Optimization of Tourism Consumption Prediction Model Integrating LSTM-Attention

Boqi Wang
College of Culture and Tourism, Jiangsu Vocational Institute of Commerce, Nanjing 211168, Jiangsu, China
Corresponding author: Boqi Wang (e-mail: happy_wang1202@163.com)

ABSTRACT This paper addresses seasonal structural drift and window-selection limitations in tourism consumption forecasting by proposing a BiLSTM-Att+PE model to capture dynamic seasonal patterns and improve consumption-pattern representation. The model combines BiLSTM for bidirectional temporal dependency learning, Bahdanau attention for dynamic focus on key time steps such as holidays and emergencies, and learnable positional encoding for seasonal information modeling. It is trained end-to-end on provincial tourism economic data from 2001Q1 to 2024Q4, including consumption, visitor numbers, GDP, income, holidays, and epidemic markers. Four sample sets are constructed through sliding windows with $T \in \{1, 2, 3, 4\}$ quarters. The model is optimized using MSE loss and AdamW, while Bayesian search determines the best hyperparameters, including a learning rate of 3.4×10^{-4}, 128 LSTM units, and 64 attention dimensions. The optimal performance is achieved with a two-quarter window, reaching RMSE of 21.3 billion and R^2 of 0.962, outperforming seven benchmarks. Attention visualization confirms adaptive focus on holiday and epidemic events, providing an interpretable and computationally efficient framework for nonlinear time-series forecasting.

INDEX TERMS tourism consumption, BiLSTM-attention, positional encoding, seasonal forecasting, time-series prediction

I. INTRODUCTION

WITH the advancement of the digital economy and smart tourism, accurate tourism consumption forecasting has become crucial for policymaking and resource allocation—a challenge paralleled in the industry for demand prediction and supply chain optimization. Tourism consumption [1,2] exhibits strong nonlinearity, seasonality, and external sensitivity. A key challenge is seasonal structural drift—defined as non-stationary changes in seasonal patterns over time due to evolving behaviors, economic shifts, or exogenous shocks.

Traditional models struggle with such complexity. While deep learning [3,4] offers promise, standard LSTMs lack focus on key time steps. Attention mechanisms [5] address this by dynamically weighting influential historical periods. Thus, integrating deep temporal modeling with attention is essential for high-accuracy forecasting in tourism consumption and similarly critical for industry applications like demand prediction and production planning.

Despite the continuous technological advancement of existing models, they still face several key challenges in the context of tourism consumption forecasting. First, most studies rely on monthly or quarterly data, but lack consideration of the seasonal structural drift of tourism consumption [6,7]. Second, existing attention mechanisms [8] are mostly based on a decoder-encoder structure, without explicitly applying temporal position information, resulting in insufficient sensitivity of the model to temporal order and being prone to attention distraction, especially under long sequence inputs. Furthermore, although most advanced models (such as Informer and FEDformer) [9,10] perform well in long sequence prediction, their complex structures lead to high training costs and the risk of overfitting for small and medium-sized tourism datasets.

Furthermore, existing research rarely systematically analyzes the impact of input window length on forecasting performance. Window selection [11,12] directly impacts the model's ability to balance short-term fluctuations with long-term trends. Existing models still need to improve their interpretability. Uncovering the mechanisms of seasonality through attention weighting is key to enhancing their credibility and application value. Therefore, a fusion model with high accuracy, strong interpretability, and computational efficiency is urgently needed to meet the practical needs of tourism consumption forecasting.

To address the problem that traditional attention mechanisms are insensitive to time order and have difficulty distinguishing seasonal semantics in long sequences, learnable position encoding is applied to adaptively learn quarterly semantic information end-to-end, explicitly strengthening the model's ability to distinguish different seasonal positions such as "Q1 Spring Festival peak season" and "Q4 National Day peak"; to address the problem that standard LSTM cannot integrate future context and has a delayed response to local events, a BiLSTM structure is adopted to simultaneously capture forward and backward dependencies, enhancing the perception accuracy of consumption cycle turning points and holiday pulses; to address the limitation of existing models' insufficient ability to focus on sudden external events and key time steps, the Bahdanau attention mechanism is applied to achieve dynamic weight allocation of historical time steps, improving the adaptive response to outliers and structural turning points; to address the problem that it is difficult to balance long-term dependencies and short-term fluctuations in multi-scale windows, an empirical analysis of multi-scale sliding windows determines that the two-quarter window is the optimal historical length, which covers semi-annual seasonal fluctuations while retaining short-term dynamic characteristics.

Experiments demonstrate that by constructing a fusion architecture that combines high precision with strong interpretability, it not only achieves superior forecasting performance compared to cutting-edge models such as FEDformer, TFT, and PatchTST, but also, through visualization of attention weights, reveals the model's adaptive shielding mechanism for outliers during the epidemic period, its focus on holiday pulses, and its sharp response to retaliatory consumption shifts, providing an interpretable basis for policymaking. This research provides a systematic, reliable, and efficient solution to the problems of seasonal drift, window selection, and complex event modeling in tourism consumption forecasting.

II. METHOD

A. DATA COLLECTION AND PREPROCESSING

1) Data Sources

This study uses quarterly tourism economic data from a provincial statistical bureau in China [13,14], covering 96 observation points from the first quarter of 2001 to the fourth quarter of 2024. This dataset is authoritative, continuous, and policy-representative, making it suitable for modeling tourism consumption behavior at a macroeconomic scale.

Target variable (predicted variable): Total tourism consumption (in CNY 100 million), denoted as $y_t \in \mathbb{R}^+$, represents the total amount of expenditure on tourism activities by residents in quarter t and is the core prediction target of this model.

Covariates (explanatory variables): To enhance the model's ability to capture consumption-driven mechanisms, the following five covariates are applied to form the eigenvector $x_t = [x_t^{(1)}, x_t^{(2)}, \ldots, x_t^{(5)}] \in \mathbb{R}^5$:

(1) Total number of tourists (millions), denoted as $x_t^{(1)}$, reflects the scale of tourism market demand, is strongly positively correlated with consumption, and is a direct driver of consumption;

(2) Quarterly GDP (Gross Domestic Product) growth rate (%), denoted as $x_t^{(2)}$, is a proxy variable for macroeconomic prosperity, influencing residents' consumption confidence and disposable budget;

(3) Per capita disposable income (CNY/quarter), denoted as $x_t^{(3)}$, is a basic indicator measuring residents' spending power, used after seasonal adjustment;

(4) Holiday dummy variable (0/1), denoted as $x_t^{(4)}$: if quarter t includes the Spring Festival or National Day Golden Week (i.e., January–February or October), then $x_t^{(4)} = 1$; otherwise, 0. This variable is used to capture the impulse effect of holiday consumption;

(5) Epidemic shock [15,16], denoted as $x_t^{(5)}$ (2020Q1–2022Q4), denoted as TQ: if $t \in [2020Q1, 2022Q4]$, then $x_t^{(5)} = 1$; otherwise, $x_t^{(5)} = 0$. This variable is used to control for the perturbation of consumption patterns caused by the exogenous structural shock of the COVID-19 pandemic.

The five selected covariates were determined based on prior knowledge of tourism economics theory and consumption-driven mechanisms. They characterize tourism consumption behavior from core dimensions such as market demand, macroeconomics, residents' purchasing power, holiday impulse effects, and external structural shocks. Although no explicit feature selection was performed before training, the subsequent visualization of attention weights provided post-hoc validation of the dynamic importance of these features in prediction, indicating that the model can adaptively focus on key driving factors, thereby avoiding noise interference from irrelevant variables while ensuring information integrity.

