Research Article

Tourism Information Data Processing Method Based on Multi-Source Data Fusion

YaoGuang Li and HeChi Gan (ID)

College of Culture and Tourism, Jeonju University, Jeonju 55069, Jeollabub-do, Republic of Korea

Correspondence should be addressed to HeChi Gan; hechigan@163.com

Urban social civilization and the quality of life of residents are gradually improved, and the development scale and trend of the leisure tourism industry have been growing. This paper constructs a multi-source data fusion model based on an ensemble learning algorithm, uses Ctrip 2020 open data set to train the model, and then obtains the tourism information data processing and prediction results. This paper takes the data of Ctrip as the training set and compares the trained model with the data of tunic and Feizhu. In this paper, sensor detection technology is used to analyze many famous scenic spots in China, including tourist type, gender, and location. The results show that tourism feature extraction results are consistent with data from trending flying bamboo, tunics, and other websites, according to the results of a multi-source fusion of tourism information. Among them, in the data of the first half of 2020, the prediction accuracy of the model after data processing is about 62%. Affected by the epidemic situation, the accuracy of the model is low. In the second half of the year, the prediction accuracy is 78%, which can be used to fuse tourism information in a short time. Therefore, the data show that the model has high learning ability and high trend prediction ability in tourism data processing, which can provide necessary information support for tourists.

1. Introduction

Tourism information processing technology usually uses a POI model related to specific human social activities to represent a group of points in arc tourism information. This paper analyzes the distribution and agglomeration characteristics of leisure tourism space and discusses how to coordinate the relationship with urban construction and development, which plays a vital role in the construction of livable cities, the development of tourism, and the sustainable development of cities.

At present, many studies refer to the urban landscape and tourism data. For example, Chen y divides the urban leisure space into four categories by analyzing the world's major parks [1]. Shang CF discusses and analyzes the research status of the concept of leisure tourism at home and abroad. As a new data source and research idea, it can further analyze the urban leisure tourism space [2]. Oberoi analyzes the distribution characteristics of tourism information services in Guangzhou by using the nearest neighbor distance method

and other spatial analysis methods [3]. In terms of coupling analysis of tourism information data and scenic spot data, Bernardi PD uses landscape pattern index, gravity model, gravity model, and coupling analysis to analyze the urban spatial distribution structure of Anhui Province Based on DMSP/OLS image, statistical data, and tourism information data [4]. PON W C selects tourism information data and NPP remote sensing data for kernel density analysis and further analyzes the characteristics of the urban spatial structure of Wuhan City [5].

In data fusion, a new artificial intelligence technology proposed by poetry y a proposes that multi-source data fusion is that a central server coordinates multiple clients to complete a learning task without disclosing data [6]. Ramadhan GR proposes a user-level differential privacy training algorithm, which effectively reduces the possibility of recovering personal information from the transmission model by adding privacy protection to the aggregation algorithm [7]. On the other hand, Huyan w proposed a differential privacy hybrid model, which partitions users by their trust

preferences, to reduce the size of the required user base [8]. Oezturan m combines gradient selection and secret sharing algorithm, which greatly improves the communication efficiency while ensuring user privacy and data security [9]. In terms of resource optimization in multi-source data fusion, Liu P considered multi-source data fusion through the wireless network and proposed the problem of optimizing energy consumption and global multi-source data fusion time [10–11]. The above research in the data source processing is not fully integrated, so the tourism data prediction cannot achieve accurate prediction.

This paper constructs a multi-source data fusion model based on an ensemble learning algorithm, uses Ctrip 2020 open data set to train the model, and then obtains the tourism information data processing and prediction results [11–12]. This white paper uses Ctrip data as a training set and compares the trained model with tunic and flying bamboo data. This treatise mainly analyzes some of China's famous scenic spots, such as tourist type, gender, and location. According to the results of the multi-source fusion of tourism information, the results of tourism feature extraction are consistent with the data of Feizhu, tunic, and other websites in the trend.

2. Tourism Data Processing and Multi-Source Data Fusion

2.1. Distributed Tourism Data Processing Method. In the information age, massive tourism data will bring many problems to the centralized data processing mode with the cloud computing model as the core [13]. First, the processing mode of uploading all data to the cloud will not only cause low efficiency but also cause additional bandwidth overhead, at the same time, the network delay will also increase [14]. Second, due to the improvement of users' privacy awareness, the data of edge devices are likely to leak in the upload link, and the security of personal privacy cannot be guaranteed [15]. The distributed data processing mode can effectively solve the delay and efficiency problems of traditional cloud computing [16]. At the same time, aiming at the problem of "data island", Google proposed a new concept of "multi-source data fusion" for the first time [17]. The model is individually trained on multiple edge devices using training samples, and sharing of multi-source information is achieved by aggregating model parameters without disclosing user privacy. In addition, due to the diversity of edge devices, the data collected by the devices exhibits diversity in annotations, semantics, and existing formats [18]. For example, for the description of the same object, there are text class, picture class, video class, and audio class data, and multi-modal data widely exists [19]. Different modal data can describe the target object from many aspects [20]. By eliminating redundant data and fusing various data sources for correlation and supplementary analysis, more valuable new information can emerge from the data, to achieve the effect of $1 + 1 > 2$. Multimedia data collected from the Internet and mobile devices are typical unstructured data, which is significantly different from the traditional structured data format [21]. Therefore, the processing of multi-source heterogeneous data collected by different edge devices becomes an urgent problem in big data research [22]. In the traditional multi-source heterogeneous data fusion algorithm, data centralized processing has the risk of data privacy leakage in practical application [23]. Therefore, there are still many problems in multi-source heterogeneous data processing without disclosing user privacy: first, due to enterprise competition and user privacy protection awareness, data exchange is blocked for a long time, and information sharing cannot be realized, so the value of heterogeneous data cannot be fully exploited. Secondly, the neural network is used to process the data, and the model designed according to the data cannot be changed once it is determined [24]. However, in edge computing, there are differences in data structure and the number of types collected by edge devices. If we design a neural network for the data of each network edge device, the workload is huge [25–26]. At the same time, the model can only be applied to a single node or the edge device with the same data characteristics as the node, the universality is not high, and it can not give full play to the value of other heterogeneous data in the Internet of things [26]. To solve the problem of multi-source heterogeneous data fusion without disclosing user privacy in edge computing, this paper proposes a multi-source heterogeneous data fusion algorithm based on multi-source data fusion [27–30]. Starting from the data structure characteristics collected by edge devices, combined with tensor tucker decomposition theory, this paper studies the adaptive processing of multi-source heterogeneous data model on different edge devices, and solves the single adaptability problem caused by the disunity of the heterogeneous data model in multi-source data fusion.

