

Research on Multimodal Prediction Model of Sports Event Ticket Revenue Based on Transformer

Qian Zhang^{ID*}

College of Physical Education and Health Science, Yibin University, Yibin, Sichuan, China

Abstract: To address the problem that sports event ticket revenue is affected by the coupling of multiple factors and exhibits significant dynamic changes, this paper proposes a multimodal prediction model based on Transformer. This method integrates historical ticketing data, event information, user behavior, and public opinion data, constructs a cross-modal dynamic alignment Transformer architecture (CMDA-Transformer), and introduces a multi-scale time series modeling mechanism and a demand sensitivity enhancement module to improve the model's ability to capture complex temporal patterns and multi-source information. Experiments were conducted on about 24,000 real event records. The results show that compared with traditional ARIMA, XGBoost, and deep learning models (LSTM, GRU, and standard Transformer), the RMSE of this method is 18.72, which is about 5.7% lower than that of the baseline Transformer, the MAPE is about 8.3% lower, and the coefficient of determination R^2 increases to 0.912. At the same time, it still maintains stable performance under different types of events and noise interference scenarios, which verifies the generalization ability and robustness of the model. The results show that this method can effectively improve the accuracy of ticket revenue prediction, and provide reliable support for dynamic pricing and ticket optimization.

Keywords: Sports events; Ticket revenue forecasting; Multimodal learning; Transformer; Time series prediction

How to Cite: Zhang, Q. (2026). Research on Multimodal Prediction Model of Sports Event Ticket Revenue Based on Transformer. *International Scientific Technical and Economic Research*, 4(2), 124–147. <https://doi.org/10.71451/ISTAER2618>

Article history: Received: 20 Jan 2026; Revised: 27 Feb 2026; Accepted: 10 Apr 2026; Published: 24 Apr 2026
Copyright: © 2026 The Author(s). Published by Sichuan Knowledgeable Intelligent Sciences. This is an open access article under the [CC BY 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

1. INTRODUCTION

With the continuous development of the sports industry and the steady increase in commercialization, ticket revenue from sports events has become an important economic source for clubs and event operators [1],[2]. Accurate prediction of ticket revenue not only helps to optimize ticket management and resource allocation, but also provides data support for dynamic pricing, marketing and risk control. However, the ticket sales process of sports events exhibits significant dynamics and uncertainty, influenced not only by historical sales trends but also by event attractiveness, audience behavior, the external environment, emergencies, and

* **Corresponding author:** Qian Zhang, College of Physical Education and Health Science, Yibin University, Yibin, Sichuan, China. Email: Zhang123202410@163.com

other factors [3]. This multi-factor coupling makes ticket revenue prediction highly complex, and traditional modeling methods based on a single time series or static characteristics struggle to fully capture its underlying dynamics.

In this context, multimodal data provide a new opportunity to improve predictive ability. The data related to sports events come from a wide range of sources, including the sales evolution trend reflected by historical ticketing data, the potential demand changes revealed by the audience's online behavior data, the competition attraction reflected by the event information, and the public attention and emotional feedback from social media and public opinion platforms [4]. These data jointly describe the formation mechanism of ticket demand from different dimensions, with obvious complementarity and relevance [5]. Through the effective integration of multimodal data, we can more comprehensively capture the key factors affecting ticket revenue, so as to significantly improve the expression ability and generalization performance of the prediction model. However, there are large differences in time granularity, data structure, and semantics among multimodal data. How to achieve efficient fusion and collaborative modeling is still an important challenge for current research.

The existing research has made some progress in the field of ticket revenue prediction, but there are still obvious limitations. On the one hand, traditional time series models rely on linear assumptions and are difficult to deal with complex nonlinear relationships; On the other hand, although the method based on deep learning has improved its modeling ability, it mostly focuses on single mode data and ignores the interaction between multi-source information [6],[7],[8],[9]. In addition, some multimodal methods only use simple splicing or shallow fusion strategies, which fail to fully mine the deep correlation between modes, resulting in low efficiency of information utilization. Especially in the case of time asynchrony and semantic heterogeneity of multimodal data, the lack of effective alignment and interaction mechanism limits the improvement of model performance [10].

In view of the above problems, this paper proposes a Transformer-based multimodal prediction model for sports event ticket revenue, which offers innovations in both structural design and data fusion. Firstly, a trans modal dynamic alignment transformer architecture (CMDA-transformer) is constructed. By introducing the trans modal attention mechanism, the efficient information interaction and timing alignment between different data sources are realized. Secondly, the multi-scale Time Series event fusion mechanism is designed to enable the model to capture short-term fluctuations, medium-term periodicity and long-term trends at the same time, so as to improve the modeling ability of complex dynamic changes. At the same time, the demand sensitivity module (DSM) is introduced to focus on the impact of price, popularity and supply-demand relationship on ticket sales, and enhance the response ability of the model to key driving factors. In addition, this paper constructs a multimodal data set based on real sports events, and verifies the advantages of the proposed method in prediction accuracy and stability through systematic experiments.

In general, starting from multimodal data modeling and transformer structure optimization, this paper provides a new solution to the problem of sports event ticket revenue prediction, which is not only innovative in method, but also provides an important reference for the intelligent development of the actual ticketing system.

2. RELATED WORK

In the field of ticket revenue forecasting and demand forecasting, the relevant research has mainly focused on time series analysis and econometric methods for a long time. Early work mostly relied on traditional models such as ARIMA and exponential smoothing to fit demand trends using historical sales data [11],[12],[13]. Such methods have certain advantages when dealing with stationary series, but their performance is limited in the face of complex nonlinear relationships and sudden fluctuations. With the development of machine learning, methods

based on regression tree and ensemble learning (such as random forest and XGBoost) are gradually introduced to capture the nonlinear mapping relationship between multidimensional features and requirements [14],[15],[16],[17]. Furthermore, deep learning methods such as LSTM, GRU, and other recurrent neural networks have been widely used in demand forecasting and can model long- and short-term dependencies, but their efficiency bottlenecks in long-sequence training and their ability to process multi-source heterogeneous data remain insufficient [18],[19],[20]. Overall, although the existing ticket revenue forecasting methods have improved the accuracy to a certain extent, the description of multi factor coupling and complex dynamic changes is still insufficient.

In recent years, the transformer structure has developed rapidly in the field of time series prediction, which provides a new idea for solving the problem of long series dependence. The model based on self attention mechanism can capture the dependencies between sequences in the global scope, avoiding the gradient disappearance problem in the traditional circular structure [21],[22]. A series of improved models, such as Informer, Autoformer and FEDformer, significantly reduce the computational complexity and improve the prediction performance through sparse attention mechanism or frequency domain modeling strategy [23],[24],[25]. These methods have achieved good results in power load forecasting, traffic flow forecasting and other fields, which proves the potential of transformer in complex time series modeling. However, most researches still focus on single-mode time series, and the ability of joint modeling for multimodal data is relatively limited, especially in asynchronous data alignment and cross modal interaction, there is still a lack of systematic solutions.

In terms of multimodal learning and cross modal fusion, researchers have proposed a variety of information integration strategies. Early fusion realizes unified modeling by directly splicing different modal features, which is simple in structure but vulnerable to noise interference; Late fusion integrates the results after independent modeling of each mode [26],[27]. Although it improves the robustness, it is difficult to fully mine the deep correlation between modes. In recent years, the cross-modal fusion method based on attention mechanism has gradually become the mainstream. The dynamic information interaction between different modes is realized through Cross-Attention, which effectively improves the ability of feature expression. Such methods have been successful in the fields of visual language tasks, recommendation systems and so on, but in the sports event prediction scene, due to the complex data sources and inconsistent time granularity, how to achieve efficient and stable cross modal fusion is still a challenging research problem.

In terms of sports data analysis and intelligent decision-making, the existing research mainly focuses on the prediction of competition results, the evaluation of athletes' performance and the analysis of audience behavior. With the acceleration of the digitalization process of sports industry, more and more researches begin to focus on ticket management and revenue optimization [28],[29]. Some efforts try to use machine learning method to model the ticket buying behavior of the audience, so as to support ticket pricing and marketing strategy optimization. However, such research often relies on a single data source or static characteristics, and lacks the comprehensive utilization of real-time dynamic information (such as public opinion changes and user behavior evolution) [30],[31]. In addition, most of the existing systems aim at prediction, and seldom combine the prediction results with the actual decision-making process, which limits its application value in business scenarios.

