

Research Article

Tourism Growth Prediction Based on Deep Learning Approach

Xiaoling Ren,¹ Yanyan Li,² JuanJuan Zhao ,³ and Yan Qiang³

¹College of Economics and Management, Taiyuan University of Technology, Jinzhong 030600, China

²College of History and Culture, Shanxi University, Taiyuan 030006, China

³College of Information and Computer, Taiyuan University of Technology, Jinzhong 030600, China

Correspondence should be addressed to JuanJuan Zhao; zhaojuanjuan@tyut.edu.cn

Received 18 February 2021; Revised 4 April 2021; Accepted 13 April 2021; Published 14 July 2021

Academic Editor: M. Irfan Uddin

Copyright © 2021 Xiaoling Ren et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The conventional tourism demand prediction models are currently facing several challenges due to the excess number of search intensity indices that are used as indicators of tourism demand. In this work, the framework for deep learning-based monthly prediction of the volumes of Macau tourist arrivals was presented. The main objective in this study is to predict the tourism growth via one of the deep learning algorithms of extracting new features. The outcome of this study showed that the performance of the adopted deep learning framework was better than that of artificial neural network and support vector regression models. Practitioners can rely on the identified relevant features from the developed framework to understand the nature of the relationships between the predictive factors of tourist demand and the actual volume of tourist arrival.

1. Introduction

Most countries depend on the tourism sector for economic growth as it creates jobs and contribute about 10.4% to the global gross domestic product (GDP) as at 2019 [1]. Prediction of tourism demand (TD) is one of the critical activities in the tourism sector due to the need for precise forecasts in making certain operational decisions, such as pricing, resources, staff, revenue, and capacity management [2, 3]. Precise TD prediction is also required by the governments to facilitate planning of destination infrastructure, operational flexibility, and environmental quality control [4]. Both quantitative and qualitative techniques are used in TD forecasting. The qualitative techniques depend on intuition, experience, and understanding of a specific destination market and exhibit poor adaptability [5]. They are used mainly in the prediction of tourist arrivals based on historical records and other determinants of tourism volume [6].

Most of the existing studies in TD prediction are based on the quantitative approach; they normally present a model using training data from historical tourist arrival volumes (TAVs) and other TD predictors [7, 8]. Web technology

advancements have made search engines an essential tool for tourists when planning their trips, especially in getting relevant information on their areas of interest. Search Intensity Indices (SIIs) have been recognized as a potential TD indicator in the destination market [9, 10] and have been examined by many researchers for TD prediction [11]. SII data are important for accurate TD prediction even though some practitioners have reported some challenges in using them with the conventional prediction models. Two of the common practical barriers that exist are as follows: (i) issues related to feature selection. Rosselló-Nadal and He [12] stated that many factors are considered potential TD predictands, such as the exchange rate, travel cost, tourism prices, and several SII data. The increasing number of the potential influential predictands reduces the availability of training datasets in the feature space, meaning nonavailability of enough data for the building of accurate models.

According to Wang et al. [13], training of most prediction models on training data with numerous explanatory factors is a tedious task; hence, feature engineering is an essential step in building a prediction model as it provides the best set of relevant features that will improve the performance of the developed model [14]. Although some

factors have established meanings, it is possible that numerous potential keywords are focused with end user in terms of tourism marketing. Selected feature to TD prediction is currently based mainly on the existing knowledge of the tourism of destination market with the selection of the effective features demand significant human efforts [15]. (ii) The second issue is related to lag order selection. Even though numerous TD prediction methods have adopted the SII data, few studies focus on the determination of the huge relation for series data of the time. Some of the current studies investigated unpredictability hypothesis using either the Granger causality test or Pearson correlation coefficients [16]; in these studies, the investigation of the less hypothesis is performed by examining the extent of the relationship between the lagged values of a factor and the volume of tourist arrival.

However, the reliability of both the Granger causality test and Pearson correlation coefficients is still in doubt in the face of nonlinear underlying relationships [17]. Therefore, the effective selection of the potential data relationships will enhance the development of highly accurate prediction models. The performance of time-series and AI models in prediction tasks is excellent as they rely less on feature selection depends on the existing information of market destination.