2.2. Multi-Source Data Fusion System Model. In this paper, we aim at the problem of heterogeneous data fusion without data exchange, consider the introduction of multi-source data fusion to edge computing, and learn the potential characteristics of multi-user without invading user privacy. The basic architecture of its intelligent sensor system is shown in Figure 1.

In this framework, the system is composed of an edge node, Internet of things, and cloud server. The edge node connects with the cloud server through the Internet of things (such as gateway and router). Multi-source data fusion is a distributed learning framework, in which the original data is collected and stored on multiple edge nodes, and the model training is performed at the nodes, and then the model is gradually optimized through the interaction between node n and cloud server h. The formula is as follows

$$H = \frac{\sum_{j=1}^{k} \sum_{h=1}^{k} \sum_{t=1}^{n_j} \sum_{r=1}^{n_h} \left| y_{ij} - y_{hr} \right|}{2n^2 u} \tag{1}$$

Based on the above framework, multi-source data fusion can use local data from multiple independent edge nodes sharing model, and use model transmission instead of data transmission to avoid the risk of user privacy leakage in the process of transmitting the collected original data from multiple edge nodes to the cloud server.
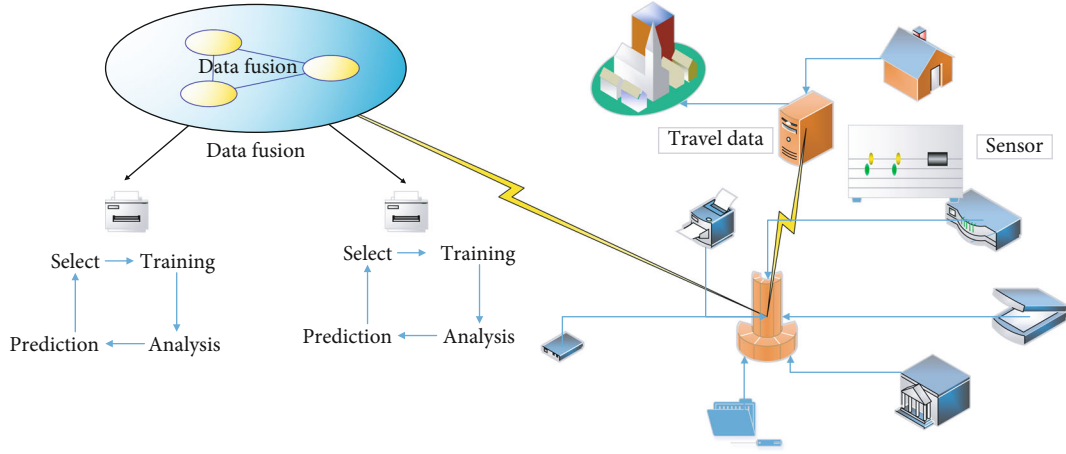
FIGURE 1: Multivariate data fusion framework diagram.

The multi-source heterogeneous data fusion algorithm proposed in this paper is mainly divided into feature extraction modules, feature fusion modules, and feature decision modules. The feature extraction module is composed of feature extraction sub-networks corresponding to various heterogeneous data. In the initialization stage, the central control node randomly initializes the network parameters of the feature extraction module, feature fusion module, and feature decision module, and sends them to the edge node T, where the fusion parameter is C.

$$\sum_T = C(\max(\sigma_i, v, 0))$$
$$E_j = \frac{1/2 u_j \sum_{i=1}^{n_j} \sum_{r=1}^{n_j} \left| y_{ji} - y_{jr} \right|}{n_j^2} \quad (2)$$

In the model training stage, after receiving the model from the central control node, the edge node selects the corresponding feature extraction module according to the data set structure of the local node and uses the local data set to train the feature extraction module, feature fusion module, and feature decision module. The termination condition of a new round of edge node training is that the number of local node training rounds exceeds the given number of training rounds. After training, the training model is returned to the central control node for model aggregation. The first practice test is to use the average aggregation algorithm for the function fusion module and the function determination module. For the feature extraction module, the average sub-module is extracted according to the corresponding feature extraction submodule to ensure that the same mode data extract features are similar. Finally, the updated model is redistributed to the edge nodes for a new round of training. This paper assumes that the heterogeneous data to be processed are audio, visual, and textual data. In the feature extraction module, according to the features of different modes, different feature extraction sub-networks are used to extract the features of audio, visual, and text information. Audio and visual feature sub network: for audio information and visual information, covered acoustic analysis framework and face facial expression analysis framework is used to sample and extract features from MoSi data sets (sampling frequency is 100 Hz and 30 Hz, respectively). In this section, it is assumed that the heterogeneous data features to be processed are audio data features

$$V_{jh} = \frac{\sum_{z=1}^{hj} \sum_{r=1}^{n_h} \left| y_{ji} - y_{hr} \right|}{n_j n_h (u_j + u_h)} \quad (3)$$

Visual data features:

$$V_{nb} = \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} \left(p_j s_h + p_h s_j \right) D_{jh} \quad (4)$$

Text data features:

$$V_0 \int_0^\infty dF_j(y) \int_0^y (y-x) dF_h(x) \quad (5)$$

After the feature fusion module, the feature output is v0. Next, taking the above assumption as the basic condition, the basic principle of the proposed heterogeneous data fusion algorithm based on tucker decomposition is described. The first mock exam module introduces a high tensor W with heterogeneous data feature space, and each mode of the tensor corresponds to spatial mapping of heterogeneous data characteristics. Therefore, when fusing the features of each heterogeneous data, the high-order tensor w can not only introduce the features of other heterogeneous data modes for correction but also memorize the ongoing heterogeneous data modal features. When the features of the heterogeneous data to be processed are P and S, the memory unit W is a third-order tensor, and the three dimensions of the tensor corresponding to the feature spaces of the three heterogeneous data features. In the heterogeneous data feature fusion proposed in this section, the memory unit with the heterogeneous data feature can be obtained by modular multiplication of the heterogeneous data feature and the feature space

corresponding to the memory unit, and further feature fusion operation can be carried out. The fusion operation is mainly divided into three stages. The memory unit W is modulo multiplied with the heterogeneous data feature AZ along the first order to obtain a new memory unit with AZ feature.

