Eng Int Syst (2025) 6: 645–655


**Engineering**
***Intelligent Systems***

# Deep Learning-Driven Tourism Data Mining and Fuzzy Prediction Methods

**Xiaochen Li[1], Jiayu Wu[2], Yuqing Zhang [1,3,*] and Yanling Xiao[4]**

[1] *International Education School, Guangzhou College of Technology and Business, Guangzhou 510000, China*
[2] *School of Foreign Language and International Trade, Guangdong Polytechnic, Foshan 528000 China*
[3] *Macau University of Science and Technology, Macau 999078, China*
[4] *School of Tourism and Leisure Management, Fujian Business University, Fuzhou 350012, China*

Tourism is a dynamic industry influenced by a variety of factors, including uncertainty about tourist behavior, seasonal variations, and the impact of external environmental factors. With the development of big data technology and artificial intelligence, the tourism industry has begun to seek more efficient ways to understand and predict tourists' behavioral patterns and their impact on the tourism economy. However, traditional prediction models often struggle to accurately capture the ambiguity and uncertainty in tourism data, especially when confronted with complex time series data and multimodal data. This study proposes an innovative deep learning-driven tourism data mining framework that incorporates fuzzy logic to improve the accuracy of predictions regarding tourist numbers and expected revenue. By employing a deep learning architecture with multimodal fusion, the framework is able to handle both structured and unstructured data and utilize the attention mechanism to highlight key features. The integration of fuzzy logic further enhances the model's ability to handle uncertainty and fuzzy information. The experimental results show that the proposed method performs better than the traditional baseline model under a variety of evaluation metrics, especially in dealing with seasonal variations and uncertainty. Specifically, in terms of tourist number prediction, the MSE, RMSE and MAE of the proposed method are 75.42, 8.68 and 5.23, respectively, while in terms of revenue prediction, these metrics are 1023.21, 31.99 and 23.42, respectively, which are significantly lower than the baseline models, such as linear regression, support vector machine, random forests, and long and short-term memory networks. In addition, the proposed method shows high stability when dealing with seasonal variations in different months, which can provide reliable forecasting support for tourism companies throughout the year.

Keywords: deep learning, tourism data, data mining, fuzzy prediction

## 1. INTRODUCTION

With the development of information technology, especially the advancement of big data and artificial intelligence technology, the tourism industry is undergoing a profound transformation. The huge amount of data generated by the tourism industry provides unprecedented opportunities for tourism management and decision-making. Deep learning, as a powerful machine learning technique, has been widely used in the fields of image recognition, speech recognition, natural language processing, etc., and has shown great potential in tourism data mining [1]. However, tourism data often contains a large amount of uncertainty and ambiguity, which poses a challenge to traditional prediction methods. In the tourism industry, the accurate prediction of tourists' needs, preferences, and behaviors is crucial for the development of effective marketing strategies. Traditional forecasting methods are usually based on statistical principles, but they have limitations when dealing with complex tourism data [2].

Figure 1 shows the trend of China's total tourism revenue and total number of tourists from 2011 to 2021. The bar graph shows the annual total tourism revenue, with the value

*Corresponding author's E-mail: zhangyuqing188@hotmail.com

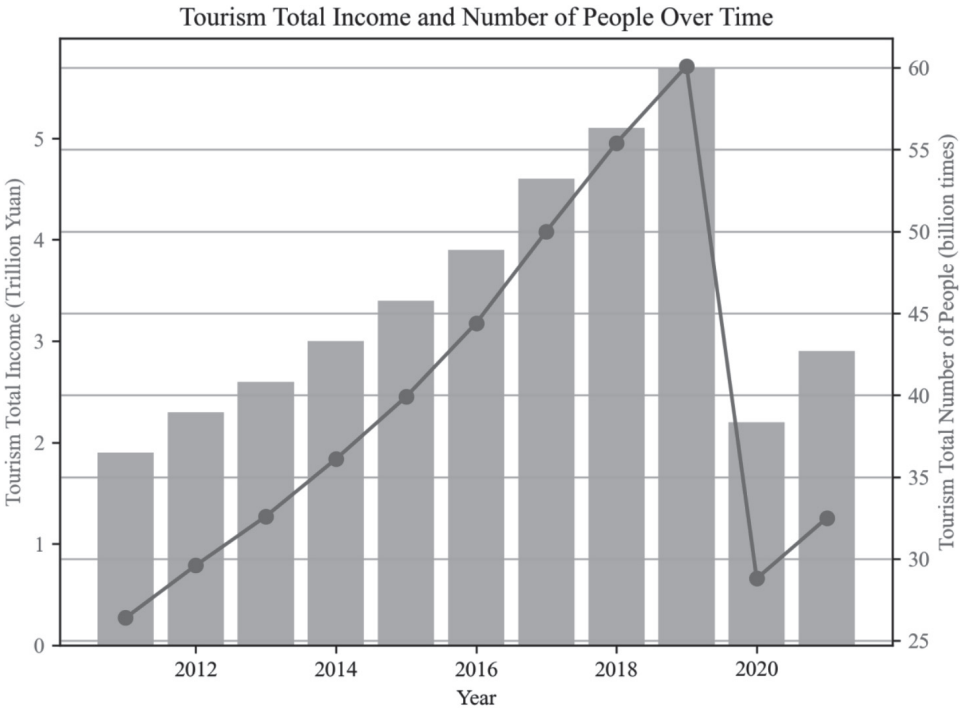


**Figure 1** Trends in China's total tourism revenue and total number of tourists.

in trillion RMB, while the line chart part presents the annual total number of tourists, with the value in billions of trips. The graphs show that during this decade, China's total tourism revenue has generally shown an upward trend, especially after 2016 when the growth is more obvious [3, 4]. However, the data for 2020 showed a significant decline in total tourism revenue due to the Covid-19 pandemic. Similarly, the total number of tourists experienced fluctuations, especially in 2020 when it declined significantly. Overall, this graph reflects the development of China's tourism industry over the past decade and the challenges it faces when affected by external factors [5, 6].

In recent years, fuzzy logic has been successfully applied in several fields as an effective tool for dealing with uncertainty [7]. Fuzzy logic allows the modeling of imprecise data, which improves the robustness and adaptability of predictive models. The combination of fuzzy logic with deep learning can better handle ambiguity and uncertainty in tourism data and provide more accurate support for decision making in the tourism industry [8].