The performance of the existing prediction methods is not satisfactory in every destination market due to the impact of numerous real-world situations, especially when adopting numerous SII data as TD indicators that may demand significant field-related expertise to restrain ambiguity. In addition, each SII parameter has different numbers of lag effective; the complexity increases in the presence of the same problem, such as bias in platform and language [18].

The emergence of deep learning techniques has been the solution to most of the above barriers as they have provided the means for achieving accurate TD prediction [19]. Deep learning is an extension of the ANN techniques of two processing layers of nonlinear relationships; they have an influence on many methods due to its built-in feature selection capability. The deep learning network is also relevant in time-series analysis owing to its advantages of being flexible and ability to discriminate nonlinear relationships. Long-term dependencies can be specifically handled and learnt by the recurrent neural network (RNN), attention mechanism, and long short-term-memory (LSTM) models. Hence, deep learning is considered an alternative solution to TD prediction. This work proposes a deep learning technique for TD prediction that will simultaneously address the earlier mentioned practical issues.

Structure of this manuscript is as follows: introduction is mentioned before, and then related work shows the existing methods. The deep learning section comes with its properties, following by difference of deep learning and normal neural network, then continuing by the proposed method and detail regarding the main method. Result and discussion section comes with analysis and the given results followed by conclusion section to summarize the research.

2. Related Work

The need for accurate TD prediction is enormous as it provides the necessary tourism-related information to researchers and practitioners; such information aids in making decisions on certain activities, such as resource allocation, risks, and opportunity identification. A review of the existing literature on TD prediction was presented in this section; deep learning technique was also reviewed as it is the selected basis for this work.

2.1. Tourism Demand Prediction. The existing TD prediction-related studies can be classified into two approaches: qualitative and quantitative. The qualitative approach mainly depends on the experience and knowledge on a specific domain; hence, they are sometimes considered “artistic,” and they exhibit low generalization potential [20]. The quantitative approach has been the method of choice for the estimation of the mount relation between many notes data in tourism. They develop models depending on historical data for prediction to potential tourist arrival volumes. The performance of the quantitative approach can be improved using two main approaches: the first plan is to introduce more related parameters that can motivate tourism-related travels and the second strategy is adoption of more complex models that can accurately generalize future trends.

The construction of TD prediction models mainly depends on input factors that are highly related to TD with no missing values. The available TD prediction models can be grouped in different ways using different criteria. They can be grouped into determinants and indicators based on the nature of their influence on TD. Determinants are the major prediction parameters. The traditional theory of economic, such as utility theory, and the theory of the behaviour of consumption estimate that TD is influenced by both qualitative and quantitative economic factors, but most of the TD prediction models do not consider the qualitative economic factors due to the difficulty of their quantification. Such models rely mainly on the quantitative economic factors because they can be measured and used feature to prediction methods. Considering TD nature, it could be stated that the consideration of only economic factors is not enough. Some works have previously focused on the impact of noneconomic determinants on travel motivations, as well as the impact of travel motivation on the destination choices. Kumar and Kumar [21], for instance, introduced qualitative noneconomic determinants such as climate index, special events, and leisure time index. These limits will have classified to pull, resistance, and push, considering their relationship with the specific domain; however, the pull factors are the only considered destination tourism market attributes [22] while the push factors have more influence on the source market [23]. The resistance factors on the other hand are comprised of factors restricting travel from source to the destination market, such as the relative prices and perceived corruption [24].

The determinants of economic theory mainly influence TD. The TD prediction accuracy can be improved by introducing some leading parameters that are seen as secondary factors into the prediction model [25]. The advancement of Web technology has made it possible for most tourists to search for important info before engaging on their trips; such information may be related to selection of the destination, hotel reservations, booking flights, and planning activities. The attention of tourist is reflected on the SII data; hence, they are effective TD indicators that need to be incorporated in TD prediction models. The study by [26] focused on the analysis of the relationship between the search words of tourists on US cities and the level of attractiveness of the cities. From the analysis, SII data were found as important tourism scale indicators in the destination market. Furthermore, Hilal et al. [27] predicted TD of Hong Kong from 9 source countries using the Google Trends Index; the study attested to the importance of SII data for such tasks.