$$W_{jh} \frac{AZ_{jh} - P_{jh}}{AZ_{jh} + P_{jh}} \qquad (6)$$

Secondly, memory unit W is modulo multiplied with heterogeneous data features along the second-order to obtain memory unit w with AZ features. Finally, the memory unit W is modularly multiplied with the heterogeneous data features along with the third order, and finally, the fusion tensor with the three features is obtained Z. The specific process can be expressed as follows:

$$E_{AZ} = \sum_{j=1}^{k} G_{jj} p_j s_j \qquad (7)$$

Among them:

$$P_j = \int_0^\infty dF_h(y) \int_0^y (y - x) dF_j(y) \qquad (8)$$

For the fused data, this chapter uses the traditional full connection layer to make decisions based on the global characteristics, including the prediction of regression model and the probability prediction of a classification model. In this module, the L1 norm loss function is used to measure the error between the target value and the predicted value. Its et expression is

$$E_T \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} \left(p_j s_h + p_h s_j \right) D_{jh} \left(1 - L1_{jh} \right) \qquad (9)$$

The expression of NL is:

$$NL_t = \tanh \left(w_c, x_t + u_c (r_t \Theta h_{t-1}) + b_c \right) \qquad (10)$$

Then, it is assumed that n edge nodes participate in the training of the shared model, and all edge nodes collect m kinds of heterogeneous data. In the initialization stage, according to the collected m kinds of heterogeneous data, the cloud designs the corresponding feature extraction module F, feature fusion module I and feature decision module C. Then the shared model G can be expressed as:

$$G = F * f/C = \frac{\sqrt{1/M \sum_{i=1}^{n} (I_{it} - I_{it})^2}}{I_{it}} \qquad (11)$$

Where * represents the model splicing operation. Specifically, in the feature extraction module F, the corresponding feature extraction sub-network MF is designed from kinds of heterogeneous data, which can be expressed as:

$$\ln \left(\frac{MF_{it}}{MF_{it} - 1} \right) = \alpha + \beta \ln MF_{it} - 1 + \varphi X_{it} - \tau_t \qquad (12)$$

Where x is the feature extraction sub-network of the I heterogeneous data. In feature fusion module I, a high-order tensor with spatial dimension characteristics of heterogeneous data is constructed. After training, the parameters of the tensor expanded along the I module can reflect the spatial dimension characteristics of the I heterogeneous data. In the feature decision-making module C, through the training of the fused heterogeneous data features, the potential relationship between heterogeneous data is mined in a deeper level, and the feature expression of the model in multi-source heterogeneous data is improved. Due to the diversity of data collected by edge devices, the data processed by each edge node is different. Therefore, in the premise of not disclosing user privacy, multi-source heterogeneous data fusion has the problem of insufficient adaptability. The size of the tensor after fusion is consistent with that of memory unit W. Therefore, when the factor matrix satisfies the square matrix constraint, there is an identity relationship between the core tensor and the original tensor in the spatial dimension. Using this feature, in the initialization phase, the global is further set, and the feature extraction sub-network fi feature graph is defined

$$\ln \left(\frac{FI_{it}}{FI_{it} - 1} \right) = \alpha + \beta \ln FI_{it} - 1 + v_i + \mathfrak{J}_t \qquad (13)$$

Thus, the problem of heterogeneous data fusion caused by the uncertainty of heterogeneous data in edge computing is solved. In the model training stage, n edge nodes participate in the training according to the heterogeneous data types they have

$$N_t z_t \Theta h_{t-1} + (1 - z_t) \Theta h_t \qquad (14)$$

The corresponding feature extraction sub-network FZ is adaptively selected for training. In the feature fusion module, because the feature graph of the feature extraction sub-network r is set to FR in the global initialization definition stage, the tensor size after heterogeneous data fusion is constrained to a fixed value

$$R = \frac{FR}{1 - FZ}$$
$$FR = -\frac{1}{T} \ln(1 + \beta) \qquad (15)$$

Suppose that all the heterogeneous data types to be processed are N, X, and H, respectively, and the three dimensions of the corresponding memory unit w correspond to the feature spaces of the three heterogeneous data. The

heterogeneous data types collected by node 1 and node n are different. In the model training stage, the features are obtained, respectively

$$f(\text{N}) = \frac{1}{Nh}\sum_{i=1}^{N} k\left(\frac{X_i - x}{h}\right)$$

$$\text{W} = k(\text{H}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{16}$$

According to the number of heterogeneous data types on the node, the feature fusion stage is divided into two parts: first, memory unit W is modulo multiplied with AF feature along the first order to get a new memory unit w with AF feature. Secondly, the memory unit w multiplies the VF features along the second-order to obtain the fusion tensor with the above two heterogeneous data features $Z_{\circ}$ The process can be expressed as follows:

$$\text{W}1_j \sum_i c_{ij} u_{(\text{AF}|\text{AV})} \tag{17}$$

W1 is the fusion of VF features based on AF features. In this process, the model first uses the memory unit to memorize the AF features, and obtains the model with AF features, and takes this as a priori condition for the fusion of VF features. Thus, in the process of model training, the memory unit can not only learn the spatial dimension features of each heterogeneous data but also capture the potential relationship between different heterogeneous data. The training mechanism on node n is similar to that on node 1. The above process can be expressed as follows:

$$k_{t1}[\text{N}] = \sum_j \cos\left(w_i^1, w_j^2\right)/\text{K} \tag{18}$$

Where n is the node K in the j-round global iteration, using the heterogeneous data collected locally, the learning rate is η The gradient descent algorithm is used to get the local model. In the model aggregation stage, because each edge node uses the adaptive selection mechanism of feature extractor to train the feature extraction module, it is necessary to merge the feature extraction sub-networks selected and trained by each edge node, and then use the average aggregation algorithm to get the shared model with global heterogeneous data features

$$u_{(j|i)} = w/\text{N} *_{ij} A_i \tag{19}$$

Where a is a shared model with global characteristics obtained by aggregating the local models on n edge nodes through the joint average algorithm. J is the number of all heterogeneous samples and the total number of all heterogeneous samples on the edge node.

3. Multivariate Fusion Data Model

3.1. Methods. This paper constructs a multi-source data fusion model based on an ensemble learning algorithm, uses Ctrip 2020 open data set to train the model, and then obtains the tourism information data processing and prediction results. This paper takes the data of Ctrip as the training set and compares the trained model with the data of tunic and Feizhu. In this paper, sensor detection technology is used to analyze many famous scenic spots in China, including tourist type (youth, adult), gender, and location.