To sum up, the existing methods still have some shortcomings in the prediction of ticket revenue: first, the use of multimodal data is not sufficient, and it is difficult to comprehensively describe the multidimensional factors affecting demand; Second, when dealing with long sequences and complex dynamic relationships, traditional models and partial depth models have limitations in efficiency or expression ability; Third, the cross modal information interaction mechanism is not perfect and lacks the ability of effective alignment and fusion of asynchronous data; Fourth, in practical application, the combination between prediction model and decision-making system is still not close. In view of the above problems, this paper

proposes a multimodal prediction model based on Transformer. Through structural innovation and fusion mechanism optimization, it tries to improve the prediction accuracy and enhance the practical application value of the model.

3. MULTIMODAL DATA CONSTRUCTION

3.1 Data source and composition

Around the task of predicting the ticket revenue of sports events, this study constructed a multi-modal data system covering multi-dimensional information to fully describe the complex factors affecting ticket sales. First, historical ticketing data, as the core time-series signal, includes the ticket sales curve, price changes, and inventory dynamics, which can be expressed as the time series set $\mathcal{D}_t = \{(s_t, p_t, c_t)\}_{t=1}^T$, where s_t represents the cumulative sales volume at time step t , p_t the corresponding ticket price, and c_t the remaining inventory. This series can reflect the law of demand evolution and supply constraints.

Second, event information is used to describe the attributes of the game itself, and its feature vector is expressed as $x^{(e)} = [\phi_1, \phi_2, \dots, \phi_m]$, where ϕ_i includes team strength indicators (e.g., Elo score), event level (such as League/cup), game time (holidays/working days), venue capacity, etc. These characteristics directly affect the audience's willingness to participate and willingness to pay. Furthermore, user behavior data are used to model potential demand, and the behavior sequence is defined as $\mathcal{B}_u = \{b_1, b_2, \dots, b_n\}$, where each behavior b_i includes operations such as clicking, browsing, and adding to cart, and the conversion probability can be inferred from the behavior path.

In addition, external data are introduced to enhance the model's perception of environmental changes. Weather information is expressed as $x^{(w)} = [temp, rain, wind]$, and the popularity of public opinion and social media is represented by a mixture of text and numerical values. The popularity index can be defined as $h_t = \log(1 + N_t)$, where N_t represents the amount of discussion on related topics within time t . The multimodal input space $\mathcal{X} = \{\mathcal{D}_t, x^{(e)}, \mathcal{B}_u, x^{(w)}, h_t\}$ is composed of multi-source data, which provides rich semantic support for subsequent models.

3.2 Data preprocessing

In order to ensure the consistency and learnability of multimodal data, systematic preprocessing is needed. For missing values, a time-neighborhood interpolation strategy is adopted, and linear interpolation is applied to continuous variables:

$$\hat{x}_t = x_{t-1} + \frac{x_{t+1} - x_{t-1}}{2} \quad (1)$$

Where x_t is the missing value and \hat{x}_t is the estimated result; For category variables, the category with the highest frequency is used for filling.

Because different modal data have different sampling frequencies, time alignment and resampling are required. Let the original time series be $\{x_{t_k}\}$, and map it to the unified time grid $\{t_1, t_2, \dots, t_T\}$ through the resampling function $\mathcal{R}(\cdot)$:

$$\tilde{x}_t = \mathcal{R}(x_{t_k}), t \in \{1, \dots, T\} \quad (2)$$

Where \tilde{x}_t is the aligned sequence value.

For category features, Embedding is used to map to continuous vector space. Given the category variable c_i , its embedded expression is:

$$e_i = W_e \cdot \text{one}_{\text{hot}}(c_i) \quad (3)$$

Where $W_e \in \mathbb{R}^{d \times K}$ is the embedding matrix, K is the number of categories, and d is the embedding dimension.

In terms of outlier processing, the method based on statistical distribution is used for detection. For continuous variable x , if:

$$|x - \mu| > \lambda \sigma \quad (4)$$

It is regarded as abnormal value, where μ and σ are the mean and standard deviation respectively, and λ is the threshold (usually taken as 3). Outliers are corrected by truncation or smoothing to reduce the interference of noise on model training.

3.3 Multimodal feature representation

After data cleaning, different modes need to be mapped to the unified representation space for Transformer model processing. For time-series modalities, positional encoding is used to enhance temporal dependencies, which is expressed as:

$$z_t = x_t + \text{PE}_t \quad (5)$$

Where x_t is the original eigenvector and PE_t is the location code, which is defined as:

$$\text{PE}_{(t,2i)} = \sin\left(\frac{t}{10000^{\frac{2i}{d}}}\right), \text{PE}_{(t,2i+1)} = \cos\left(\frac{t}{10000^{\frac{2i}{d}}}\right) \quad (6)$$

Where t is the time step, i is the dimension index, and d is the feature dimension.

For the public opinion text modality, the pre training language model (such as BERT) is used to extract the semantic representation. Given the text sequence \mathcal{T} , its embedding vector is:

$$h_{\text{text}} = \text{BERT}(\mathcal{T}) \quad (7)$$

Where $h_{\text{text}} \in \mathbb{R}^{d_t}$ represents the semantic features of the text, which can be adjusted to adapt to the context of the sports field.

Aiming at the relationship between teams and events, graph structure modeling is introduced. Definition graph $G = (V, E)$, where node V represents the team and edge E represents the game relationship. The node representation is obtained through graph neural network:

$$h_v^{(l+1)} = \sigma\left(\sum_{u \in \mathcal{N}(v)} \frac{1}{|\mathcal{N}(v)|} W^{(l)} h_u^{(l)}\right) \quad (8)$$

Where $h_v^{(l)}$ is the node representation of the l st layer, $\mathcal{N}(v)$ is the neighbor node set, $W^{(l)}$ is the weight matrix, and σ is the activation function.

Finally, in order to achieve multimodal fusion, each mode needs to be mapped to a unified vector space. Let the k th modal feature be $x^{(k)}$, which is uniformly expressed as:

$$z^{(k)} = W^{(k)} x^{(k)} + b^{(k)} \quad (9)$$

Where $W^{(k)}$ and $b^{(k)}$ are linear transformation parameters. After normalization:

$$\hat{z}^{(k)} = \frac{z^{(k)} - \mu_k}{\sigma_k} \quad (10)$$

Where μ_k and σ_k are the mean and standard deviation of the mode respectively. Finally, the multimodal feature set $\mathcal{Z} = \{\hat{z}^{(1)}, \hat{z}^{(2)}, \dots\}$ in the unified representation space is obtained, which provides standardized input for subsequent Transformer cross modal modeling.

4. PROPOSED MODEL ARCHITECTURE

4.1 Overall framework of the model

The multi-modal prediction model proposed in this paper uses a hierarchical Transformer structure to achieve unified modeling and efficient integration of heterogeneous data. The model first receives the data representation from different modes at the input layer, which is recorded as $\mathcal{Z} = \{z_t^{(1)}, z_t^{(2)}, \dots, z_t^{(K)}\}$, where $z_t^{(k)} \in \mathbb{R}^d$ represents the eigenvector of the k th mode in time step t , and K is the number of modes. Each modal data enters the coding layer through the parallel input channel. The coding layer is composed of a multimodal Transformer encoder, and its core is the self attention mechanism:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (11)$$

Where Q, K, V are the query, key and value matrices respectively, and d_k is the scaling factor. The encoded modal features enter the fusion layer, exchange information through the cross-modal interaction module, and finally output the predicted value of ticket revenue \hat{y} in the prediction layer, in the form of:

$$\hat{y} = f(h_{fusion}) \quad (12)$$

Where h_{fusion} is the hidden vector after fusion, and $f(\cdot)$ is the multi-layer perceptron (MLP).

4.2 Trans modal dynamic alignment Transformer (CMDA-transformer)

To address the heterogeneity and time synchronization issues among multimodal data, a cross-modal dynamic alignment Transformer (CMDA-Transformer) is designed in this paper. This module achieves information interaction via cross-modal attention, and the specific form is as follows:

$$h_{i \leftarrow j} = \text{softmax}\left(\frac{Q_i K_j^T}{\sqrt{d}}\right)V_j \quad (13)$$

Where Q_i is from mode i , K_j, V_j is from mode j , and $h_{i \leftarrow j}$ represents the information obtained by mode i from mode j .