Yang et al. [28] stated that data of SII portray choices to tourist and offer rapid data, which represents right on time variations to the choices of people. The relevance data of SII to prediction of motel residence were also presented by Li et al. [29]. These attributes made them better than the variable time-series of individual models as they can solve the problems associated with abrupt variations in econometric patterns [30].

Search engine selection is mainly based on their popularity in the targeted tourism market. Google and Baidu provide the SII data commonly used in the existing literature [31, 32]. However, the Baidu Index provides daily search volumes while Google company trends provide and normalize the index every week/month. Several prediction models are available in the field of TD prediction. In [8], the authors suggested that these methods are classified into time-series, AI, and econometric models, but TD prediction is usually done using time-series and econometric models [33].

Most of the common prediction methods are extensions of the AMA model [12] while the complex ones, such as Bayesian model, Markov-switching model, and generalized dynamic factor model, are developed for better performance [22, 24]. The existing techniques in this group rely on historical time-series patterns to determine the relationships between TD predictands and the tourist arrival volumes. When building a predictive model, the main task is to introduce the best set of features that will reduce the prediction errors based on the measured performance metrics, such as mean absolute errors (MAEs), root mean squared errors (RMSEs), or mean absolute percentage error (MAPE).

The AI models are the soft computing and machine learning methods used in TD prediction. The study by Law and Au [34] relied on multivariate regression analyses to build models for the identification of nonlinear relationships using neural networks. Furthermore, Zhang et al. [16] presented the improvement of the comprehensibility of TD prediction models using the rough set approach. According to Chakrabarty [35], evolutionary computing methods can be easily used to predict the monthly arrival of tourists at the

Balearic Islands of Spain. Effectiveness and efficiency proved by machine learning were used to evaluate the distribution of people during tourism [34]. Hybrid models that integrate different models have recently been shown to provide better results [20, 36]. However, the “No Free Lunch” concept provides that no single technique can perform well in all scenarios as all methods are associated with certain limitations in certain scenarios [37]. Econometrics and time-series models normally depend on stabilizer economic structure and patterns of historical while AI models depend on the size with quality from the training data available.

2.2. Deep Learning (DL). Successful prediction of TD using AI models, such as SVR and neural networks, has been reported. The study by Chen et al. [38] succeeded in training deep network models through greedy layer-wise pretraining for wide range practical applications. Since the development of deep learning, it has found numerous practical applications, ranging from pattern recognition to natural language processing and image recognition [39]. It has also been used in predicting sequential data problems [40]. This section discussed two common deep network techniques that have shown great efficiency in time-series prediction.

The reviewed deep learning techniques are LSTM and RNN with mechanism of attention. Regarding the RNN, its popular deep networking [41] processes data elements through selective passage of information across the time steps. This is an important attribute for its suitability in TD prediction as its structure is included in the time-series data series provides important information of context. Presented in Figure 1, either of the x for input or y for output of RNN, it can be a singular point of data, but both are time-series data. The memory of the RNN is preserved in the neuron of the hidden layer that captures all information that has been previously processed. The neuron output is generated based on the state of the neuron of the previous hidden layer and the current input via a feedback loop mechanism. The RNN can establish relation between a range of loop elements; it is also being found efficient in data series of nontime [42].

LSTM was developed as an extension of the RNN; it was built with not just a recurrent learning unit, but with several gates for the capturing of the longer and shorter states from the starting and last units, respectively. This feature has enabled the use of LSTM in solving time-series prediction problems. Regarding the mechanism of attention, its feature selection technique works together beside other deep learning (DL) models. The model can assign different weights to different inputs to learn the relevance of the input without the need of doing it prior to model fitting.

The incorporated attention mechanism of LSTM made it ideal for TD prediction because it offers solution to both prediction and feature selection problems. Figure 2 shows the structure example of LSTM.

2.3. Rationale of This Work. Much progress has been made in the prediction of TD, but these developments are yet to reflect in the performance of feature selection processes for TD prediction. Although the prediction performance of TD

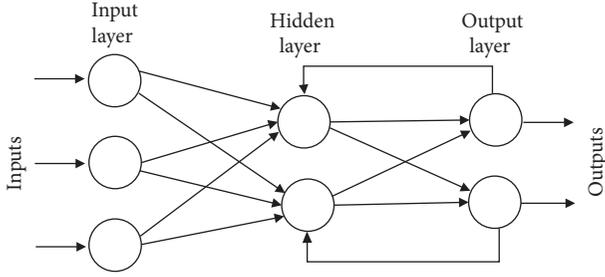


FIGURE 1: RNN structure.