3.2. Data Acquisition and Processing. Through a comprehensive comparison of other online tourism service platforms such as tunics and Hitake, this paper has the highest number of online comments on http://Trip.com/'s Saikei Wetland, and Ctrip's online comments are rich in content and have interference factors. Few, positive and negative comments, and most of them are the true feelings of tourists. Therefore, it is common and effective to choose Ctrip online game reviews as your data source. Through Python technology, this paper grabs the online comments of tourists in Xixi Wetland scenic spot on Ctrip from May 1, 2020, to May 1, 2021, removes the invalid online comments that deviate from the research theme, and collates 3052 effective online comments, and makes specific research and analysis of the collected online comments. In addition, in cooperation with Xixi Wetland Tourism Development, the research group carried out an offline questionnaire survey on tourists' comprehensive satisfaction of Xixi National Wetland Park during the long holidays of May Day and national day in 2020, and obtained 102 valid questionnaires, which fully investigated the customer source market and tourists' satisfaction of Xixi National Wetland Park, as a supplementary means of network big data survey, this paper puts forward more relevant and targeted suggestions for the future development of Xixi National Wetland Park.

According to the text data of tourists' online reviews and the comprehensive satisfaction offline questionnaire, the research results of Xixi Wetland Park's tourism image perception are relatively consistent. When analyzing the influencing factors of tourism perception experience, there are many overlapping problems, that is, the important concerns when improving the tourism brand image. For example, given the insufficient setting of public toilets, rest chairs, and free drinking water points, drinking water, and lunch supply points should be added accordingly; For the lack of a single tour line of battery car and battery boat in the park and long waiting time for tourists to enter and leave the park, we should provide parking guidance service; Continuously improve service awareness and management level, and focus on ticket and catering prices, safety management of cruise terminal, queuing guidance, increase the number of sanitation personnel, increase recreational equipment, strengthen the integration of tourism formats and products, and improve the effective supply of products.

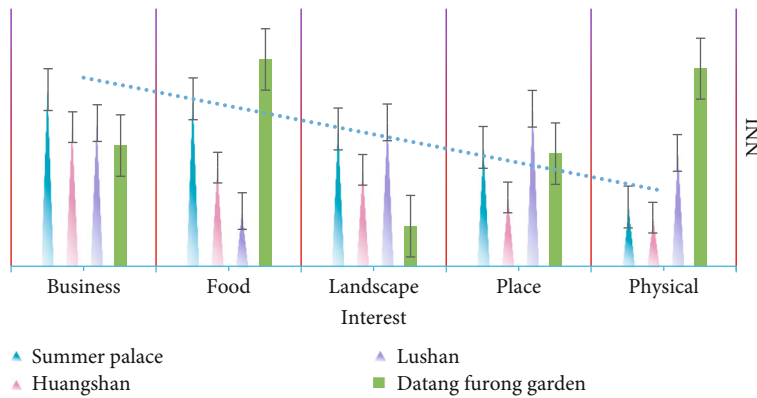Figure 2: Word frequency statistics for comment texts of famous scenic spots.

Figure 3: Overall NNI of leisure and tourism space.

4. Results and Discussion

4.1. Multi-Source Information Fusion Tourism Information Processing Method. Use Python to perform word frequency statistics on the comment texts of scenic spots and famous scenic spots in China, and extract the top 50 high-frequency feature words, as shown in Figure 2. Key terms such as "travel" and "attractions" are ranked high, indicating that projects that attract tourists to Tiger Beach are primarily special venues and ropeways across the ocean. The most attractive to tourists is the third-ranked "performance", such as "dolphin show". In addition, the characteristic words such as "getting tickets", "buying tickets" and "tickets" indicate that tourists pay great attention to the purchase of tickets, and the innovation and improvement of tourist ticket purchase methods in scenic spots should be highly valued. Compared with China's famous scenic spots, "shows" ranked first, indicating that watching performances are the main activity of tourists in Shengya.

As shown in Figure 3, the overall NNI (neural network intelligence) of leisure tourism space is less than 1, and the NNI of different types of space is also less than 1, among which catering services and commercial services are lower, indicating that these two types are more inclined to cluster

development, and the cluster scale is larger. Catering and business services have a strong economic effect on agglomeration and are affected by factors such as economic level, population density and convenience of transportation. The NNI of sports leisure and cultural entertainment services were 0.35 and 0.52, respectively, which also showed a general trend of agglomeration. The NNI of scenic spot service is the largest, which is 0.46. Because of the influence of natural ecological conditions and urban open space, the agglomeration degree of all scenic spots is relatively low.

As shown in Figure 4, it has passed the significance test at the level of 5%, and the influence coefficient is positive, indicating that the innovation and technological progress of the tourism industry will play a promoting role in the growth of local tourism consumption without considering other factors. As the main platform of scientific and technological innovation in the industry, tourism colleges and research institutes, represented by coastal areas, have formed a certain scale. A large number of tourism colleges and universities have formed mature curriculum and perfect talent training systems, with rich research results, which are applied to the development and design of tourism creative products and tourism service supply management.
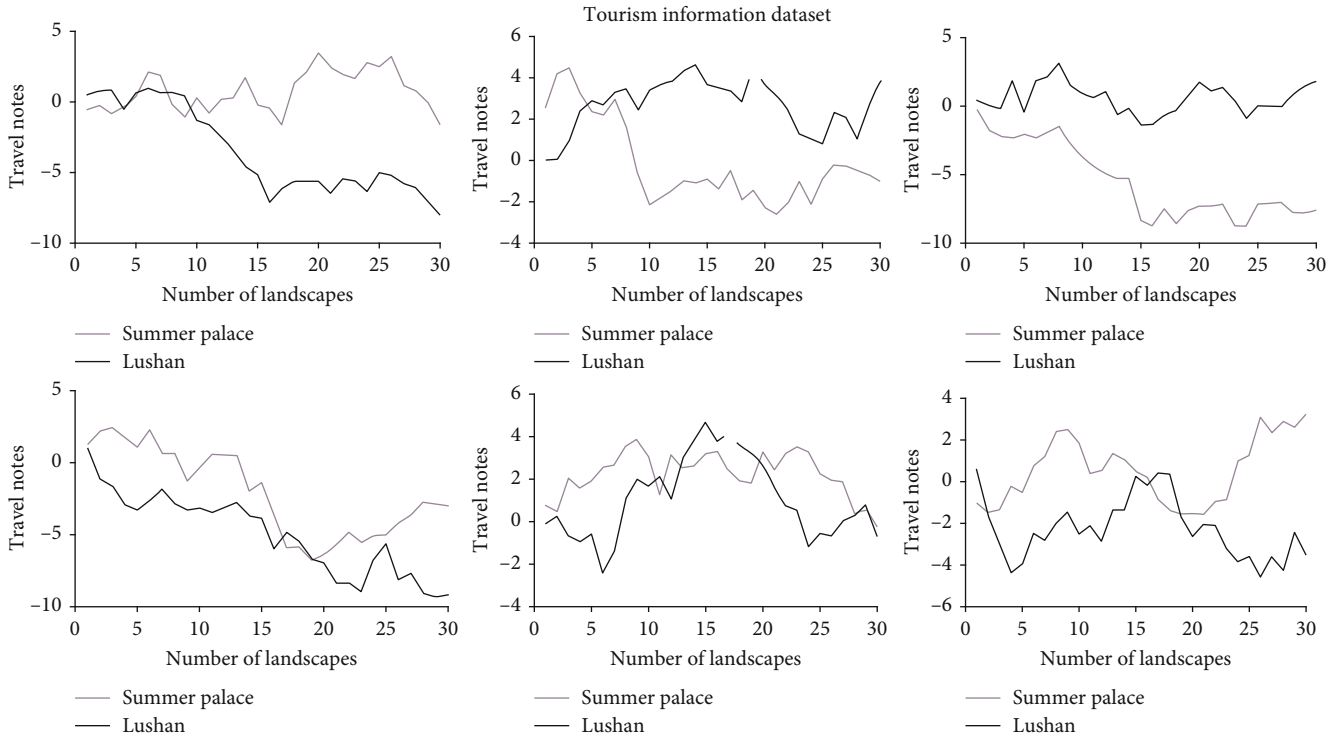
Figure 4: Tourism Information Dataset.