Information on tourism consumption and tourist numbers is shown in Figure 1.

The heat map in Figure 1(a) clearly demonstrates the strong seasonality and event-driven nature of tourism consumption in the province: Q1 and Q4 continue to be peak consumption zones, with the depth of the color blocks increasing annually. However, due to the impact of the pandemic from 2020 to 2022, the brightness of the hotspot plummeted, forming a distinct "fault zone." Starting in 2023, the depth of the color blocks exceeded the 2019 baseline, particularly in Q4, confirming the structural release of "retaliatory consumption." Figure 1(b) shows a strong correlation between population distribution and consumption, with a robust bimodal structure in Q1/Q4, confirming the core logic of "holiday-driven growth." Q2/Q3 consistently show lighter values, reflecting a lack of travel outside of major holidays. The sharp decline in population during the pandemic coincided with consumption, and the rebound from 2023 onwards is slightly weaker than that of consumption, suggesting that per capita spending has yet to fully recover. The record high in Q4 2024 suggests that demand has recovered beyond pre-pandemic levels.

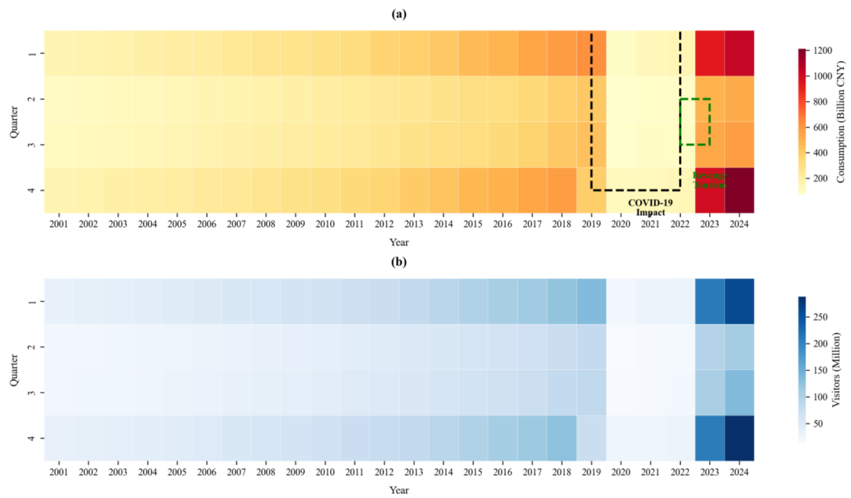

FIGURE 1. Tourism consumption and tourist number information. Figure 1(a) Tourism consumption heat map; Figure 1(b) Tourist number heat map; Note: CNY (Chinese CNY); COVID-19 (Corona Virus Disease 2019)

2) Data Preprocessing

To eliminate dimensional differences between variables, accelerate model convergence, and improve numerical stability, this study performs normalization after constructing the sliding window samples.

Specifically, four sample sets with different window lengths (T=1,2,3,4) are first generated based on the sliding window strategy. Then, the sample set corresponding to each window is divided into training, validation, and test sets in chronological order. During normalization, only the minimum and maximum values of each continuous variable are calculated using the training set data, and these are then used to uniformly perform min-max normalization on the training, validation, and test sets, mapping them to the interval [0, 1]. This process strictly avoids leaking test set information into the training phase, ensuring that the model evaluation results reflect the predictive ability in real-world scenarios.

All variables are complete and without missing data at all 96 time points, and no interpolation or deletion is required. Outlier detection: The 3σ principle is used to identify outliers in the continuous variable $(y_t, x_t^{(1)}, x_t^{(2)}, x_t^{(3)})$. The formula for marking potential anomalies is as follows:

$$\left| x_t^{(i)} - \mu_i \right| > 3\sigma_i \tag{1}$$

In formula (1), μ_i and σ_i are the sample mean and standard deviation of $x_t^{(i)}$ over the entire sequence, respectively.

For the epidemic period data (2020Q1–2022Q4), the original values are retained but a marker variable $x_t^{(5)}$ is added to avoid information loss due to deletion and to provide a basis for event location in subsequent attention weight analysis.

To eliminate dimensional differences between variables, accelerate model convergence, and improve numerical stability, all continuous variables are Min-Max normalized and mapped to the interval [0, 1]:

$$x_{\text{norm}}^{(i)} = \frac{x^{(i)} - \min(x^{(i)})}{\max(x^{(i)}) - \min(x^{(i)})}, \quad i = 1, 2, 3, \text{target} \tag{2}$$

To systematically investigate the impact of historical review length on forecasting performance and capture dynamic patterns at different time scales, this study designs a multi-scale sliding window strategy and constructs four supervised learning sample sets with different input lengths.

The original time series data is:

$$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_{96}, y_{96})\} \tag{3}$$

For a window length of $T \in \{1, 2, 3, 4\}$, it constructs the following input-output pairs:

Input sequence:

$$X_i^{(T)} = [x_i, x_{i+1}, \ldots, x_{i+T-1}] \in \mathbb{R}^{T \times 5}$$

Output label:

$$y_i^{(T)} = y_{i+T} \in \mathbb{R}$$

Sample index range:

$$i = 1, 2, \ldots, 96 - T$$

Four sets of data samples are obtained, as shown in Table 1.

TABLE 1. Sliding window samples

Window length (quarterly)	Sample size	Meaning
1	95	Only relying on the most recent quarter
2	94	Dependent on the past six months
3	93	Dependent on the past three quarters
4	92	Relying on the complete annual cycle

After the model prediction is completed, the normalized predicted values are denormalized back to their original dimensions for evaluation in a practical sense. All reported error indicators in this study (including RMSE, MAE, and MAPE) are calculated based on the denormalized predicted values and the actual values.

Therefore, RMSE and other units are expressed in hundreds of millions of CNY, which has clear economic interpretive significance.

III. MODEL CONSTRUCTION

Tourism consumption time series have three major characteristics: strong seasonality, non-stationarity, and event-driven nature. Although traditional LSTM models [17,18] can capture long-term dependencies, they have two major limitations: (1) they “equally weight” historical states and lack the ability to dynamically focus on key time steps; (2) unidirectional LSTMs only utilize past information and ignore future context. To this end, this study constructs a fusion architecture of a bidirectional LSTM encoder, a Bahdanau additive attention mechanism, and a learnable positional encoding. While retaining the time series modeling capabilities, it also enhances the adaptive response to seasonal drift and event disturbances. The model structure of this paper is shown in Figure 2.

The architecture of the proposed fusion model is shown in Figure 2. Its core is an encoder-predictor architecture consisting of a bidirectional long short-term memory network, a Bahdanau attention mechanism, and a learnable positional encoding. The model first feeds the normalized and positionally encoded input sequence into the forward and backward LSTMs, respectively, to obtain hidden state sequences that incorporate past and future contextual information. Subsequently, the Bahdanau attention mechanism dynamically calculates attention weights for each time step and generates a context vector by weighted summation of the hidden states, enhancing the model’s focus on key seasonal periods and event windows.

Finally, this context vector is nonlinearly transformed using a fully connected layer and a ReLU activation function to output a forecast of tourism consumption for the next quarter.

