A BACK-PROPAGATION NEURAL NETWORK WITH DELAY AND SHIFT WINDOW FOR TOURISM DEMAND FORECASTING

Thanh-Nghi Doan[1,2]
[1]Faculty of Information Technology, An Giang University, An Giang, Vietnam
[2]Vietnam National University Ho Chi Minh City, Vietnam
dtnghi@agu.edu.vn

Abstract

This article studies machine learning techniques and factors that affect tourism demand to develop a predictive model for tourism demand in the coming years. The model was developed using the back-propagation neural network approach and expert knowledge for analyzing factors affecting tourist satisfaction. The data used in the study were collected over a ten- year period and comprised information on the local economic and social situation, as well as specialized tourism data. In addition, survey results evaluating tourism in An Giang province in 2019 were included. The study results demonstrate that the developed model has successfully captured the underlying patterns in the An Giang tourism data, enabling the prediction of the necessary tourism indicators for the future. The model achieved a high level of accuracy with an RSME of 0.04. Furthermore, our approach showed several advantages when compared to other classical statistical methods. Based on our research findings, we proposed policies to support businesses, planning, and management units in forecasting and investing in the development of tourism in each specific locality more effectively.

***Keywords*: An Giang, Back-propagation neural network, Tourism demand forecast.**

1. Introduction

In the tourism industry, forecasting tourism demand plays a vital role as it enables tourism managers and business owners to plan and execute appropriate business policies based on actual conditions. This is a complex task due to its multi-sectoral and interdependent nature, which is heavily reliant on related fields' mechanisms, policies, and operating efficiencies. The accuracy of tourism demand forecasting is often impacted by various factors such as fluctuations in the global economy, oil prices, exchange rates, electricity, water, telecommunication costs, infrastructure, hotel room rates, environmental risks, natural disasters, epidemics, etc. Several forecast indicators, including market demand, customer trends, spending capacity, and length of stay, are considered dependent variables that change over time and space. These variables are dependent on many related factors and have a consistent relationship with each other, serving as the basis for forecasting. The variables needed for forecasting, such as the total number of guests, international tourists, domestic tourists, length of stay, day-trip visitors, and sold tour programs, are referred to as dependent variables. The factors used to serve forecasting, such as time, economic indicators, demographic indicators, transportation costs, tourism taxes, and market indicators, are known as independent variables. Several methods are available for performing forecasting, ranging from simple forms such as consulting expert opinions to various complex calculation methods. However, depending on the type of variable to be forecasted, the purpose, the required time frame, and the availability of resources to support forecasting activities such as funding, technology, skills, and expert teams, each agency or organization can choose to apply appropriate forecasting methods. This paper presents the construction of a tourism demand forecasting model that combines an artificial neural network model with an expert method. This approach uses both quantitative and qualitative methods, each with certain advan- tages and limitations. Quantitative methods provide specific and clear results from statistical data, while expert methods rely on experience. Combining these two approaches can produce a more accurate forecasting model that takes into account both objective data and subjective expertise.

Precise prediction is of utmost importance in the tourism sector as it is heavily relied upon by tourism administration units and government departments that deal with inbound tourism population forecasts. To ensure the precision and reli- ability of forecasts, it is essential to implement a scientific and reasonable forecasting model. In [1], the authors utilized the artificial neural network theory to develop a Back-Propagation (BP) network forecast model, which has been compared with traditional prediction methods such as the Logistic Model,

© The Author(s) 2023. This is an open access article published under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Exponential Smoothing Model, and Autoregressive Model. Results show that the BP neural network forecast model outperforms these traditional prediction methods in terms of accuracy and effectiveness. Therefore, the implementation of this model can offer tourism administration units and govern- ment departments a significant reference to plan and execute suitable policies for inbound tourism management. In a related study [2], tourism time series modulation was utilized for forecasting purposes. The researchers utilized data on Tourism Revenue and Total Overnights collected from hotels in the North region of Portugal for their models. They experimented with various feed-forward Artificial Neural Networks (ANN) models with different input features and hidden nodes to fore- cast the Tourism time series. The empirical results revealed that Dedicated ANN models outperformed models with multiple outputs. The inclusion of the previous 12 values of the same time series was found to be crucial for generating high-quality forecasts. To predict Tourism Revenue, Foreign Overnights and GDP of contributing countries were deemed relevant variables. The model achieved a 4.7% error and a Pearson correlation of 0.98. The Total Overnights forecast had a 6.0% error and a Pearson correlation of 0.98. Domestic Overnights were found to be more predictable than Foreign Overnights. There has been a growing interest in using machine learning methods to improve the precision of tourism prediction in recent years. In line with this trend, the study [3] proposed a novel paired neural network model that combines insights from tourism academia and industry to achieve accurate predictions. The model decomposes tourist arrival data into long-term trends and short-term seasonal components using two low-pass filters, which are then modeled by a pair of autoregressive neural network models in a parallel structure. The model is evaluated using data on tourist arrivals to the United States from twelve source markets, and empirical results show that the proposed paired neural network model outperforms the selected bench- mark models in all error measures and different horizons. Similarly, the authors in [4] built an accurate forecasting model for Indonesian tourism demand by employing Gross Domestic Product (GDP), Consumer Price Index (CPI), and exchange rates from five major visitor countries of Indonesia as independent variables to predict the number of tourist arrivals. Non-linear relationships and high fluctuations in data are among the major concerns in forecasting due to the seasonal and fragile nature of tourism. The study utilizes artificial neural network backpropagation, which adapts to changes in input data and presents a convenient alternative to econometric and time-series forecasting models. The study produces a monthly forecasting model for tourist arrivals in Indonesia and reaches an optimal configuration with a single hidden layer and 31 hidden neurons. In another work, the researcher of [5] suggested a new method to predict tourism arrivals by using a compound web-search index generated from relevant search terms as an additional input for autoregressive prediction. Comparing the performance of two methods, autoregressive integrated moving average (ARIMA) models and artificial neural network (ANN), the study found that incorporating Google Trends data significantly improves tourist arrival pre- diction compared to using past arrivals alone. The study also shows that ANN outperforms ARIMA models in predicting tourist arrivals, highlighting the potential of machine learning techniques for tourism forecasting research. Another important aspect that has been highlighted in [6] is the integration of information and internet technology in the tourism industry. The study emphasizes that integrating cultural and tourism forms creates an industrial linkage, and e-commerce tech- nology is an important path for rapid development. This research is specifically focused on the design of a modern tourism intelligent system that is based on computer network technology. The study proposes the use of artificial intelligence to guide tourism, underscoring the significant potential of technology in enhancing the tourism industry. In its analysis of the tourism industry in Vietnam, the authors in [7] sheds light on the growth of the industry, which has emerged as one of the top tourist destinations in Southeast Asia, creating numerous employment and income opportunities. However, the COVID-19 pandemic has dealt a severe blow, causing a 78.7% year-on- year decrease in international tourist arrivals in 2020. Despite this downturn, the study is optimistic, predicting that the number of international tourists will recover to pre-pandemic levels in the next few years. To make such predictions, the study employs artificial neural network (ANN) methodology to forecast the monthly number of international tourists to the country, even during lockdown periods. By demonstrating the accuracy of the best ANN models, the study shows that these models could be a valuable tool for policymakers and firm managers in making better investment and strategic decisions. Meanwhile, in [8], machine learning models with multisource Internet data predictors, including temporal factors, posts and comments, search queries index, and previous tourist arrivals records, were utilized to predict international tourist arrivals in Indonesia during the COVID-19 pandemic. Model accuracy is assessed using RMSE, MAE, and MAPE, and the results indicate that multisource data predictors outperform single- source data or other predictors. Furthermore, utilizing more training sets enhances the model’s accuracy, and RF models exhibit superior performance to ANN and SVR models. This pioneering research provides valuable insights into the dynamics of Indonesia’s tourism demand during the pandemic. The authors in [4] developed a precise forecasting model for the demand for Indonesian tourism. To predict the number of tourist arrivals in Indonesia, they utilize Gross Domestic Product (GDP), Consumer Price Index (CPI), and exchange rates from five key visitor countries. One of the primary concerns when forecasting is the nonlinear relationships and high volatility of the data. Tourism is inherently seasonal and fragile, with frequent periods of boom and decline that can endanger industry survival. To address this challenge, we employ artificial neural network backpropagation, which can adapt to changes in input data. This method is a favorable alternative to econometric and time-series forecasting models due to its adaptability. They construct a monthly tourism arrivals forecasting model for Indonesia and obtain an optimal configuration with a single hidden layer and 31 hidden neurons. The authors in [9] investigated the optimal number of lagged inputs (NLIs) in forecasting international tourism demand. The performances of eight machine learning models are evaluated using different NLIs with data on international tourist arrivals in ten European countries. The findings show that as NLIs increase, the error of most machine learning models decreases rapidly at first and then becomes stable (or fluctuates around a certain value) when NLIs reach a cutoff point related to 12 and its multiples. This trend remains consistent regardless of the size of the test set. Moreover, it is recommended to select one cycle of data as the NLIs for nonlinear and ensemble models, while multiple cycles are better suited for linear models. Lastly, the study demonstrates that different categories of models produce significantly different prediction results when the optimal NLIs are utilized. The authors in [10] highlighted the importance of forecasting Chinese cruise tourism demand as a crucial aspect of investment decision-making and planning. To improve the accuracy of forecasting, the authors propose a least squares support vector regression model with a gravitational search algorithm (LSSVR-GSA) that utilizes big data, specifically search query data (SQD) from Baidu and economic indexes. The proposed model optimizes hyper-parameters of the LSSVR model using GSA. By comparing the performance of various models, the study finds that LSSVR-GSA, utilizing selected mobile keywords and economic indexes, achieves the highest forecasting accuracy. These findings suggest that the proposed methodology is effective and that big data can be useful predictors in forecasting Chinese cruise tourism demand. The authors in [7] provided an overview of Vietnam’s tourism industry, with a particular emphasis on international tourists. Following that, the study evaluates the predictive power of artificial neural network (ANN) methodology using data on international tourists’ visits to Vietnam from 2008 to 2020. Several ANN models are tested to forecast the monthly number of international tourists to Vietnam, even during periods of lockdown due to the COVID-19 pandemic. The findings indicate that the most accurate ANN models, selected with appropriate architectures and data from the preceding 12 months, can predict the next month’s tourist numbers with a mean absolute percentage error (MAPE) ranging from 7.9% to 9.2%. Given the method’s forecasting accuracy, it can provide a valuable tool for Vietnam’s policymakers and business managers to enhance their investment and strategic decision-making processes. The authors in [11] proposed a machine learning-based tourism demand forecasting model us- ing a stacked autoencoder and long short-term memory hybrid model. The model utilizes a hierarchical greedy pretraining method to improve the performance of deep learning models. The study employs monthly search engine strength data and relevant influencing factors from a city to evaluate the model’s performance, and the results show that the proposed model outperforms the LSTM model. The study provides evidence of the effectiveness of the proposed LSTM-based unsupervised pretraining method for tourism demand forecasting. Currently, there are not many studies on forecasting tourism demand in An Giang using artificial neural networks and expert methods.