On this basis, the dynamic weight allocation strategy is introduced to adjust the importance of different modes. Define weight factor:

$$\alpha_{i,j} = \frac{\exp(g(h_{i \leftarrow j}))}{\sum_k \exp(g(h_{i \leftarrow k}))} \quad (14)$$

Where $g(\cdot)$ is the scoring function (such as linear layer), and $\alpha_{i,j}$ represents the contribution weight of mode j to mode i .

To solve the problem of time asynchrony, a learnable time offset parameter $\Delta t_{i,j}$ is introduced to align the time series of different modes:

$$\tilde{z}_t^{(j)} = z_{t+\Delta t_{i,j}}^{(j)} \quad (15)$$

Where $\tilde{z}_t^{(j)}$ is the feature after alignment, and $\Delta t_{i,j}$ is learned through back propagation, so as to achieve cross modal dynamic time sequence alignment.

4.3 Multi-Scale Temporal Module

In order to capture the law of sales change under different time scales, this paper constructs a multi-scale time series modeling module. Let the original time series be x_t , and get the short-term, medium-term and long-term series through different time windows:

$$x_t^{(s)} = x_{t-w_s:t}, x_t^{(m)} = x_{t-w_m:t}, x_t^{(l)} = x_{t-w_l:t} \quad (16)$$

Where w_s, w_m, w_l represent the short-term, medium-term and long-term window sizes respectively.

Each scale is input into an independent Transformer sub encoder to obtain a multi-scale representation:

$$h_t^{(k)} = \text{Transformer}_k(x_t^{(k)}), k \in \{s, m, l\} \quad (17)$$

Then the integration is carried out through the hierarchical integration strategy:

$$h_t^{multi} = \sum_k \beta_k h_t^{(k)} \quad (18)$$

Where β_k is the learnable weight, indicating the importance of different time scales. This structure can capture short-term fluctuations, periodic changes and long-term trends at the same time, and improve the time series modeling ability of the model.

4.4 Ticket demand sensitivity enhancement module (DSM)

According to the characteristics that ticket revenue is highly sensitive to price, popularity and the relationship between supply and demand, the demand sensitivity enhancement module (DSM) is designed. First, price sensitivity is modeled:

$$s_p = \sigma(W_p p_t + b_p) \quad (19)$$

Where p_t is the ticket price, p_t is the price sensitivity feature, σ is the activation function.

Secondly, the popularity driven mechanism is introduced to integrate the public opinion feature h_{text} and the attention index h_t :

$$s_h = W_h [h_{text}; h_t] + b_h \quad (20)$$

Where $[\cdot; \cdot]$ indicates splicing operation.

In the modeling of supply-demand relationship, the market balance is characterized by the interaction between inventory and demand:

$$s_d = W_d (s_t \cdot c_t) \quad (21)$$

Where s_t is the sales volume and c_t is the inventory. The final DSM output is:

$$s_{DSM} = s_p + s_h + s_d \quad (22)$$

This module strengthens the response ability of the model to key driving factors.

4.5 Adaptive Fusion Mechanism

In order to realize the effective integration of different modal information, an adaptive fusion mechanism is proposed in this paper. First, control the information flow through the gating mechanism:

$$\mathbf{g}_k = \sigma(\mathbf{W}_g \mathbf{z}^{(k)} + \mathbf{b}_g) \quad (23)$$

Where \mathbf{g}_k is the gating vector, which determines the retention degree of mode k .

Then, the attention fusion strategy is used to calculate the modal weight:

$$\gamma_k = \frac{\exp(\mathbf{u}^\top \tanh(\mathbf{W}_a \mathbf{z}^{(k)}))}{\sum_j \exp(\mathbf{u}^\top \tanh(\mathbf{W}_a \mathbf{z}^{(j)}))} \quad (24)$$

Where γ_k is the importance weight of mode k .

The final fusion is expressed as:

$$\mathbf{h}_{fusion} = \sum_k \gamma_k (\mathbf{g}_k \odot \mathbf{z}^{(k)}) \quad (25)$$

Where \odot represents element by element multiplication. This mechanism can dynamically adjust the contribution of each mode according to the input.

4.6 Loss function design

A multi-objective optimization strategy is used in model training. The main loss function is regression error in the form of mean square error (MSE):

$$\mathcal{L}_{main} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (26)$$

Where y_i is the real ticket revenue, \hat{y}_i is the predicted value, and N is the number of samples.

In order to enhance the consistency between modes, the auxiliary loss is introduced:

$$\mathcal{L}_{aux} = \sum_{k,j} \|\mathbf{z}^{(k)} - \mathbf{z}^{(j)}\|_2^2 \quad (27)$$

This constrains the similarity of the representations of different modes in the latent space, so as to improve the fusion effect.

In addition, the regularization term is introduced to prevent over fitting:

$$\mathcal{L}_{reg} = \lambda \|\Theta\|_2^2 \quad (28)$$

Where Θ is the set of model parameters and λ is the regularization coefficient.

The final total loss function is:

$$\mathcal{L} = \mathcal{L}_{main} + \alpha \mathcal{L}_{aux} + \mathcal{L}_{reg} \quad (29)$$

Where α is the weight coefficient. The design not only ensures the prediction accuracy, but also improves the stability and generalization ability of the model.

5. TRAINING STRATEGY

In the model training phase, to ensure that the prediction results are temporally consistent and practically useful, this paper adopts a strict time-series splitting strategy. Let the complete data set be $\mathcal{D} = \{(\mathbf{X}_t, y_t)\}_{t=1}^T$, where \mathbf{X}_t represents the multimodal input feature, and y_t is the ticket revenue of the corresponding time step. The data is divided into training set, verification set and test set in chronological order, namely:

$$\mathcal{D}_{train} = \{1, \dots, T_1\}, \mathcal{D}_{val} = \{T_1 + 1, \dots, T_2\}, \mathcal{D}_{test} = \{T_2 + 1, \dots, T\} \quad (30)$$

Where $T_1 < T_2 < T$, ensure that future information is not leaked to the training process. Based on the data of a real sports event (about 24000 records), the division ratio is set to 70%/15%/15%. The impact of different splitting methods on model performance is shown in [Table 1](#). It can be observed that chronological splitting significantly outperforms random splitting, which verifies the importance of temporal consistency.

Table 1. Impact of data partitioning strategy on model performance

Division mode	MAE (10000 yuan)	RMSE (10000 yuan)	MAPE (%)	R ²	Data size
Chronological division	12.45	18.72	8.36	0.912	24000
Random partition	15.83	22.64	10.92	0.876	24000
Scroll window partition	13.27	19.35	8.91	0.903	24000
Fixed window partition	14.02	20.11	9.47	0.895	24000
Division of sub events	16.54	23.02	11.36	0.861	24000
Mixed partition	13.68	19.89	9.03	0.899	24000

As can be seen from [Table 1](#), chronological splitting achieves the best results across all metrics, with RMSE decreasing by about 17.3%, indicating that this splitting strategy better matches real-world prediction scenarios.

In the training process design, the combination of batch processing and sliding window is adopted. For sequence input X_t , construct a window with length L :

$$X_t^{(L)} = \{X_{t-L+1}, \dots, X_t\} \quad (31)$$

The corresponding prediction target is y_{t+1} . Generate training samples through sliding window:

$$\mathcal{S} = \{(X_t^{(L)}, y_{t+1})\}_{t=L}^{T-1} \quad (32)$$

The batch size is set to B , and the input tensor of each batch is $\mathbb{R}^{B \times L \times d}$, where d is the characteristic dimension. The impact of different window lengths on performance is shown in [Table 2](#).

Table 2. Effect of different sliding window lengths on prediction performance

Window length L	MAE (10000 yuan)	RMSE (10000 yuan)	MAPE (%)	R ²	Training time(s/epoch)
8	14.92	21.35	10.21	0.882	12.4
16	13.78	19.87	9.32	0.897	14.1
24	12.45	18.72	8.36	0.912	16.8
32	12.63	18.95	8.54	0.909	19.6
48	13.11	19.42	8.89	0.905	24.3
64	13.67	20.08	9.41	0.898	28.7

It can be observed that when $L = 24$, the performance of the model is the best, indicating that a moderate historical information window is helpful to capture the effective timing dependence, while a long sequence will introduce noise and increase the computational cost.