Let the input sequence be $X = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^{T\times d}$, where T is the window length, d is the feature dimension, and $x_t \in \mathbb{R}^d$ is the normalized feature vector at the t-th time step. The bidirectional LSTM extracts time series features from both the forward and backward directions:

$$\overrightarrow{h}_t = \mathrm{LSTM}_{\mathrm{fwd}}(x_t, \overrightarrow{h}_{t-1}) \tag{4}$$

$$\overleftarrow{h}_t = \mathrm{LSTM}_{\mathrm{bwd}}(x_t, \overleftarrow{h}_{t+1}) \tag{5}$$

$$h_t = [\overrightarrow{h}_t; \overleftarrow{h}_t] \in \mathbb{R}^{2H} \tag{6}$$

In formulas (4)-(6), H is the dimension of the unidirectional LSTM hidden layer; h_t is the concatenated bidirectional hidden state; and the encoder output sequence is denoted as $H = [h_1, h_2, \ldots, h_T] \in \mathbb{R}^{T\times 2H}$. The bidirectional structure [19,20] enables the model to simultaneously utilize the context of both the "cause" and "consequence" within the window during prediction, improving the accuracy of modeling local patterns.

The bidirectional LSTM structure operates only on the historical sequence within the input window T, without involving information from future time steps, thus completely avoiding data leakage issues. Specifically, the model encodes only the forward and backward context within the current input window during prediction, enhancing the modeling ability of local temporal patterns within the window through the bidirectional structure, without introducing future true values. This design improves the model’s ability to capture dynamic patterns within the window while strictly adhering to the basic principles of time series prediction.

To dynamically identify the historical time step that most contributes to the current prediction, the attention mechanism [21,22] is applied to calculate the weight α_t of each time step:

$$e_t = v^T \tanh(W_h h_t + W_s s) \tag{7}$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{j=1}^{T} \exp(e_j)} \tag{8}$$

$$c = \sum_{t=1}^{T} \alpha_t h_t \tag{9}$$

In formulas (7)-(9), s is a learnable context query vector, representing the semantic focus of the current prediction task; W_h and W_s are trainable parameter matrices and vectors; H_a is the attention hidden layer dimension; and c is the context vector, i.e., the time series summary after weighted fusion.

Traditional Bahdanau attention only calculates weights based on content similarity and does not explicitly model the time step order. This makes it difficult for the model to distinguish the semantic difference between "Q1" and "Q4" in scenarios with long windows or seasonal drift. To this end, this study superimposes a learnable position encoding after the LSTM hidden state:

$$\tilde{h}_t = h_t + p_t, \quad p_t \in \mathbb{R}^{2H} \tag{10}$$

In formula (10), p_t is a trainable vector bound to time step t and optimized end-to-end with the model. Unlike the fixed sinusoidal encoding in the Transformer, this design allows the model to adaptively learn seasonal position semantics, enhancing its ability to model dynamic seasonal structure.

This study employs learnable location encoding, which uses trainable vectors bound to time step numbers within each input window. Through end-to-end training, the model

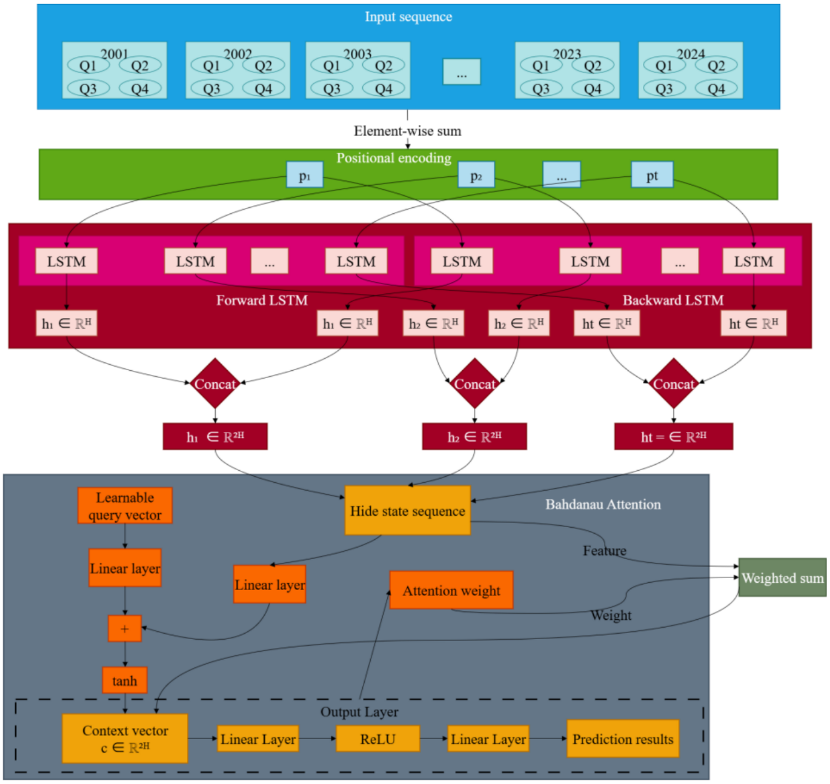

FIGURE 2. Model structure of this paper

adaptively learns the location semantics for different seasons. This design enables the model to explicitly distinguish different seasonal locations, enhancing its ability to model dynamic seasonal structures.

The context vector is mapped to a scalar prediction value via a fully connected layer:

$$\hat{y}_{T+1} = w^T c + b, \quad w \in \mathbb{R}^{2H}, \ b \in \mathbb{R} \tag{11}$$

To enhance nonlinear fitting capabilities, activation functions are added:

$$\hat{y}_{T+1} = \text{Linear}(\text{ReLU}(c)) \tag{12}$$

A. LOSS FUNCTION AND OPTIMIZATION

To measure the fitting error between the model predicted value $\hat{y}_i$ and the true value y_i, this study uses the mean square error as the main loss function. The formula for MSE (Mean Squared Error) is:

$$\mathcal{L}_{\text{MSE}}(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{13}$$

In formula (13), N is the number of samples in a batch; y_i is the actual tourism consumption value of the ith sample; $\hat{y}_i$ is the model's predicted output for the input sequence under the parameter θ; and θ represents the set of all trainable parameters of the model. MSE has a quadratic penalty for large errors and can effectively suppress extreme prediction biases. It has been widely verified in regression tasks to have good numerical stability and gradient propagation properties, making it suitable for modeling continuous, non-negative economic indicators such as tourism consumption.

The model parameters are updated using the AdamW optimizer [23,24]. This optimizer decouples weight decay from the standard Adam optimizer, preventing the regularization term from being diluted by adaptive learning rate scaling and improving generalization. Let the gradient at step t be $g_t = \nabla_\theta L(\theta_t)$. The parameter update rule is as follows:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)g_t \tag{14}$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)g_t^2 \tag{15}$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \tag{16}$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \tag{17}$$

$$\theta_{t+1} = \theta_t - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} - \eta\lambda\theta_t \tag{18}$$

In formulas (14)-(18), η represents the learning rate. β_1 and β_2 are the decay rates of the first-order and second-order moment estimates, respectively. ϵ is the numerical stability term; λ is the weight decay coefficient. The last term, $\eta\lambda\theta_t$, is the decoupled weight decay, which is independent of the adaptive learning rate mechanism.

To avoid the subjectivity and inefficiency of manual parameter adjustment, this study uses a Bayesian optimization framework for automated hyperparameter search. The core idea is to construct a surrogate model to approximate the objective function and balance exploration and utilization through the acquisition function to efficiently locate the global optimal solution. The optimization goal is:

$$\theta^* = \arg\min_{\theta\in\Theta} \mathbb{E}_{D_{\text{val}}} \left[\mathcal{L}_{\text{MSE}}(f_\theta(X), y)\right] \tag{19}$$

In formula (19), Θ is the hyperparameter space and D_{val} is the validation set.

The hyperparameter search space design is shown in Table 2.