Under the special evaluation content, the target layer, criterion layer, and index layer of urban physical examination are established, and the index system reflecting the urban natural background and the operational signs of tourism information is constructed, as shown in Table 1. At the same time, according to the national standards of relevant indicators, the vertical establishment of planning expectation value and the horizontal comparison of the level of cities at the same level, the standard value and reference interval is determined, and the evaluation and calculation are carried out according to the positive and negative direction of the indicators. According to the data sources and evaluation contents of each index, the calculation methods can be divided into two categories: tourism information spatial analysis and statistical data analysis. For the smaller-scale spatial data of urban districts and counties and below, the methods of tourism information computing geometry, buffer analysis and overlay analysis can be used, such as calculating the coverage of public service facilities through buffer analysis and overlay analysis. Using the statistical data, we can summarize the static data of the basic elements of the city, calculate the average value, per capita value, proportion, number statistics, and coverage analysis.

As shown in Figure 5, the distribution and total amount of tourism information are always increasing. To analyze the change direction and trend of the center of gravity of urban space and leisure tourism space more intuitively, the migration trajectory of the center of gravity of urban space is drawn by using tourism information. It shows that the overall development of urban space and leisure tourism space is first to the northwest, then to the southwest, and the spatial growth is mainly in the southwest cities.

Table 1: The Index System of Tourism Information Operation Signs.

Item	Summer palace	Huangshan	Lushan	Datang Furong garden
Business	4.8	3.78	3.89	3.28
Food	4.55	2.68	1.5	5.62
Landscape	3.73	2.62	3.91	1.09
Place of interest	3.23	1.87	4.28	3.06
Physical education	1.61	1.32	3.08	5.39

Among the famous scenic spots in China, there are "dinosaur museum", "Ocean Museum" and "undersea tunnel". The reason why the "dinosaur museum" ranks high is that it is far away from the other four museums, which is also one of the important reasons why "walking" ranks high. As shown in Figure 6, the second-highest frequency word in the two scenic spots is "children", which indicates that the two scenic spots have a strong attraction for children and are the choice of many families for parent-child travel. In addition, "a lot of people", which ranks third among the famous scenic spots in China, ranks 50th among the characteristic words of tiger beach. The main reason is that Tiger Beach covers a large area and tourists are more evacuated. China's famous scenic spots are more attractive to tourists.

As shown in Table 2, tourists have a higher perception of tourism attractions and experience of tiger beach and Shengya, but a lower perception of tourism environment and facilities and services, and there are obvious differences
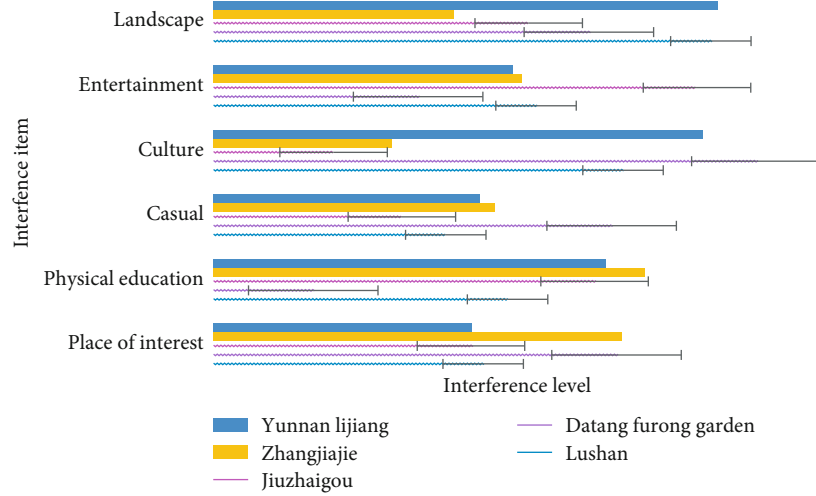
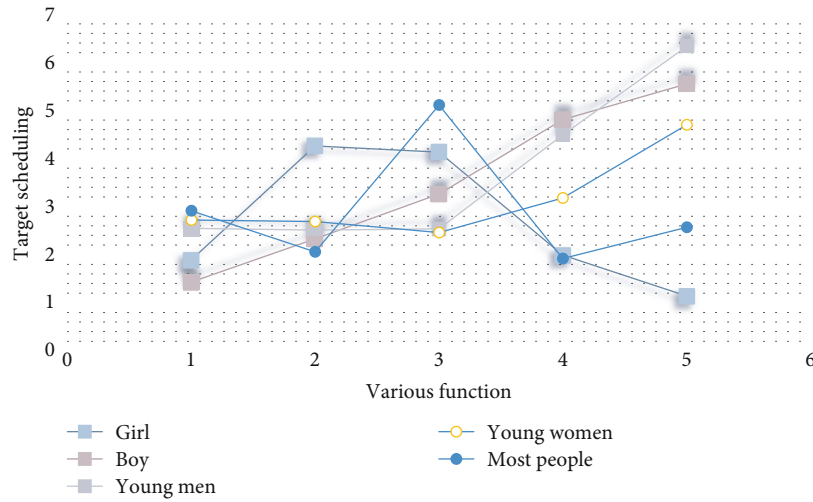Figure 5: Distribution and Total Change Trend of Tourism Information.