Therefore, forecasting tourism demand is crucial for sustainable tourism industry development, but there is a significant gap in the literature regarding An Giang province. This study aims to fill this gap by providing a comprehensive forecasting model that incorporates both objective data and subjective expertise, helping tourism managers and government departments formulate appropriate policies. The success of the tourism industry in An Giang province depends on accurately forecasting demand and designing appropriate policies, underscoring the significance of this study. The main contributions of the paper are as follows:

- Develop a tourism demand forecasting model by utilizing an artificial neural network model in conjunction with an expert method.
- Create an interactive Geographic Information System (GIS) map of An Giang province’s tourism sector that supports tourists, tourism agencies, and government offices and is compatible with multiple mobile devices for convenient access to information.
- Construct a comprehensive GIS database containing in-depth data on natural and cultural resources, tourist routes, products, services, and infrastructure.

2. Materials and Methods

2.1. *A. Research Model of Tourist Satisfaction*

The Service Quality (SERVQUAL) model, initially pro- posed by [12], is widely used to evaluate the overall quality of tourism services. This model comprises 22 indicators categorized under five criteria, namely, Reliability, Responsiveness, Assurance, Empathy, and Tangibility. Moving on, the authors in [13] proposed a sustainable development index for tourist destinations, consisting of 13 evaluation criteria, with the fourth criterion reflecting the level of satisfaction with tourism activities, measured using a questionnaire tool that includes six indicators. This sustainable development index was later applied and supplemented by [14] to evaluate tourism activities on Phu Quoc Island. The survey results indicated that the roads were easy to navigate (27.7%), tourists felt that the money spent was worth it (85.0%), thought that tourism facilities handled customer complaints well (35.7%), and would return (91.9%).

To investigate the factors that affect tourist satisfaction with the quality of tourism services in Kien Giang, the authors in [15] conducted a survey of 295 tourists, using a stratified random sampling method by location. They constructed a theoretical model based on five groups of influencing factors, including tourism scenery, technical infrastructure, transportation means, tour guides, and accommodation facilities. Exploratory factor analysis revealed that tourist satisfaction was related to all five components, with the attitude of tour guides being the strongest influencing factor, and the convenience of accommodation facilities being the lowest factor.

In another study, the authors in [16] aimed to assess the contentment of local travelers regarding the quality of service at the historical village of Duong Lam in Vietnam. To accomplish this, the researchers formulated a service model derived from the SERVQUAL model. After conducting an exploratory factor analysis, the study found seven categories of factors that influence satisfaction, including tourism service capacity, pricing of goods and services, cultural aspects, infrastructure, traditional crafts, traditional festivals, and cuisine. The results showed that the service quality at Duong Lam had initially met the needs of tourists, as roughly 80% of them reported being satisfied.

Thus far, there have been numerous studies concerning tourist satisfaction. However, investigating tourist satisfaction at specific tourist destinations, particularly based on the criteria and evaluation framework proposed by [13], remains an issue requiring further clarification. Most recently, the authors in [17] conducted a study on the model of tourist satisfaction at tourist destinations in Ho Chi Minh City, including eight component criteria with 15 evaluation indicators. After conducting a comprehensive review of tourism forecasting models, this paper proposes a tourism forecasting model with variables as depicted in Fig. 1.

Fig. 1: The tourism forecasting model for indicators in An Giang province.

The variables in this model can be categorized into two groups, dependent variables and independent variables. The dependent variables in this context refer to the indicators that need to be forecasted concerning tourism demand. In this model, we will forecast the following six indicators: the number of annual tourist arrivals, international tourist arrivals, domestic tourist arrivals, tourism expenditure, hotel occupancy rate, and tourism revenue. On the other hand, independent variables are the factors utilized to facilitate the forecasting process. In this model, eight component criteria consisting of 17 evaluation indicators will be utilized. These indicators include (1) accessibility to tourist destinations (comprehensive and easy-to-find information, convenient access to tourist destinations), (2) safety (secure transportation, public security, food hygiene, environmental hygiene), (3) infrastructure and technical facilities (completeness and consistency), (4) tourism resources (unique factors, attractiveness), (5) cost and service prices (reasonable service prices, price increase, average income per capita of the local area), (6) tourism human resources (friendliness, understanding of the responsible field), (7) service readiness (continuous availability, timely and fast service), and (8) hospitality (friendliness of the community).

Accessibility: It encompasses physical and informational access to tourist destinations. Comprehensive and easily searchable information about tourist attractions is essential to attract potential tourists. Additionally, evaluating physical accessibility, such as assessing obstacles and port conditions, and identifying touring routes connecting multiple destinations, is crucial to ensure optimal accessibility.

Security: Political security and social safety are priorities for tourist destinations to attract and satisfy tourists. Tourists prefer visiting places with no frequent attacks, violence, terrorism, wars, or territorial disputes. Safety concerns encompass life, property, and dignity, and should address natural disasters, epidemics, traffic accidents, and social evils like theft, robbery, murder, fraud, prostitution, drugs, food safety, and healthcare facilities. The natural environment should also be safe, with no dangerous risks, wild animals, or environmental pollution.

Infrastructure and Technical Facilities: Infrastructure encompasses fundamental necessities that fulfill the minimum requirements for tourists, including a secure and efficient transportation system, sufficient healthcare facilities, and a reliable supply of energy and fuel sources. A contemporary and productive communication network is also imperative to provide visitors with accurate and timely information about their destinations. Technical facilities in tourism encompass lodgings, dining alternatives, entertainment venues, shopping centers, and other tourism services. These factors are vital in determining the satisfaction level of tourists. Therefore, it is imperative to have all-inclusive, synchronized, diverse, modern and updated technical facilities that keep up with global tourism development trends.

Tourism Resources: Tourism resources are the key factors that determine the destinations tourists will visit and what kinds of tourism activities they will engage in. It is challenging to attract visitors and conduct tourism activities without tourism resources. Tourism resources comprise both natural and cultural resources. Cultural resources encompass historical and cultural sites, festivals, handicrafts and traditional villages, ethnic cultures (culinary culture, clothing, language, housing, life cycle rituals, etc.), cultural and sporting events, museums, exhibitions, contemporary architectural works, and the local community's cultural and social life. Natural resources consist of natural phenomena, landscapes, and ecosystems. Tourism resources play a critical role in determining the content of tourist visits. Additionally, tourism activities at the destinations must be based on the potential tourism resources available.

Cost and Service: Tourists prioritize cost when making travel decisions. They expect services that meet or exceed their standards, and the price is crucial to their satisfaction. Tourists compare prices between destinations and expect services and products worth their money. Unsatisfactory experiences with services or products will result in tourist dissatisfaction.

Tourism Human Resources: Tourism staff is essential to tourist satisfaction. They should possess professionalism, passion for the job, guest-friendliness, honesty, and integrity. Based on the service attitude of tourism personnel, tourists decide whether to return to the tourist destination or the tourism service facility.

Service Readiness: This pertains to the uninterrupted, expedited, and timely provision of services that cater to the tourists' needs during their journey. The lack of assurance of basic tourism services upon their arrival can cause dissatis- faction and a lack of interest in prolonging their stay in the tourist. destination.

Hospitality: The local residents' hospitality holds immense sway over the tourists' affective experiences. The tourists feel at ease, secure, and embraced in places where the local residents exhibit amiability and warmth. The locals' hospitality facilitates tourists' accessibility to the indigenous cultural elements and helps them foster sanguine impressions of the destination, despite the incomplete tourism services. The hospitality exuded by the locals constitutes a pivotal determinant of tourists' attachment, retention, and revisitation to the tourism destination.