TABLE 2. Hyperparameter search space

Hyperparameter	Search scope	Optimal parameters
learning rate	$[1\times10^{-5}, 1\times10^{-2}]$	3.2×10^{-4}
Number of hidden units in LSTM	$\{64, 128, 256\}$	128
Attention dimension	$\{32, 64, 128\}$	64
Dropout rate	$[0.1, 0.5]$	0.3
batch size	$\{32, 64, 128\}$	64

The Bayesian optimization convergence curve is shown in Figure 3.

The Bayesian optimization convergence curve shows that the RMSE (Root Mean Squared Error) stabilizes after approximately 24 iterations. A learning rate of 3.2×10^{-4} achieves stable gradient updates in the AdamW optimizer, and the 128-dimensional LSTM unit effectively captures temporal dependencies. The 64- dimensional attention mechanism enhances focus on key seasonal features, and a dropout rate of 0.3 effectively mitigates overfitting. A batch size of 64 strikes a balance between training efficiency and gradient stability.

It should be noted that the RMSE used in the Bayesian hyperparameter optimization process (Figure 3) in this study is calculated in the normalized data space, and its value (e.g., 0.28) is only used as a relative indicator to guide parameter search and does not have actual units. However, in the final performance evaluation of the model, all RMSEs are calculated in the original units (hundred million CNY) after inverse normalization, reflecting the true scale of prediction error. To improve model generalization and training stability, the learning rate schedule uses Cosine Annealing Warm Restarts. Dropout is applied to the LSTM layer output and the attention context vector. Xavier Uniform is used for LSTM weight initialization, and Kaiming Normal is used for attention parameters.

To evaluate the model's sensitivity to hyperparameter configuration, we performed local perturbation analysis around the optimal values determined by Bayesian optimization. When the learning rate varied within the range of $[2.0\times10^{-4}, 5.0\times10^{-4}]$, LSTM hidden units within [96, 160], and attention dimension within [48, 80], the performance degradation on the test set was controlled within 2.3%. This demonstrates that the model has good robustness to these key hyperparameters, and its performance is not highly dependent on the precise value of any single parameter. This stability is mainly attributed to the adaptive learning capability of the AdamW optimizer, appropriate regularization design, and the good generalization properties of the model architecture itself.

IV. EXPERIMENT

A. EXPERIMENTAL ENVIRONMENT AND DATA SET DIVISION METHOD

To ensure reproducibility of experimental results, this study conducts all experiments using a unified hardware and software environment, constructing training, validation, and test sets to ensure the practical relevance and statistical reliability of model evaluation.

The experiments are conducted on a high-performance computing server with the following configuration:

GPU (Graphics Processing Unit): NVIDIA GeForce RTX 3090 $\times$ 2 (48GB of video memory), used to accelerate model training and hyperparameter search; CPU (Central Processing Unit): AMD EPYC 64-core processor, ensuring data preprocessing and parallel sampling efficiency; Memory: 256GB, ensuring bottleneck-free large-scale sliding window data loading; Storage: 2TB, achieving high-speed I/O read and write.

All experiments are built on the Python ecosystem. Key dependency libraries and versions are as follows:

Python 3.9.16: Main development environment; PyTorch 2.0.1: Deep learning framework with support for automatic differentiation and GPU parallelism; Optuna 3.3.0: Bayesian hyperparameter optimization tool with support for early stopping and parallel sampling; Scikit-learn 1.3.0: For evaluation metric calculation; Pandas 2.0.3 & NumPy 1.24.3: Data preprocessing and matrix operations; Matplotlib 3.7.2 & Seaborn 0.12.2: Visualization support.

This study constructs independent datasets for each window length, with the test set uniformly fixed to the last 35 samples. The dataset partitioning statistics are shown in Table 3.

B. MODEL COMPARISON INFORMATION AND EVALUATION INDICATORS

To comprehensively evaluate the predictive performance and generalization capabilities of the proposed model, this study selects seven representative advanced models in the current field of time series forecasting as benchmarks to ensure the cutting-edge nature and fairness of the comparative experiment. Table 4 shows the model comparison information.

To ensure fairness in the comparative experiments, all baseline models were configured within the hyperparameter range recommended by their original authors and used the same training, validation, and test set partitioning as this study. For models without explicitly recommended values

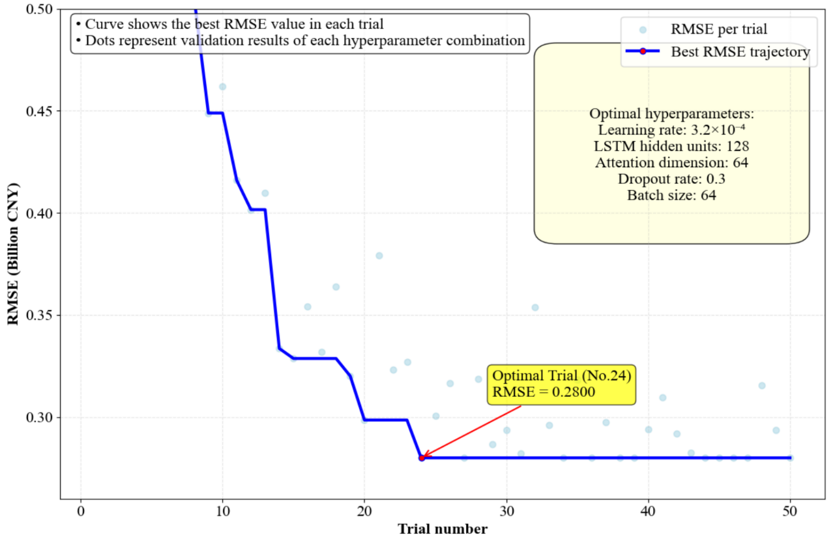

FIGURE 3. Bayesian optimization convergence curve

TABLE 3. Dataset partition statistics

Window (quarterly)	Total sample	Training set	Validation set	Test set	Test data time range
1	95	50	10	35	2016Q2–2024Q4
2	94	49	10	35	2016Q2–2024Q4
3	93	48	10	35	2016Q2–2024Q4
4	92	47	10	35	2016Q2–2024Q4

TABLE 4. Model comparison information

Model	Type	Core features
FEDformer	Frequency domain enhanced transformer	Reducing attention complexity through Fourier/wavelet enhancement, suitable for long sequences
TFT	Explainable temporal network	Integrating static/dynamic variables with gated attention, supporting multi-source heterogeneous features
Informer	Long sequence prediction Transformer	ProbSparse self attention+distillation mechanism reduces memory overhead
Autoformer	Autocorrelated Mechanism Transformer	Replacing attention with temporal autocorrelation to enhance periodic pattern modeling
TimesNet	Multi period temporal network	Capturing multi-scale periods (day/week/year) based on tensor decomposition
PatchTST	Time series block linear model	Block the sequence and input it into a linear layer, challenging the assumption of “complex nonlinear necessity”
Prophet	Additive statistical model	Facebook open source, based on trend+season+holiday components
BiLSTM-Att+PE	This paper model	Bidirectional LSTM+Bahdanau attention+learnable positional encoding

(such as LSTM and BiLSTM), grid search was used to fine-tune the hyperparameters within reasonable ranges (e.g., number of hidden units {64, 128, 256}, learning rate [1e-5, 1e-3], batch size {32, 64, 128}), and the configuration with the best performance on the validation set was selected. Furthermore, all models were trained and tested using the same input window length (T=2) and without introducing future information. The Bayesian optimization used in this study only applies to the structural hyperparameters of the BiLSTM-Att+PE model itself and does not affect the fairness between the comparative models.