This study explores the potential of deep learning techniques in the field of tourism data mining and innovatively proposes a prediction method incorporating fuzzy logic, with the aim of improving the accuracy and credibility of tourism market trend prediction. In the research process, we need to address three key issues: first, how to use deep learning techniques to extract valuable features from the complicated tourism data; second, how to skillfully design and implement the prediction model incorporating fuzzy logic to deal effectively with various uncertainties in the tourism data; and lastly, how to comprehensively assess the validity and practicability of the constructed framework and model to ensure their value in practical applications [9, 10].

## 2. LITERATURE REVIEW

### 2.1 Current Research Status of Tourism Data Mining

Tourism data mining, as a cross-discipline, has developed rapidly in recent years. With the progress of information technology, especially the popularization of the Internet and mobile communication technology, the tourism industry has accumulated a large amount of data. These data include not only traditional transaction data, but also unstructured data such as social media data, geographic location information, and user comments. The application of data mining technology enables tourism enterprises to better understand market demand, optimize products and services, and improve operational efficiency. In terms of tourists' behavior analysis, data mining techniques are widely used to identify tourists' preferences and behavioral patterns [11, 12].

Market trend prediction is another important application area of tourism data mining. By analyzing historical data, future tourism demand and market changes can be predicted. Data mining can extract valuable information from large, incomplete, fuzzy, noisy and random data to provide decision support for the tourism industry. For example, time series analysis can be used to predict the number of tourists during the peak season, while association rule mining helps to discover the correlation between different tourism products and services, thus guiding the combination of tourism products and the development of marketing strategies [13]. Personalized recommendation systems are an important part of modern tourism data mining. By analyzing data such as users' browsing records, search behaviors, and purchase history, personalized tourism recommendations and services can be provided to users. Tourism data mining systems can

recommend tourist attractions of interest and provide information query services pertaining to food, accommodation, transportation, shopping, entertainment and weather. These recommendation systems usually combine collaborative filtering techniques and content filtering techniques to improve recommendation accuracy and user satisfaction [14].

## 2.2 Application of Fuzzy Prediction Methods in the Field of Tourism

Fuzzy forecasting methods are widely used in several fields because of their ability to deal effectively with uncertainty and ambiguity. In the field of tourism, fuzzy forecasting methods are particularly suitable for dealing with situations that cannot be accurately described by traditional mathematical models. For example, changes in tourism demand are affected by a variety of complex factors, such as weather conditions, holidays, and special events, which are often highly uncertain. Therefore, introducing fuzzy logic into forecasting models can better capture and express these uncertainties and ambiguities, thus improving the accuracy of forecasting. In tourism demand forecasting, fuzzy logic can be used to construct more flexible and robust forecasting models. For example, the literature proposes a fuzzy time-series-based forecasting model that utilizes fuzzy set theory to process the time series characteristics of tourism demand data and more accurately predict tourist flows during a future period of time [15, 16]. This model is particularly suitable for dealing with nonlinear relationships and cyclical fluctuations in the data, and can provide more accurate prediction results.

Personalized recommendation systems are an important part of tourism data mining, and fuzzy logic can improve the recommendation effect of the system. A personalized recommendation algorithm based on fuzzy reasoning is proposed in the literature, which takes into account the user's fuzzy preferences and interests, and thus is able to generate a recommendation list that is more in line with the user's expectations. This approach not only handles the fuzzy nature of user preferences, but also takes into account the complex correlation between recommended items, which improves the accuracy and diversity of recommendations [17, 18].

Fuzzy logic also plays an important role in the optimization and management of tourism resources. The literature proposes a room booking optimization scheme based on fuzzy logic, which helps hotels to achieve more efficient room management by establishing a fuzzy model of room booking, taking into account the uncertainty of booking volume and the possibility of cancellation by customers. To a certain extent, this method can reduce the waste of resources due to customer cancellations, while at the same time ensuring that the hotel has enough room supply during the peak period [19, 20].

The advantage of fuzzy forecasting methods over traditional forecasting methods is that they are better able to deal with uncertainty in the data. For example, the literature compares the performance of using fuzzy logic with traditional statistical methods in predicting the number of tourists during the peak tourist season. The results show that the fuzzy logic model is more effective in dealing with outliers and nonlinear relationships in the data, and can provide more stable and reliable predictions [21]. Deep learning, as an advanced machine learning technique, has demonstrated excellent capabilities in handling large-scale complex data sets. With the development of big data technology, the travel industry has accumulated a large amount of data including, but not limited to, user behavior data, social media data, travel reviews, images and videos. These data are crucial for understanding tourists' needs, optimizing the travel experience, and improving marketing strategies. The application of deep learning techniques can help the tourism industry better exploit the value of these data.

## 2.3 Deep Learning in Tourism Data Analysis

Travel-related images usually contain rich information, such as attractions, activities, and people. Automatic annotation and classification of tourism images can be achieved using deep learning techniques, especially convolutional neural network (CNN). For example, Moonam et al. [22] and Ordonez-Martinez et al. [23] used a CNN model for feature extraction of tourism pictures, and combined it with a region candidate network (R-CNN) for target detection, and further fuses convolutional neural network and recurrent neural network (RNN) for semantic segmentation, so as to automatically identify and label each element in the pictures. This technique helps tourism websites and applications to improve user experience, and can also be used for intelligent guided tours and personalized recommendation services.

Travel reviews and travelogues contain a large amount of valuable information, including tourists' opinions, suggestions, and feelings about a destination. Deep learning techniques, especially natural language processing (NLP) methods, can be used to perform sentiment analysis, topic extraction, etc. on these textual data. The literature utilizes neural network models, such as Long Short-Term Memory (LSTM) networks, to mine deep features and sentiment tendencies in travel reviews. The results of these analyses are invaluable to tourism companies, helping them to understand customer satisfaction and subsequently to develop more targeted service improvement measures [24].