2.2. *Artificial Neural Network for tourism demand forecasting*

An artificial neural network [18] is a computer program that simulates the information processing of a biological neural network. It is made from a large number of elements (called artificial neurons or processing elements) connected through links (each link has a weight value called a link weight) that work as a unity to solve a specific issue. In practice, many ANNs are good tools for modeling nonlinear statistical data. They can be used to model complex relationships between input and output data. An ANN is configured for a specific application (pattern recognition, data classification, etc.) through a learning process from a set of training samples. Learning is, in substance, the process of calibrating the link weights between neurons. There are three common machine learning methods: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is the most popular method, which is typically the back-propagation algorithm.

a) Back-propagation Algorithm

The back-propagation algorithm (BP) [19] is often applied to multi-layer feed-forward neural networks consisting of processing elements with activated and continuous functions. Given the training set consisting of p samples, the input-output pairs are as follows:

$$\{x^{(k)}, d^{(k)}\}, k = 1, 2, \ldots, p \qquad (1)$$

The back-propagation algorithm provides procedures for altering the weight values in back-propagation neural networks to properly classify given input samples. The basis of this algorithm is the reduced gradient method [20]. For a given input-output pair, the back-propagation algorithm performs two phases of the data stream. Firstly, the input pattern $x^{(k)}$is propagated from *input* layer to *output* layer and it generates the actual output signal as the forward output. Next, the error, the difference between $d^{(k)}$ and $y^{(k)}$, will be backpropagated from the *output* layer to the *input* layers in order to help them to adjust their weight values. For example, consider a three-layer network as shown in Fig. 2 with *n* nodes in the input layer, *n* nodes in the hidden layer, and *m* nodes in the output layer. Solid lines indicate the forward spread of the signals; the dashed lines indicate the backward propagation of the errors.

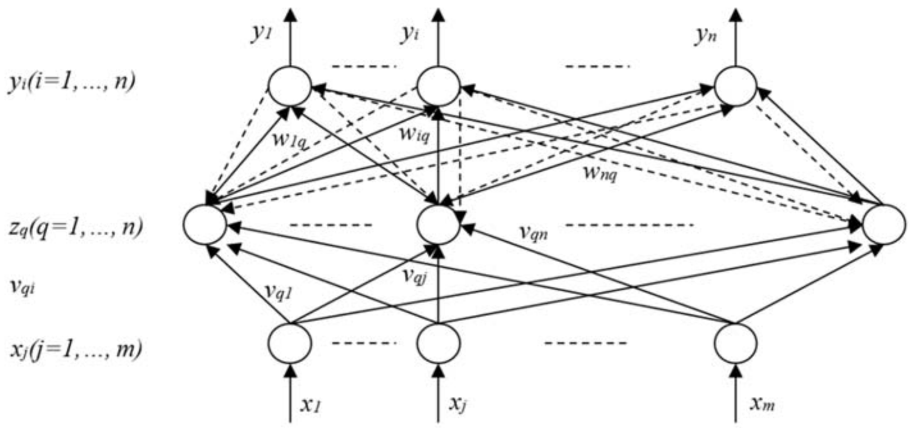

Fig. 2: A back-propagation neural network with three layers.

Consider a pair of training data points *(x, d)*. With the given input pattern *x*, the node *q* in the hidden layer receives the network's input:

$$net_q = \sum_{j=1}^{m} v_{qj} x_j \qquad (2)$$

and generates the *output*

$$Z_q = a(net_q) = a\left(\sum_{j=1}^{m} v_{qj} x_j\right) \qquad (3)$$

The network's *input* for the i^{th} node of the *output* layer is as follows:

$$net_i = \sum_{q=1}^{l} w_{iq} z_q = \sum_{q=1}^{l} w_{iq}\, a\left(\sum_{j=1}^{l} w_{qj} x_j\right) \qquad (4)$$

And generates the *output*

$$y_i = a(net_i) = a\left(\sum_{q=1}^{l} w_{iq} z_q\right) = a\left(\sum_{q=1}^{l} w_{iq} a\left(\sum_{j=1}^{m} v_{qj} x_j\right)\right) \qquad (5)$$

The above equations determine the forward spread of *input* signals through the layers.

Cost function is defined by equation

$$E(w) = \frac{1}{2}\sum_{i=1}^{n}(d_i - y_i)^2 = \frac{1}{2}\sum_{i=1}^{n}[d_i - a(net_i)]^2 = \frac{1}{2}\sum_{i=1}^{n}\left[d_i - a\left(\sum w_{iq} z_q\right)\right]^2 \qquad (6)$$

By the reduced gradient method, the weights are updated as follows:

$$\Delta w_{iq} = -\eta \frac{\partial E}{\partial W_{iq}} \qquad (7)$$

We have

$$\Delta w_{iq} = -\eta \left[\frac{dE}{\partial y_i}\right]\left[\frac{\partial y_i}{\partial net_i}\right]\left[\frac{\partial net_i}{\partial w_{iq}}\right] = \eta[d_i - y_i][a'(net)][z_q] = \eta \delta_{oi} z_q \qquad (8)$$

Here, δ_{oi} is the error signal, and double index *oi* determines the i^{th} node in the *output* layer. The error signal is defined as:

$$\partial_{oi} = -\frac{\partial E}{\partial net_i} = -\left[\frac{\partial E}{\partial y_i}\right]\left[\frac{\partial net_q}{\partial net_i}\right] = [d_i - y_i][a'(net_i)] \qquad (9)$$

The net_i is the i^{th} node's *input* in the *output* layer

And $a'(net_i) = \delta a(net_i)/\delta net_i$

The weights on *input* connections and hidden classes are updated as follows:

$$\Delta v_{qj} = -\eta \left[\frac{\partial E}{\partial v_{qj}}\right] = -\eta \left[\frac{\partial E}{\partial net_q}\right]\left[\frac{\partial net_q}{\partial v_{qj}}\right] = -\eta \left[\frac{\partial E}{\partial z_q}\right]\left[\frac{\partial z_q}{\partial net_q}\right]\left[\frac{\partial}{\partial v_{qj}}\right] \tag{10}$$

Thus,

$$\Delta v_{qj} = \eta \sum\left[(d_i - y_i)a'(net_i)w_{iq}\right] a'(net_q)x_j \tag{11}$$

Deducing that

$$\Delta v_{qj} = \eta \sum\left[\delta_{oi} w_{iq}\right] a'\left(net_q\right)x_j = \eta \delta_{lq} x_j \tag{12}$$

Here, δ_{lq} is the error signal of the q^{th} node in the hidden layer and is defined as:

$$\delta_{lq} \equiv -\frac{\partial E}{\partial net_q} = -\left[\frac{\partial E}{\partial z_q}\right]\left[\frac{\partial z_q}{\partial net_q}\right] = a'(net_q)\sum_{i=1}^{n} \delta_{oi} w_{iq} \tag{13}$$

The net_q is the q^{th} node's hidden *input*.

In the case of general networks with arbitrary classes, back-propagation has the form as:

$$\Delta w_{ij} = \eta \delta_i x_j = \eta \delta_{output-i} x_{input-j} \tag{14}$$

Here, the *'output-i'* and the *'input-j'* determine two connections from the j^{th} node to the i^{th} node; x_j is the appropriate input, an active point from a hidden node or from external input, δ_i is the learning signal.

When the activated function is a *sigmoid* function, we have:

$$\delta_{oi} = \frac{1}{2}\left(l - y_i^2\right)[d_i - y_i] \tag{15}$$

And

$$\delta_{hq} = \frac{1}{2}\left(l - z_q^2\right)\sum \delta_{oi} w_{iq} \tag{16}$$

b) Back-Propagation Neural Network with Delay and Shift Window

Typically, real-time data are used as both the output data and the input sample data for prediction problems. Real-time inference refers to data from both the present and the past. The "before-after" relationship is the centerpiece of learning from real-time data. The most challenging aspect of learning these patterns is actually how to identify, define, and maintain those relationships. To address this issue, a back-propagation neural network is transformed into a time-delay neural network and implements a shift window concept, as illustrated in Fig. 3.

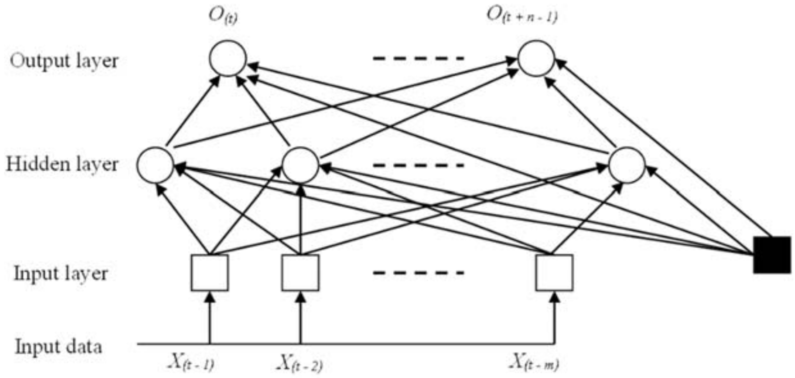

Fig. 3. Back-propagation neural network on real-time data with a shift window.

If the input data has x bits and is delayed with m shift windows, there will be $m*x$ input units to encode the input sample. When new data is inserted, it will be placed at the input node at a certain end of the network. The older data will be shifted by one unit on the network's entry nodes, the same as the shift register. Based on the data, predict how many times the lookback windows will shift by a corresponding amount. Correspondingly, the larger the prediction need, the larger the look ahead windows. An example of a network's "input-output" relation that performs prediction with real-time data can be formulated as follows:

$$Y_{(t)} = f_{nn}(X_{(t-1)}, X_{(t-2)}, \dots, X_{(t-m)}) \tag{17}$$

where m is the input shift window and also the number of the network's nodes, the output shift window is equal to 1 and also the network's output nodes. The relationship shows that the network will predict Y at the next time t, with the present time being $t\text{-}1$, knowing that its values and its dependent values are in the past and present time.

3. Results And Discussion

3.1. *Experimental setup*

All the experiments were carried out on a computer featuring an Intel Core i5 CPU, 8GB of RAM, and the Microsoft Windows 10 operating system. The model for predicting tourism, which employed back-propagation neural networks and genetic algorithms, was implemented in the programming language C++.