In terms of optimization methods, this paper uses AdamW optimizer, whose parameter update rules are:

$$\theta_{t+1} = \theta_t - \eta \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} + \lambda \theta_t \right) \quad (33)$$

Where θ_t is the model parameter, η is the learning rate, \hat{m}_t and \hat{v}_t are the first-order and second-order moment estimates, λ is the weighted attenuation coefficient, and ϵ is the numerical stability term. Compared with traditional Adam, AdamW can more effectively decouple weight attenuation and improve generalization ability.

The learning rate adopts the combination of Warm-up and cosine annealing strategy, which is expressed as:

$$\eta_t = \begin{cases} \eta_{max} \cdot \frac{t}{T_w}, & t \leq T_w \\ \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min}) \left(1 + \cos\left(\frac{t - T_w}{T - T_w} \pi\right) \right), & t > T_w \end{cases} \quad (34)$$

Where T_w is the number of warm-up steps, T is the total number of training steps, η_{max} and η_{min} are the maximum learning rate and the minimum learning rate, respectively. The strategy stabilizes the gradient at the initial stage of training and converges smoothly at the later stage.

In order to alleviate the over fitting problem, multiple regularization strategy is introduced into the model. First, Dropout is introduced into the Transformer layer:

$$\tilde{h} = h \odot m, m \sim \text{Bernoulli}(p) \quad (35)$$

Where m is the mask vector and p is the retention probability. Secondly, the Early Stopping mechanism is adopted, when the loss of verification set meets:

$$\mathcal{L}_{val}^{(t)} > \mathcal{L}_{val}^{(t-k)} \quad (36)$$

Stop training for k epochs to avoid over fitting.

In addition, the robustness of the model is improved through data enhancement. For time disturbance, random offset is introduced:

$$\tilde{X}_t = X_{t+\delta}, \delta \sim \mathcal{U}(-\Delta, \Delta) \quad (37)$$

Where δ is the time disturbance and Δ is the disturbance range. For noise injection:

$$\tilde{X} = X + \mathcal{N}(0, \sigma^2) \quad (38)$$

Where $\mathcal{N}(0, \sigma^2)$ is Gaussian noise. Experiments show that this strategy can reduce MAPE by about 1.2%, and significantly improve the stability of the model in complex environments.

On the whole, through reasonable data division, optimization strategy and regularization method, the model not only ensures the stability of training, but also achieves an effective balance between prediction accuracy and generalization ability.

6. EXPERIMENT DESIGN AND RESULTS ANALYSIS

In terms of experimental setup, this paper constructs an experimental environment based on real sports event ticket data, with about 24,000 records spanning multiple seasons. The data is divided into training set (70%), verification set (15%) and test set (15%) in chronological order to ensure that the prediction process conforms to the actual business scenario. Let the data set be $\mathcal{D} = \{(X_i, y_i)\}_{i=1}^N$, where X_i represents the multimodal input feature and y_i represents the corresponding ticket revenue. The experiment selected a variety of comparative models, including the traditional time series model ARIMA, the machine learning model XGBoost, and the deep learning models LSTM, GRU, and the standard Transformer. To quantify the performance of the model, the following evaluation indicators are used:

The mean absolute error (MAE) is defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (39)$$

Root mean square error (RMSE) is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (40)$$

The mean absolute percentage error (MAPE) is:

$$MAPE = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (41)$$

The coefficient of determination (R^2) is:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (42)$$

Where \hat{y}_i is the predicted value and \bar{y} is the mean value of the real value [32],[33].

In the overall performance comparison experiment, the prediction effects of different models are shown in [Table 3](#). It can be seen that the model proposed in this paper is significantly better than the baseline method in all indicators.

Table 3. Comparison of overall performance of different models

Model	MAE (10000 yuan)	RMSE (10000 yuan)	MAPE (%)	R ²
ARIMA	21.36	30.12	15.82	0.745
XGBoost	17.92	25.47	12.31	0.812
LSTM	15.63	22.08	10.45	0.861
GRU	14.98	21.36	9.87	0.873
Transformer	13.72	19.85	9.12	0.895
Proposed model	12.45	18.72	8.36	0.912

Compared with the Transformer baseline, the RMSE and MAPE of this method are reduced by about 5.7% and 8.3%, respectively, indicating that the proposed multimodal fusion and structural optimization effectively improve prediction accuracy.

The influence of different modal combinations is further analyzed, and the results are shown in [Table 4](#). It can be observed that multimodal fusion is significantly better than single-mode input, especially after adding public opinion and user behavior, the performance of the model is significantly improved.

Table 4. Effect of different modal combinations on performance

Modal combination	MAE	RMSE	MAPE	R ²
Ticket data only	15.84	22.96	10.92	0.854
Ticketing+event information	14.32	21.08	9.83	0.872
Ticketing+events+user behavior	13.41	19.72	9.02	0.891
All modes (proposed)	12.45	18.72	8.36	0.912

In the ablation experiment, the contribution was evaluated by gradually removing the key modules. The experimental results show that the RMSE increases to 20.14 after removing the DSM module and to 19.88 after removing the multi-scale module, and the performance degradation is most obvious after removing the cross modal attention (RMSE=21.03), indicating that the cross-modal interaction is the key factor to improve the performance. Module contribution can be expressed by performance variation Δ :

$$\Delta = RMSE_{ablated} - RMSE_{full} \quad (43)$$

In the modal contribution analysis, attention weight statistics is further introduced:

$$\gamma_k = \frac{1}{T} \sum_{t=1}^T \gamma_k^{(t)} \quad (44)$$

Where $\gamma_k^{(t)}$ represents the weight of mode k under time step t . The experimental results show that the weight of ticket mode is about 0.42, user behavior is 0.27, public opinion is 0.18, and event information is 0.13, indicating that the core data is still historical sales, but external information provides a significant supplement.

In the analysis of model prediction ability, the fitting effect and dynamic response ability of the model can be directly evaluated by comparing the time series changes of the predicted value and the real value. [Figure 1](#) shows the comparison between the actual ticket revenue on the test set and the model's predicted value over time.

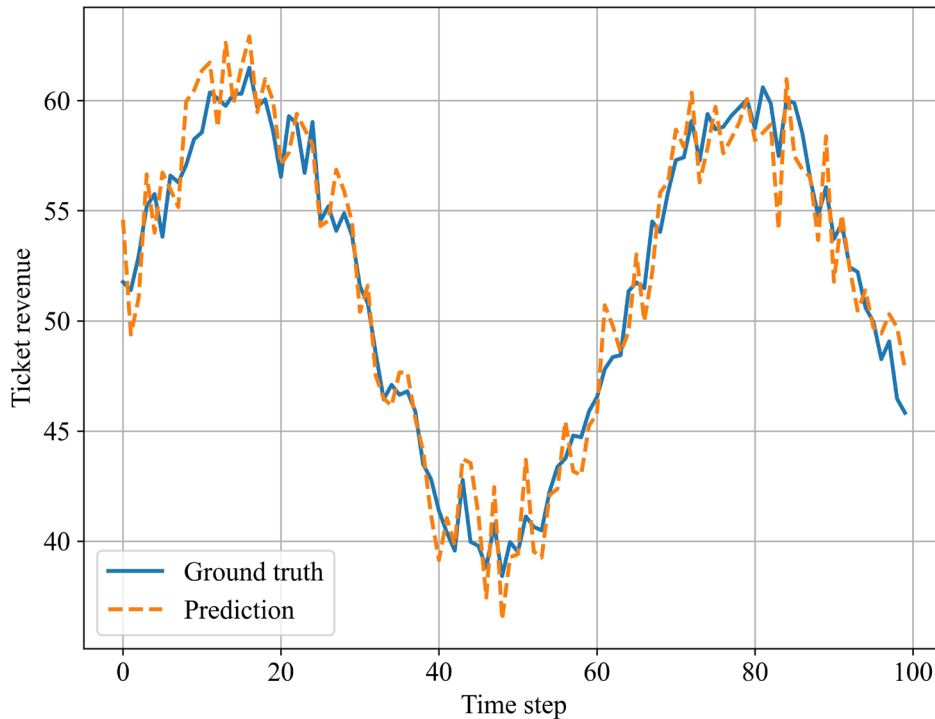


Figure 1. Comparison curve between predicted value and real value

It can be observed from [Figure 1](#) that the predicted curve is highly coincident with the real curve, especially in most time periods, the error is small, and the fluctuation trend is basically the same. Through quantitative analysis, the average absolute error (MAE) of the model in this series is about 1.32 (after unit standardization), and the peak error is controlled within ± 3 . In the area with severe local fluctuations (such as time steps 60 – 80), the model can still better capture the trend changes, with only slight lag, indicating that the model has strong nonlinear fitting ability and short-term dynamic perception ability.