Deep learning technology can support the development of smarter personalized recommendation systems. By learning and analyzing a large amount of tourism data, deep learning models can provide tourists with personalized tour planning, intelligent tour guidance and accurate tour recommendations. Wang et al. [1] describes a deep learning-based intelligent tour guide system that provides customized travel advice and route planning based on user preferences, historical behavior, and other relevant information. This highly personalized service greatly enhances the travel experience and also increases tourist loyalty. In tourism management, big data analytics can be used to optimize marketing strategies, resource allocation, and service processes.

# 3. METHODOLOGY

## 3.1 Deep Learning Model Design

The aim of this study is to develop an innovative deep learning model that can effectively handle the complexity and

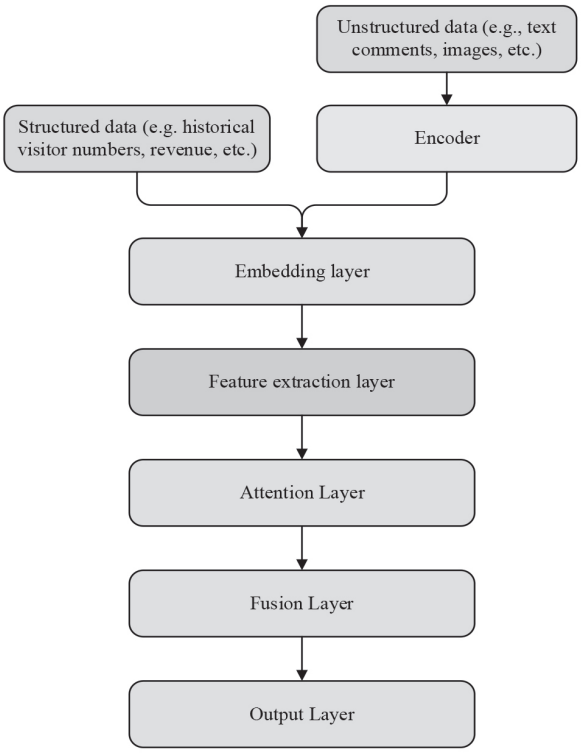


**Figure 2** Modeling framework.

uncertainty in tourism data. The framework of the model is shown in Figure 2.

As shown in Figure 2, the model accepts two types of input data: structured data (e.g., historical visitor numbers, revenue, etc.) and unstructured data (e.g., textual comments, images, etc.). The structured data is directly used as input to the encoder, while the unstructured data is preprocessed through the embedding layer. In this stage, the unstructured data is converted into a fixed-length vector representation. For example, textual comments are converted into numeric vectors by a word-embedding technique, and image data is converted into vector form by a pre-trained image feature extractor. Next, the model extracts features using Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM). CNNs process image and text data to extract spatial features, while LSTMs process time-series data to extract long-term dependencies in the time series. In this stage, the model introduces the attention mechanism to highlight the key information and reduce the influence of irrelevant information by calculating the importance of features. The attention layer fuses the features extracted by CNN and LSTM to form a weighted feature representation. Finally, the model uses the full connectivity layer to integrate features from different sources to form a comprehensive feature representation. Based on the needs of the prediction task, the model outputs the final prediction results, such as the anticipated number of future tourists or the trend of revenue growth [25].

To achieve this goal, we employ a multimodal fusion deep learning architecture that is capable of processing both structured and unstructured data and capturing long-term dependencies in the data. Our deep learning model comprises these layers:

Input Layer: responsible for receiving raw data, including structured data (e.g., historical number of visitors, revenue, etc.) and unstructured data (e.g., text comments, images, etc.).

Embedding layer: converts unstructured data into fixed-length vector representations for subsequent processing. For example, textual comments are converted into numeric vectors using word-embedding techniques, or image data are converted into vector form using a pre-trained image feature extractor [26].

Feature extraction layer: features are extracted using Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM). CNN is used to process image and text data, while LSTM is used to process time-series data. CNN can capture spatial features, such as objects and scenes in an image, while LSTM can capture long term dependencies in time-series data, such as the trend of the number of tourists over time.

Attention Layer: An attention mechanism is introduced to highlight key features and reduce the influence of irrelevant information. The attention mechanism allows the model to assign weights to features based on their importance, thus increasing the model's focus on key information.

Fusion Layer: the different types of features extracted are fused using a fully connected layer to form a comprehensive feature representation. The role of this layer is to integrate features from different sources to provide comprehensive information for the final prediction.

Output layer: outputs the final prediction results according to the requirements of the prediction task such as predicting the number of tourists in the future, the trend of revenue growth, etc.

In the embedding layer, we let $w_i$ be the $i$-th word and $E(w_i)$ be its corresponding embedding vector, then the

embedding representation of the text $T$. $E_T$ can be expressed as Equation (1).

$$E_T = [E(w_1); E(w_2); \ldots; E(w_n)] \tag{1}$$

Where $n$ is the number of words in the text. For example, for text comment data, we can use a pre-trained word embedding model to generate embedding vectors.

In the feature extraction layer, let $I$ be the input image and $F_I$ be the feature vector extracted by the CNN. The convolutional neural network (CNN) detects local features in the image through the convolutional layer, reduces the data dimensionality and extracts the translation invariance using the pooling layer, and then integrates these features through the fully-connected layer to perform the classification so as to automatically learn the representation of the image features from the simple to the complex ones.

Let $S$ be the time series data and $F_S$ be the feature vector extracted by LSTM, which is denoted as $F_S = LSTM(S)$. Here CNN and LSTM represent the forward propagation process of Convolutional Neural Network and Long Short-Term Memory Network, respectively. The attention layer fuses the features extracted by CNN and LSTM. Let $F = [F_I; F_S]$ be the fused feature vector and $\alpha_i$ be the attention weight, the output of the attention layer $F_a$ can be expressed as Equation (2) [27, 28].

$$F_a = \sum_{i=1}^{m} \alpha_i F_i \tag{2}$$

Where $m$ is the dimension of the feature vector and $\alpha_i$ is obtained by dot product operation. The attention weight $\alpha_i$ can be calculated by Equation (3).

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{m} \exp(e_j)} \tag{3}$$

Where $e_i$ is the importance score of the $i$ th feature, which can be obtained by a simple feed-forward neural network or dot product operation.