3.2. *Data Collection*

Collecting data is a challenging task in tourism research that demands substantial effort and costs. Hence, this study engaged volunteers to collect data across diverse locations, utilizing various methods. Two distinct approaches were employed: (i) referring to readily available documents at functional units; and (ii) conducting field data collection with the assistance of 10 volunteers (consisting of IT and tourism students). Our approach was tested on the tourism resources dataset of An Giang province, which comprised various indicators such as economic, social, demographic, commercial, and tourism factors. These indicators included the price index, average income per capita, service costs, website traffic, annual tourist arrivals, tourist expenditure, room capacity, survey results on accessibility to tourist destinations, tourism resources, infrastructure, human resources, hospitality, security, and community friendliness. The dataset was provided by legally responsible functional agencies, such as the Department of Culture, Sports and Tourism, Statistical Office of An Giang province, An Giang Meteorological and Hydrological Forecasting Center, and other reliable reference materials. To ensure the completeness of the training data and increase the forecasting model's reliability, we used data collected over a 10-year period. The primary data on tourist evaluations has a total of 23 surveys, multiplied by 5, resulting in 115 samples.

In addition, we undertook visits to tourist destinations to obtain real-world GPS coordinates, ensuring a high level of accuracy. This enabled us to verify document information and supplement and enhance existing information sources. During the implementation phase, our primary focus was on collecting spatial data (geographic coordinates) of travel resources in An Giang by utilizing a Garmin GPS device (GPSmap 62sc) [21]. We also conducted a survey to gather information on the status and resources of tourism in the province, which included eight factors such as natural and cultural tourism resources, tourist attractions, tourism products and services, material and technical facilities for tourism, and other relevant information. The methods employed for conducting the survey and gathering data include the methods of information synthesis and collection, as well as the field investigation method, which entails the use of questionnaires and in-depth interviews, and Google Maps for the implementation of a Geographic Information System [22]. The process of collecting geographic information on tourism destinations is depicted in Fig. 4. After three months of gathering information on tourism resources in An Giang's 11 districts, we obtained 2630 tourism objects with detailed descriptions, 10 survey forms for managing tourism information, and 2630 high-quality photos, which also include fly-cam video clips of famous destinations. Then all these data are processed and shown on the website and online map at the address http://agtravellive.com and http://map.agtravellive.com/diadiem/bando. To ensure the information is always up to date, we have delivered a mobile application to enable a variety of users to regularly update the information on travel objects. The main users are tourists, reporters, editors, editorial boards, tourism business companies, the Department of Culture, Sports and Tourism, content administrators, system administrators, and statistical management users. Fig. 5 shows the statistics of tourism objects that have been collected and shown on the tourism information system.

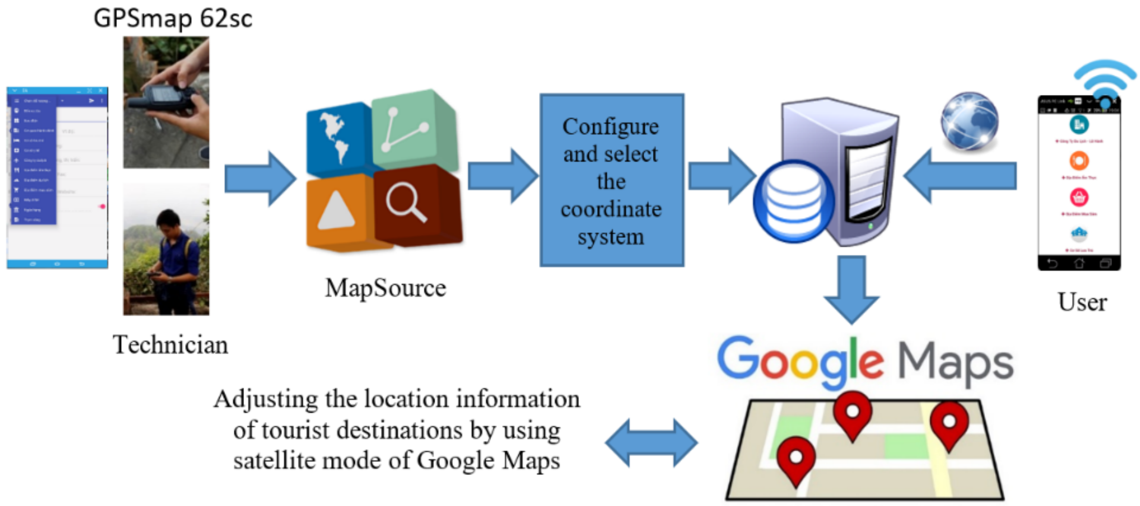

Fig. 4: The process of collecting geographic information of tourist destinations.

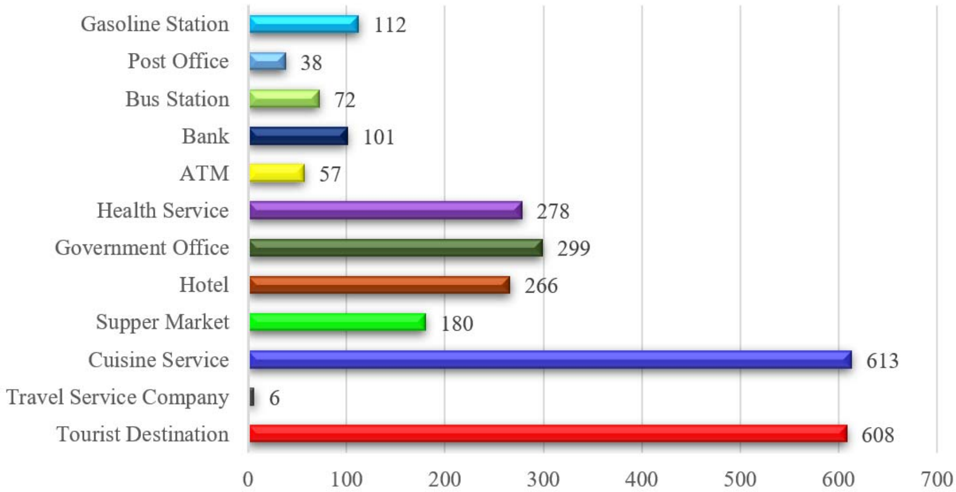

Fig. 5: The statistics of collected tourism objects in An Giang.

3.3. *Neural Network Training*

To construct the tourism demand forecasting system, six separate neural networks have been trained, with each one dedicated to forecasting one of the six tourism indicators: *tourist arrivals, international tourist arrivals, domestic tourist arrivals, tourism expenditure, room capacity usage*, and *tourism revenue*. All six networks undergo identical training procedures; therefore, we only present the training process of a typical neural network, which predicts the number of Tourist Arrivals to An Giang province in the upcoming years. The general model is presented in Fig. 6. The steps for training the network are carried out as follows:

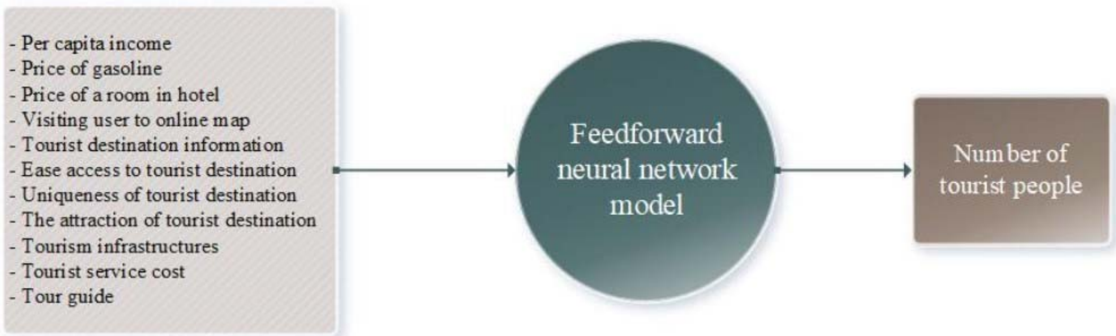

Fig. 6. The model of Neural Network for predicting the number of tourist people.

Step 1 - Data preprocessing: The dataset is preprocessed by normalization, which is essential before inputting the data samples into the neural network for training. The input and output values are normalized to ensure that they have values within the range of [0, 1] or [-1, 1].

In this forecasting model, the entire training and testing dataset was normalized to values within the range of [0, 1]. To normalize the data, the following formula was used:

$$X_n = X/D_{max} \tag{18}$$

where, X represents the value that needs to be normalized, X_n represents the normalized value, and D_{max} represents the maximum value of the dataset.

Step 2 - Neural Network Design: The neural network is designed with three layers, including an input layer with 23 nodes, a hidden layer with 24 nodes, and an output layer with 1 node.

The 23 input nodes represent different variables such as average income per capita (in million VND), USD exchange rate (in thousand VND), gasoline price (in thousand VND), average room rate (in thousand VND), website visits (number of times), number of tourists in the previous year (number of tourists), complete information of the tourist attraction (%), ease of access to the tourist attraction (%), uniqueness factor of the destination (%), the attractiveness of the destination (%), complete tourism value chain development (%), synchronized tourism value chain development (%), reasonable service prices (%), price adjustment status (%), friendliness of tourism staff (%), staff's understanding of their assigned field (%), continuous readiness to serve tourism (%), timely responsiveness to tourism (%), safety in transportation (\%), public security (%), food hygiene (%), environmental hygiene (\%), and community friendliness. Additionally, there is an output node that predicts the number of tourist arrivals in the following year. The network is trained using randomly generated weights and is configured as specified in Table 1.

Learning type: Back-propagation algorithm	Learning rate: 0.1
Lowest error: 0.01	Iterations number: 1,000
Alpha: 0.9	Sigmoid Output Layer: Yes
Momentum: 0.8	Weight: Load

Table 1: Configuring neural networks when training by back-propagation algorithm.