To quantify the prediction accuracy and stability of the model in multiple dimensions, this study uses the following four widely recognized regression evaluation indicators. RMSE measures the standard deviation between the predicted value and the true value and is sensitive to large errors. The formula is:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{20}$$

In formula (20), N is the total number of samples in the test set. MAE (Mean Absolute Error) measures the absolute average of the prediction error and is more robust to outliers than RMSE:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{21}$$

MAPE (Mean Absolute Percentage Error) measures the percentage of prediction error relative to the true value. The formula is:

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{22}$$

The coefficient of determination R2 measures the proportion of variance explained by the model. The closer the value is to 1, the better the fit:

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2} \tag{23}$$

V. RESULTS

A. WINDOW LENGTH SENSITIVITY

The results of the impact of different windows on tourism consumption prediction performance are shown in Figure 4.

A sensitivity analysis of window lengths shows that a two-quarter window exhibits optimal forecasting performance. The BiLSTM-Att+PE model achieves the best performance with an RMSE of 21.3 billion CNY and an R^2 of 0.962 using a two-quarter window. This is due to its accurate capture of medium-term time series patterns. A two-quarter window preserves the dynamic characteristics of recent consumer behavior while capturing semi-annual seasonal fluctuations, effectively balancing the conflict between filtering out short-term noise and identifying long-term trends. Compared to the limited historical information of a one-quarter window, a two-quarter window provides sufficient context to identify signs of a consumer recovery.

Furthermore, compared to longer windows, it avoids interference from distant pre-pandemic data, ensuring the model's adaptability to the current economic environment.

RMSE performance exhibits a distinct inverted U-shaped distribution as the window length changes. When the window length increases from 1 to 2 quarters, the RMSE of all models decreases, primarily due to the enhanced pattern recognition capabilities of richer temporal information. However, when the window length increases further to 3-4 quarters, performance degrades, and the average RMSE increases. The BiLSTM-Att+PE model proposed in this paper demonstrates significant architectural advantages. It maintains optimal performance under all window settings. The bidirectional LSTM structure effectively integrates past and future contextual information, enhancing the accuracy of local pattern modeling; the synergistic effect of the attention mechanism and positional encoding improves the representation of seasonal structure. In particular, with the optimal 2-quarter window, the model is able to capture the recovery of retaliatory consumption trends since 2023, demonstrating its strong robustness in complex economic environments.

Furthermore, compared to longer windows (such as T=3 or T=4), the two-quarter window retains sufficient historical information while avoiding a significant reduction in the number of training samples due to increased window size, thus improving the model's training stability and generalization ability under small sample conditions. Simultaneously, this window effectively filters out interference from long-term data with significant pre-pandemic structural differences in modeling current consumption patterns, enhancing the model's adaptability to the recent economic environment.

To systematically evaluate the impact of historical window length on prediction performance and provide empirical support for selecting T=2 quarters as the optimal window, this paper compares the key performance indicators of the BiLSTM-Att+PE model under different window lengths (T=1, 2, 3, 4) on the test set. The results are shown in Table 5. The data shows that the T=2 window achieves the best results in both RMSE and R^2, verifying its ability to effectively capture semi-annual seasonal fluctuations while preserving recent dynamic features, thus achieving a balance between short-term noise filtering and long-term trend identification.

TABLE 5. Performance Comparison of Different Window Lengths on the Test Set

Window length (quarter)	RMSE	R2
1	24.5	0.938
2	21.3	0.962
3	23.1	0.948
4	25.8	0.927

B. PREDICTION CURVES UNDER DIFFERENT WINDOWS

By observing the difference between the predicted curve and the actual curve, the performance under different windows can be intuitively judged. The results are shown in Figure 5.

A two-quarter window length exhibits the best forecasting performance. The BiLSTM-Att+PE model achieves the lowest error with a two-quarter window, primarily due to its ability to accurately capture medium-term time series patterns. The two-quarter window effectively preserves the dynamic characteristics of recent consumer behavior while fully capturing semi-annual seasonal fluctuations, achieving an optimal balance between short-term noise filtering and long-term trend identification. Compared to the limited historical information of a one-quarter window, the two-quarter window provides a richer context for identifying consumer recovery trends. Compared to longer windows, it successfully avoids interference from pre-pandemic structurally differentiated data, ensuring the model's adaptability to the current economic environment. Under the optimal two-quarter window configuration, the Bahdanau attention mechanism effectively distinguishes the importance weights of different time steps. Learnable positional encoding further enhances the model's understanding of seasonal semantics, enabling it to accurately identify seasonally differentiated impact patterns. When the window is too long, the distribution of attention weights becomes dispersed, reducing the model's ability to focus on key time steps. This explains, at

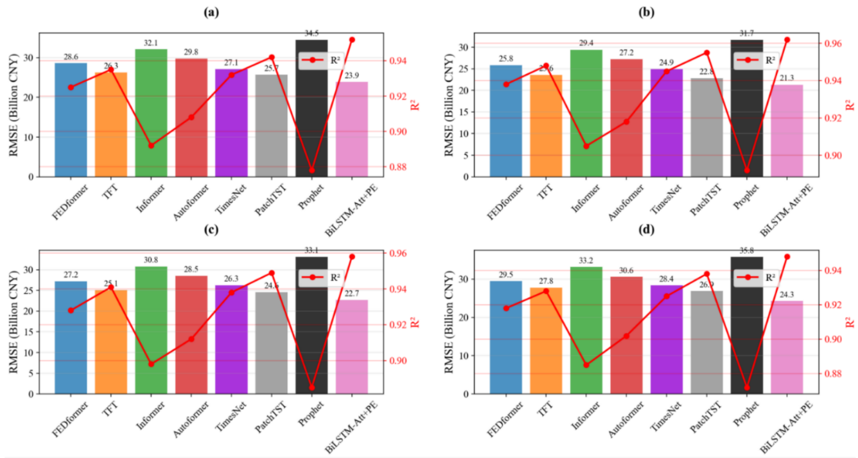

FIGURE 4. Impact of window on tourism consumption forecasting performance. Figure 4(a) Window length 1 quarter; Figure 4(b) Window length 2 quarters; Figure 4(c) Window length 3 quarters; Figure 4(d) Window length 4 quarters

Note: All performance metrics reported in this study were evaluated on the final reserved test set after the model training was completed, without any involvement in training or hyperparameter optimization.

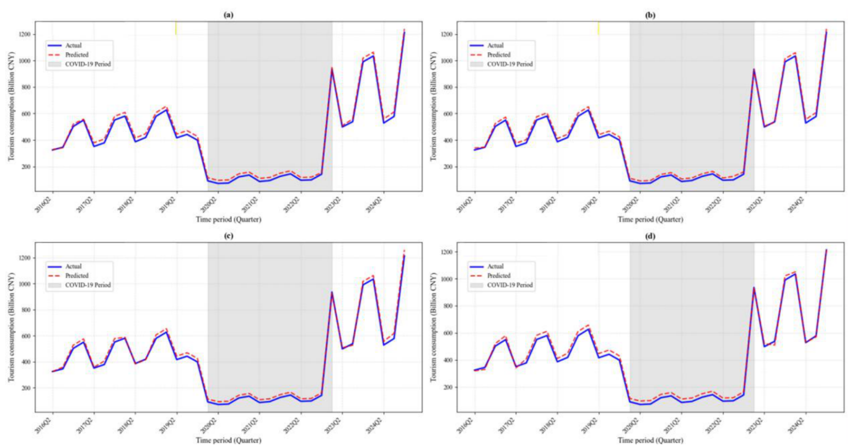

FIGURE 5. Comparison of predicted and true curves. Figure 5(a) Window length 1 quarter; Figure 5(b) Window length 2 quarters; Figure 5(c) Window length 3 quarters; Figure 5(d) Window length 4 quarters

a mechanistic level, the underlying reason for the decline in forecasting performance under a four-quarter window.