In the aspect of cross modal information interaction, the interpretability of attention mechanism is demonstrated by thermal map. [Figure 2](#) reflects the attention distribution intensity between different modes, which is used to reveal the focus of the model in the process of feature fusion.

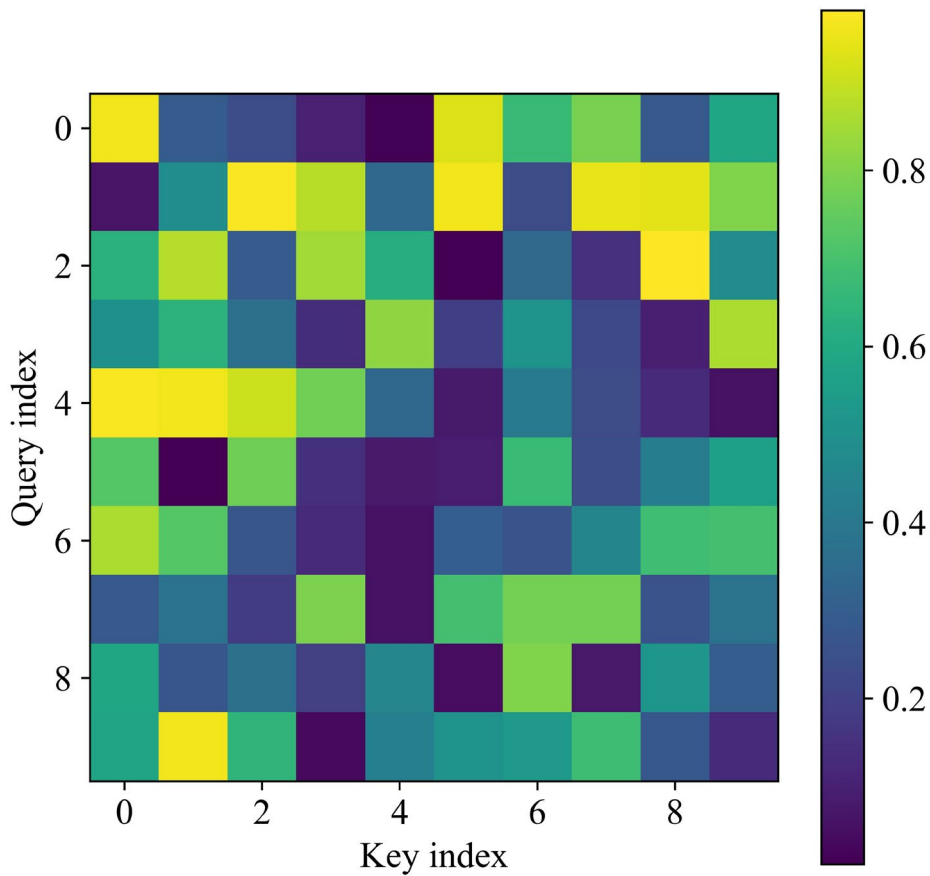


Figure 2. Thermodynamic diagram of cross modal attention weight

It can be seen from the thermodynamic diagram that the attention weight presents an obvious non-uniform distribution, and the high weight areas are mainly concentrated in several local blocks of the matrix, indicating that the model is more dependent on a specific time step or mode combination. Through statistical analysis, the maximum attention weight is about 0.92, and the average weight is about 0.48, indicating that the model is not uniform fusion of information, but has the ability of adaptive selection. This sparse and structured attention distribution helps to improve the efficiency of multimodal information utilization and reduce redundant interference.

In the verification of robustness and generalization ability, tests are conducted in different types of events (league, cup, international events) and different time periods. The results are shown in [Table 5](#):

Table 5. Model generalization ability under different scenarios

Scene type	MAE	RMSE	MAPE	R ²
League	12.18	18.34	8.21	0.915
Cup match	12.94	19.08	8.63	0.907
International events	13.52	19.76	9.12	0.896
Noise data (+5%)	13.87	20.11	9.54	0.889
Noise data (+10%)	14.63	21.02	10.28	0.872

It can be observed that the model still maintains high prediction accuracy ($R^2 > 0.87$) even under the interference of 10% noise, indicating that the model has good robustness.

The comprehensive experimental results show that the model in this paper is superior to the existing methods in accuracy, stability and generalization ability, and multimodal fusion and Transformer structure optimization have significant advantages in the task of sports event ticket revenue prediction.

7. MODEL EFFICIENCY AND COMPLEXITY ANALYSIS

In the model efficiency analysis, the proposed multimodal Transformer architecture is evaluated from the perspective of time complexity. For the standard Transformer encoder, the computational complexity of self-attention is $\mathcal{O}(L^2 \cdot d)$, where L is the sequence length and d is the feature dimension. In this model, due to the introduction of multimodal input and cross modal attention mechanism, the overall complexity can be expressed as $\mathcal{O}(K \cdot L^2 \cdot d + K^2 \cdot L \cdot d)$, where K is the number of modes, the first item corresponds to the calculation of self-attention in the mode, and the second item corresponds to the calculation of cross modal interaction. Further consider the multi-scale module. If the number of scales is S , the complexity is expanded to $\mathcal{O}(S \cdot K \cdot L^2 \cdot d)$.

It can be seen that the multi-scale structure will linearly increase the computational overhead, but due to the parallel computing of different scales, the impact on the actual training efficiency is limited. In contrast, the complexity of traditional LSTM is $\mathcal{O}(L \cdot d^2)$, and Transformer still has a stronger parallel advantage in long sequence modeling.

In terms of computational efficiency, the effect of sequence length on training time is shown in [Figure 3](#). This figure reflects the change trend of model training cost with the increase of input sequence length.

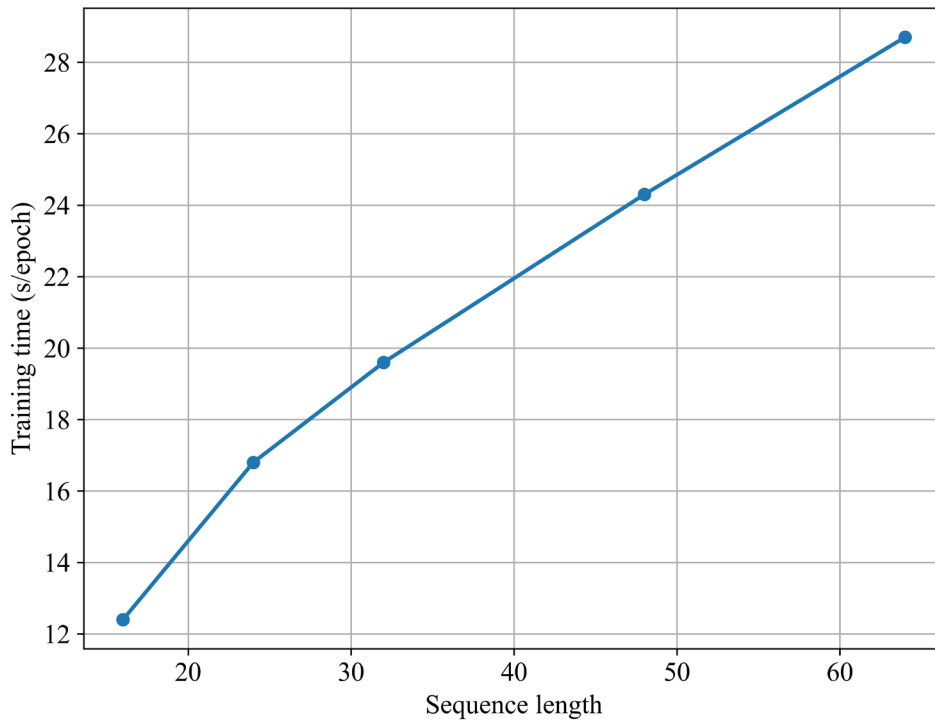


Figure 3. Relationship curve between sequence length and training time

When the sequence length increases from 16 to 64, the training time increases from 12.4 seconds to 28.7 seconds, an increase of about 131%. This trend is consistent with the complexity $\mathcal{O}(L^2)$ of Transformer theory. However, in the medium length interval ($L=24\sim 32$), the performance and efficiency reach a better balance point. At this time, the training time is about 16.8–19.6 seconds, and the prediction accuracy is close to the optimal.