3.4. *Measurement of Forecasting Accuracy*

Predictions are always subject to error because the situation with tourism is complicated and dependent on many factors. Forecasting error will be a measure of how close the predicted value is to the actual value. In fact, this error is the difference between the actual value (d_t) and the predicted value (y_t) and is calculated as follows:

$$e_t = |d_t - y_t| \qquad (19)$$

A forecasting model is considered good if the forecasting error on testing data is quite small. In fact, if the model was built properly, the fluctuations of the prediction error would not go in any direction. The fluctuations in prediction errors are frequently caused by unforeseeable external phenomena. This means that the random oscillations of e_t in each period are purely random oscillations around the predicted value y_t, so the total prediction error will be zero. At present, the two most popular methods of calculating forecasting errors are absolute prediction error and relative prediction error. In fact, the calculation of forecasting errors is quite good in cases where the problem is complex and the forecasting model has a large error. Therefore, to measure the prediction accuracy of the rice disease prediction model, we use the relative error, namely the Root Mean Square Error (RMSE).

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}} \qquad (20)$$

where e_t is the prediction error in t stage, n is the number of forecasting observations.

3.5. *Resuts and Discussion*

After training the back-propagation neural network model with delay and sliding window for 1000 generations, we present the results in Fig. 7, Fig. 8 and Fig. 9.

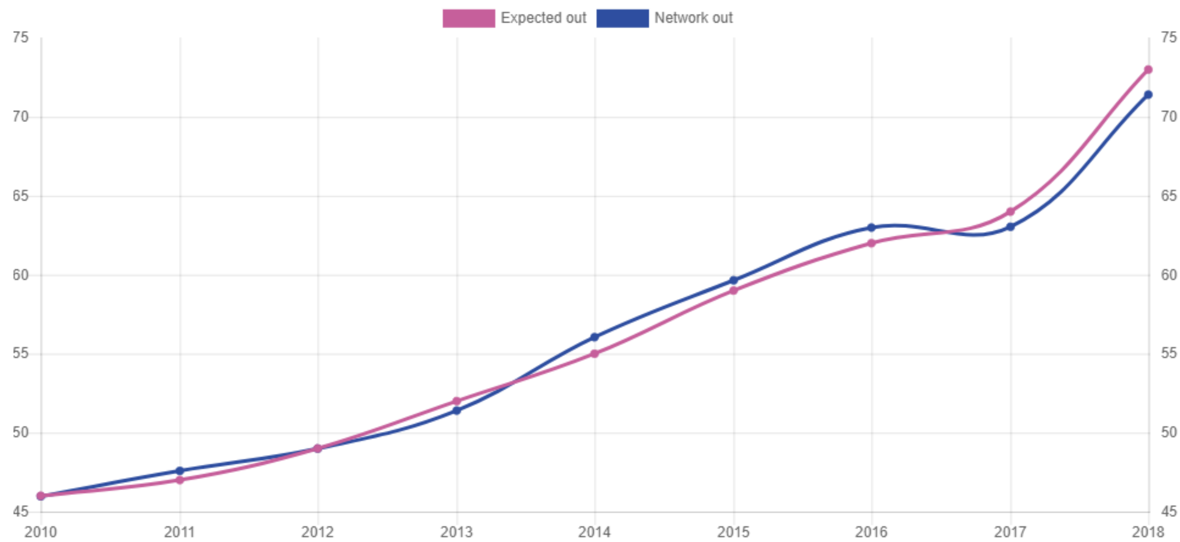

Fig. 7: Training results of the neural network for predicting the number of tourist arrivals.

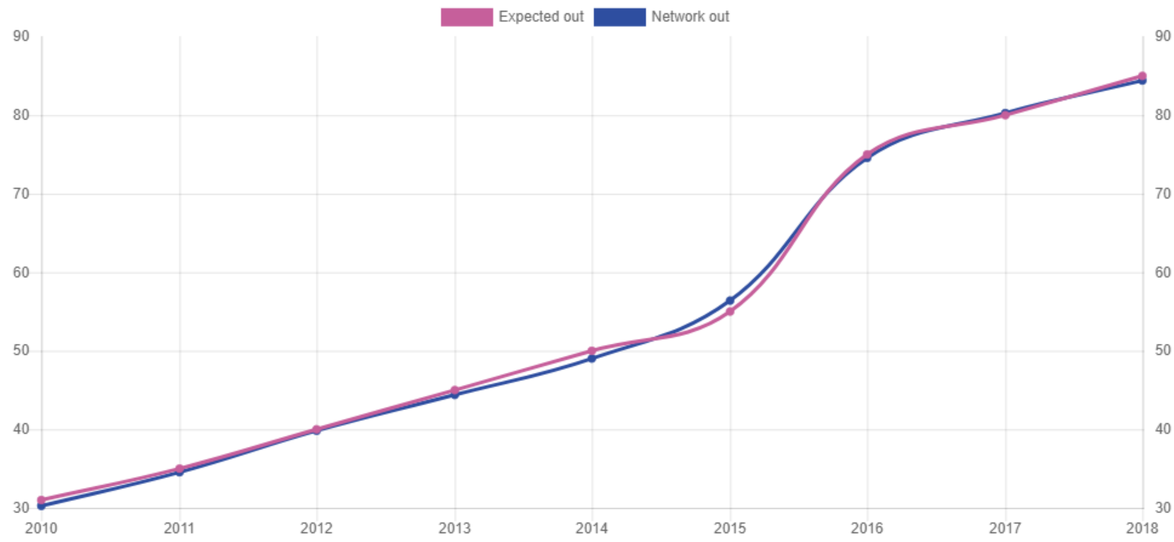

Fig. 8: Training results of the neural network for predicting the number of international tourist arrivals.

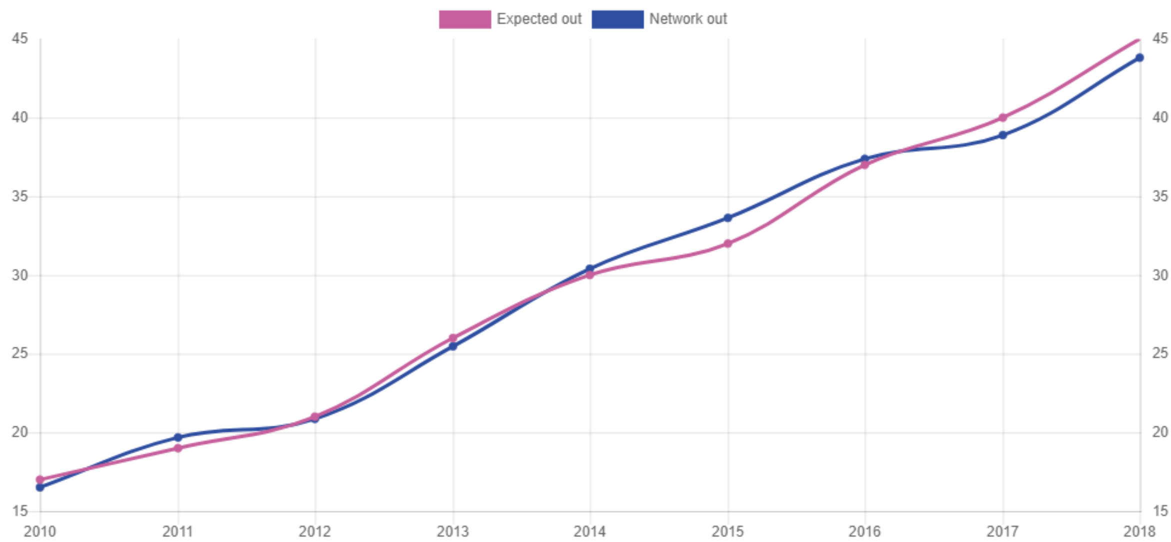

Fig. 9: Training results of a network for predicting tourism revenue.

In terms of the training dataset, it is evident from these graphs that the output curve of the network (depicted in light color) has closely approximated the actual output curve (depicted in dark color). This implies that the network has effectively modeled the entire dataset of the number of tourists visiting An Giang over the years.

Moving on to the test dataset, we observe that the network is capable of predicting previously unlearned data, as evidenced by the output curve of the neural network being in close agreement with the actual output curve, and the RMSE prediction error on the test dataset being 0.04 (Table 2). This demonstrates the network's ability to generalize well to unlearned data during the training process.