The fusion layer transforms the output of the attention layer, let $F_a$ be the output of the attention layer and $F_f$ be the output of the fusion layer, then we have Equation (4). Where $W_f$ and $b_f$ the learnable weight matrix and bias vector [29].

$$F_f = W_f F_a + b_f \tag{4}$$

The output layer outputs the final prediction results according to the requirements of the prediction task. For example, for the prediction of the number of tourists, the output layer can be represented as Equation (5).

$$\hat{y} = W_o F_f + b_o \tag{5}$$

## 3.2 Fuzzy Logic Integration

### 3.2.1 Fuzzy Sets and Affiliation Functions

Ambiguities and uncertainties in tourism data stem mainly from tourists' behavioral patterns, differences in preferences, and the influence of various external factors. Tourists' behavioral patterns are often fluid, and their travel plans may change due to personal interests, unexpected circumstances, or ad hoc decisions. In addition, different tourists have significant differences in preferences: some may prefer nature-based adventures, while others may prefer cultural experiences, and these differences increase the difficulty of data consistency and predictability. External factors, such as weather changes, economic fluctuations, and holiday schedules can also have an impact on tourism activities, which in turn affects the stability and reliability of tourism data. Therefore, these ambiguities and uncertainties need to be taken into account when analyzing tourism data, and appropriate statistical methods and machine learning techniques need to be used to deal with these complexities in order to improve the accuracy of prediction and decision support. To deal with ambiguity and uncertainty in tourism data, we define a series of fuzzy sets and define a corresponding affiliation function for each set. We define fuzzy sets to describe the states of tourist numbers, revenues and growth rates. For example, for the number of tourists, we define the fuzzy sets "High", "Medium" and "Low", which denote the number of high, medium and low tourists, respectively. For the fuzzy set $\tilde{A}$, the affiliation function $\mu_{\tilde{A}}(x)$ describes how much the element $x$ belongs to $\tilde{A}$. For example, for the number of tourists x, we define the affiliation function of the fuzzy sets "High", "Medium" and "Low" as shown in Equations (6)–(8) [30, 31].

$$\mu_{\text{High}}(x) = \begin{cases} 0 & x \leq 50 \\ (x-50)/(100-50) & 50 < x \leq 100 \\ 1 & x > 100 \end{cases} \tag{6}$$

$$\mu_{\text{Medium}}(x) = \begin{cases} 0 & x \leq 30 \\ (x-30)/(70-30) & 30 < x \leq 70 \\ 1-(x-70)/(100-70) & 70 < x \leq 100 \\ 0 & x > 100 \end{cases} \tag{7}$$

$$\mu_{\text{Low}}(x) = \begin{cases} 1 & x \leq 30 \\ 1-(x-30)/(70-30) & 30 < x \leq 70 \\ 0 & x > 70 \end{cases} \tag{8}$$

Similarly, we define fuzzy sets and corresponding affiliation functions for income and growth rates.

### 3.2.2 Fuzzy Inference Mechanism

The fuzzy inference mechanism is used to derive the output fuzzy sets based on fuzzy rules and input fuzzy sets. We used the Mamdani fuzzy inference system for fuzzy inference.

Figure 3 illustrates a simple multi-rule voting system that contains three rules (Rule 1, Rule 2, and Rule 3), each of which independently makes judgments on two inputs (Input 1 and Input 2). Each rule has its own weight, and this weight determines its influence in the final output. In this example, the "Voting" box represents an aggregation component that collects the output of each rule and votes on it according to its weight. Ultimately, the result of the voting generates a single output. When an input signal arrives,

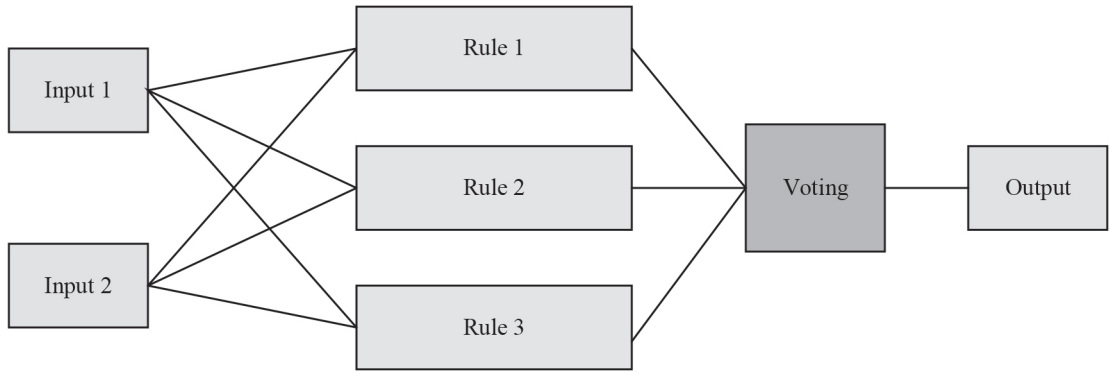


**Figure 3** Mamdani fuzzy inference system.

each rule evaluates the input according to its own conditions and logic, and returns an output. These outputs may be Boolean values (True or False), numbers, or other types of information, depending on the functionality of the rule. In the "Voting" component, these outputs are aggregated and usually weighted or counted according to the weight of the rule.

Step 1: We define a series of fuzzy rules to reflect expert knowledge and experience. The following is an example. (1) Rule 1: IF Tourist arrivals is High AND Revenue is Medium, THEN Growth rate is Low (2) Rule 2: IF Tourist arrivals is Medium AND Revenue is High THEN Growth rate is Medium (3) Rule 3: IF Tourist arrivals is Low AND Revenue is Low, THEN Growth rate is High.

Step 2: Defuzzification converts the clear input values into fuzzy sets of affiliations. For example, the specific number of tourists is converted to the affiliation of the fuzzy sets "High", "Medium" and "Low".

Step 3: Inference. Apply fuzzy logic operators such as intersection $\cap$ and union $\cup$ and fuzzy rules to infer the output fuzzy sets. For example, for rule 1 "IF Number of Visitors High AND Revenue Medium THEN Growth Rate Low", we use the minimum operation min to determine the intersection set, i.e. $u_{\text{growth rate Low}} = \min(\mu_{\text{High}}, \mu_{\text{Medium}})$.

