focus on algorithmic performance metrics without sufficient consideration of practical integration into municipal planning workflows, where computational efficiency, regulatory transparency, and stakeholder interpretability significantly influence real-world applicability.

Our work addresses these gaps by introducing MADENet, a novel neural architecture that integrates multi-head attention mechanisms with adaptive dropout regularization, specifically designed for bike-sharing demand forecasting. By embedding explainable AI frameworks (SHAP and LIME) directly into the predictive pipeline, we provide both theo-retical foundations and empirical validation of the approach's effectiveness in delivering transparent, actionable insights for sustainable urban mobility management.

3.Preliminary

3.1 Data Overview

As shown in Table 1, this study selects data provided by Capital Bicycle, which in-cludes information such as date, season, and weather conditions.

Table1: Data Field Description.

Field Name	Description
year	Year of observation
month	Month of observation
day	Day of observation
hour	Hour of observation
weekday	Day of the week
season	Season of observation
holiday	Whether it is a holiday
workingday	Whether it is a working day
weather	Weather condition
temp	Temperature in Celsius
atemp	Body temperature in Celsius
humidity	Humidity of the environment
windspeed	Wind speed in m/s
casual	Number of casual users
registered	Number of registered users
count	Total number of bike rentals
day_type	Type of the day (weekday/weekend)

3.2 Data processing

Data preprocessing was a critical step to ensure the dataset was clean, well-structured, and suitable for accurate prediction. The initial dataset included 10,886 records for training and 6,493 records for testing, with 12 and 9 features, respectively. First, missing data were examined and no missing values were identified in either the training or test datasets.

Outliers were then identified through the use of statistical methods. Variables like humidity, wind speed, and the target variable count exhibited irregularities. Specifically, wind speed contained numerous zero values, which were treated as potential missing data. To handle such cases, outlier removal was performed based on the criterion of values deviating more than 3 standard deviations from the mean.

For outlier correction in the atemp (apparent temperature), a linear regression model was employed to predict the correct values based on temperature, as defined by the formula:

$$atemp_{predicted} = \beta_0 + \beta_1 \cdot temperature \tag{1}$$

where β_0 and β_1 are the coefficients derived from a linear regression model.

In the case of the target variable count, extreme outliers were eliminated using the following criteria:

$$|count - \mu_{count}| > 3\sigma_{count} \tag{2}$$

where μ_{count} is the mean of the count variable and σ_{count} is its standard deviation. Data points falling outside of this range were removed from the dataset.

A logarithmic transformation was then applied to the count variable to stabilize variance, as this transformation is effective in reducing skewness and making the data more normally distributed:

$$count_{log} = log(count + 1) \tag{3}$$

Temporal features were also extracted from the timestamp column, including the day of the week, month, and hour, which were treated as categorical variables to capture time-dependent patterns.

These preprocessing operations ensured that the data were ready for model training, mitigating issues like outliers and skewed distributions while enriching the dataset with additional time-related features for improved predictive performance.

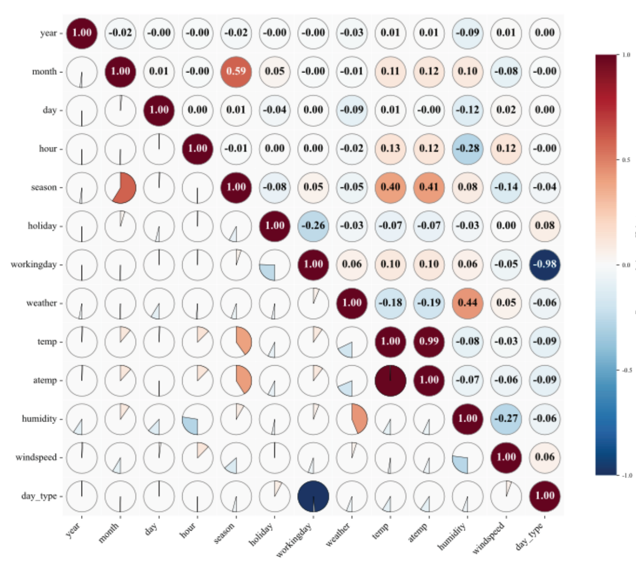
3.3 Correlation analysis

The data analysis aimed to identify key factors influencing shared bicycle demand, leveraging correlation analysis and trend evaluations. The correlation matrix (Figure 1) was calculated to explore the relationships between different features and the target variable, count. The correlation coefficient (r) between two variables, x and y, was computed using the formula:

$$\mathrm{r} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \, (y_i - \bar{y})^2}} \tag{4}$$

Where: x_i and y_i are the individual data points for variables x and y, $\bar{x}$ and $\bar{y}$ are the mean values of x and y, respectively, the summation is taken over all data points in the dataset.

Figure 1. Correlation Matrix of Bike-sharing Dataset Features.



As shown in Figure 1, this correlation matrix illustrates the strength and direction of linear relationships between bike-sharing system variables. The analysis reveals several notable patterns among temporal, environmental, and operational factors. Temperature (temp) and apparent temperature (atemp) demonstrate a strong positive correlation (0.99), indicating their close relationship in weather conditions. Environmental variables show moderate correlations, with temperature exhibiting positive associations with seasonal patterns (0.40-0.44). Temporal variables display expected relationships, such as the moderate correlation between working days and day type (-0.98), reflecting the distinction between weekdays and weekends. Holiday patterns show negative correlations with working days (-0.26), confirming the inverse relationship between holidays and regular work schedules. Wind speed and humidity demonstrate relatively weak correlations with most other variables, suggesting their more independent influence on system usage patterns.

3.4 Influence of Time Conditions on Rental Demand

Figure 2. Monthly Bike Rental Trends.

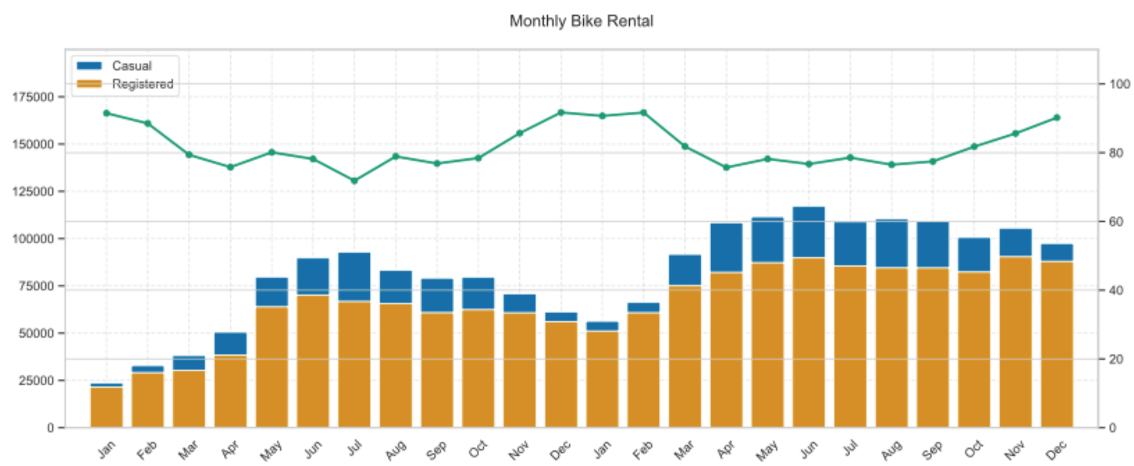


Figure 2 demonstrated a general upward trajectory in shared bicycle usage from 2 years, with notable seasonal fluctuations. Demand peaked during the summer and autumn months, with a significant decline observed in spring and winter. Registered users consistently represented over 75% of the total rentals each month, although their proportion slightly decreased in the warmer months, indicating a shift in user behavior. Non-registered users exhibited a preference for rentals during hotter months, further supporting the notion that shared bicycles are often used for short-term, seasonal needs.