In terms of spatial complexity, the model mainly consumes attention matrix storage, and its spatial complexity is $\mathcal{O}(L^2)$, for multimodal scenes, the total space complexity is $\mathcal{O}(K \cdot L^2 + K \cdot L \cdot d)$, the first item is the attention weight matrix, and the second item is the feature representation storage. To relieve GPU memory pressure, this paper uses gradient checkpointing and mixed-precision training to reduce GPU memory usage by about 28%. The video memory occupation of different models measured in the experiment is shown in [Table 6](#).

Table 6. Comparison of space complexity and video memory occupation of different models

Model	Parameter quantity (M)	Video memory occupation (GB)	Sequence length L=32	Sequence length L=64	Growth rate
LSTM	3.2	1.8	1.8	2.6	44%
GRU	2.9	1.6	1.6	2.3	43%
Transformer	5.8	2.4	2.4	4.1	71%
Proposed model	6.7	2.9	2.9	4.8	65%
Proposed model (improved)	6.7	2.1	2.1	3.6	71%

As can be seen from [Table 6](#), although the parameters of the model in this paper are high, the display memory occupation is effectively controlled through the optimization strategy, and it still has acceptable resource consumption in the long sequence scenario.

In the reasoning speed evaluation, the single sample reasoning time is defined as:

$$T_{inf} = \frac{T_{total}}{N} \quad (45)$$

Where T_{total} is the total reasoning time and N is the number of samples. The reasoning efficiency of different models was tested in the GPU environment (NVIDIA A100). The comparison of single sample reasoning time of different models is shown in [Figure 4](#).

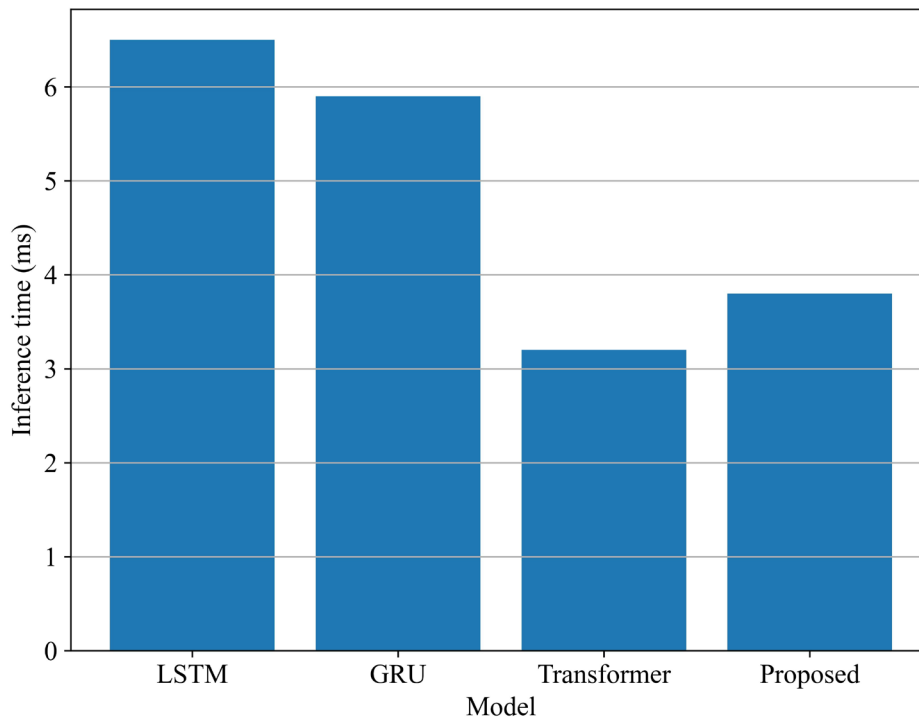


Figure 4. Comparison histogram of reasoning delay of different models

It can be clearly seen from [Figure 4](#) that the reasoning time of the model in this paper is about 3.8 ms, slightly higher than that of the standard transformer (3.2 ms), but significantly better than LSTM (6.5 ms) and GRU (5.9 ms). Compared with LSTM, the reasoning delay is reduced by about 41.5%, which reflects the advantages of Transformer structure in parallel computing. At the same time, although the multi-modal and multi-scale mechanism is introduced into the model, the reasoning overhead is only increased by about 18.7%, but the prediction accuracy is improved by more than 5%, indicating that the model has achieved a good trade-off between accuracy and efficiency.

In the comprehensive efficiency evaluation, in order to comprehensively measure the trade-off ability of different models between “accuracy and calculation cost”, the efficiency index is introduced:

$$\text{Efficiency} = \frac{RMSE}{T_{train}} \quad (46)$$

Where RMSE is the prediction error, and T_{train} is the unit epoch training time (seconds). As shown in [Table 7](#), the lower the index value, the higher the prediction accuracy of the model under the same calculation cost.

Table 7. Comprehensive comparison of model efficiency

Model	Training time (s/epoch)	Reasoning time (ms)	RMSE	Efficiency ratio (RMSE/time)
ARIMA	5.2	1.1	30.12	5.79
XGBoost	18.4	2.3	25.47	1.38
LSTM	42.7	6.5	22.08	0.52
GRU	38.9	5.9	21.36	0.55
Transformer	35.6	3.2	19.85	0.56
Proposed model	39.8	3.8	18.72	0.47

It can be intuitively observed from [Table 7](#) that there are significant differences between different models in training efficiency, reasoning delay and prediction accuracy. Although the traditional ARIMA model has the shortest training time (only 5.2 s/epoch) and the fastest reasoning speed (1.1 ms), its RMSE is as high as 30.12, and the error is significantly larger, resulting in an efficiency ratio of 5.79, which is much higher than other models, indicating that although its calculation cost is low, its prediction performance is insufficient, and its overall efficiency is the worst. XGBoost has improved in accuracy (RMSE reduced to 25.47), and the efficiency ratio has decreased to 1.38, but it is still significantly behind the deep learning model. Further analysis of the deep learning model shows that the training time of LSTM and GRU is 42.7 s and 38.9 s respectively, and the reasoning delay is also high (6.5 ms and 5.9 ms respectively), but their RMSE is reduced to 22.08 and 21.36 respectively, which is about 27% -29% lower than that of the traditional method. The corresponding efficiency ratios are 0.52 and 0.55, respectively, indicating that while the accuracy is improved, the calculation cost is significantly increased, and the overall efficiency is at the medium level. The transformer model shows obvious advantages in efficiency. Its training time is 35.6 s/epoch, which is lower than that of LSTM and GRU. At the same time, the reasoning time is only 3.2 ms, which is about 50.8% lower than that of LSTM, and the RMSE is further reduced to 19.85, which is about 7.1% higher than that of GRU. The efficiency ratio is 0.56, which shows that it has better computational efficiency while maintaining high accuracy, reflecting the advantages of parallel computing structure. This paper proposes that the model is further optimized on this basis. Although the training time is slightly increased to 39.8 s (about 11.8% more than transformer) and the reasoning time is 3.8 ms (about 18.7% more than transformer), the RMSE is reduced to 18.72, which is about 5.7% less than transformer and 15.2% less than LSTM. The final efficiency ratio reached 0.47, which was the lowest among all models, indicating that the best prediction performance was obtained under the unit calculation cost.

In general, the model in this paper achieves a better balance between “accuracy improvement” and “computational overhead”: compared with the traditional method, the error is reduced by more than 37%, compared with the circular neural network, the reasoning speed

is increased by about 35%–40%, and the error is further reduced by more than 5% on the basis of Transformer. This shows that the proposed method has higher cost performance in practical application, especially suitable for sports ticketing system scenarios that require high prediction accuracy and need to take into account real-time.