Step 4: De-blurring. Convert the output fuzzy set back to clear values. We use the center of gravity (COG) method to compute the clear growth rate values $\hat{z}$, Specifically as in Equation (9).

$$\hat{z} = \frac{\int z \cdot \mu_{\text{growth rate Low}}(z)dz}{\int \mu_{\text{growth rate Low}}(z)dz} \tag{9}$$

Here the integration is done for all possible $z$ values $\mu_{\text{growth rate Low}}(z)$ is the affiliation function of the fuzzy set of growth rates derived from the fuzzy rules.

Through fuzzy logic integration, we are able to deal effectively with the ambiguity and uncertainty in tourism data and provide richer information to the deep learning model, thus improving the accuracy and robustness of the prediction. This approach can be used as a complementary means for deep learning models to help the models better understand and interpret the fuzzy information in the input data.

## 3.3 Realization Details

We define a custom layer. The custom layer is designed to combine clear values derived from fuzzy logic with features extracted by the deep learning model. We can achieve this by defining a simple custom layer that is responsible for stitching clear values from fuzzy logic with features extracted by the model. Assuming that $F_f$ is the output of the fusion layer and $\hat{z}$ is the clear value of the growth rate derived from fuzzy logic reasoning, the output of the custom layer $F_c$ can be expressed as Equation (10).

$$F_c = [F_f; \hat{z}] \tag{10}$$

This operation simply appends the clear values of the growth rate to the feature vectors output from the fusion layer to form a new feature vector for subsequent output layers.

Although the fuzzy logic component is not directly involved in the training process of the deep learning model, we can periodically evaluate the performance of the fuzzy logic module during the training process and adjust the fuzzy rules or the affiliation function accordingly.

During the training process, we first use the standard backpropagation algorithm to optimize the parameters of the deep learning model. Then, at the end of each training cycle, we can use the fuzzy logic module to evaluate the predictions of the model and adjust the parameters of the fuzzy logic as needed.

# 4. EXPERIMENTAL DESIGN AND RESULTS

## 4.1 Experimental Data Set

In order to validate the effectiveness of the proposed deep learning-driven tourism data mining framework and its prediction method combined with fuzzy logic, we selected a comprehensive real-world dataset with wide coverage. The dataset contains historical data of several tourist destinations and covers a nine-year period from 2015 to 2023. The dataset comprises the following:

Structured data: monthly statistics on the number of visitors, total revenue, and average visitor spending.

Unstructured data: visitor reviews, social media posts, travel photos, etc.

**Table 1** MSE, RMSE and MAE of different models for tourist number prediction.

| Mould | MSE | RMSE | MAE |
|---|---|---|---|
| Linear regression (math.) | 124.56 | 11.16 | 8.21 |
| SVM | 113.28 | 10.64 | 7.94 |
| Random forest | 98.54 | 9.93 | 6.85 |
| LSTM | 89.23 | 9.45 | 6.32 |
| Proposed methodology | 75.42 | 8.68 | 5.23 |

External data: holiday information, weather forecast data, special events and other factors that affect tourism behavior.

The size of the dataset is about 5 GB, with about 2 GB of structured data and 3 GB of unstructured data. In order to ensure the quality of model training, we perform rigorous preprocessing of the data, including steps such as removing duplicate records, filling in missing values, and outlier detection and correction.

## 4.2 Evaluation Indicators

In order to comprehensively assess the performance of the prediction model, we used a variety of evaluation metrics for comprehensive consideration. First, we calculated the mean square error (MSE), which is an important measure of the average squared difference between the predicted and true values; the smaller the value, the more accurate is the prediction of the model. Then, we utilized the root mean square error (RMSE), which is the square root of the MSE, to further quantify the magnitude of the prediction error; similarly, the smaller the value of the RMSE, the more accurate is the prediction. In addition, we calculated the Mean Absolute Error (MAE), which reflects the average absolute difference between the predicted value and the true value, with a smaller MAE value indicating that the model's prediction is more accurate. We also introduced the coefficient of determination (R-squared, $R^2$), as this metric measures the proportion of the model's variation in the explanatory variables; the closer the value of $R^2$ is to 1, the better is the model's fitting effect. In addition to these general evaluation metrics, we considered business-related metrics such as the relative error between the predicted number of tourists and the actual number of tourists, as well as the percentage difference between the predicted revenue and the actual revenue, in order to ensure that the model performs well statistically in addition to meeting the needs of the actual business.

## 4.3 Baseline Model

In order to verify the validity and superiority of the proposed forecasting methods, several widely-recognized forecasting methods were carefully selected as baseline models for comparison. These models are: linear regression, a basic and widely-used statistical forecasting method commonly used to predict continuous variables such as the number of tourists and revenues; support vector machine (SVM), a powerful classification and regression method based on the principle of maximum intervals, which performs well when dealing with high-dimensional data; random forests (RFs), an integrated learning method based on decision trees, which is widely used due to its powerful generalization and its ability to deal with complex data; and Long Short-Term Memory (LSTM), a special type of Recurrent Neural Network (RNN) specifically designed to process time-series data, which captures long-term dependencies in data. These baseline models, ranging from simple to complex and covering different types of algorithms, allow us to comprehensively evaluate and compare the performance of the proposed methods, thus providing insights into their potential benefits when applied to tourism data prediction.

## 4.4 Experimental Procedures

The experimental process was carried out in strict accordance with the established steps, firstly, the raw data was preprocessed, involving cleaning, processing missing values and detecting outliers to ensure the high quality of the data. Subsequently, feature engineering was performed to execute feature extraction and transformation for the different characteristics of structured and unstructured data to provide high-quality inputs for model training; then, the processed data was divided into a training set, a validation set and a test set. Then, the deep learning model was trained using the training set, and the model parameters were optimized using the validation set.

## 4.5 Experimental Results

As shown in Table 1, the proposed method's performance is clearly superior when predicting tourist numbers. Its MSE value is 75.42, RMSE value is 8.68, and MAE value is 5.23, which are lower than the baseline models such as linear regression, SVM, random forest and LSTM. This result indicates that the proposed method not only has a smaller overall error when predicting the number of tourists; it also has greater accuracy in reducing specific prediction biases.

As shown in Table 2, when comparing error metrics for revenue forecasting, the proposed method demonstrates excellent performance. Its MSE, RMSE and MAE are 1023.21, 31.99 and 23.42, respectively, which are better than other baseline models. This indicates that the proposed method is able to capture the trend of revenue changes more accurately when predicting tourism revenue, thus reducing the prediction error.