Figure 3. Daily Bike Rental Distribution by Weekday and Day Type.

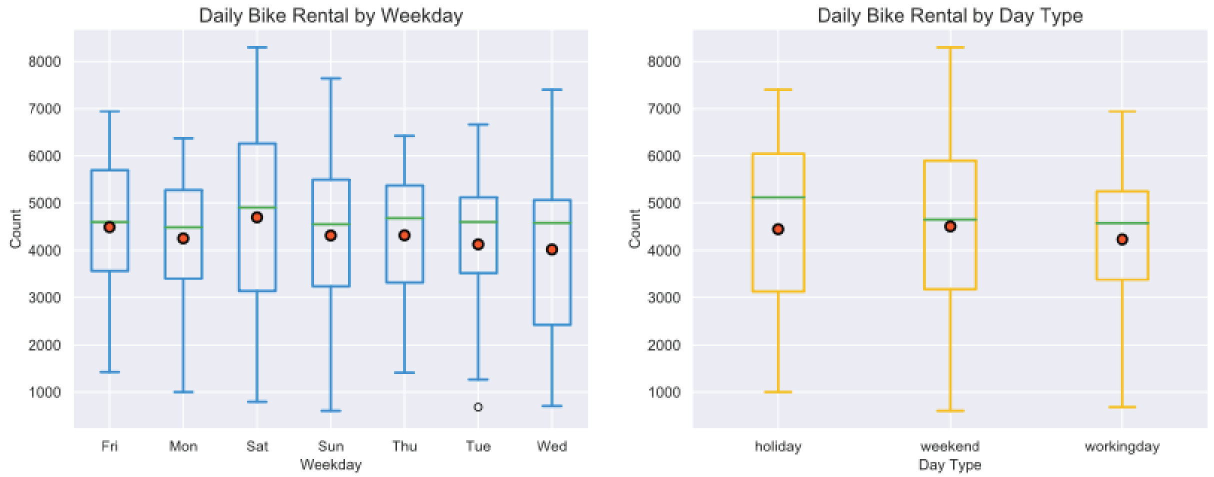


Figure 3 revealed that Saturdays experienced the highest rental volumes, likely driven by leisure activities, while Sundays saw slightly reduced demand, possibly due to lower mobility. Rentals on weekdays were more consistent, with Fridays exhibiting the lowest demand, albeit still higher than other weekdays. The impact of holidays on demand was more variable, with seasonal effects influencing demand peaks, particularly lower usage in the spring and winter months.

Figure 4. Hourly Bike Rental Patterns.

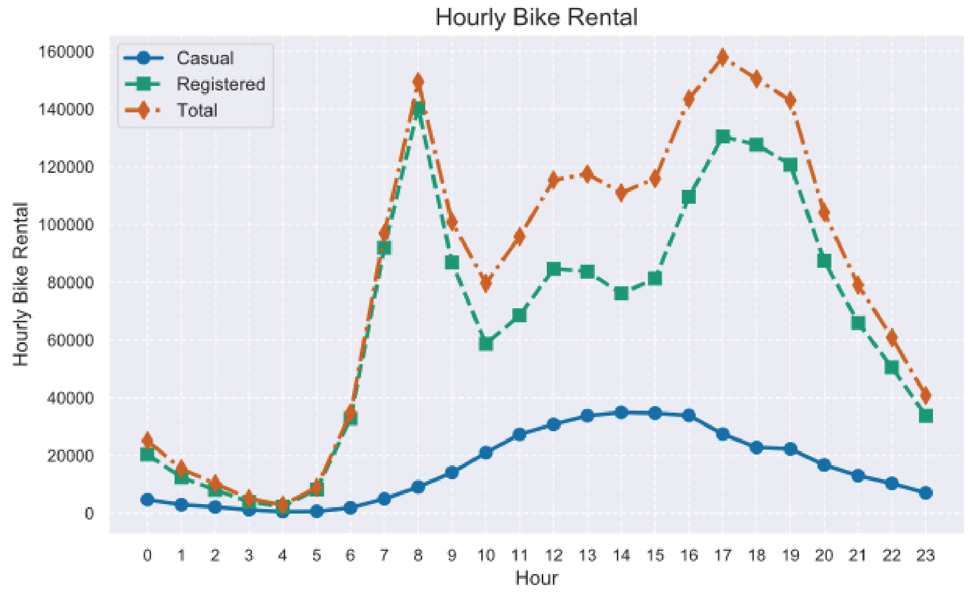
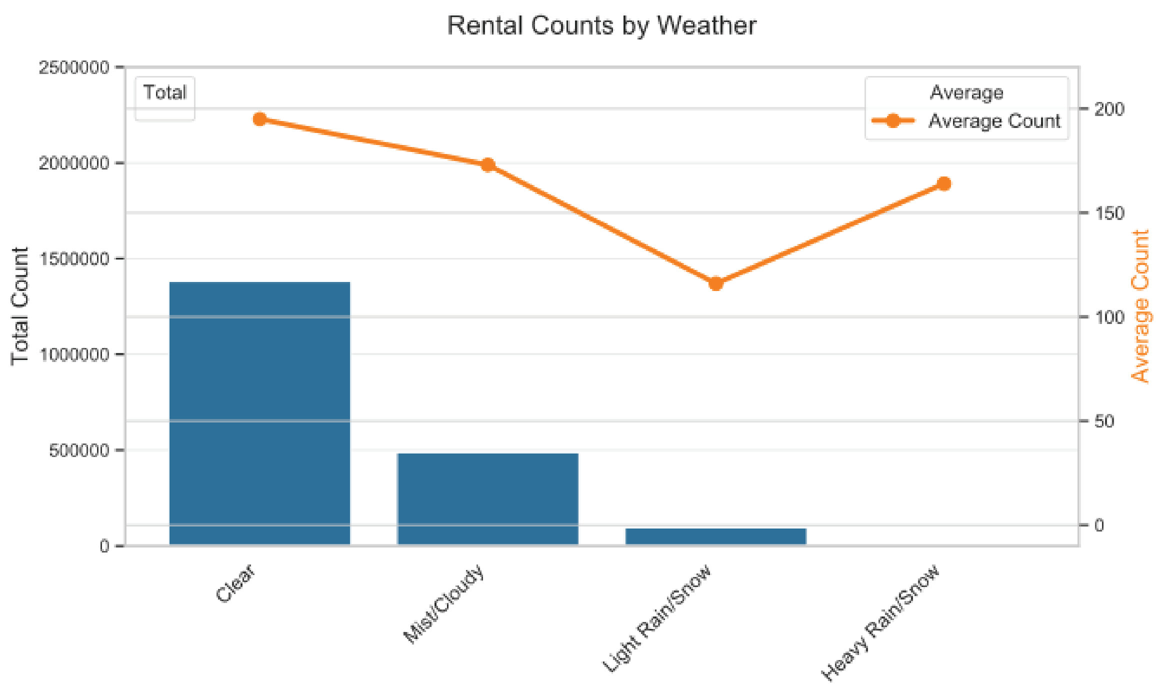


Figure 4 showed that registered users exhibited relatively stable demand throughout the day, while casual users demonstrated peak demand during morning and evening rush hours (7:00–8:00 and 17:00–18:00), reflecting their use of bicycles for commuting. These findings highlight the importance of considering time-related factors such as time of day, workdays, weekends, and holidays when planning for bicycle distribution and management.