Figure 6: Projects that attract tourists among China's famous attractions.

Table 2: Perception of tourist attractions and tourist experience.

Item	Landscape	Place of interest	Physical education	Casual
Summer palace	0.8	0.12	1.62	0.86
Huangshan	2.06	3.46	2.95	3.87
Lushan	5.23	3.86	3.47	3.19
Datang Furong	4.1	5.08	4.55	2.36
Jiuzhaigou	1.64	1.76	3.54	4.43
Zhangjiajie	4.66	2.31	2.96	5.06

Table 3: Tourists' cognitive evaluation and emotions.

Item	Datang Furong garden	Jiuzhaigou	Zhangjiajie	Yunnan Lijiang
Place of interest	4.96	3.17	5.02	3.18
Physical education	1.23	4.69	5.31	4.83
Casual	4.9	2.32	3.46	3.28
Culture	6.68	1.48	2.19	6.02
Entertainment	2.52	5.95	3.79	3.68

between the two scenic spots. In the overall perception of scenic spots, tiger beach and Shengya have little difference. The two scenic spots belong to marine theme scenic spots, in which leisure and entertainment are all around the ocean, which is a romantic ocean trip for tourists; Children are

attracted by all kinds of interesting marine creatures and performances in the scenic area, which is also a good choice for parent-child travel. Among the tourist attractions, tiger beach and Shengya are quite different, which indicates that tourists pay more attention to the tourist attractions, and
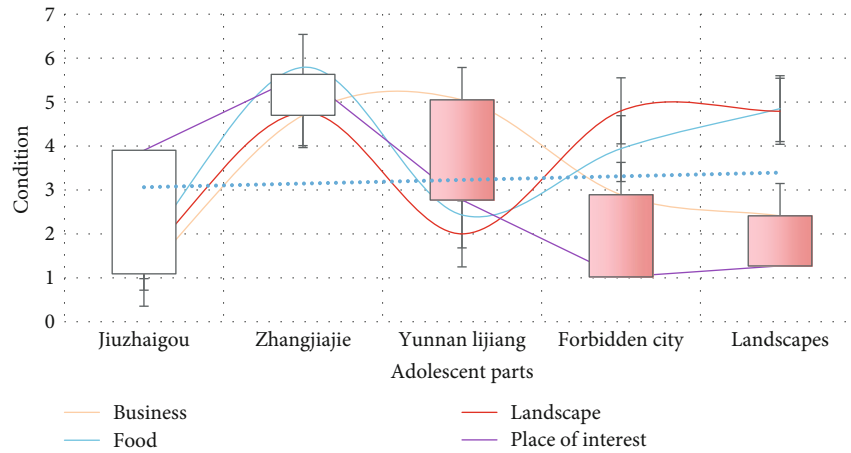
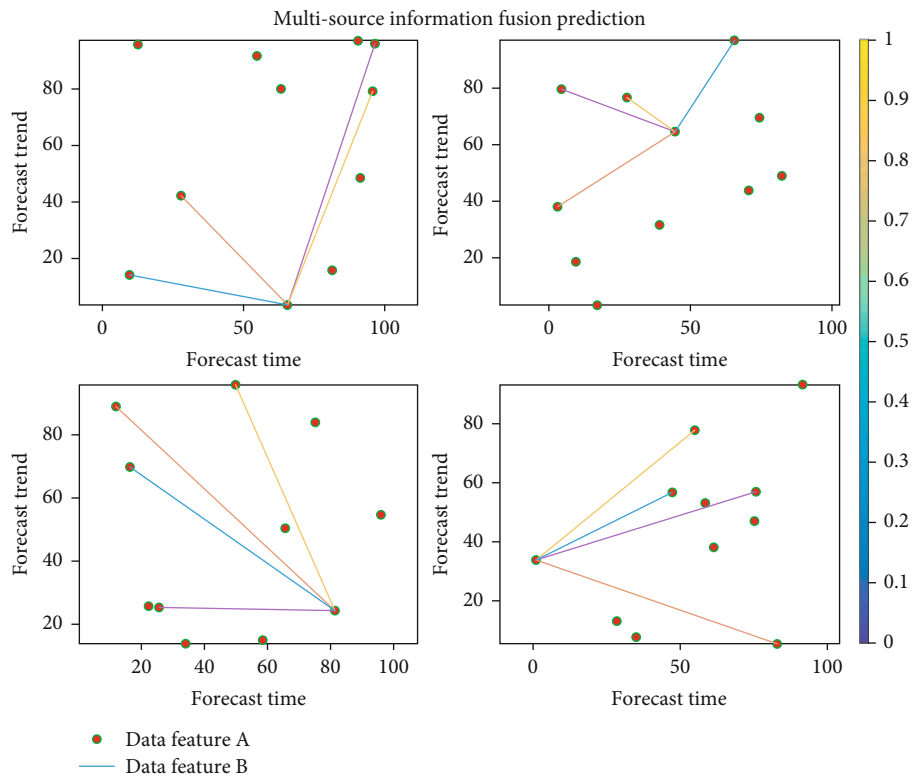FIGURE 7: The overall trend of tourism information.

FIGURE 8: Multi-source information fusion prediction.

Shengya is more attractive. Tourists have a higher perception of Shengya theme venues and special performances.

As shown in Table 3, among the famous scenic spots in China, tourists' cognitive evaluation is more positive and there are not too many negative emotions. This shows that tourists are more satisfied with the service of famous scenic spots in China. Tourists' perception of the tourism environment and facilities of tiger beach and Shengya is weak. In the level of tourism environment perception factors, there are obvious differences in social environment perception. Among them, China's famous scenic spots are located in urban areas and are very close to Xinghai Square, which makes "Xinghai Square" stand out. Dalian Xinghai Square

itself is one of the scenic spots that Dalian must visit. Tourists will take two scenic spots as a one-day tour plan, while Tiger Beach Ocean Park is relatively remote, this is also one of the important reasons why tourists sometimes choose Shengya when they choose two marine theme scenic spots.