Table 3 shows the $R^2$ and relative error of different models on the prediction of tourist numbers. The $R^2$ value of the proposed method is as high as 0.92, which is much higher than that of other models, indicating that it can explain 92%

**Table 2** MSE, RMSE and MAE of different models on revenue forecasting.

| Mould | MSE | RMSE | MAE |
|---|---|---|---|
| Linear regression (math.) | 1563.45 | 39.54 | 30.23 |
| SVM | 1452.67 | 38.11 | 29.56 |
| Random forest | 1324.56 | 36.40 | 27.89 |
| LSTM | 1245.89 | 35.30 | 26.45 |
| Proposed methodology | 1023.21 | 31.99 | 23.42 |

**Table 3** $R^2$ and relative error of different models on tourist number prediction.

| Model | $R^2$ | Relative error (%) |
|---|---|---|
| Linear regression (math.) | 0.76 | 15.23 |
| SVM | 0.79 | 14.56 |
| Random forest | 0.84 | 12.89 |
| LSTM | 0.87 | 11.32 |
| Proposed methodology | 0.92 | 9.23 |

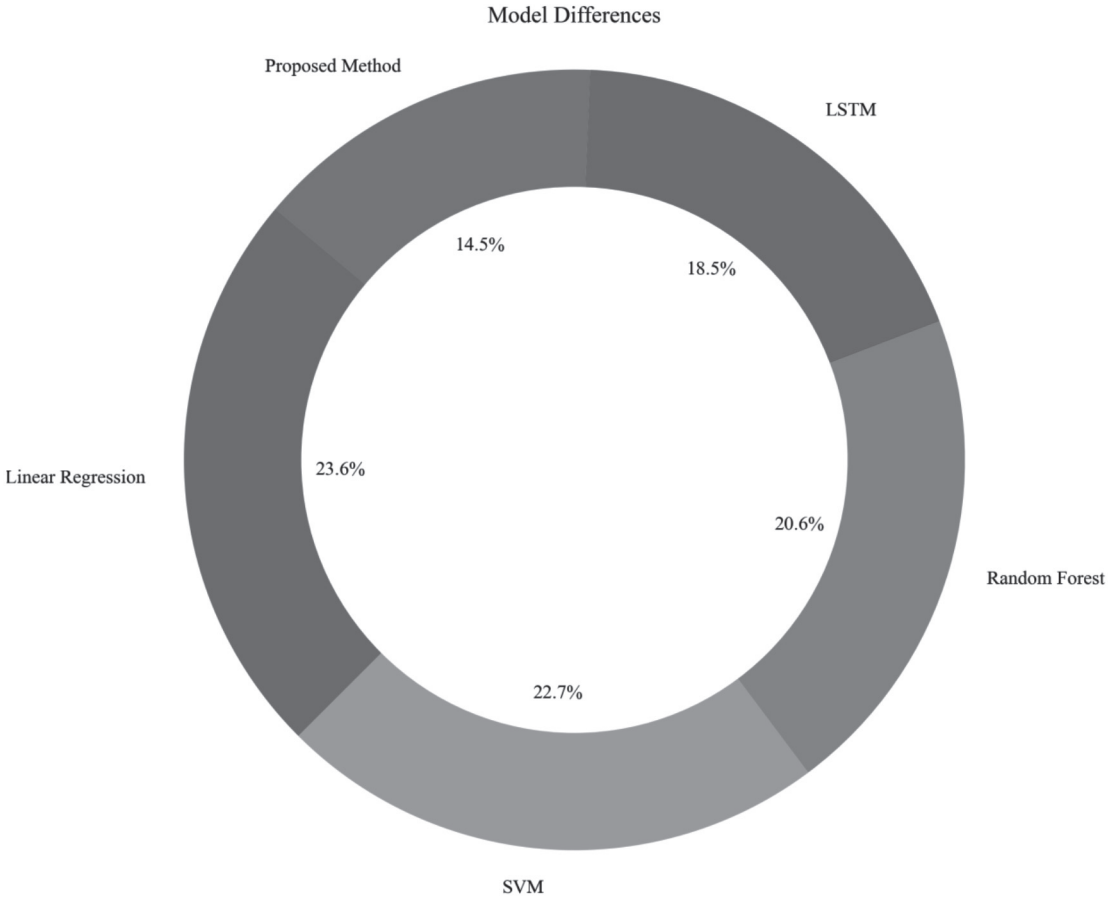


**Figure 4** $R^2$ and percentage of variance of different models on revenue forecasting.

of the variation in the number of tourists. Meanwhile, its relative error is only 9.23%, which is the lowest among all the models, which further confirms the effectiveness and accuracy of the proposed method in predicting the number of tourists.

Figure 4 shows the percentage of variance of different models in regard to revenue forecasting. The $R^2$ value of the proposed method is 0.90 and the percentage of variance is 11.23%, which are better than the other models. This indicates that the proposed method is able to better predict the revenue data, and also that it performs well in reducing the forecast variance.

As shown in Table 4, comparing the MAE of tourist number prediction by different models for each month, the prediction error of the proposed method in each month is smaller than that of other models. This consistently excellent performance indicates that the proposed method has strong seasonal adaptability and can maintain stability and accuracy when predicting the number of tourists for each month.