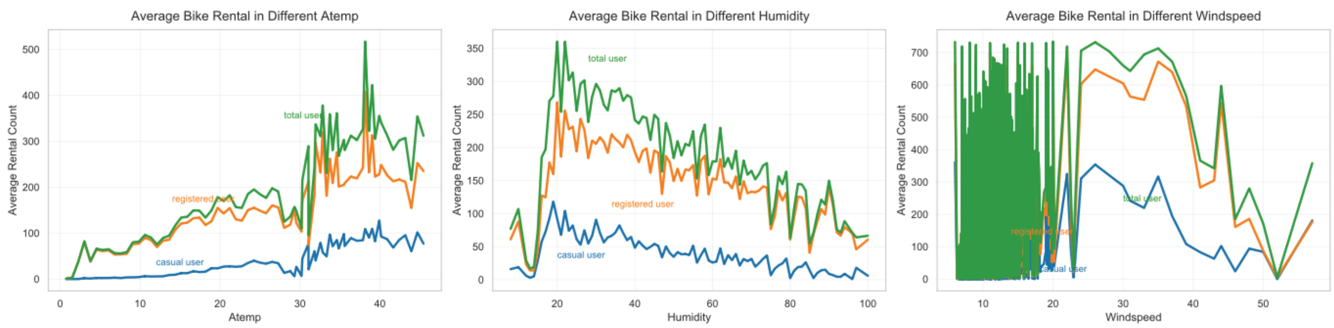
3.5 Influence of Weather on Rental Demand

Figure 5. Impact of Weather Conditions on Bike Rental Demand.



Weather conditions emerged as a significant factor influencing rental demand. The analysis of weather conditions (Figure 5) demonstrated a clear correlation between weather type and rental volumes. Interestingly, despite initial expectations, demand remained relatively high even during extreme weather events, such as heavy rain and snow. However, the distribution of weather-related data indicated that severe weather conditions, such as storms, led to a dramatic decline in rentals, with total demand during such events being just 1/9000th of the demand during clear weather, confirming a strong negative correlation between weather severity and demand.

Figure 6. Environmental Factor Influence on Average Bike Rental.



Environmental factors, including temperature, humidity, and windspeed (Figure 6), significantly influenced rental demand. Demand was lowest around 4°C, increasing with temperature up to 36°C, beyond which it declined, indicating a preference for moderate weather. Humidity showed a negative correlation, with demand highest at 20% humidity, which is more favorable for outdoor activities. Windy speed also impacted demand. Rentals were stable at wind speeds between 10 and 40 km/h, but sharply declined above 40 km/h. Interestingly, a brief rebound in demand was observed during high wind speeds around 17:00, likely due to commuter patterns during the evening rush hour. This suggests that while high wind speeds generally reduce rentals, commuter demand during peak hours can still drive usage.

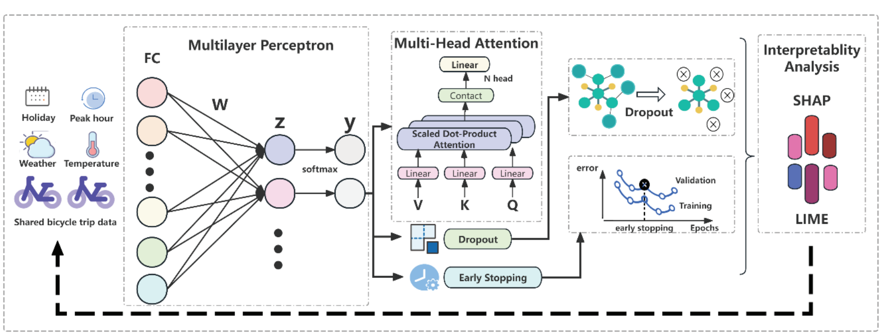
These findings demonstrate the significant impact of both temporal and environmental conditions on shared bicycle rentals, with time-related factors revealing distinct hourly, weekly, and seasonal patterns, while environmental variables exhibit complex nonlinear relationships including temperature optima (4°C-36°C), humidity thresholds, and wind speed effects that interact dynamically across different temporal contexts. These empirical insights directly inform our MADENet architecture design, where the observed temporal periodicity patterns necessitate attention mechanisms for dynamic feature weighting, the nonlinear environmental relationships require multilayer perceptron components to capture complex meteorological-demand interactions, and the identified threshold effects guide our regularization strategy to ensure systematic integration of these multifaceted spatio-temporal and environmental factors without overfitting. The following methodology section details how these empirical findings are systematically integrated into our proposed framework.

4.Methodology

Chapter 4 introduces the MADENet model for bicycle-sharing demand forecasting, designed to address random user behavior and dynamic external factor that leads to supply–demand imbalance. MADENet integrates a multi-head attention mechanism with adaptive dropout regularization and early-stopping within a multilayer perceptron framework to strengthen key spatio-temporal feature representations while preventing overfitting. The architecture consists of five sequential components—Input, Attention, Dropout, MLP and Output layers—where the attention module assigns probabilistic weights to critical regional and temporal signals, the adaptive dropout adapts its rate during training, and the MLP captures complex nonlinear dependencies.

Based on the described architecture and functional components, the proposed MADENet framework is structurally illustrated in Figure 7, which visually elucidates the synergistic integration of its multi-head attention mechanisms, adaptive dropout layers, and hierarchical feature processing pathways.

Figure 7. MADENet Framework Architecture.



4.1 Multilayer Perceptron

The Multilayer Perceptron is a fundamental component of MADENet. It consists of an input layer, one or more hidden layers, and an output layer. Each neuron in a layer is connected to all neurons in the next layer:

The output of a neuron y_j in the l-th layer is calculated as:

$$y_j^{(l)} = f\left(\sum_{i=0}^{n_{l-1}} E\left[w_{ji}^{(l)} x_i^{(l-1)} + b_j^{(l)}\right]\right) \tag{5}$$

Where:

n_{l-1} is the number of neurons in the $(l-1)$-th layer,

$w_{ji}^{(l)}$ is the weight between the i -th neuron in the $(l-1)$-th layer and the j-th neuron in the l-th layer,

$x_i^{(l-1)}$ is the output of the i-th neuron in the $(l-1)$-th layer,

$b_j^{(l)}$ is the bias of the j-th neuron in the l-th layer,

and f is the activation function, such as the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$ or the ReLU function $f(x) = max(0, x)$,

$E[\cdot]$ signifies the expectation operator applied to the weighted input.