Year	Input	Expected output (d_t)	Network output (y_t)	Error (e_t)
2010	1:18.50, 2:19.50, 3:21.00, 4:20.00, 5:32.10, 6:70.90, 7:80.22, 8:54.39, 9:76.24, 10:72.43, 11:52.61, 12:54.90, 13:63.70, 14:53.63, 15:43.63, 16:59.98, 17:53.54, 18:75.65, 19:54.81, 20:39.57, 21:40.92, 22:66.92, 23:42.00	46.00	45.97	0.03
2011	1:19.50, 2:21.00, 3:22.00, 4:25.00, 5:32.50, 6:62.26, 7:71.07, 8:74.21, 9:54.73, 10:77.68, 11:54.30, 12:63.37, 13:76.24, 14:55.66, 15:59.38, 16:59.30, 17:55.40, 18:69.21, 19:54.47, 20:43.46, 21:35.84, 22:60.15, 23:46.00	47.00	47.58	0.58
2012	1:21.45, 2:22.50, 3:22.00, 4:26.00, 5:32.90, 6:50.00, 7:50.00, 8:50.00, 9:50.00, 10:50.00, 11:50.00, 12:50.00, 13:50.00, 14:50.00, 15:50.00, 16:50.00, 17:50.00, 18:50.00, 19:50.00, 20:50.00, 21:50.00, 22:50.00, 23:47.00	49.00	48.99	0.01
2013	1:23.08, 2:24.60, 3:22.00, 4:27.00, 5:33.30, 6:73.61, 7:55.66, 8:48.80, 9:36.60, 10:75.82, 11:55.66, 12:63.87, 13:81.66, 14:73.61, 15:52.95, 16:62.69, 17:54.90, 18:85.98, 19:63.62, 20:37.03, 21:44.82, 22:60.32, 23:49.00	52.00	51.40	0.6
2014	1:27.82, 2:24.60, 3:22.00, 4:28.00, 5:33.70, 6:79.37, 7:56.67, 8:68.96, 9:57.95, 10:73.95, 11:55.49, 12:77.42, 13:58.11, 14:70.39, 15:55.66, 16:61.33, 17:44.23, 18:74.63, 19:53.46, 20:45.66, 21:48.54, 22:57.27, 23:52.00	55.00	56.05	1.05
2015	1:32.60, 2:24.90, 3:22.00, 4:29.00, 5:34.10, 6:75.82, 7:56.67, 8:74.21, 9:55.40, 10:77.68, 11:52.61, 12:69.12, 13:54.73, 14:71.07, 15:54.81, 16:72.34, 17:54.56, 18:79.20, 19:56.84, 20:62.26, 21:52.78, 22:61.50, 23:55.00	59.00	59.65	0.65
2016	1:33.21, 2:24.90, 3:22.00, 4:30.00, 5:34.50, 6:78.36, 7:60.23, 8:63.70, 9:53.03, 10:78.02, 11:56.00, 12:66.08, 13:54.05, 14:77.68, 15:60.91, 16:76.92, 17:55.07, 18:75.82, 19:54.13, 20:60.23, 21:45.33, 22:57.61, 23:59.00	62.00	62.99	0.99
2017	1:36.00, 2:24.90, 3:22.00, 4:31.00, 5:34.90, 6:80.05, 7:60.91, 8:79.63, 9:57.27, 10:78.69, 11:59.05, 12:66.75, 13:74.21, 14:67.18, 15:56.67, 16:77.42, 17:54.90, 18:73.27, 19:58.03, 20:53.80, 21:57.86, 22:66.92, 23:62.00	64.00	63.04	0.01
2018	1:37.50, 2:24.90, 3:22.00, 4:32.00, 5:35.50, 6:80.22, 7:67.85, 8:78.27, 9:65.23, 10:82.93, 11:63.45, 12:56.76, 13:72.68, 14:74.46, 15:65.82, 16:63.37, 17:54.05, 18:70.23, 19:61.76, 20:52.78, 21:56.67, 22:61.16, 23:64.00	73.00	71.41	1.59
			RSME	**0.04**
2019	1:40.50, 2:24.90, 3:22.00, 4:32.00, 5:50.40, 6:82.30, 7:70.14, 8:80.01, 9:70.05, 10:82.93, 11:70.02, 12:56.76, 13:72.68, 14:80.01, 15:70.02, 16:66.02, 17:55.02, 18:72.05, 19:66.02, 20:53.01, 21:58.04, 22:63.02, 23:73.00		74.94	

Table 2: Training results of the neural network for tourist arrivals.

After undergoing training, the predictive model was employed to forecast tourism indicators for the year 2019 in An Giang province. These indicators encompassed the quantity of *domestic tourist arrivals, international tourist arrivals, tourism expenditure, room occupancy, and tourism revenue*. The findings of this study are presented in Fig. 10 to Fig. 19. The figures clearly demonstrate that the neural network successfully modeled the tourism data over a period of 10 years in An Giang province, and effectively predicted the forecasted values for the following years, particularly in 2019.

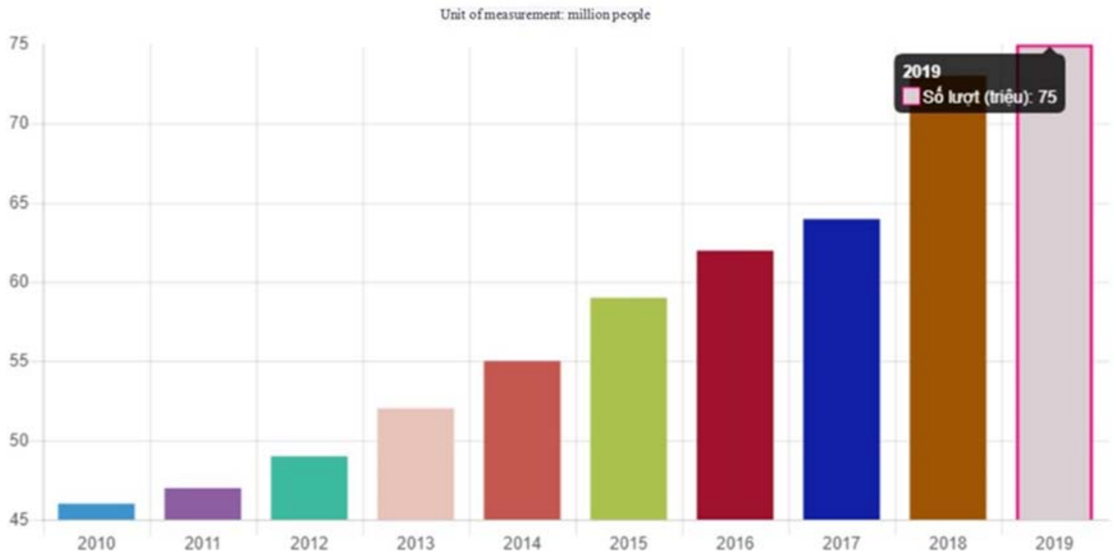

Fig. 10: Statistical and forecasting chart of number of tourist arrivals by year.

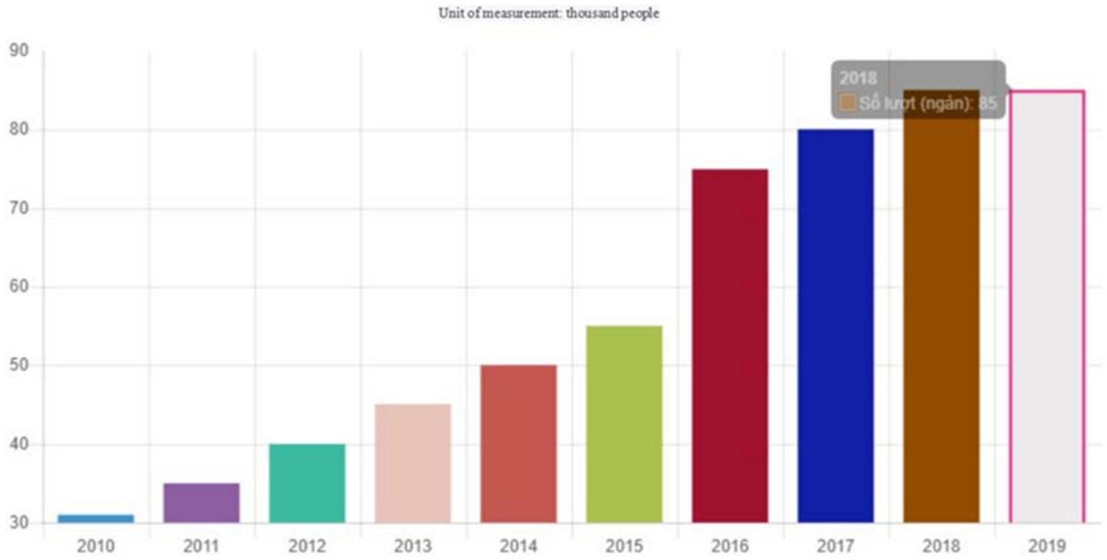

Fig. 11: Statistical and forecasting chart of a number of international tourist arrivals by year.

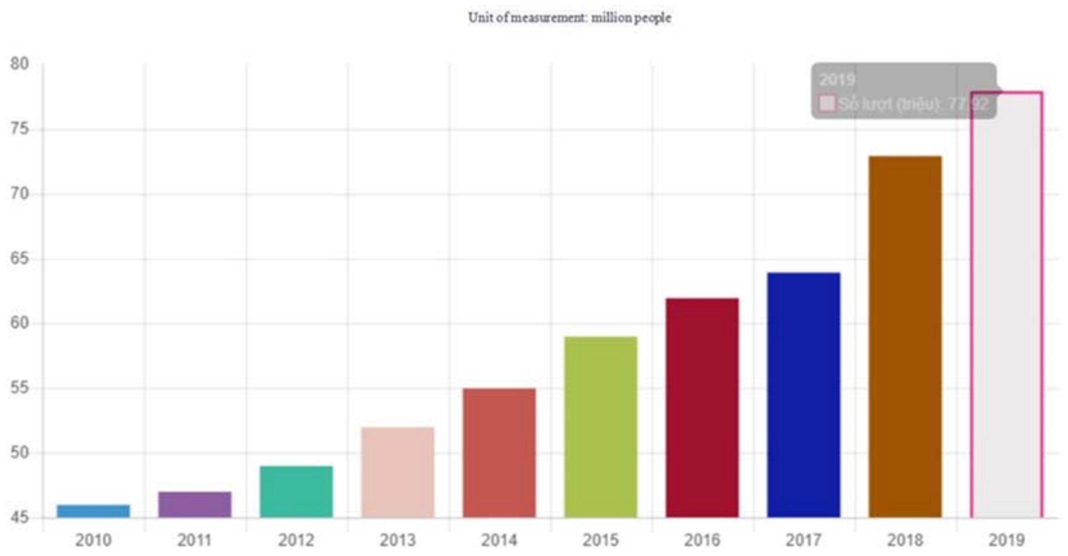

Fig. 12: Statistical and forecasting chart of a number of domestic tourist arrivals by year.