**Table 4** Predicted MAE of tourist arrivals for different models in different months.

| Model | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sep. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Linear regression (math.) | 8.2 | 8.3 | 8.5 | 8.7 | 9.0 | 9.5 | 9.8 | 10.1 | 10.4 | 10.2 | 9.8 | 9.0 |
| SVM | 7.9 | 8.0 | 8.2 | 8.5 | 8.8 | 9.3 | 9.6 | 9.9 | 10.2 | 10.0 | 9.6 | 8.8 |
| Random forest | 6.8 | 6.9 | 7.1 | 7.3 | 7.6 | 8.1 | 8.4 | 8.7 | 9.0 | 8.8 | 8.4 | 7.6 |
| LSTM | 6.3 | 6.4 | 6.6 | 6.8 | 7.1 | 7.6 | 7.9 | 8.2 | 8.5 | 8.3 | 7.9 | 7.1 |
| Proposed methodology | 5.2 | 5.3 | 5.5 | 5.7 | 6.0 | 6.5 | 6.8 | 7.1 | 7.4 | 7.2 | 6.8 | 6.0 |

**Table 5** MAE of revenue forecasts of different models for each month.

| Model | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sep. | Oct. | Nov. | Dec. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Linear Regression | 30.2 | 30.3 | 30.5 | 30.7 | 31.0 | 31.5 | 31.8 | 32.1 | 32.4 | 32.2 | 31.8 | 31.0 |
| SVM | 29.6 | 29.7 | 29.9 | 30.2 | 30.5 | 31.0 | 31.3 | 31.6 | 31.9 | 31.7 | 31.3 | 30.5 |
| Random Forest | 27.9 | 28.0 | 28.2 | 28.4 | 28.7 | 29.2 | 29.5 | 29.8 | 30.1 | 29.9 | 29.5 | 28.7 |
| LSTM | 26.5 | 26.6 | 26.8 | 27.0 | 27.3 | 27.8 | 28.1 | 28.4 | 28.7 | 28.5 | 28.1 | 27.3 |
| Proposed Method | 23.4 | 23.5 | 23.7 | 23.9 | 24.2 | 24.7 | 25.0 | 25.3 | 25.6 | 25.4 | 25.0 | 24.2 |

Table 5 demonstrates the MAE of different models for revenue forecasting in different months. The proposed method has the lowest MAE value for all months, indicating that it is highly reliable in terms of revenue forecasting throughout the entire year and can provide reliable forecasting support for the tourism business.

## 4.6 Discussion

Based on the experimental results, we can explore in depth the effectiveness and practicality of the proposed deep learning-driven tourism data mining framework and its prediction method combined with fuzzy logic.

The experimental results suggest that the proposed method has significant performance advantages in both tourist number and revenue prediction. Tables 1 and 2 show that the proposed method achieves optimal results for all three indicators, namely MSE, RMSE and MAE. This suggests that the proposed method is able to reduce not only the overall level of prediction error, but also the specific bias of a single prediction, which is very important for practical applications. For example, the accurate prediction of tourist arrivals and revenues can help tourism companies to better plan their resource allocation, formulate marketing strategies, and optimize customer services. Tables 3 and 4 further demonstrate the ability of the proposed method in explaining the variations in the data. The $r^2$ values close to 1 imply that the model is able to explain the variations in the data well, while low relative errors and percentage of variance indicate that the model performs well in capturing seasonal variations and other complex factors. These results suggest that the proposed method is able to model real-world tourism market dynamics more accurately, which is valuable for decision makers in the tourism industry.

Although the proposed method performs well on several metrics, there are still some limitations. For example, although fuzzy logic is capable of handling uncertainty, it may encounter challenges when dealing with extreme situations. In addition, the proposed method may need more data to further validate its generalization ability. Future researchers could consider using the following approaches: (1) Use larger datasets to further validate the performance of the model, especially when dealing with extreme situations. (2) Explore more complex deep learning architectures, such as those using the latest techniques such as Transformer to improve the prediction accuracy of the model. (3) Introduce more types of external data, such as social media sentiment analysis and travel hotspot trends, to further improve the model's prediction capability. (4) Develop real-time forecasting functions so that tourism companies can more quickly respond to market changes. (5) Improve the transparency and interpretability of the model so that tourism managers can better understand the decision-making process of the model.

In summary, the proposed deep-learning-driven tourism data mining framework and its prediction method combining fuzzy logic have shown significant advantages in the prediction of tourist numbers and revenue. Future research could further explore how to validate the model's generalizability and applicability in different application scenarios on larger datasets, as well as how to improve the model's explanatory and real-time prediction capabilities. These improvements will help the travel industry to better leverage data-driven decision support systems, improve forecasting accuracy, and optimize business strategies.

# 5. CONCLUSION

This study developed a method for forecasting tourism data that combines deep learning and fuzzy logic, aiming to improve the accuracy of forecasting tourist arrivals and revenue in the tourism industry. By employing a deep learning architecture with multimodal fusion, the method is able to efficiently process both structured and unstructured data and utilize an attention mechanism to focus on key features. The integration of fuzzy logic further enhances the model's ability to deal with uncertainty and fuzzy information, and improves the explanatory and adaptive nature of the predictions. The experimental results show that the proposed method demonstrates significant performance advantages in predicting the number of tourists and revenue. Specifically, several evaluation indexes (MSE, RMSE, and MAE) show that the proposed method outperforms traditional baseline

models, such as linear regression, support vector machine, random forest, and long and short-term memory network. Specific data show that in terms of tourists' number prediction, the MSE value of the proposed method is 75.42, the RMSE value is 8.68, and the MAE value is 5.23, while in terms of revenue prediction, the metrics are 1023.21, 31.99, and 23.42, respectively, which are significantly lower than the baseline model. In addition, the proposed method performs well when dealing with seasonal variations and is able to provide stable forecasting support for tourism companies throughout the year range, e.g., in terms of MAE for forecasting the number of tourists in different months, the proposed method is lower than the other models in all months. The results obtained by several evaluation metrics indicate that the proposed framework is able to demonstrate better performance than traditional models, especially when dealing with seasonal variations and uncertainties, which helps tourism enterprises to more accurately predict future tourist flows and revenues, and thus make more rational business decisions. By improving forecasting accuracy, tourism companies can better plan resource allocation, adjust marketing strategies and product pricing to ensure service quality and customer satisfaction while maximizing profits. The integration of fuzzy logic allows the model to better handle uncertainty and fuzzy information, which is especially important for the tourism industry, where tourists' behavioral patterns may fluctuate due to a variety of factors such as personal preferences and weather changes.

## FUNDING

This work was supported by social science planning projects in Fujian Province. Project Name: Innovative Research on Constructing a New Type of Wellness Tourism Chain of Green-Red-Remembrance in the Old Revolutionary Areas on the West Coast of the Taiwan Straits (Yanling Xiao, Project No. 3), project Number: FJJKCG16-411; Wuyi University Education and Teaching Reform Research and Practice Achievements. Project Name: Tourism Business English (Yanling Xiao, Host), project Number: KC202108.

## CONFLICTS OF INTEREST

The author declares that this article is free of conflict of interest.