4.2 Multi-Head Attention Mechanism

The multi-head attention mechanism in MADENet is used to dynamically capture cross-regional demand dependencies and temporal periodicity patterns.

The scaled dot-product attention is calculated as:

$$Attention(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{6}$$

Where: Q is the query matrix, K is the key matrix, V is the value matrix, and d_k is the dimension of the keys.

The multi-head attention is composed of multiple parallel attention heads. The output of the multi-head attention is:

$$MultiHead(Q, K, V) = \text{Concat}(head_1, \cdots, head_h)W^O \tag{7}$$

Where: $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$, W_i^Q, W_i^K, W_i^V are the weight matrices for the i-th head, and W^O is the output weight matrix.

4.3 Dropout

Dropout is a regularization technique used in MADENet to prevent overfitting. During training, neurons in a layer are randomly "dropped out" with a probability p.

Let x be the input to a layer. The output y after applying dropout is:

$$y = \frac{r}{1-p} \odot x \tag{8}$$

Where: r is a binary mask vector of the same length as x, and each element r_i is randomly set to 0 with probability p and 1 with probability $1-p$.

4.4 Early Stopping

Early stopping is another technique to prevent overfitting. It monitors the performance of the model on a validation set during training.

Let $E_{val}(t)$ be the error on the validation set at the t-th training epoch. The training stops when the following condition is met:

$$E_{val}(t) > E_{val}(t-k) + \epsilon \tag{9}$$

Where: k is a predefined patience parameter, and ϵ is a small positive constant. This ensures that the model does not over-train on the training data and generalizes well to new data.

5.Experiment

5.1 Experimental Configuration and Setup

The experimental evaluation was conducted using the Capital Bicycle dataset comprising over 17,000 hourly demand records with comprehensive temporal and environmental features including weather conditions, temperature, humidity, and seasonal variables. The dataset was partitioned using an 8:2 stratified split for training and testing to ensure representative sampling of underlying demand patterns and variability.

MADENet was implemented using Python and TensorFlow framework, with training executed on a high-performance computing cluster using optimized parallel processing. Input data was formatted as 3D tensors (samples, 1, features) to accommodate the multi-head attention mechanism requirements. The model architecture employed 4 attention heads with key dimension of 32, multilayer perceptron structure of [512, 256, 128] neurons, and adaptive dropout rates (0.2-0.3). Training optimization utilized the Adam optimizer (learning rate 0.001) with mean squared error loss function, batch size of 64, and early stopping strategy (patience=20) monitoring validation loss to prevent overfitting. Hyperparameter optimization was conducted through systematic grid search across attention heads, hidden layer configurations, dropout rates, L2 regularization strengths (0.0, 0.001, 0.01), and batch normalization settings, with training limited to 100 epochs maximum.

5.2 Performance Metrics

Model performance was evaluated using the following metrics:

R^2 measures how well the model explains the variance in the target variable and indicates the proportion of the total variation that is captured by the model. It is defined as:

$$1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2} \tag{10}$$

MAE measures the average absolute difference between predicted and actual values, providing a straightforward interpretation of prediction accuracy. It is defined as:

$$\frac{\sum_{i=1}^{n}|y_i-\hat{y}_i|}{n} \tag{11}$$

MSE calculates the average of the squared differences between the predicted and actual values, offering a measure of how far predictions deviate from actual observations. It is defined as:

$$\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n} \tag{12}$$

RMSE provides the square root of MSE, offering an interpretable estimate of the average magnitude of prediction error in the same units as the target variable. It is defined as:

$$\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}} \tag{13}$$

MAPE expresses the prediction error as a percentage of the actual values, providing an intuitive and scale-independent measure of model accuracy. It is defined as:

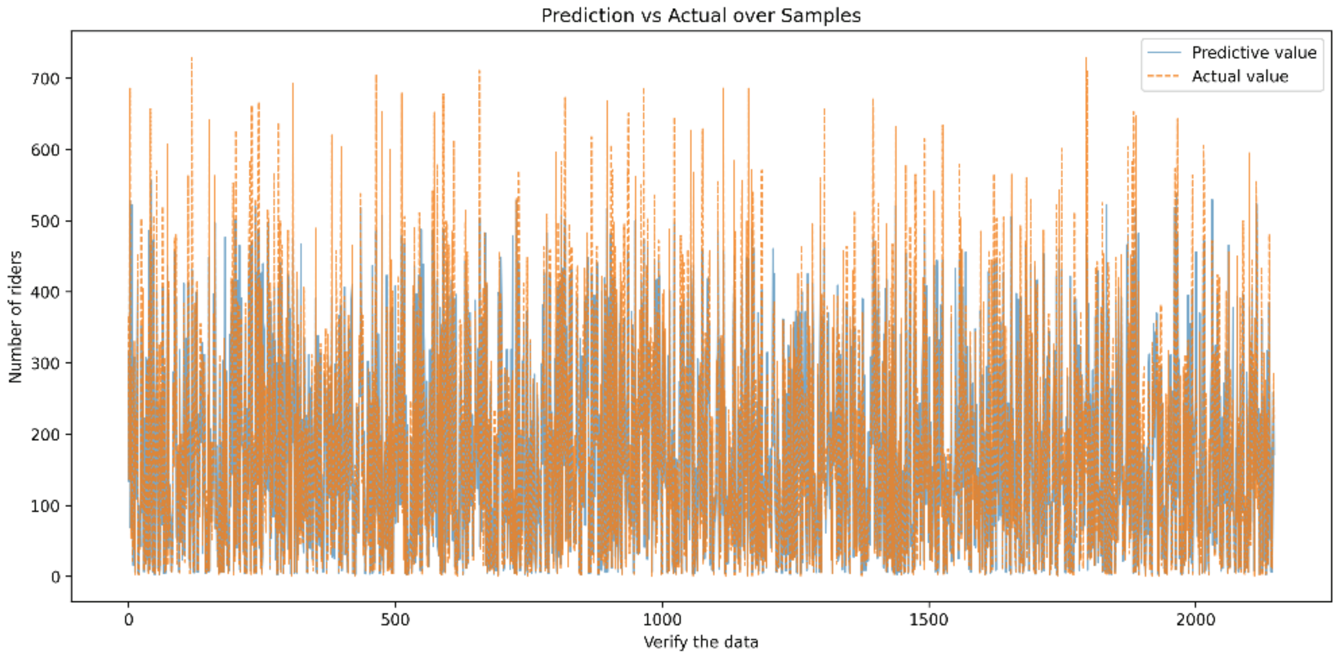
$$\frac{\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|}{n} \tag{14}$$

5.3 Comparative Analysis of Forecast Results

To comprehensively assess the performance of the MADENet model, Figure. 8 visualizes the predicted hourly bike-sharing demand alongside the actual recorded values across the test dataset. The continuous blue line represents the model's predicted values, while the orange dashed line denotes the ground truth. The graph clearly shows that MADENet captures both the short-term fluctuations and long-term periodic patterns in user demand with high fidelity.