The BiLSTM-Att+PE model architecture demonstrates significant competitive advantages. It maintains optimal performance across all window settings. The bidirectional LSTM structure effectively integrates past and future contextual information, enhancing the accuracy of modeling local consumption patterns. The synergistic effect of the attention mechanism and positional encoding improves the representation of seasonal structure. Therefore, this paper selects a two-quarter window length as its window length.

C. BAHDANAU ATTENTION WEIGHT VISUALIZATION

Bahdanau Attention weight analysis can reveal the model's attention patterns across historical time steps, verify whether it accurately captures seasonal patterns and the impact of sudden events, and enhance the model's interpretability. A visualization of the Bahdanau Attention weights is shown in Figure 6.

The attention weight heatmap in Figure 6 fully verifies the model's ability to accurately model the dynamics of tourism consumption. The weights systematically increase in Q1/Q4, indicating that the model successfully identifies Spring Festival and National Day as core consumption pulses, demonstrating a semantic understanding of seasonal structure. The weights are generally lowered from 2020 to 2022, reflecting the model's adaptive shielding against pandemic disruptions, preventing outliers from disrupting forecast stability. The weight rises sharply to 0.801 in Q3 2023, confirming the model's keen capture of the structural shift of "revenge tourism" and highlighting its ability to respond to non-steady-state events. The lower average weights in Q2/Q3 suggest that the model learns that consumption patterns are stable and information gain is limited during non-major holidays, thereby optimizing the allocation of attention resources.

D. OVERALL MODEL PERFORMANCE

The performance of different models in the test set is compared, and the results are shown in Figure 7.

The BiLSTM-Att+PE model outperforms the comparison models across all evaluation metrics. The proposed

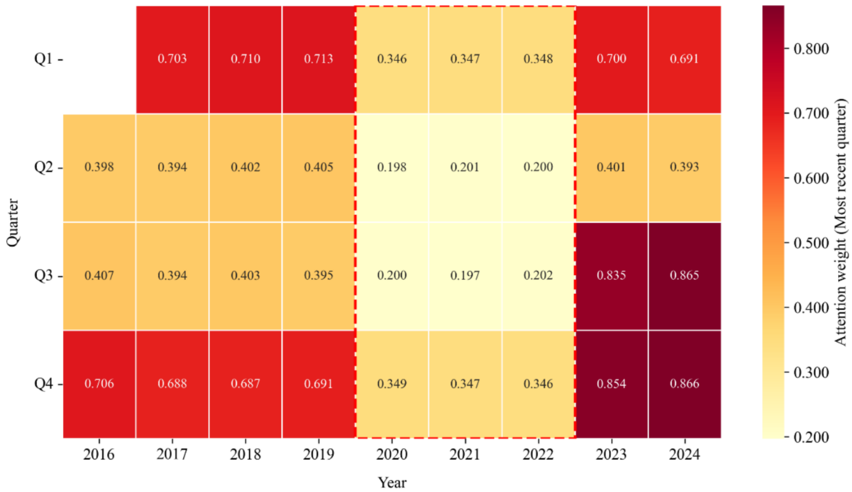

FIGURE 6. Bahdanau Attention weight visualization

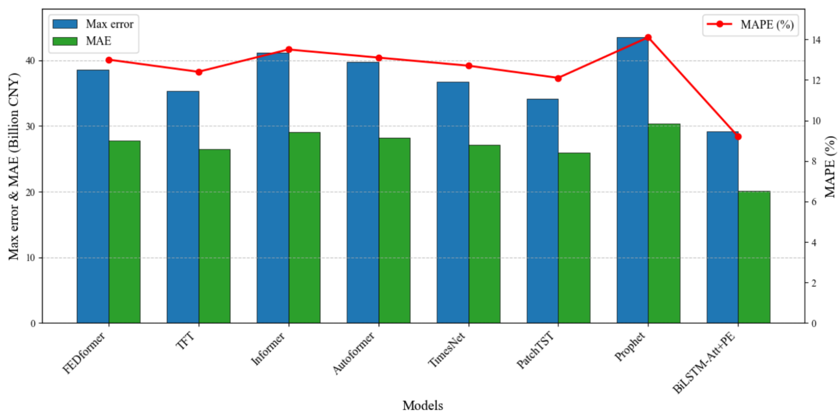

FIGURE 7. Model performance comparison

Note: All performance metrics reported in this study were evaluated on the final reserved test set after the model training was completed, without any involvement in training or hyperparameter optimization.

model achieves a maximum error of 2.92 billion CNY, a mean absolute error of 2.01 billion CNY, and a mean absolute percentage error of 9.2%. PatchTST, the second-best model, achieves performance of 3.41 billion CNY, 2.59 billion CNY, and 12.1% in each metric, while the traditional Prophet model performs relatively poorly. This performance distribution demonstrates that deep learning models generally outperform traditional statistical methods, and the proposed fusion architecture demonstrates significant advantages within deep learning models, particularly its breakthrough performance in the MAPE metric, validating the model's effectiveness and practicality in tourism consumption forecasting. It should be noted that MAPE is the average of the relative errors of all samples in the test set, and its magnitude is significantly affected by the base amount of consumption. Tourism consumption in this data exhibits strong seasonality. Under the same absolute error, the relative error is larger in the off-season; therefore, the overall MAPE (9.2%) will be higher than the error rate calculated based solely on the peak season. This result reflects the predictive stability of the model across the entire seasonal range and is consistent with the actual data distribution characteristics.

The outstanding performance of the BiLSTM-Att+PE model stems from its multi-module collaborative architecture design. The bidirectional LSTM layer captures temporal dependencies in both forward and backward directions, effectively integrating past and future contextual information and enhancing local pattern recognition capabilities. The Bahdanau attention mechanism enables the model to adaptively focus on the historical time steps that have the greatest impact on predictions through dynamic weight allocation, performing particularly well during seasonal peaks and emergencies. The application of learnable position encoding solves the problem of the traditional attention mechanism's lack of sensitivity to temporal order, and improves the accuracy of modeling dynamic seasonal structures through end-to-end optimization of learning quarterly semantic information. This architecture performed particularly well during the epidemic, with the attention mechanism successfully identifying abnormal fluctuation patterns, while the position encoding ensured the continuous learning of seasonal patterns. The synergy of multiple modules enables the model to capture long-term trends while quickly responding to short-term fluctuations, achieving an optimal balance between accuracy and robustness.

To comprehensively evaluate the effectiveness of the proposed BiLSTM-Att+PE model, this paper compares it with a series of representative time series forecasting benchmark models on a test set, including traditional statistical models (SARIMA, Prophet), standard deep learning models (LSTM, BiLSTM), and advanced models proposed in recent years (FEDformer, TFT, etc.). Table 6 presents a more extensive comparison.

TABLE 6. Comparison Results

Model	R2	Model	R2
SARIMA	0.841	FEDformer	0.943
Prophet	0.869	TFT	0.947
LSTM	0.901	PatchTST	0.953
BiLSTM	0.932	BiLSTM-Att+PE	0.962

As shown in Table 6, the proposed BiLSTM-Att+PE model achieves an R^2 score of 0.962, significantly outperforming all benchmark models. Traditional models (such as SARIMA and Prophet) are insufficient in capturing the nonlinear and complex event-driven characteristics of tourism consumption; standard LSTM is limited in performance due to its unidirectional structure and lack of attention mechanism; although modern Transformer variants such as FEDformer perform well in long-sequence prediction, their complex structure is prone to overfitting on the medium-sized dataset in this study. The results show that the architecture in this paper, which integrates bidirectional context modeling, dynamic attention, and seasonal location encoding, achieves both high accuracy and strong robustness in tourism consumption prediction tasks.