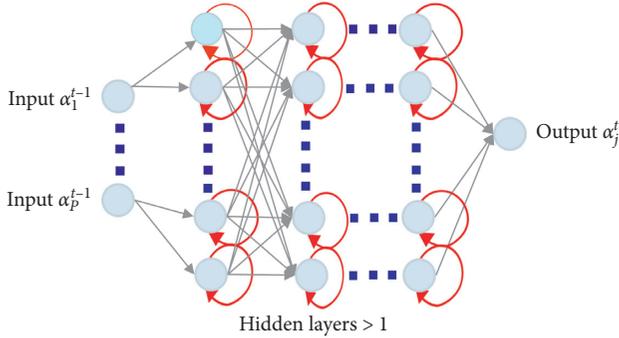


FIGURE 2: Structure example of LSTM.

prediction models is dependent on the selected features used as the training data, TD prediction is still facing two practical limitations; the first issue is associated with feature selection. For secondary TD prediction factors, feature selection is mainly aimed at query selection with the aim of collecting search engine keywords that are related to tourism. However, the issue of long tail in SII data means there are numerous search density small search queries, thereby reflecting travel experiences unique [22]. Some irrelevant and redundant features can be removed using common senses to arrive at a set of ideal subqueries related, but this demands much human effort. The second limitation is related to lag order selection. This lag order determines the nature of the relationship between time-series data. Hence, lag order selection is an important step of preliminary in TD prediction; it is done using methods such as Granger causality test [27].

Most times, some tests do not capture nonlinear relationships or explain the underlying effects of confounding [29]. A wrong lag order selection process will invalidate the subsequent steps of building a prediction model.

Deep learning is an aspect of artificial intelligence that is considered a potential alternative to the existing TD prediction models due to its two unique properties which are (i) ability to naturally learn from highly nonlinear correlations and (ii) ability to automatically select appropriate features at different network layers due to its built-in feature selection mechanism. Moreover, lag selection can be reduced by exploiting the temporally local correlation between TD and its predictands; this can carefully select the best feature of the raw data of input.

The properties made deep learning (DL) ability solution of the over-reliance in the field of expertise. Hence, the aim

of the proposed work is to develop a model of deep learning for TD prediction via using the deep network framework for autoextraction of the relevant features with suitable lag orders from various potential features.

With the proposed method, deep learning used features extracted and structured as the neural network with variable nodes and layers. Taken features control the number of layers and nodes and control the system accordingly.

2.4. Deep Learning vs. Neural Network. The neural network works on the basis of entering the information that must be processed into the input layer and then into the hidden layers that are limited in number in the neural network. Also, the number of nodes is fixed in a normal neural network. As for deep learning, the number of hidden layers that are responsible for the process is variable. What distinguishes deep learning is the change in the number of hidden layers as well as the number of nodes in the concerned layer. As for the proposed method, deep learning also contains feedback between the hidden layers and the output layer as well.

3. Methodology

A deep learning-based conceptual framework for TD prediction is proposed in this study; the following subsections provide the detail of the adopted deep learning framework for addressing the abovementioned problems. The proposed conceptual model was used for TD prediction.

The proposed deep learning-based model for TD prediction was developed according to the steps shown as follows:

- (1) The first step is the identification of the search engine platforms. While planning for travels, tourists normally use several search engines to get information about their potential destination. Hence, it is expected that different source marketing platforms have various search engines. So, that is why it is important to identify the search engine platforms first. Google, for instance, is the commonest search engine in most of the English-speaking regions of the world, while Baidu and Yandex are common in places such as Russia and China [29].
- (2) The second step is the data collection phase. There is a need to get the volume of monthly tourist arrival volume from reliable sources. Most TD prediction variables can be sourced from various resource depends on data that are available. The collection of other factors, such as data of SII, involves the following substeps:
 - (i) Identification of the initial search terms of the potential destination of the market and maximization of potential set of keywords searched in advance that will reflect the points of interest to tourist on the end user market using Google related to trends queries.