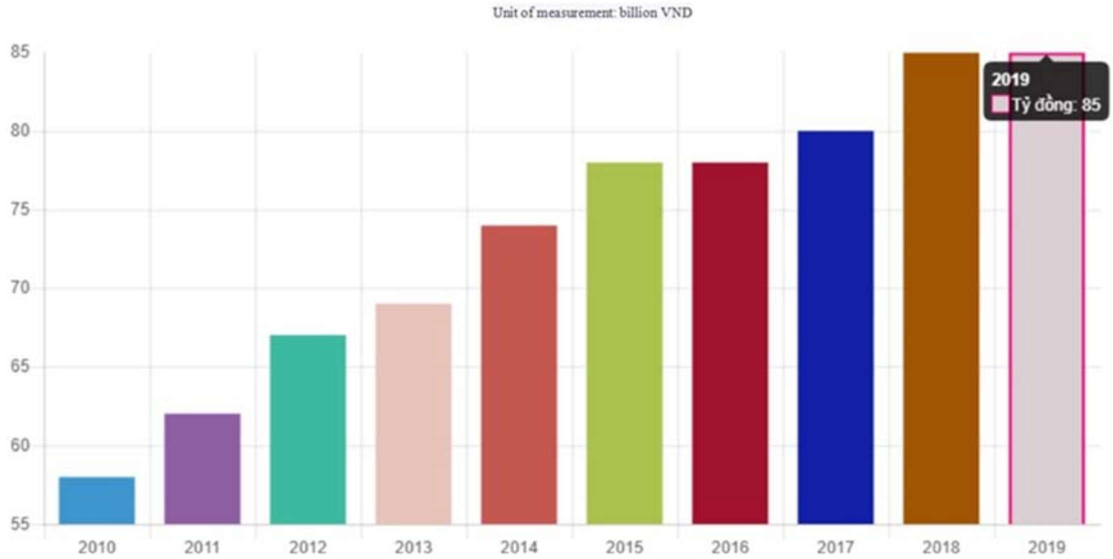

Fig. 13: Statistical and forecasting chart of tourism expenditure by year.

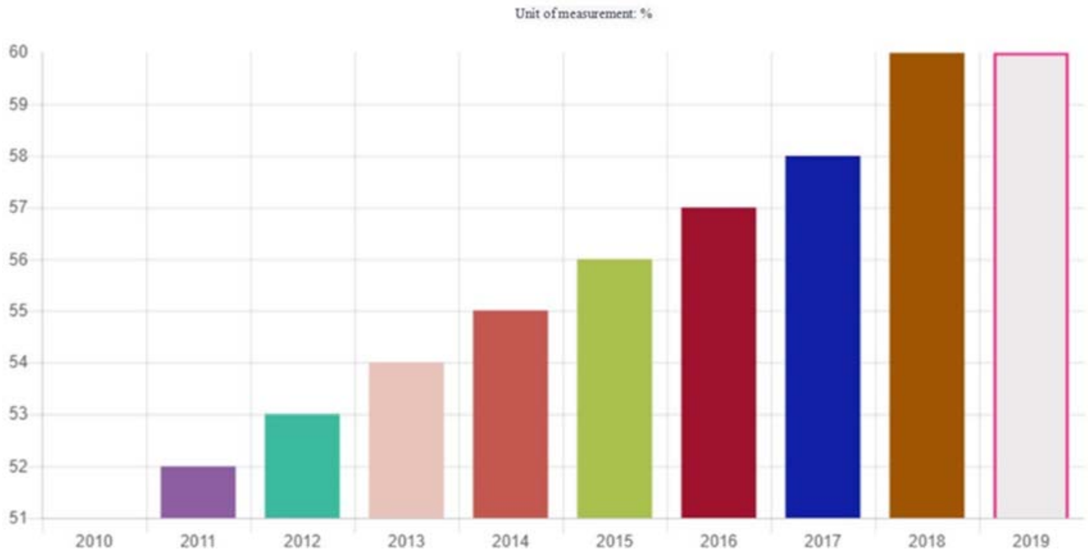

Fig. 14: Statistical and forecasting chart of room occupancy by year.

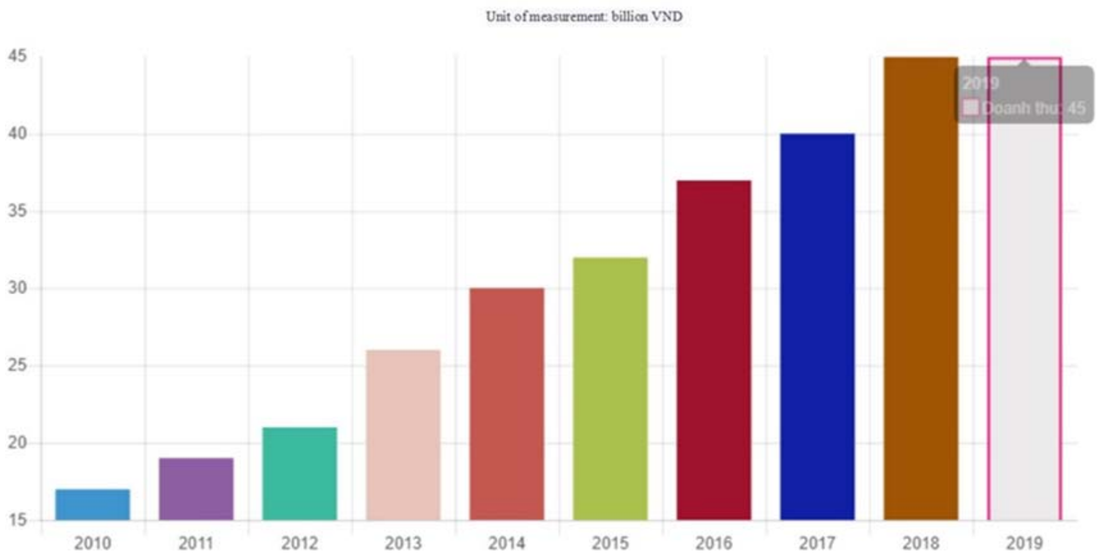

Fig. 15: Statistical and forecasting chart of tourism revenue by year.

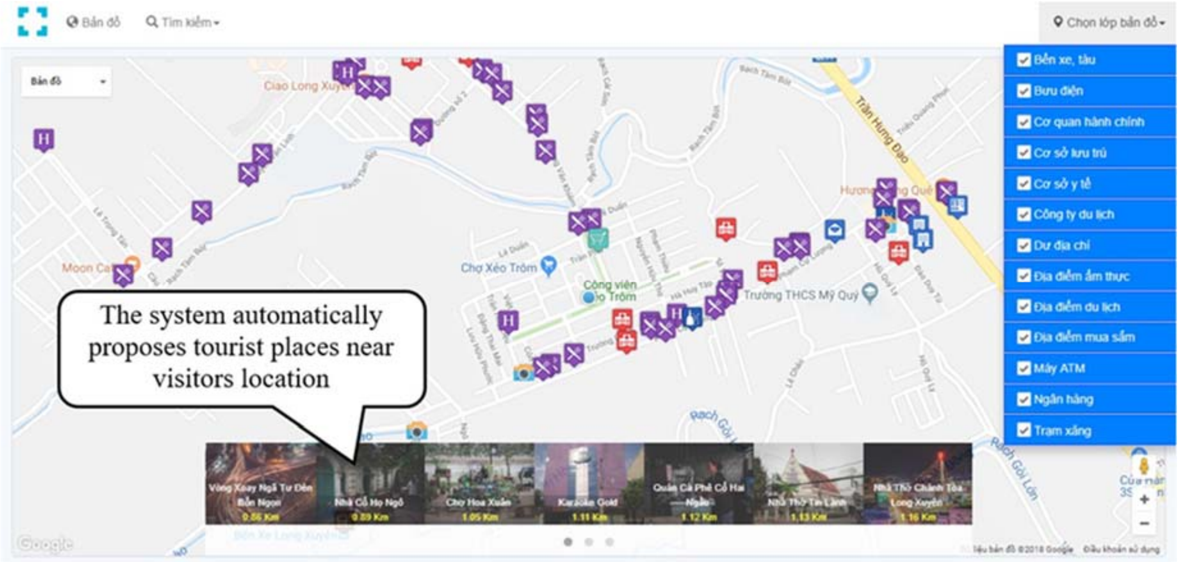

Fig. 16: The main screen of online smart tourism destination map.

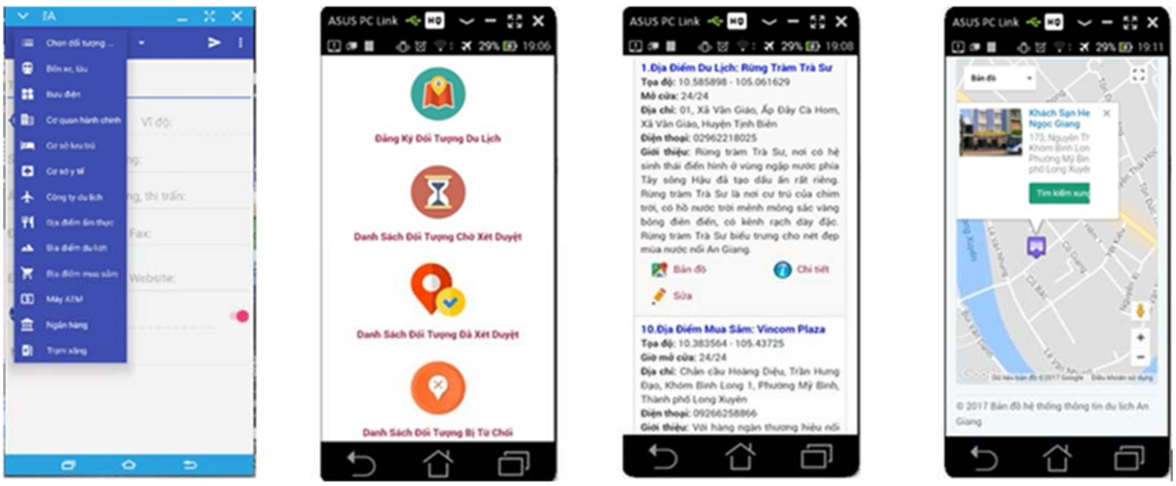

Fig. 17: Mobile device application for collecting information about tourist destinations.