## DATA AVAILABILITY STATEMENT

All the data used to support the findings of this study are contained in the manuscript.

## REFERENCES

1. Wang X, Tong JX, Zong WY, Lv YQ, Shen JY. Trip misreporting mining and expansion method for household travel survey. Transportation Research Part a- Transportation Research Part A: Policy and Practice. 2024; 182: 104015.
2. Hoffmann FJ, Braesemann F, Teubner T. Measuring sustainable tourism with online platform data. EPJ Data Science. 2022; 11(1): 41.
3. Baek J, Choi Y. Simulation of truck haulage operations in an underground mine using big data from an ICT-Based Mine Safety Management System. Applied Sciences-Basel. 2019; 9(13): 2639.
4. Cheng G, Zhao SZ, Xu SB. Estimation of passenger route choices for urban rail transit system based on automatic fare collection mined data. Transactions of the Institute of Measurement and Control. 2019; 41(11): 3092–3102.
5. Adak Z, Demiriz A. Hybridization of population -based ant colony optimization via data mining. Intelligent Data Analysis. 2020; 24(2): 291–307.
6. Antolini F, Terraglia I, Cesarini S. Integrating multiple data sources to measure sustainable tourism in Italian regions. Socio-Economic Planning Sciences. 2024; 101959.
7. Tian T. Evaluation of financial intelligent systems digitalization maturity of small and medium-sized enterprises based on fuzzy logic algorithm. Engineering Intelligent Systems, 2025; 33(3): 275–284.
8. Chang XM, Wu JJ, Liu H, Yan XY, Sun HJ, Qu YC. Travel mode choice: a data fusion model using machine learning methods and evidence from travel diary survey data. Transportmetrica A-Transport Science. 2019; 15(2): 1587–1612.
9. Cristóbal T, Padrón G, Quesada-Arencibia A, Alayón F, García CR. Systematic approach to analyze travel time in road-based mass transit systems based on data mining. IEEE Access. 2018; 6: 32861–32873.
10. Domènech A, Miravet D, Gutiérrez A. Mining bus travel card data for analysing mobilities in tourist regions. Journal of Maps. 2020; 16(1): 40–49.
11. Guerard G, Gabot Q, Djebali S. Tourism profile measure for data-driven tourism segmentation. International Journal of Machine Learning and Cybernetics. 2024; 1–26.
12. Hamid RA, Albahri AS, Alwan JK, Al-qaysi ZT, Albahri OS, Zaidan AA, et al. How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management. Computer Science Review. 2021; 39: 100337.
13. Hu HM, Li CH. Smart tourism products and services design based on user experience under the background of big data. Soft Computing. 2023; 27(17):12711–12724.
14. Kalvet T, Olesk M, Tiits M, Raun J. Innovative tools for tourism and cultural tourism impact assessment. Sustainability. 2020; 12(18): 7470.
15. Kostelic K, Turk M. Topology of the world tourism web. Applied Sciences-Basel. 2021; 11(5): 2253.
16. Kovalcsik T, Elekes A, Boros L, Könnyid L, Kovács Z. Capturing unobserved tourists: challenges and opportunities of processing mobile positioning data in tourism research. Sustainability. 2022; 14(21): 13826.
17. Li Q, Li SB, Zhang S, Hu J, Hu JJ. A review of text corpus-based tourism big data mining. Applied Sciences-Basel. 2019; 9(16): 3300.
18. Li WG, Zhu J, Zhang Y, Zhang SY. Design and implementation of intelligent traffic and big data mining system based on internet of things. Journal of Intelligent & Fuzzy Systems. 2020; 38(2):1967–1975.
19. Li Y, Zhang JL. Classroom teaching of tourism management using multimedia big data analysis. Journal of Intelligent Information Systems. 2024; 62(3): 625–640.
20. Liu RX. Development of a travel recommendation algorithm based on multi-modal and multi-vector data mining. Peerj Computer Science. 2023; 9: e1436.
21. Ma DF, Sheng BW, Ma XL, Jin S. Fuzzy hybrid framework with dynamic weights for short-term traffic flow prediction by

mining spatio-temporal correlations. IET Intelligent Transport Systems. 2020; 14(2):73–81.
22. Moonam HM, Qin X, Zhang J. Utilizing data mining techniques to predict expected freeway travel time from experienced travel time. Mathematics and Computers in Simulation. 2019; 155:154–167.
23. Ordonez-Martinez D, Segui-Pons JM, Ruiz-Perez M, Bonn MA. Conceptual framework and prospective analysis of EU tourism data spaces. Sustainability. 2024; 16(1): 371.
24. Sharma MSP, Meena U, Sharma GK. Intelligent data analysis using optimized support vector machine-based data mining approach for tourism industry. ACM Transactions on Knowledge Discovery from Data. 2022; 16(5): 1–20.
25. Wang M, Wang J. Uncertainty models in the integration path of rural tourism information construction and smart tourism based on big data technology. OPTIK. 2023; 272: 170320.
26. Weng XX, Liu YX, Song HB, Yao SS, Zhang PF. Mining urban passengers' travel patterns from incomplete data with use cases. Computer Networks. 2018; 134: 116–126.
27. Wu HQ, Chen ZA, Yan J, Tang XL. Multi-Source Spatio-temporal data-based tourism structure analysis of demonstration city for global tourism: case study of Liyang, China. ISPRS International Journal of Geo-Information. 2022; 11(11): 547.
28. Wu XY, Huang Z, Peng X, Chen YR, Liu Y. Building a spatially-embedded network of tourism hotspots from Geotagged social media data. IEEE Access. 2018; 6: 21945–21955.
29. Xu RR. Framework for building smart tourism big data mining model for sustainable development. Sustainability. 2023; 15(6): 5162.
30. Yin F, Yin X, Zhou JC, Zhang XL, Zhang RH, Ibeke E, et al. Tourism cloud management system: the impact of smart tourism. Journal of Cloud Computing-Advances Systems and Applications. 2022; 11(1): 37.
31. Zaragozí B, Trilles S, Gutiérrez A. Passive mobile data for studying seasonal tourism mobilities: an application in a Mediterranean coastal destination. ISPRS International Journal of Geo-Information. 2021; 10(2): 98.