Throughout more than 2,000 test samples, the predicted curves closely track the empirical values without significant lag or deviation, even in regions with high vari-ance and sharp peaks. This alignment indicates that the model not only learns overall trends but also adapts well to abrupt changes likely caused by external factors such as holidays, weather anomalies, or peak commuting hours.

Figure 8. Comparison of predicted and true values.



As shown in Table 2, to evaluate the predictive performance and generalization ability of the proposed MADENet model, a comprehensive comparative analysis was conducted against a variety of baseline models, including traditional machine learning methods and advanced deep learning architectures. These models included IrConv-LSTM [32], CNN-LSTM [33], Bi-LSTM [9], DeepAR [34], Decision Tree Regression [10], LSTM [12], RNN, AdaBoost Regression, K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Elastic Net, Bayesian Ridge, Ridge Regression, Linear Regression, and Lasso Regression.

Table 2. Model Evaluation Results.

Model	**R^2**	**MAE**	**MSE**	**RMSE**	**MAPE**
MADENet	0.9515	0.2320	0.1046	0.3234	7.3083
IrConv-LSTM	0.9363	0.2515	0.1263	0.3554	8.2922
CNN-LSTM	0.9205	0.2990	0.1600	0.4000	9.1285
Bi-LSTM	0.8281	0.4492	0.3437	0.5862	12.4637
DeepAR	0.9182	0.3028	0.1653	0.4066	9.4387
Decision Tree	0.9088	0.2933	0.1781	0.4220	8.8363
LSTM	0.8416	0.4416	0.3073	0.5544	12.7009
RNN	0.8046	0.4692	0.3924	0.6264	16.1000
AdaBoost	0.7829	0.5320	0.4301	0.6558	15.3739
KNN	0.7673	0.5050	0.4611	0.6790	16.4689
SVR	0.6367	0.6140	0.7198	0.8484	22.1008
Elastic Net	0.4742	0.7968	1.0267	1.0133	25.5929
Bayesian Ridge	0.4742	0.7959	1.0267	1.0133	25.5597
Ridge Regression	0.4740	0.7958	1.0271	1.0135	25.5543
Linear Regression	0.4739	0.7958	1.0273	1.0135	25.5526
Lasso Regression	0.4705	0.8014	1.0340	1.0169	25.7206

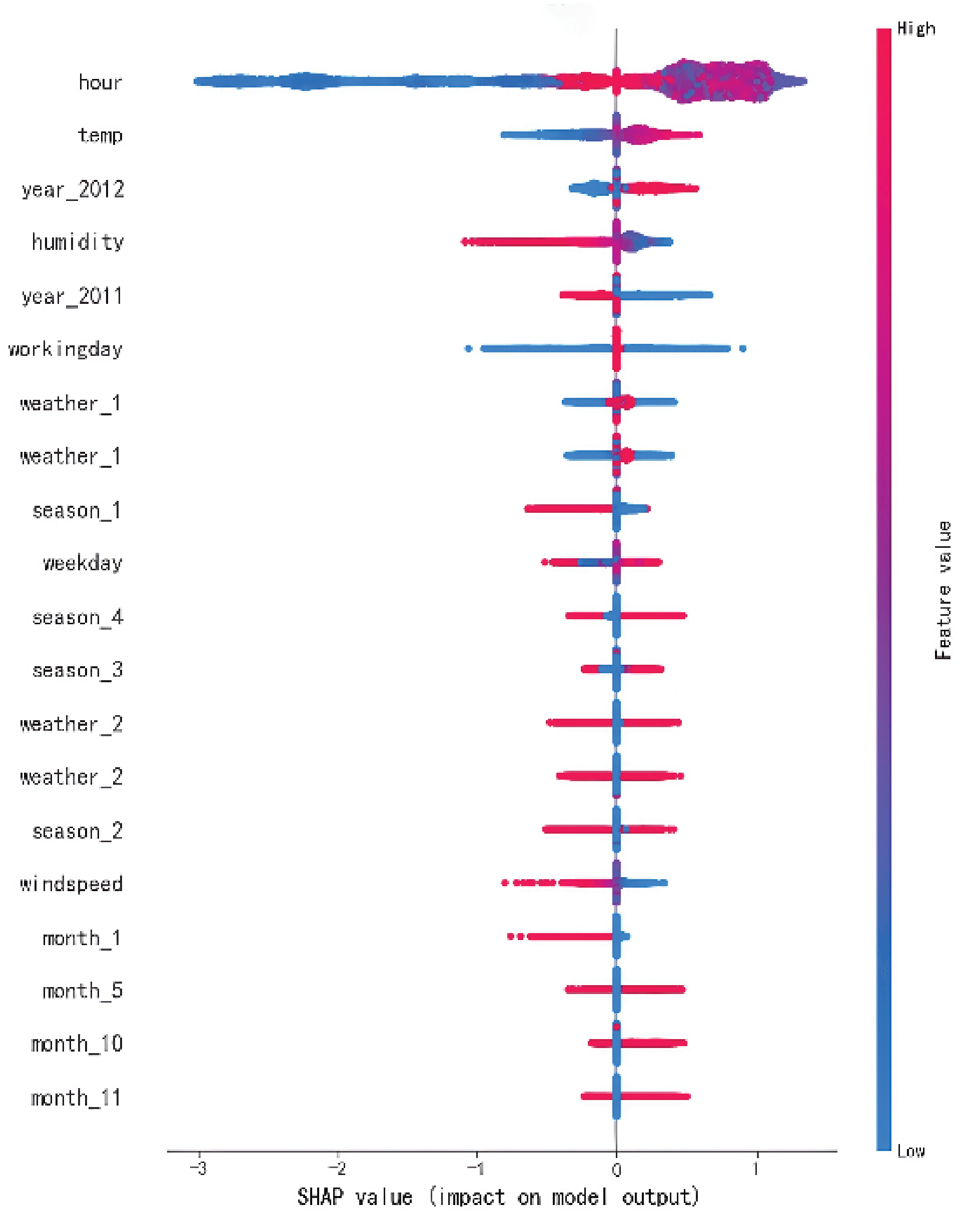
MADENet achieves superior performance across all evaluation metrics, demonstrating exceptional predictive accuracy with R^2=0.9515, MAE=0.2320, MSE=0.1046, RMSE=0.3234, and MAPE=7.31%. This comprehensive performance establishes MADENet as the leading method among all tested approaches.

Compared to advanced deep learning architectures, MADENet significantly outperforms CNN-LSTM with 22.4% MAE reduction and 34.6% MSE reduction, while achieving notable improvements of 7.7% in MAE over IrConv-LSTM. MADENet demonstrates substantial advantages over recurrent architectures, with 47.4% and 50.2% MAE improvements compared to Bi-LSTM and LSTM, respectively. Notably, MADENet surpasses the state-of-the-art probabilistic DeepAR model (R^2=0.9182) by 3.6% in variance explanation while achieving 23.4% improvement in MAE, demonstrating superior effectiveness in capturing complex spatio-temporal demand patterns.