As shown in Figure 7, through the coupling analysis of tourism information data and scenic spot data, it can be found that the overall tourism information is gradually decreasing from the central urban area to the surrounding areas, and there is a high coupling between the two data, and the areas showing coupling differences can also reflect the buildings with the same nature, such as stations, airports, industrial development zones, and other large areas. There is
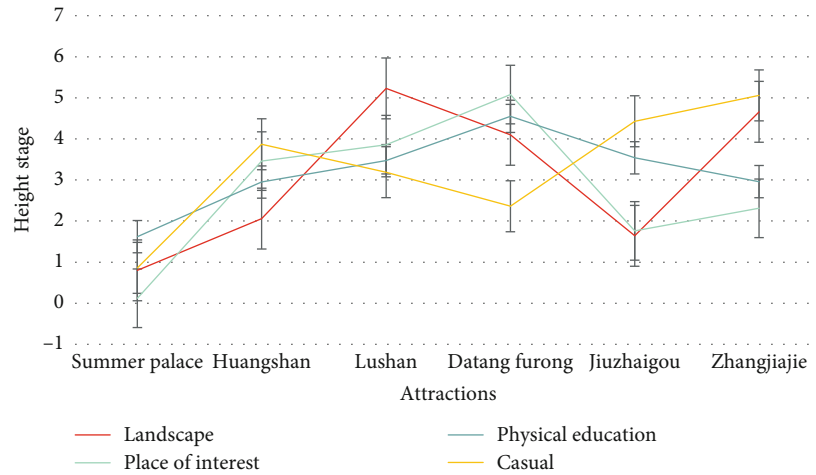
FIGURE 9: The center of gravity of urban space and leisure tourism space.

a wide range of low low coupling around the central city because the economic construction of suburban towns is relatively backward, so the value of tourism information and various leisure tourism service facilities are low, which is in line with the reality.

The center of gravity of urban space and leisure tourism space is shown in Figure 8. This paper finds that with the increase of human activity intensity and urban construction, the area of tourism information increases year by year, and the direction of gravity shift is from southwest to northwest. The results show that the multi-method comprehensive analysis based on big data can better describe the spatial distribution attributes and spatiotemporal evolution characteristics of the region. For example, this paper objectively and comprehensively reveals the agglomeration characteristics of leisure tourism space and provides indispensable reference information for land space planning. But at the same time, there are some limitations. The analysis of the characteristics of regional spatial agglomeration in this paper is more about the results, and the analysis of its driving factors or influencing factors is still lacking.

As shown in Figure 9, according to the results of the multi-source fusion of tourism information in this model, the results of tourism feature extraction are consistent with the data of Feizhu, tunic, and other websites in the trend. Among them, in the data of the first half of 2020, the prediction accuracy of the model after data processing is about 62%. Affected by the epidemic situation, the accuracy of the model is low. In the second half of the year, the prediction accuracy is 78%, which can be used to fuse tourism information in a short time. Therefore, the data show that the model has high learning ability and high trend prediction ability in tourism data processing, which can provide necessary information support for tourists.

4.2. Discussion. Given the target redundancy in the reconstruction results of specific scenic spots, this paper constructs a multi-source data fusion model based on an ensemble learning algorithm. In the field of 3D reconstruction, the irrelevant objects in a specific scene are eliminated by data fusion to achieve 3D scene reconstruction! Firstly, the lightweight algorithm is used to extract and match the features of different types of feature points, and the point clouds at different times are fused to complete the reproduction of the point cloud map. Then, for the irrelevant targets that may exist in the constructed point cloud map, with the help of multi-source sensor data and deep learning application technology in the field of computer vision, the target detection and elimination are carried out in three-dimensional space. For the two different processes of point cloud map modeling and target detection, the point cloud registration method is used to fuse them, and finally, the scene reappearance in the scenic spot environment is completed! The experimental results show that the method based on multi-source data fusion can effectively combine the two processes of 3D modeling and target detection, and complete the construction of a point cloud map without redundant targets in scenic spots.

By comparing the online comment text data of tourists and the offline comprehensive satisfaction survey questionnaire, it is found that the survey results of the tourism image perception of West Wetland Park are relatively consistent. Based on multi-source data integration of leisure travel space based on tourist information data and tourist attraction data, it is found that the leisure travel space in the central city area is distributed in a clear central multipoint layout, and the overall leisure travel is there. In the urban area, there is a trend of high agglomeration of scenic spots, which shows that scenic spots are the most concentrated areas of leisure tourism, while the agglomeration effect of peripheral counties is weak. In addition, the overall distribution pattern of leisure tourism space shows obvious characteristics of denseness and sparseness and a trend from northwest to southeast.

5. Conclusions

The multi-source fusion of tourism information of this model shows that the results of this model are consistent with the data of Feizhu, tunic, and other websites in the trend. Among them, in the data of the first half of 2020, the

prediction accuracy of the model after data processing is about 62%. Affected by the epidemic situation, the accuracy of the model is low. In the second half of the year, the prediction accuracy is 78%, which can be used to fuse tourism information in a short time. Therefore, the data show that the model has high learning ability and high trend prediction ability in tourism data processing, which can provide necessary information support for tourists. Based on the online comments on Ctrip, combined with the survey results of tourists' comprehensive satisfaction in the same period, the online tourism big data and offline questionnaire verify each other, comprehensively analyze the tourism image perception of Xixi National Wetland Park, and innovate the means and methods of tourism image perception research, However, there are still some deficiencies in the number and period of tourists' online comments. In the follow-up research, it is necessary to further increase the number of online tourism big data and offline questionnaire survey samples, and enrich the correlation analysis of tourists' attributes, temporal and spatial behavior characteristics and tourism image perception, try to reduce the error caused by the difference of online and offline survey samples.

Data Availability

The data underlying the results presented in the study are available within the manuscript.

Conflicts of Interest

The author(s) declare (s) that they have no conflicts of interest.

References

[1] Y. Chen, "A survey on industrial information integration 2016-2019," *Journal of Industrial Integration and Management*, vol. 5, no. 4, pp. 56–59, 2020.

[2] C. F. Shang, Y. F. Wang, and J. L. Du, "Information integration for motor generation," *Current Opinion in Physiology*, vol. 8, no. 4, pp. 116–120, 2019.

[3] A. Oberoi, M. Arora, and V. K. Garg, "A novel approach for dynamic information integration," *International Journal of Reasoning-based Intelligent Systems*, vol. 13, no. 2, pp. 76-77, 2021.