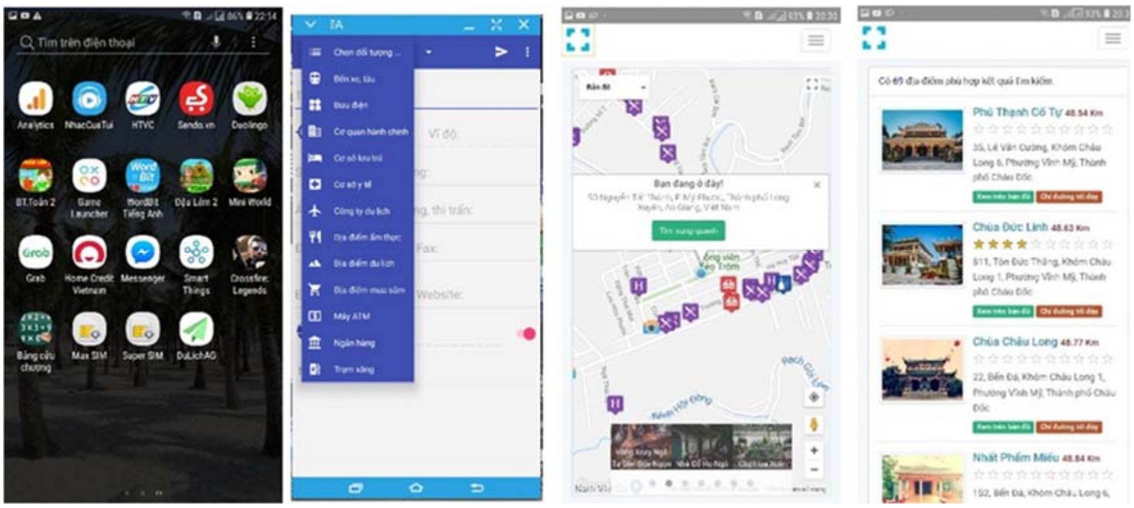

Fig. 18: The group of screens of user mobile application.

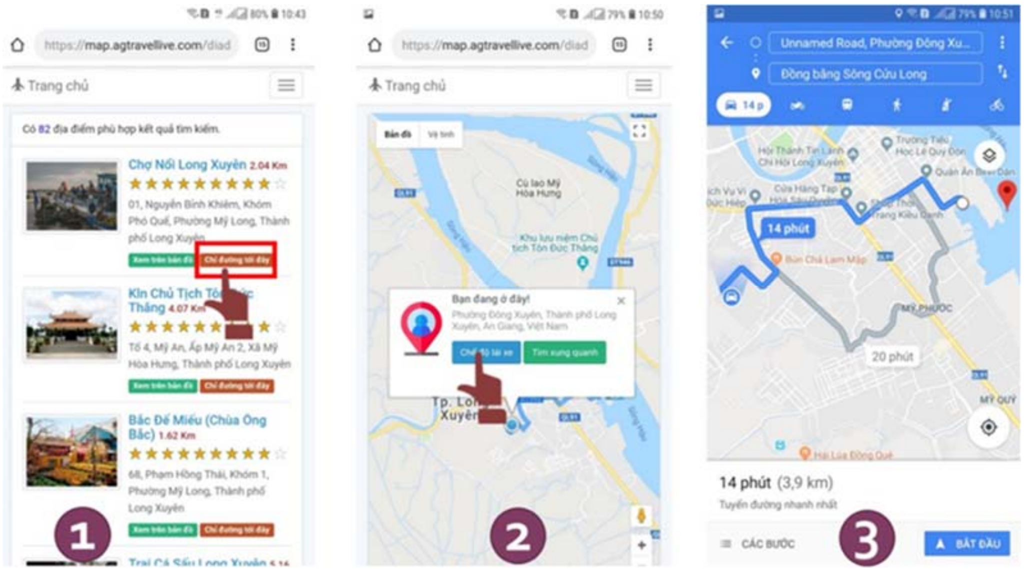

Fig. 19: The group of screens that shows how to search and find the direction to the destinations.

4. Conclusion And Future Research Work

The paper has proposed a novel tourism demand forecasting model utilizing artificial neural networks and expert methods. We investigated and analyzed the factors affecting tourism prediction based on expert knowledge in the tourism industry, as well as conducted a survey on visitor satisfaction levels during their travels to An Giang. All of this data was then modeled using techniques in the field of artificial intelligence. The resulting model was utilized to forecast tourism indicators for the upcoming years and has already produced promising results in meeting the complex and difficult-to-control demand for tourism forecasting. The model achieved a high level of accuracy with an RSME of 0.04. A new online smart tourism destination map of An Giang province has been introduced with detailed information on tourism resources. The functions of the map make it easy for travelers to find information online quickly. In addition, the map is designed to be highly interactive, allowing travelers to rate, comment and rank travel destinations and other services online. A complete GIS software and database system including spatial and non-spatial data of tourism were built. This allows individuals, agencies, and government officers to manage and evaluate the whole tourist activities in the province. Future research will explore additional machine-learning techniques to improve the accuracy of the model.

Conflicts of Interest: The authors have no conflicts of interest to declare that we have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Deng Zu-tao and Lu Yu-qi. The Application of BP Neural Network in Forecasting China's Inbound Tourist Population. Tourism Science, 2006.
[2] J P Teixeira and P O Fernandes. Tourism time series forecast with artificial neural networks. Te´khne - Review of Applied Management Studies, 12(1):26–36, 2014.
[3] Y Yao, Yi Cao, Xuemei Ding, Jia Zhai, Junxiu Liu, Yuling Luo, Shuai Ma, and Kailin Zou. A paired neural network model for tourist arrival forecasting. Expert Syst. Appl., 114:588–614, 2018.
[4] Andry Alamsyah and Putu Bella Ayastri Friscintia. Artificial neural network for indonesian tourism demand forecasting. In 2019 7th International Conference on Information and Communication Technology (ICoICT), pages 1–7, 2019.
[5] Wolfram H o¨pken, Tobias Eberle, Matthias Fuchs, and Maria Lexhagen. Improving Tourist Arrival Prediction: A Big Data and Artificial Neural Network Approach. Journal of Travel Research, 60:998–1017, 2020.
[6] Ying An. The Design of Computer Network-based Intelligent Tourism System. 2022 6th International Conference on Intelligent Computing
[7] Le Quyen Nguyen, Paula Odete Fernandes, and Joao Paulo Teixeira. Analyzing and Forecasting Tourism Demand in Vietnam with Artificial Neural Networks. Forecasting, 4(1):36–50, 2022.
[8] Dinda Thalia Andariesta and Meditya Wasesa. Machine learning models for predicting international tourist arrivals in Indonesia during the COVID-19 pandemic: a multisource Internet data approach. Journal of Tourism Futures, ahead-of-p(ahead-of-print), jan 2022.
[9] Jianwei Bi, Tian Han, and Hui Li. International tourism demand forecasting with machine learning models: The power of the number of lagged inputs. Tourism Economics, 28:621 – 645, 2020.
[10] Gang Xie, Yatong Qian, and Shouyang Wang. Forecasting chinese cruise tourism demand with big data: An optimized machine learning approach. Tourism Management, 82:104208, 2021.
[11] Nan Yu and Jiaping Chen. Design of machine learning algorithm for tourism demand prediction. Computational and Mathematical Methods in Medicine, 2022, 2022.
[12] A Parasuraman, V Zeithaml, and Leonard L Berry. SERVQUAL:A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64:12–40, 1988. and Control Systems (ICICCS), pages 769–772, 2022.

[13] T W Manning, Eduardo Fayos-sola`, B Weiss, and Isidro del Rio. Indicators of sustainable development for tourism destinations. WTO, 2005.
[14] Truong Thi Kim Chuyen. Develop a strategy for sustainable tourism development in Phu Quoc island district, Kien Giang province until 2020, 2008.
[15] Nguyen Hong Giang Luu Thanh Duc Hai. Analysis of factors affecting the satisfaction of tourists when traveling in Kien Giang. Can Tho University, Journal of Science, 19:85–96, 2011.
[16] Tran Huu Cuong Le Thi Tuyet , Nguyen Anh Tru, Vu Thi Hang Nga, Tran Thi Thuong. Study on Domestic Tourists' Satisfaction with Tourism Service Quality at Duong Lam Ancient Village. Science and Technology Development, 12:620–634, 2014.
[17] Hoang Trong Tuan. Factors affecting tourist satisfaction at tourist attractions in Ho Chi Minh City. Journal of science , Ho Chi Minh Universtiy of eduaction, pages 87–97, 2015.
[18] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4):115–133, 1943.
[19] Stuart E. Dreyfus. Artificial neural networks, backpropagation, and the kelley-bryson gradient procedure. Journal of Guidance Control and Dynamics, 13:926–928, 1990.
[20] Elijah Polak. Optimization: Algorithms and consistent approximations. 1997.
[21] American Garmin Ltd. GPSMAP® 62sc. https://www.garmin.com/en- us/p/89556, 2020.
[22] Shunfu Hu. Online map service using google maps api and other javascript libraries: An open source method. 2012.

Authors Profile

Thanh-Nghi Doan received his doctorate degree in computer science from University of Rennes 1, France, in 2013. He has worked as a Ph.D. candidate in TEXMEX Research Team, IRISA, France. He is currently working at An Giang University, Vietnam National University Ho Chi Minh City, Vietnam. His research is mainly focused on machine learning, data mining and high-performance computing in computer vision, agricultural applications including insect pest image classification systems, prediction of the damage of rice diseases. Especially he made a major contribution to 2D and 3D image understanding systems in agriculture in Mekong Delta.