Against traditional machine learning approaches, MADENet exhibits remarkable performance gains. Compared to Decision Tree regression, MADENet achieves 4.7% higher R^2 and 20.9% lower MAE. The superiority extends to ensemble methods, with MADENet outperforming AdaBoost by 56.4% in MAE reduction and achieving 21.8% improvement over KNN. When evaluated against linear regression methods, MADENet demonstrates exceptional advancement with over 100% improvement in R^2 (from 0.47 to 0.9515) and approximately 70% reduction in prediction errors across all metrics.

5.4 Factor Analysis Based on SHAP Values

Figure 9. SHAP Anaylsis of the importance of impact factors.

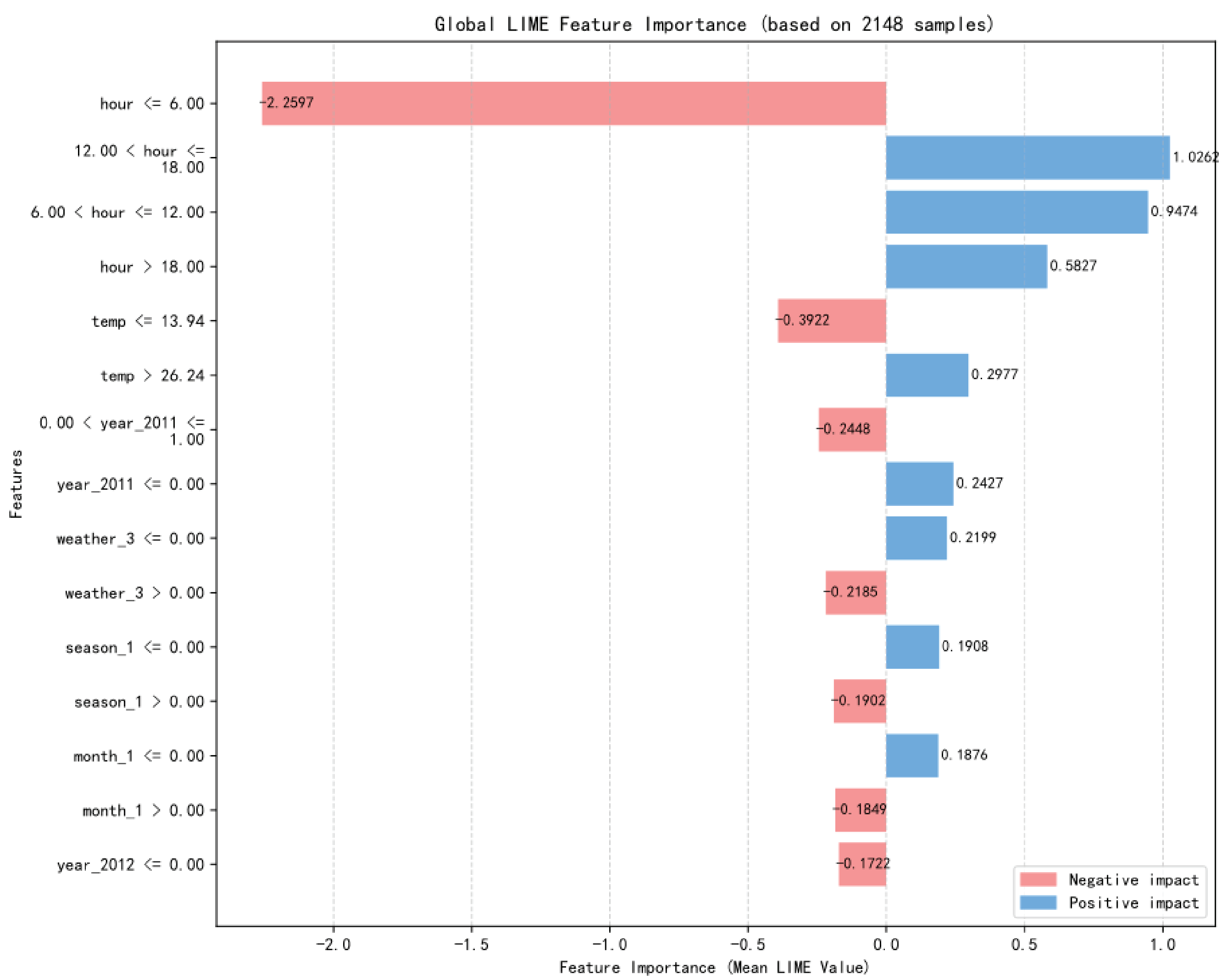


To enhance interpretability and understand MADENet's internal decision logic, SHAP values were employed to evaluate feature contributions to model outputs. Figure 9 presents the SHAP summary plot where each point represents a single SHAP value, with color gradients denoting feature values from low (blue) to high (red). The analysis reveals that temporal indicators, particularly hour-of-day variables, exert the greatest influence on model predictions, with the "hour" feature demonstrating both high density and wide spread, indicating its dominant role in shaping ridership forecasts and reflecting strong regularity in daily urban mobility patterns.

Climatic and seasonal features such as temperature and humidity exhibit substantial but nuanced impacts, with warmer temperatures generally corresponding to increased predicted demand while humidity demonstrates complex nonlinear effects. Calendar context variables including working days, weekdays, and seasons present moderate SHAP influence with values clustering near zero, suggesting their effects depend primarily on interactions with other features rather than exerting strong independent influence, while environmental indicators like wind speed show relatively limited marginal contributions in the current model configuration.

5.5 Factor Analysis Based on LIME Values

Figure 10. LIME Anayisis of the importance of impact factors.



LIME values were utilized to further investigate feature importance through global feature impact quantification. Figure 10 shows the global LIME feature importance plot where positive values represent positive contributions and negative values indicate negative impacts. The analysis confirms that time-of-day variables dominate the predictive power, with early morning hours exhibiting the highest positive impact, strongly suggesting commuter-driven demand patterns, while midday periods also demonstrate substantial positive influence, reinforcing the critical importance of specific temporal windows in demand forecasting.

Temperature emerges as another crucial factor with consistent positive LIME values, indicating that warmer conditions generally increase bike-sharing demand and aligning with observed weather-behavior relationships in urban cycling environments. Notably, negative impacts are observed for certain months and seasonal variables, suggesting weather-dependent demand variations, while some features display minimal contribution with LIME values near zero, indicating marginal influence and highlighting the model's ability to distinguish between critical and peripheral predictive factors.

6.Conclusions

This study successfully addresses the dual challenge of achieving high-precision and interpretable demand forecasting in urban bike-sharing systems through MADENet, a novel neural architecture combining multi-head attention, adaptive dropout, and early stopping mechanisms. The framework achieves 95.1% prediction accuracy while providing systematic transparency through integrated SHAP and LIME analysis, identifying hour-of-day, temperature, and humidity as dominant drivers. These interpretable insights empower urban planners with actionable guidance for station placement optimization, dynamic fleet rebalancing, and evidence-based policy formulation, ultimately supporting sustainable urban mobility through more effective resource allocation and environmental responsiveness.