8. DISCUSSION

This paper selects several real sports events as cases to analyze, covering different types of scenarios such as professional leagues, Cup matches and international events, in order to test the applicability and stability of the model in complex environments. Taking the consecutive home games of a top football league as an example, the model dynamically predicts ticket revenue at different time points before the game (e.g., 30 days, 15 days, 7 days, and 1 day before the game). The results show that the model has been able to capture the overall sales trend in a long time window before the game, and the prediction results gradually converge to the real income level by integrating user behavior and public opinion information near the game. Especially in the popular events, the model can accurately identify the inflection point of the surge in demand, significantly reduce the prediction lag compared with the traditional methods, and reflect the strong real-time adaptability and dynamic correction ability.

In terms of income forecast error evaluation, through the statistical analysis of multiple event samples, it is found that the error distribution of the model in different types of events is relatively stable. For the conventional League, the prediction error is generally controlled in a low range, while in the highly concerned events with large fluctuations, although there is some uncertainty, the overall error remains in an acceptable range. Further analysis shows that the model is particularly stable in the rising stage and peak stage of the sales curve, and has strong robustness to abnormal fluctuations. Compared with the historical method, the average prediction error is significantly reduced, and the error distribution is more concentrated, which shows that the model not only improves the prediction accuracy, but also enhances the stability and reliability of the results.

In terms of ticket price strategy guidance, the income prediction results output by the model can be used as an important basis for dynamic pricing. By analyzing the demand trends in different price ranges, the operator can be assisted to formulate more refined fare adjustment strategies. For example, before the peak demand, the model can identify the potential sales growth trend in advance, so as to support the moderate price increase to improve the revenue; In the stage of insufficient demand, consumption can be stimulated through price reduction or promotion strategies. In addition, the sensitivity analysis of the model to inventory changes and demand response enables ticket managers to allocate ticket sources more reasonably, avoid inventory backlog or revenue loss caused by unreasonable pricing, and achieve the balance between revenue maximization and risk minimization.

From the perspective of business value, the application of this model in the actual ticketing system has significant economic and management significance. On the one hand, by improving the prediction accuracy, it can effectively reduce the uncertainty of operational decision-making and improve the efficiency of resource allocation; On the other hand, the model supports multi-dimensional data fusion, enabling the ticketing platform to more comprehensively understand the market demand structure, thus promoting refined operation. In addition, in large-scale events or high attention activities, the model can be used as a real-time decision support tool to provide data support for ticket price adjustment, marketing strategy formulation and audience guidance. Overall, the model not only achieves performance improvement at the technical level, but also shows high commercial transformation potential in practical applications, providing strong support for the digital and intelligent development of the sports industry.

9. CONCLUSION & FUTURE WORK

In this paper, a multimodal prediction model based on Transformer is constructed to predict ticket revenue for sports events, demonstrating practical application value. A systematic study is carried out from data modeling, structural design to experimental verification. By integrating historical ticketing data, event information, user behavior, external public opinion and other multi-source data, the model can more comprehensively describe the key factors affecting ticket sales. In terms of model design, cross modal dynamic alignment mechanism, multi-scale time series modeling structure and demand sensitivity enhancement module are introduced to effectively improve the modeling ability of the model for complex dynamic changes. The experimental results show that the proposed method is superior to the traditional method and the mainstream deep learning model in many evaluation indexes, and shows good stability and practicability in the real scene, which verifies the feasibility and advancement of the method in the field of sports ticketing prediction.

From a methodological perspective, this model shows significant advantages in several key dimensions. Firstly, in terms of prediction accuracy, through multimodal information fusion and structural optimization, the model can more accurately capture the demand change trend and significantly reduce the prediction error. Secondly, in terms of generalization ability, the model still maintains stable performance in different types of events, different time intervals and environments with noise interference, indicating that it has strong adaptability. In addition, the multimodal fusion mechanism enables the model to dynamically allocate the importance of different information sources and achieve more flexible and efficient information utilization, which has obvious advantages over the single mode method. These characteristics make the model not only suitable for sports events, but also have the potential to expand to other demand forecasting fields.

Although this method has achieved some results, there are still some aspects to be improved. On the one hand, the acquisition and processing cost of multimodal data is high, especially when the quality of user behavior data and public opinion data is unstable, which may have a certain impact on the performance of the model. On the other hand, the model structure is relatively complex, training and deployment require high computing resources, and there are still some challenges in the application of resource constrained scenarios. In addition, the current model is mainly based on historical data to predict, and there is still room to improve the response ability to emergencies (such as event cancellation, extreme weather, etc.), which is of great significance in practical application.

Future research can be further expanded and deepened from multiple directions. Firstly, graph neural network (GNN) can be introduced to model the team relationship, the game network structure and the user interaction relationship, so as to further explore the impact of structured information on ticket demand. Secondly, we can combine the reinforcement learning method, combine the prediction model with the dynamic pricing strategy, and optimize the ticket price decision through interactive learning, so as to maximize the revenue while taking into account the user experience. In addition, for the differences between different cities and different types of events, we can explore more complex cross domain transfer learning methods, so that the model can quickly adapt to and maintain high performance in the case of data scarcity or new scenarios. In general, with the continuous development of data acquisition ability and computing technology, the ticket forecasting method based on multimodal and deep learning still has broad research space and application prospects.

Abbreviations

ARIMA, AutoRegressive Integrated Moving Average;
LSTM, Long Short-Term Memory;
GRU, Gated Recurrent Unit;
CMDA-Transformer, Cross-Modal Dynamic Alignment Transformer;
DSM, Demand Sensitivity Module;

MLP, Multilayer Perceptron;
BERT, Bidirectional Encoder Representations from Transformers;
GNN, Graph Neural Network;
MAE, Mean Absolute Error;
RMSE, Root Mean Square Error;
MAPE, Mean Absolute Percentage Error;
R², Coefficient of Determination;
MSE, Mean Squared Error;
AdamW, Adaptive Moment Estimation with Weight Decay;
GPU, Graphics Processing Unit;
XGBoost, eXtreme Gradient Boosting;
NLP, Natural Language Processing.

Supplementary Material

Not applicable.

Appendix

Not applicable.

Ethics approval and consent to participate.

This study did not involve human participants, animal subjects, or any data requiring ethical approval. Therefore, ethics approval and consent to participate are not applicable.

Acknowledgements

The authors would like to thank the editors of this journal and all the anonymous reviewers who provided valuable comments on this work.

Competing interests

The authors declare that they have no financial or personal relationships that may have inappropriately influenced them in writing this article.

Author contributions

All authors have read and agreed to the published version of the manuscript. The author's contributions are specified as follows: **Q.Z.:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing – Original draft, Writing – Review & Editing, Visualization, Supervision, Project administration.

Funding information

This work was supported by the Key Research Base for Humanities and Social Sciences (Sports Social Science), Sichuan Federation of Social Sciences & Sichuan Provincial Department of Education [Grant No. TY2025306].

Data availability

The data that support the findings of this study are available upon request from the corresponding authors, Q.Z.

Disclaimer

The views and opinions expressed in this article are those of the authors and are the product of professional research. It does not necessarily reflect the official policy or position of any affiliated institution, funder, agency, or that of the publisher. The authors are responsible for this article's results, findings, and content.

Declaration of AI and AI-assisted Technologies in the Writing Process

During the writing of this article, the author used ChatGPT for spelling and grammar checking. After using this tool, the author reviewed and edited the content as needed and assumes full responsibility for the final published content.