[4] P. de Bernardi, A. Bertello, and R. Shams, "Logics hindering digital transformation in cultural heritage strategic management: an exploratory case study," *Tourism Analysis*, vol. 24, no. 3, pp. 315–327, 2019.

[5] W. C. Poon and K. Y. Koay, "Hong Kong protests and tourism: Modelling tourist trust on revisit intention," *Journal of Vacation Marketing*, vol. 4, no. 4, pp. 1356–1359, 2021.

[6] Y. A. Poetra, "Upaya Pemerintah Dalam Mengkomunikasikan Tradisi Malamang Menjadi Objek Pariwisata Budaya Di Kabupaten Padang Pariaman," *Jurnal Pustaka Budaya*, vol. 5, no. 2, pp. 52–61, 2018.

[7] G. R. Ramadhan and I. Buchori, "Strategi Integrasi Sistem Transportasi Umum Dalam Menunjang Pariwisata Kota Yogyakarta," *Jurnal Pengembangan Kota*, vol. 6, no. 1, pp. 84-85, 2018.

[8] W. Huyan and J. Li, "Research on rural tourism service intellectualization based on neural network algorithm and optimal classification decision function," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 4, pp. 1–21, 2021.

[9] M. Özturan, M. Mutlutürk, B. Çeken, and B. Sarı, "Evaluating the information systems integration maturity level of travel agencies," *Information Technology & Tourism*, vol. 21, no. 2, pp. 237–257, 2019.

[10] P. Liu, H. Zhang, J. Zhang, Y. Sun, and M. Qiu, "Spatial-temporal response patterns of tourist flow under impulse pre-trip information search: from online to arrival," *Tourism Management*, vol. 73, no. 8, pp. 105–114, 2019.

[11] M. Zhou, Y. Li, M. J. Tahir, X. Geng, Y. Wang, and W. He, "Integrated statistical test of signal distributions and access point contributions for Wi-fi indoor localization," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 5, pp. 5057–5070, 2021.

[12] G. Xiao, Q. Cheng, and C. Zhang, "Detecting travel modes using rule-based classification system and Gaussian process classifier," *IEEE Access*, vol. 7, pp. 116741–116752, 2019.

[13] X. Guangnian and Z. Wang, "Empirical Study on Bikesharing Brand Selection in China in the Post-Sharing Era," *Sustainability*, vol. 12, p. 3125, 2020.

[14] T. Arenas, M. N. Martínez, H. Xu, O. Morales, and M. Chávez, "Integrating VSM and network analysis for tourism strategies-case: Mexico and the Chinese outbound market," *Systemic Practice and Action Research*, vol. 32, no. 3, pp. 315–333, 2019.

[15] R. Strulak-Wójcikiewicz, N. Wagner, A. Łapko, and E. Hącia, "Applying the business model canvas to design the E-platform for sailing tourism," *Procedia Computer Science*, vol. 176, no. 5, pp. 1643–1651, 2020.

[16] M. Shiraishi, H. Ashiya, A. Konno et al., "Development of real-time collection, integration, and sharing Technology for Infrastructure Damage Information," *Journal of Disaster Research*, vol. 14, no. 2, pp. 333–347, 2019.

[17] K.-Y. Kwon, "A study on the path-dependent in Korea and China's early cultural industry policy," *The Journal of Chinese Cultural Research*, vol. 39, no. 6, pp. 88-89, 2018.

[18] D. E. Setyowati, Y. Antariksa, E. Haryati, H. Haryono, and I. P. P. Salmon, "Local tourism? Why not! Integrating tourism geographic spatial in Ngawi Regency, Indonesia," *Modern Applied Science*, vol. 14, no. 9, pp. 1–8, 2020.

[19] V. V. Yavorska, I. V. Hevko, V. A. Sych, O. I. Potapchuk, and K. V. Kolomiyets, "Features of application of information technologies in modern tourism," *Journal of Geology Geography and Geoecology*, vol. 28, no. 3, pp. 591–599, 2019.

[20] W. Zhu, Y. Hou, E. Wang, and Y. Wang, "Design of Geographic Information Visualization System for marine tourism based on data mining," *Journal of Coastal Research*, vol. 103, no. 1, pp. 30–34, 2020.

[21] K. Y. Chen and S. Y. Yang, "A cloud information monitoring and recommendation multi-agent system with friendly interfaces for tourism," *Applied Sciences*, vol. 9, no. 20, pp. 43–48, 2019.

[22] I. Ramos-Soler, "AM Martínez-Sala, Campillo-Alhama C. ICT and the sustainability of world heritage sites. Analysis of senior Citizens' use of tourism apps," *Sustainability*, vol. 11, no. 11, pp. 32-33, 2019.

[23] S. L. Ratnasari, E. N. Susanti, W. Ismanto, R. Tanjung, D. C. Darma, and G. Sutjahjo, "An experience of tourism development: how is the strategy?," *Journal of Environmental Management and Tourism*, vol. 11, no. 7, pp. 1877–1886, 2020.

[24] T. Melikh, D. Voit, and D. Archybisova, "Aquacultural Integration in recreational tourism: features of development and management of coastal territories," *Baltic Journal of Economic Studies*, vol. 5, no. 5, pp. 84–95, 2020.

[25] M. N. Rifa'I, "Integrasi Pariwisata Halal di Kota Malang," *FALAH Jurnal Ekonomi Syariah*, vol. 4, no. 2, pp. 194–196, 2019.

[26] G. Manyara, "Regional tourism in inter-governmental authority on development: a comparative policy and institutional best practice approach," *International Journal of Tourism Policy*, vol. 9, no. 1, pp. 50–52, 2019.

[27] M. Zhou, Y. Long, W. Zhang et al., "Adaptive genetic algorithm-aided neural network with channel state information tensor decomposition for indoor localization," *IEEE Transactions on Evolutionary Computation*, 2021.

[28] M. Abd, "The integration time-driven activity-based costing (TDABC) and events approach their role in decision-making and their effect on tourism," *African Journal of Hospitality Tourism and Leisure*, vol. 8, no. 2019, pp. 51–53, 2019.

[29] R. Peng, Y. Lou, M. Kadoch, and M. Cheriet, "A human-guided machine learning approach for 5G smart tourism IoT," *Electronics*, vol. 9, no. 6, pp. 947–949, 2020.

[30] Z. Lv, D. Chen, R. Lou, and Q. Wang, "Intelligent edge computing based on machine learning for smart city," *Future Generation Computer Systems*, vol. 115, pp. 90–99, 2021.