Despite these advances, several limitations constrain the model's broader applicability. The evaluation on a single operator's dataset (Capital Bicycle) limits generalization across different operators, urban contexts, and regional characteristics with varying infrastructure, user behaviors, and operational constraints. Additionally, the reliance on static historical data rather than real-time dynamic data sources may reduce model effectiveness in rapidly evolving urban environments where demand patterns shift due to unexpected events, policy changes, or seasonal disruptions. These constraints highlight the need for more comprehensive validation frameworks that account for cross-regional variability and real-time operational dynamics.

Future research should prioritize expanding data sources through multi-operator collaborations and cross-city validation to enhance model generalizability and robustness across diverse urban environments. Critical development areas include federated learning architectures for privacy-preserving multi-operator training, online learning capabilities for real-time adaptability, and integration of unstructured data streams such as event schedules and social media indicators. Furthermore, extending MADENet's framework to other shared mobility modes (car-sharing, scooter-sharing) and developing scalable deployment strategies for different city sizes and infrastructure levels will enhance the model's practical utility for comprehensive multimodal transportation ecosystem management, fostering data-driven governance in smart city development.

Funding

This research was funded by the National Natural Science Foundation of China (62462019,62172350), Guangdong Basic and Applied Basic Research Foundation (2023A1515012846), Guangxi Science and Technology Major Program (AA24263010), The Key Research and Development Program of Guangxi (AB24010085, AB23026120), Key Laboratory of Equipment Data Security and Guarantee Technology, Ministry of Education (GDZB2024060500), Natural Science Foundation of Guangxi Province (2025GXNSFBA069410), and Basic Scientific Research Ability Improvement Project for Young and Middle-aged Teachers of Guangxi Higher Education Institutions (2024KY0233, 2025KY0243).

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Reference

[1] Afsari, M., Dastmard, M., Bresciani Miristice, L. M., & Gentile, G. (2024). Utilizing P-median and Machine Learning for Bike-Sharing Demand Prediction. 2024 IEEE International Conference on Environment and Electrical Engineering and 2024 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), 1–5. https://doi.org/10.1109/EEEIC/ICPSEurope61470.2024.10751398

[2] Andor, M. A., Gerster, A., Gillingham, K. T., & Horvath, M. (2020). Running a car costs much more than people think—

Stalling the uptake of green travel. Nature, 580(7804), 453–455. https://doi.org/10.1038/d41586-020-01118-w

[3] Biswas, Md. E., Hossain, Md. D., Huh, E.-N., & Sultana, T. (2025). Spatiotemporal Data-Driven Hourly Bike-Sharing Demand Prediction Using ApexBoost Regression. International Journal of Data Science and Analytics. https://doi.org/10.1007/s41060-025-00820-0

[4] Chen, P., Hsieh, H., Su, K., Sigalingging, X. K., Chen, Y., & Leu, J. (2020). 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

[5] Cheng, C., & Yue, W. (2024). Construction of a Green and Low-Carbon Travel Order Prediction Model Based on Shared Bicycle Big Data. International Journal of Computational Intelligence Systems, 17(1), 130. https://doi.org/10.1007/s44196-024-00519-3

[6] Chengula, T. J., Mwakalonge, J., Comert, G., Sulle, M., Siuhi, S., & Osei, E. (2024). Enhancing advanced driver assistance systems through explainable artificial intelligence for driver anomaly detection. Machine Learning with Applications, 17, 100580. https://doi.org/10.1016/j.mlwa.2024.100580

[7] Collini, E., Nesi, P., & Pantaleo, G. (2021). Deep Learning for Short-Term Prediction of Available Bikes on Bike-Sharing Stations. IEEE Access, 9, 124337–124347. https://doi.org/10.1109/ACCESS.2021.3110794

[8] E, S. V., Park, J., & Cho, Y. (2020). Using data mining techniques for bike sharing demand prediction in metropolitan city. Computer Communications, 153, 353–366. https://doi.org/10.1016/j.comcom.2020.02.007

[9] Feng, J., Liu, H., Zhou, J., & Zhou, Y. (2024). A Spatial-Temporal Aggregated Graph Neural Network for Docked Bike-sharing Demand Forecasting. ACM Transactions on Knowledge Discovery from Data, 18(9), 1–27. https://doi.org/10.1145/3690388

[10] Galatoulas, N.-F., Genikomsakis, K. N., & Ioakimidis, C. S. (2020). Spatio-Temporal Trends of E-Bike Sharing System Deployment: A Review in Europe, North America and Asia. Sustainability, 12(11), 4611. https://doi.org/10.3390/su12114611

[11] Gao, C., & Chen, Y. (2022). Using Machine Learning Methods to Predict Demand for Bike Sharing. In J. L. Stienmetz, B. Ferrer-Rosell, & D. Massimo (Eds.), Information and Communication Technologies in Tourism 2022 (pp. 282–296). Springer International Publishing. https://doi.org/10.1007/978-3-030-94751-4_25

[12] Giner Fabregat, G., Fonseca I Casas, P., & Rivero Martínez, A. (2025). Intelligent Optimization of Bike-Sharing Systems: Predictive Models and Algorithms for Equitable Bicycle Distribution in Barcelona. Sustainability, 17(10), 4316. https://doi.org/10.3390/su17104316

[13] Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. Science Robotics, 4(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

[14] Guo, H., Li, K., & Rou, Y. (2024). Optimized Demand Forecasting for Bike-Sharing Stations Through Multi-Method Fusion and Gated Graph Convolutional Neural Networks. IEEE Access, 12, 174017–174027. https://doi.org/10.1109/ACCESS.2024.3501572

[15] Guo, H., Zhao, S., Ren, Y., Li, J., & Xu, S. (2025). An XGBoost-Based Three-Stage Prediction Approach for True User Demand of Bike-Sharing Systems Based on Spatio-Temporal Analysis. IEEE Transactions on Intelligent Transportation Systems, 1–12. https://doi.org/10.1109/TITS.2025.3586747

[16] Hu, Y., Sun, Z., Li, W., & Pei, L. (2022). Forecasting public bicycle rental demand using an optimized eXtreme Gradient Boosting model. Journal of Intelligent & Fuzzy Systems, 42(3), 1783–1801. https://doi.org/10.3233/JIFS-211202

[17] Lee, Y., Son, H., Ahn, J., & Cho, S. (2024). Bike-sharing Demand Prediction based on Artificial Intelligence Algorithm Using Weather Data. 2024 IEEE International Conference on Consumer Electronics (ICCE), 1–6. https://doi.org/10.1109/ICCE59016.2024.10444462

[18] Li, X., Xu, Y., Chen, Q., Wang, L., Zhang, X., & Shi, W. (2022). Short-Term Forecast of Bicycle Usage in Bike Sharing Systems: A Spatial-Temporal Memory Network. IEEE Transactions on Intelligent Transportation Systems, 23(8), 10923–10934. https://doi.org/10.1109/TITS.2021.3097240