- (ii) Translation of the tourism-related keywords into languages that can be carried by the other search engine platforms using Google Translate.
- (iii) Extraction of the data per month of SII that corresponds to entered keywords comes from the final stage of a certain engine search. Monthly data can be obtained from Google Trends while the Baidu Index and other search engines may only provide daily data that require conversion into monthly data. Some search platforms may also not have the SII data of some relevant keywords.

Although the feature selection step is not required in the deep learning framework that was applied in the subsequent step, it is important to remove most of the irrelevant factors that relate loosely with the size of tourist arrivals; this will be done automatic by the deep learning framework. Regarding the limitation of the linear Pearson correlation coefficients (PCCs), factors with minor associations can be prefiltered using the maximal information coefficient (MIC). This MIC is based on the concept that if there are two related features, a scatter plot of the grid that divides the data can be made to understand the relation of related features. Provide the MIC generally:

$$MIC - \rho^2, \quad (1)$$

where ρ represents the PCC and can be used as a natural nonlinearity measure. For a high value of MIC, large equation (1) implies a nonlinear relationship while a linear relationship is denoted by small equation (1) [21].

- (3) The training of the deep learning (DL) model is the third step. Being that deep learning techniques are developed with a built-in feature selection mechanism, this study proposes a deep network architecture for autoselection of the relevant factors and determination of the order delay time sequences of series.
- (4) The model interpretation considers the last step. The trained model captures the relationship between the TD predictands and the tourist arrival volume. The most influencing factors on the lag orders can be determined from the weights of neural connections and degrees of attention. The proposed framework in this study requires no manual feature selection process as the deep learning model automatically performs the feature selection process.

3.1. Deep Network Structure. The proposed deep network architecture and its articulation with historical time-series TD data are presented in this section. The integration of the attention mechanism into the LSTM network is also detailed; the mechanism of attention gives the degrees of interest in the various factors. Tourism demand prediction is mainly aimed at predicting the tourist arrival volume based on historical multivariate factors.

The input is formally portrayed as the fully observed set of feature vector as follows:

$$(x_l)_{l=1^L} = x_1, x_2, \dots, x_t. \quad (2)$$

The corresponding tourist arrival volume is represented as

$$(y_l)_{l=1^L} = y_1, y_2, \dots, y_t, \quad (3)$$

where T represents the total length of time in steps, and say the week numbers or months with the gathered database. y_t represents the tourist arrival volume (TAV) with time of step t , while x_t of vector of multivariate factors.

The prediction of TD requires the use of the time-series of multivariate parameters $(x_l)_{l=1^L}$ and the real TAV $(y_l)_{l=1^L}$ as model inputs for the construction of a model F for the prediction of y at future time steps:

$$(\hat{y}_L)_{l=T+1}^{L+\Delta} = F((x_l)_{l=1}^L, (y_l)_{l=1}^L). \quad (4)$$

This expression differs from those of autoregressive models where the availability of $(x_l)_{l=L+1}^{L+\Delta}$ is normally assumed when predicting $(y_l)_{l=L+1}^{L+\Delta}$ as both are designed for the modelling of the relationship between conditions and their consequences. Long-term dependencies can be handled by the RNN, but the training is sensitive with the RNN of changing in gradient. So, the proposed LSTM and RNN can address this issue through the provision of block memory in their current connections. Cell memory is contained in each block for the storage of the temporal states of the network; there are three gates that control the network, which are called on the basis of data flow, and they are as follows: forget, remember, and inference; they ensure that weak signals are blocked from flowing through the network. Figure 3 depicts the LSTM framework.