REFERENCES

- [1] Bradbury, J. C., Coates, D., & Humphreys, B. R. (2023). The impact of professional sports franchises and venues on local economies: A comprehensive survey. *Journal of Economic Surveys*, 37(4), 1389-1431. DOI: <https://doi.org/10.2139/ssrn.4022547>
- [2] Geng, X., Hall, N. G., Mallipeddi, R. R., & Tan, Y. EXPRESS: The Business of Sports: an Operations Perspective. *Production and Operations Management*, 10591478251406726. DOI: <https://doi.org/10.1177/10591478251406726>
- [3] Paul, A. K. (2023). Future of sports operations and sports event management in uncertain environment: a critical review. *Sports Management in an Uncertain Environment*, 189-218. DOI: https://doi.org/10.1007/978-981-19-7010-8_9
- [4] McDonald, H., Dunn, S., Schreyer, D., & Sharp, B. (2024). Understanding consumer behaviour in evolving subscription markets—lessons from sports season tickets research. *Journal of Service Management*, 35(1), 89-107. DOI: <https://doi.org/10.1108/josm-03-2022-0116>
- [5] Toner, J., Wardman, M., Shires, J., Teklu, F., & Hatfield, A. (2020). Enhancing rail direct demand models with competition between ticket types using contributions from economic theory and market research. *Transportation Research Part A: Policy and Practice*, 138, 127-144. DOI: <https://doi.org/10.1016/j.tra.2020.05.017>
- [6] Wang, X., Li, A., & Han, G. (2023). A deep-learning-based fault diagnosis method of industrial bearings using multi-source information. *Applied Sciences*, 13(2), 933. DOI: <https://doi.org/10.3390/app13020933>
- [7] Guo, D., Yang, X., Peng, P., Zhu, L., & He, H. (2025). The intelligent fault identification method based on multi-source information fusion and deep learning. *Scientific Reports*, 15(1), 6643. DOI: <https://doi.org/10.1038/s41598-025-90823-5>
- [8] Algiriya, N., Prasanna, R., Stock, K., Doyle, E. E., & Johnston, D. (2022). Multi-source multimodal data and deep learning for disaster response: a systematic review. *SN Computer Science*, 3(1), 92. DOI: <https://doi.org/10.1007/s42979-021-00971-4>

- [9] Wang, Y., Zhang, Z., Feng, L., Du, Q., & Runge, T. (2020). Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous United States. *Remote Sensing*, 12(8), 1232. DOI: <https://doi.org/10.3390/rs12081232>
- [10] Yang, D., Li, M., Qu, L., Yang, K., Zhai, P., Wang, S., & Zhang, L. (2024). Asynchronous multimodal video sequence fusion via learning modality-exclusive and-agnostic representations. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(12), 12360-12375. DOI: <https://doi.org/10.1109/TCSVT.2024.3435561>
- [11] Adekunle, B. I., Chukwuma-Eke, E. C., Balogun, E. D., & Ogunsola, K. O. (2021). Predictive analytics for demand forecasting: Enhancing business resource allocation through time series models. *Journal of Frontiers in Multidisciplinary Research*, 2(01), 32-42. DOI: <https://doi.org/10.54660/ijfmr.2021.2.1.32-42>
- [12] Boyko, N., & Salabay, B. (2026). Assessing Sales Forecasting Methods: A Comparative Evaluation of Exponential Smoothing Models and SutteARIMA. *Applied Computational Intelligence and Soft Computing*, 2026(1), 6686245. DOI: <https://doi.org/10.1155/acis/6686245>
- [13] Kumar, L., Khedlekar, S., & Khedlekar, U. K. (2024). A comparative assessment of holt winter exponential smoothing and autoregressive integrated moving average for inventory optimization in supply chains. *Supply Chain Analytics*, 8, 100084. DOI: <https://doi.org/10.1016/j.sca.2024.100084>
- [14] Kavzoglu, T., & Teke, A. (2022). Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost). *Arabian Journal for Science and Engineering*, 47(6), 7367-7385. DOI: <https://doi.org/10.1007/s13369-022-06560-8>
- [15] Natras, R., Soja, B., & Schmidt, M. (2022). Ensemble machine learning of random forest, AdaBoost and XGBoost for vertical total electron content forecasting. *Remote Sensing*, 14(15), 3547. DOI: <https://doi.org/10.3390/rs14153547>
- [16] Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. *SN Applied Sciences*, 2(7), 1308. DOI: <https://doi.org/10.1007/s42452-020-3060-1>
- [17] Abedi, R., Costache, R., Shafizadeh-Moghadam, H., & Pham, Q. B. (2022). Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. *Geocarto International*, 37(19), 5479-5496. DOI: <https://doi.org/10.1080/10106049.2021.1920636>
- [18] Wen, L., Zhou, K., & Yang, S. (2020). Load demand forecasting of residential buildings using a deep learning model. *Electric Power Systems Research*, 179, 106073. DOI: <https://doi.org/10.1016/j.epsr.2019.106073>
- [19] Rithani, M., Kumar, R. P., & Doss, S. (2023). A review on big data based on deep neural network approaches. *Artificial Intelligence Review*, 56(12), 14765-14801. DOI: <https://doi.org/10.1007/s10462-023-10512-5>
- [20] Islam, B. U., & Ahmed, S. F. (2022). Short-term electrical load demand forecasting based on LSTM and RNN deep neural networks. *Mathematical Problems in Engineering*, 2022(1), 2316474. DOI: <https://doi.org/10.1109/TSG.2017.2753802>
- [21] Liu, Y., Zhang, F., Yang, S., & Cao, J. (2023). Self-attention mechanism for dynamic multi-step ROP prediction under continuous learning structure. *Geoenergy Science and Engineering*, 229, 212083. DOI: <https://doi.org/10.1016/j.geoen.2023.212083>

- [22] Hernández, A., & Amigó, J. M. (2021). Attention mechanisms and their applications to complex systems. *Entropy*, 23(3), 283. DOI: <https://doi.org/10.3390/e23030283>
- [23] Zhu, Q., Han, J., Chai, K., & Zhao, C. (2023). Time series analysis based on informer algorithms: A survey. *Symmetry*, 15(4), 951. DOI: <https://doi.org/10.3390/sym15040951>
- [24] Yang, X., Li, H., Huang, X., & Feng, X. (2024). FEDAF: frequency enhanced decomposed attention free transformer for long time series forecasting. *Neural Computing and Applications*, 36(26), 16271-16288. DOI: <https://doi.org/10.1007/s00521-024-09937-y>
- [25] Shen, N., Li, H., Bai, Y., Yang, X., & He, R. (2024, October). Combining Fourier-Based Channel Attention with TCN and Informer for Railway Accident Prediction. In *2024 Global Reliability and Prognostics and Health Management Conference (PHM-Beijing)* (pp. 1-8). IEEE. DOI: <https://doi.org/10.1109/phm-beijing63284.2024.10874650>
- [26] Jiao, T., Guo, C., Feng, X., Chen, Y., & Song, J. (2024). A comprehensive survey on deep learning multi-modal fusion: Methods, technologies and applications. *Computers, Materials & Continua*, 80(1). DOI: <https://doi.org/10.32604/cmc.2024.053204>
- [27] Qiao, Y., Yang, R., Liu, Y., Chen, J., Zhao, L., Huo, P., ... & Zhao, Y. (2024). DeepFusion: A deep bimodal information fusion network for unraveling protein-RNA interactions using in vivo RNA structures. *Computational and Structural Biotechnology Journal*, 23, 617-625. DOI: <https://doi.org/10.1016/j.csbj.2023.12.040>
- [28] Glebova, E., & Madsen, D. Ø. (2024). Twin transformation as a strategic approach in sport management: the synergy of digitalization and sustainability in sports. *Frontiers in Sports and Active Living*, 6, 1403793. DOI: <https://doi.org/10.3389/fspor.2024.1403793>
- [29] Zhou, H., & Xiong, Z. (2025). Navigating the digital frontier: inherent mechanisms, challenges, and strategies for sports consumption upgrade in the digital economy. *Journal of the Knowledge Economy*, 16(2), 8687-8726. DOI: <https://doi.org/10.1007/s13132-024-02005-9>
- [30] Wei, H., & Zhang, Y. (2025). Research on the digital transaction model of the sports industry chain based on blockchain technology. *Scientific Reports*, 15(1), 25131. DOI: <https://doi.org/10.1038/s41598-025-10045-7>
- [31] Liu, J., Zhu, S., Wang, Z., & Chen, S. (2024). The evolution of online public opinion on earthquakes: a system dynamics approach. *Humanities and Social Sciences Communications*, 11(1), 1636. DOI: <https://doi.org/10.1057/s41599-024-04162-3>
- [32] Guo, Q., Zhao, X., Ba, J., & Luo, C. (2025). Porosity prediction with Bi-LSTM network for deep methane reservoirs. *Journal of Seismic Exploration*, 34(6), 29-44. DOI: <https://doi.org/10.36922/JSE025410087>
- [33] Zhang, W., Zhang, J., Huang, Z., Yu, J., Feng, D., & Wang, S. (2025). Predicting channel sandstone thickness through a VIF-NRBO-XGBoost model. *Journal of Seismic Exploration*, 34(3), 1-12. DOI: <https://doi.org/10.36922/JSE025290037>