[19] Li, X., Xu, Y., Zhang, X., Shi, W., Yue, Y., & Li, Q. (2023). Improving short-term bike sharing demand forecast through an irregular convolutional neural network. Transportation Research Part C: Emerging Technologies, 147, 103984. https://doi.org/10.1016/j.trc.2022.103984

[20] Li, Y., & Zheng, Y. (2020). Citywide Bike Usage Prediction in a Bike-Sharing System. IEEE Transactions on Knowledge and Data Engineering, 32(6), 1079–1091. https://doi.org/10.1109/TKDE.2019.2898831

[21] Li, Y., Zhu, Z., Kong, D., Xu, M., & Zhao, Y. (2019). Learning Heterogeneous Spatial-Temporal Representation for Bike-Sharing Demand Prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), 1004–1011. https://doi.org/10.1609/aaai.v33i01.33011004

[22] Liang, Y., Ding, F., Huang, G., & Zhao, Z. (2023). Deep trip generation with graph neural networks for bike sharing system expansion. Transportation Research Part C: Emerging Technologies, 154, 104241. https://doi.org/10.1016/j.trc.2023.104241

[23] Liang, Y., Huang, G., & Zhao, Z. (2024). Cross-Mode Knowledge Adaptation for Bike Sharing Demand Prediction Using Domain-Adversarial Graph Neural Networks. IEEE Transactions on Intelligent Transportation Systems, 25(5), 3642–3653. https://doi.org/10.1109/TITS.2023.3322717

[24] Lim, H., Chung, K., & Lee, S. (2022). Probabilistic Forecasting for Demand of a Bike-Sharing Service Using a Deep-Learning Approach. Sustainability, 14(23), 15889. https://doi.org/10.3390/su142315889

[25] Lin, H., He, Y., Li, S., & Liu, Y. (2024). Insights into Travel Pattern Analysis and Demand Prediction: A Data-Driven Approach in Bike-Sharing Systems. Journal of Transportation Engineering, Part A: Systems, 150(2), 04023132. https://doi.org/10.1061/JTEPBS.TEENG-8137

[26] Pribadi, A. P., Rauf, A. U., Rahman, Y. M. R., & Haq, Z. F. (2025). Air Quality and Urban Sustainable Development-Current Issues and Future Directions. In J. A. Parray, N. Shameem, & A. K. Haghi (Eds.), Sustainable Urban Environment and Waste Management (pp. 23–51). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-1140-9_2

[27] Qian, H., Wang, J., Chen, Y., Zheng, S., & Wei, Z. (2025). CGA-STNet: A dockless shared bicycle demand prediction model considering multiple spatial features and time periodicity. Expert Systems with Applications, 265, 126100. https://doi.org/10.1016/j.eswa.2024.126100

[28] Roscher, R., Bohn, B., Duarte, M. F., & Garcke, J. (2020). Explainable Machine Learning for Scientific Insights and Discoveries. IEEE Access, 8, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199

[29] Sajid, M., Midhun, V., Zakkariya, K. A., Surira, M. D., & Vishnu, K. P. (2023). Pedaling towards sustainability: A mixed-method study of the drivers and barriers to bike-sharing adoption. Management of Environmental Quality: An International Journal, 34(6), 1580–1606. https://doi.org/10.1108/MEQ-12-2022-0343

[30] Schnieder, M. (2023). Ebike Sharing vs. Bike Sharing: Demand Prediction Using Deep Neural Networks and Random Forests. Sustainability, 15(18), 13898. https://doi.org/10.3390/su151813898

[31] Shi, Y., Zhang, L., Lu, S., & Liu, Q. (2023). Short-Term Demand Prediction of Shared Bikes Based on LSTM Network. Electronics, 12(6), 1381. https://doi.org/10.3390/electronics12061381

[32] Shi, Z., Jiang, Z., Xu, M., Zhou, Y., & Zhu, Z. (2025). Enhancing Hybrid Bike-Sharing Systems Through Long-Term Inventory Management: A Generative-Model-Informed Reinforcement Learning Approach. IEEE Transactions on Computational Social Systems, 1–13. https://doi.org/10.1109/TCSS.2025.3557911

[33] Wang, T., Hu, S., & Jiang, Y. (2021). Predicting shared-car use and examining nonlinear effects using gradient boosting regression trees. International Journal of Sustainable Transportation, 15(12), 893–907. https://doi.org/10.1080/15568318.2020.1827316

[34] Wang, Z., & Ren, F. (2025). Developing a decision support system for sustainable urban planning using machine learning-based scenario modeling. Scientific Reports, 15(1), 13210. https://doi.org/10.1038/s41598-025-90057-5

[35] Wang, Z., Yu, D., Zheng, X., Meng, F., & Wu, X. (2025). A Model-Data Dual-Driven Approach for Predicting Shared Bike Flow near Metro Stations. Sustainability, 17(3), 1032. https://doi.org/10.3390/su17031032

[36] Wei, X., Ren, Y., Shen, L., & Shu, T. (2022). Exploring the spatiotemporal pattern of traffic congestion performance of large cities in China: A real-time data based investigation. Environmental Impact Assessment Review, 95, 106808. https://doi.org/10.1016/j.eiar.2022.106808

[37] Xiang, Z., Zeng, F., Liu, L., Wu, J., Mumtaz, S., & Leung, V. C. M. (2025). Bike-Sharing Demand Prediction Based on Dynamic Time Warping and Spatio-Temporal Graph Attention Network. IEEE Transactions on Intelligent Transportation Systems, 1–15. https://doi.org/10.1109/TITS.2025.3570009

[38] Xiao, G., Wang, R., Zhang, C., & Ni, A. (2021). Demand prediction for a public bike sharing program based on spatio-temporal graph convolutional networks. Multimedia Tools and Applications, 80(15), 22907–22925. https://doi.org/10.1007/s11042-020-08803-y

[39] Xu, F., Uszkoreit, H., Du, Y., Fan, W., Zhao, D., & Zhu, J. (2019). Explainable AI: A Brief Survey on History, Research Areas, Approaches and Challenges. In J. Tang, M.-Y. Kan, D. Zhao, S. Li, & H. Zan (Eds.), Natural Language Processing and Chinese Computing (Vol. 11839, pp. 563–574). Springer International Publishing. https://doi.org/10.1007/978-3-030-32236-6_51

[40] Xu, M., Liu, H., & Yang, H. (2020). A Deep Learning Based Multi-Block Hybrid Model for Bike-Sharing Supply-Demand Prediction. IEEE Access, 8, 85826–85838. https://doi.org/10.1109/ACCESS.2020.2987934

[41] Yu, L., Feng, T., Li, T., & Cheng, L. (2023). Demand Prediction and Optimal Allocation of Shared Bikes Around Urban Rail Transit Stations. Urban Rail Transit, 9(1), 57–71. https://doi.org/10.1007/s40864-022-00183-w

[42] Zhu, B., Hu, S., Kaparias, I., Zhou, W., Ochieng, W., & Lee, D.-H. (2024). Revealing the driving factors and mobility patterns of bike-sharing commuting demands for integrated public transport systems. Sustainable Cities and Society, 104, 105323. https://doi.org/10.1016/j.scs.2024.105323