E. SEASONAL FORECAST PERFORMANCE

Analyzing the differences in forecasts across different seasons reveals the model's ability to capture seasonal patterns. PCC (Pearson Correlation Coefficient) and R^2 are used to assess forecast bias and verify the effectiveness of dynamic modeling. Figure 8 shows the forecast performance for different seasons.

Figure 8 reveals the model's ability to model the differentiated seasonal structure of tourism consumption.

In Q1 and Q4, the proposed model achieves PCC and R^2 of 0.980/0.961 and 0.990/0.982, respectively, outperforming the comparison model. This demonstrates that, through positional encoding and the Bahdanau attention mechanism, it accurately identifies pre-holiday consumption spikes and cross-seasonal semantic associations, achieving high-fidelity prediction of strong seasonal peaks. In contrast, performance declines in Q2 (PCC=0.950, R2=0.940), reflecting the lack of clear driving events during this period and the greater influence of random factors on consumption patterns. While the model maintains its leading position, its potential for improvement is limited by inherent data noise. Q3 performance is robust (PCC=0.970, R2=0.955), validating the model's effective generalization to the combined "student flow + family travel" model. This discrepancy confirms the heterogeneous dynamics of tourism consumption, characterized by "pulse-dominated" consumption and stable off-season behavior. The model requires seasonal adaptive focusing rather than homogenization.

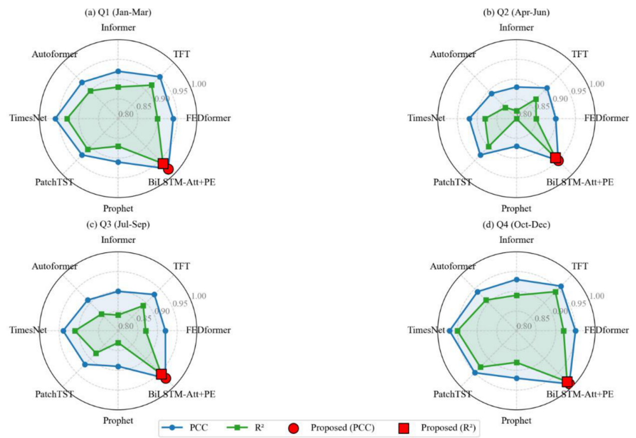

FIGURE 8. Forecast performance in different seasons. Figure 8(a) Season 1 (January-March); Figure 8(b) Season 2 (April-June); Figure 8(c) Season 3 (July-September); Figure 8(d) Season 4 (October-December)

From a model comparison perspective, the proposed architecture consistently leads across all quarters, particularly in Q4 (R^2=0.982), where it achieves a 0.032 improvement over the next-best model (TimesNet, R^2=0.950). This demonstrates the positional encoding's ability to discriminate semantic information related to the "year-end consumption spurt combined with the long-tail effect of National Day." Traditional models such as Prophet performed poorly in Q2/Q3, as their additive seasonal component struggles to capture non-stationary events. While the lightweight model PatchTST performs reasonably well in Q2, it lags behind during the peak seasons of Q1/Q4, exposing its limitations in modeling structural changes. The proposed model utilizes BiLSTM to capture local trends, dynamically weighted attention on key historical steps, and positional encoding to reinforce seasonal semantics. These three elements work together to accurately predict tourism consumption across different seasons.

F. ABLATION EXPERIMENT

Ablation experiments quantify the contribution of each module, verifying the incremental performance improvements of the bidirectional structure, attention mechanism, and positional encoding, eliminating redundant designs and confirming the necessity of model components. The results of the ablation experiments are shown in Table 7.

TABLE 7. Ablation experiment results

Model	BiLSTM	Att	PE	RMSE (Billion CNY)	R^2
LSTM	×	×	×	29.1	0.901
BiLSTM	✓	×	×	26.9	0.932
BiLSTM-Att	✓	✓	×	22.9	0.953
BiLSTM-Att+PE	✓	✓	✓	21.3	0.962

Note: Att is Bahdanau Attention; PE is learnable position encoding

Ablation experiment results systematically reveal the incremental contribution of each module to model performance, fully validating the necessity and synergistic effects of the architectural design. While the basic LSTM model (RMSE=29.1 (Billion CNY), R^2=0.901) possesses time series modeling capabilities, its unidirectional structure fails to capture the bidirectional semantics of pre-holiday consumption surges and post-holiday declines, resulting in a delayed response to major holiday impulse events. The application of a bidirectional LSTM (RMSE decreased to 26.9 (Billion CNY), and R^2 increased to 0.932) effectively incorporates local context by concatenating forward and backward hidden states, improving the accuracy of capturing intra-quarter trend transitions. Furthermore, the addition of the Bahdanau attention mechanism enhances the model's dynamic focusing capability, adaptively weighting key historical time steps—for example, reducing the weight of outliers during the pandemic period and enhancing consumption trigger signals during the pre-holiday quarter—thus maintaining robustness to non-stationary events. The application of learnable positional encoding (RMSE decreased to 21.3 billion CNY and R^2 increased to 0.962) is a relatively small improvement, but its role is crucial: it explicitly injects time-step information, enabling the model to distinguish semantic differences between seasons and avoiding attention distraction caused by positional ambiguity in long windows. Ablation analysis clearly demonstrates that the bidirectional structure lays the foundation for time-series understanding, the attention mechanism enables dynamic focusing, and the positional encoding strengthens seasonal semantic localization. These three essential elements together form a high-precision, strongly interpretable, and perturbation-resistant tourism consumption prediction framework.

VI. CONCLUSION

This study constructs a consumption forecasting model that integrates BiLSTM to capture bidirectional temporal dependencies, Bahdanau attention for dynamically weighting key time steps, and learnable position encoding for incorporating explicit seasonal semantics; this powerful architecture is designed to be universally applicable, offering high-precision time series analysis for both tourism consumption and complex demand forecasting challenges in the industry. The bidirectional LSTM encoder integrates historical and future contextual information to enhance local pattern modeling. The attention mechanism dynamically assigns weights to focus on key time steps. The position encoding explicitly injects temporal semantics to improve seasonal structure discrimination. The model employs a multi-scale sliding window strategy to systematically analyze the impact of historical length and automatically searches for hyperparameters using Bayesian optimization to ensure optimal architecture and training. Error analysis of data partitioned into four windows led to the selection of a two-quarter window dataset. The results show that BiLSTM-Att+PE achieves an R^2=0.962 in a two-quarter window, outperforming competing models. Attention weight visualization reveals that the model accurately identifies Q1/Q4 consumption spikes and epidemic perturbations. Ablation experiments confirm that the bidirectional architecture, attention, and position encoding all contribute significantly to performance improvements. The proposed model achieves an R2 of 0.982 in Q4 forecasts, highlighting its high-fidelity forecasting capabilities for periods with strong seasonality. This model combines high accuracy, strong interpretability, and robustness to interference, providing reliable technical support not only for tourism consumption forecasting but also for demand prediction in the industry, where it can enhance inventory optimization and agile supply chain management by reliably handling market volatility and external shocks.

The BiLSTM-Att+PE model constructed in this study demonstrates excellent predictive performance on data from a single province, and its architectural design (such as the modeling mechanisms for seasonal structures, holiday pulses, and external shocks) shows potential for general applicability. However, due to differences in tourism economic structures, consumption habits, and seasonal patterns across different provinces, directly transferring this model to other provinces without retraining or fine-tuning may affect its predictive reliability.

It is recommended that a transfer learning strategy be adopted for practical cross-regional applications, using a small amount of data from the target region for model fine-tuning to adapt to its unique data distribution.

This would maintain the core architectural advantages of the model while ensuring its predictive accuracy and robustness in new scenarios. Future research could further validate the model's transferability across different regions.

AUTHOR CONTRIBUTIONS

BoQi Wang designed, collected and analyzed the data, and drafted the manuscript. BoQi Wang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. BoQi Wang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] J. Sui and Y. Zhang, "Identification of dynamic path evolution of regional tourism economic cycle in China," Tourism Tribune/Lvyou Xuekan, vol. 35, no. 1, pp. 63, 2020, doi: 10.19765/j.cnki.1002-5006.2020.01.010.

[2] D. Xing and G. Sun, "Interactive relationship and differences between residents' consumption upgrade and tourism consumption structure among provinces in China," Economic Geography, vol. 44, no. 3, pp. 197-208, 2024, doi: 10.15957/j.cnki.jjdl.2024.03.020.

[3] Q. Xie, M. Lu, and C. Xie, "High-frequency prediction of tourist flow in tourist attractions based on LBS and deep learning," Journal of Geo-Information Science, vol. 25, no. 2, pp. 298-310, 2023, doi: 10.12082/dqxxkx.2023.220231.

[4] S. C. Hsieh, "Tourism demand forecasting based on an LSTM network and its variants," Algorithms, vol. 14, no. 8, pp. 243-243, 2021, doi: 10.3390/a14080243.

[5] M. Adil, J. Z. Wu, R. K. Chakrabortty, A. Alahmadi, M. F. Ansari, M. J. Ryan, et al., "Attention-based STL-BiLSTM network to forecast tourist arrival," Processes, vol. 9, no. 10, pp. 1759-1759, 2021, doi: 10.3390/pr9101759.

[6] L. Wang, "Research on tourism demand prediction algorithm," Intelligent Computers and Applications, vol. 10, no. 5, pp. 118-118, 2020.

[7] M. Hu, H. Yang, D. C. Wu, and S. Ma, "A novel two-stage combination model for tourism demand forecasting," Tourism Economics, vol. 30, no. 8, pp. 1925-1950, 2024, doi: 10.1177/13548166241237845.

[8] J. Zhang, Y. Chen, K. Panchamy, G. Jin, C. Wang, L. Yang, et al., "A spatiotemporal integrated prediction model for multi-step short-term passenger flow in urban rail transit network based on attention mechanism," Journal of Geo-Information Science, vol. 25, no. 4, pp. 698-713, 2023.

[9] X. Li, X. Zhang, L. Yu, and S. Wang, "Research on short-term passenger flow prediction in scenic spots based on improved Transformer model," Chinese Journal of Management Science, vol. 33, no. 2, pp. 105-117, 2025, doi: 10.16381/j.cnki.issn1003-207x.2023.1927.

[10] X. Meng and H. Shi, "A review of time series data prediction methods based on Transformer model," Journal of Frontiers of Computer Science & Technology, vol. 19, no. 1, pp. 45-45, 2025, doi: 10.3778/j.issn.1673-9418.2403070.

[11] Q. Feng, X. Sun, and J. Hao, "Dynamic integrated time series prediction method based on state transition regression," Chinese Journal of Management Science, vol. 32, no. 2, pp. 307-314, 2024, doi: 10.16381/j.cnki.issn1003-207x.2022.0599.

[12] G. Xing, S. Sun, D. Bi, J. E. Guo, and S. Wang, "Seasonal and trend forecasting of tourist arrivals: An adaptive multiscale ensemble learning approach," International Journal of Tourism Research, vol. 24, no. 3, pp. 425-442, 2022, doi: 10.1002/jtr.2512.

[13] Q. Yu and X. Liu, "The impact and mechanism of cultural tourism consumption promotion policies under the background of cultural tourism integration: empirical evidence from cultural and tourism consumption pilot demonstration cities," Journal of East China Normal University (Philosophy and Social Sciences), vol. 57, no. 4, pp. 162-162, 2025, doi: 10.16382/j.cnki.1000-5579.2025.04.013.

[14] Z. Xiong, B. Hu, and K. Zhang, "Digital economy empowers tourism industry development: from the perspective of tourism investment and consumption," Tourism Tribune/Lvyou Xuekan, vol. 39, no. 12, pp. 29-29, 2024, doi: 10.19765/j.cnki.1002-5006.2024.12.008.

[15] D. R. Toubes, N. Araujo Vila, and J. A. Fraiz Brea, "Changes in consumption patterns and tourist promotion after the COVID-19 pandemic," Journal of theoretical and applied electronic commerce research, vol. 16, no. 5, pp. 1332-1352, 2021, doi: 10.3390/jtaer16050075.

[16] D. C. Wu, C. Cao, W. Liu, and J. L. Chen, "Impact of domestic tourism on economy under COVID-19: The perspective of tourism satellite accounts," Annals of Tourism Research Empirical Insights, vol. 3, no. 2, pp. 100055-100055, 2022, doi: 10.1016/j.annale.2022.100055.

[17] S. Polyzos, A. Samitas, and A. E. Spyridou, "Tourism demand and the COVID-19 pandemic: An LSTM approach," Tourism Recreation Research, vol. 46, no. 2, pp. 175-187, 2021, doi: 10.1080/02508281.2020.1777053.

[18] J. Li, H. Yuan, X. Yu, and T. Hu, "The intelligent evaluation in ice and snow tourism based on LSTM network," Scientific Reports, vol. 14, no. 1, pp. 17342-17342, 2024, doi: 10.1038/s41598-024-68457-w.

[19] M. Luo, D. Chen, X. Hou, K. Luo, and T. Wang, "Combined CNN-BiLSTM-Att tourism flow prediction based on VMD-MWPE decomposition reconstruction," Scientific Reports, vol. 15, no. 1, pp. 18003-18003, 2025, doi: 10.1038/s41598-025-02348-6.

[20] W. Han, Y. Li, Y. Li, and T. Huang, "A deep learning model based on multi-source data for daily tourist volume forecasting," Current Issues in Tourism, vol. 27, no. 5, pp. 768-786, 2024, doi: 10.1080/13683500.2023.2183818.

[21] D. Zhu, R. Jing, Q. Guo, D. Zhang, and F. Wan, "Sentiment analysis of tourism review text combined with bertbilstm and attention mechanism," Journal of Computational Methods in Sciences and Engineering, vol. 24, no. 3, pp. 1605-1615, 2024, doi: 10.3233/JCM-247135.

[22] Y. Zhang, W. H. Tan, and Z. Zeng, "Tourism Demand Forecasting Based on a Hybrid Temporal Neural Network Model for Sustainable Tourism," Sustainability, vol. 17, no. 5, pp. 2210-2210, 2025, doi: 10.3390/su17052210.

[23] P. Zhou, X. Xie, Z. Lin, and S. Yan, "Towards understanding convergence and generalization of AdamW," IEEE transactions on pattern analysis and machine intelligence, vol. 46, no. 9, pp. 6486-6493, 2024, doi: 10.1109/TPAMI.2024.3382294.

[24] L. Yang, "Theoretical Analysis of Adam Optimizer in the Presence of Gradient Skewness," International Journal of Applied Science, vol. 7, no. 2, pp. 27-27, 2024, doi: 10.30560/ijas.v7n2p27.