Assuming the time-series as an input, this input is encoded by the LSTM into a set of hidden states $(y_l)_{l=1}^L$. LSTM is based on the concept that a few gates are implemented at each time step from the regulation of the flow of information through the sequences; this enables accurate capturing of any long-range dependency. For any time of step l in LSTM, the capturing of long-range dependencies requires updating of the hidden state H_l by the fixed data with the same current of time step x_l , at the preceding time step H_{l-1} , at the input gate z_l , at the forget gate f_l , at the output gate o_l , and at the memory cell c_l using the following equations [17]:

$$z_l = \sigma(W_{xi}X_l + W_{hi}h_{l-1} + b_z), \quad (5)$$

$$f_l = \sigma(W_{xi}X_l + W_{hi}h_{l-1} + b_f), \quad (6)$$

$$o_l = \sigma(W_{xi}X_l + W_{hi}h_{l-1} + b_o), \quad (7)$$

$$c_l = f_l * c_{l-1} + z_l * \tanh(W_{xi}X_l + W_{hi}h_{l-1} + b_c), \quad (8)$$

$$h_l = o_l * \tanh(c_l), \quad (9)$$

where σ and \tanh are frequent functions, while X indicates sage multiplication. The b and W are used with the LSTM as the parameters though framework training. The result of

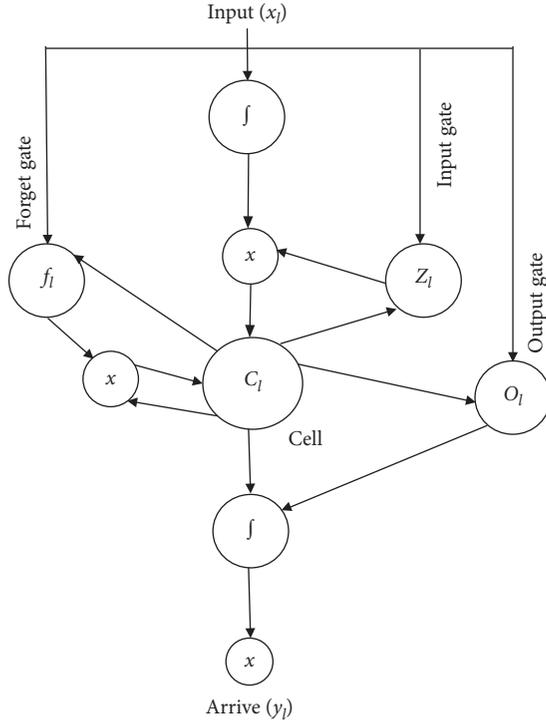


FIGURE 3: The LSTM framework.

equation (9) is utilized to estimate the results of the linear retreating layer:

$$\hat{y}_i = W_r h_i^L. \quad (10)$$

As mentioned in equation (7), W_r is indicated as the linear weight retreating layer.

The proposed method deep network (DN) structure of TD is seen in Figure 4. The framework used the LSTM parameters along with the concern technique. Identify the lag or lead relations among data series at time is important in TD expectation due to advantage effect varies depending on the periods of delay. The LSTM is used for the framework of long-term dependency in data series of the time, with attention technique which is indicated in which parts of the unit sequentially units of the framework are effective. This structure gives ability to authors to take two critical parts of data in TD expectation, called as follows:

- (i) The relationship of temporal among different demand and factors
- (ii) Significance of the factors depending on weights on TD

Therefore, the temporal of long-term dependency among different arrivals of tourist volumes and factors can be detected automatically via utilizing the LSTM of the alertness technology. Size of the input is $(m * d * n)$, while $(m * d)$ considers size of choosing data in training. The (n) is large tardiness arrangement specified by the users. All layers that are connected (dense) consider bullied according to attention technology with each attention. Then, the common relevant information is chosen in the lead series as seen in the following equation:

$$e_l = (W_e^1 \tan h W_e^2 (h_{l-1}; s_{l-1}) + W_e^3 x_l h_{l-1}), \quad (11)$$

where W_e^1 , W_e^2 , and W_e^3 are the weighing to be used by the framework. While the vector (e_l) represents as weighing that calculates the significance of feature of lead series on time l (x_l), and it was normalized e_l . Thus, lead series x_l is multiplied by attention weighing a_l : $\bar{x}_l * a_l$.

The LSTM utilizes \bar{x}_l and h_{l-1} as input and updates the concealing case at time h_l . The vector c_l is presented by calculating up the multiplication:

$$c_l = \sum_{i=1}^L h_i w_i. \quad (12)$$

The final result is generated by the linear layer:

$$y_l = w_y^1 (w_y^2 c_l + b_1) + b_2. \quad (13)$$

The LSTM and the identifier can be trained one by one.

4. Results and Discussion

4.1. Empirical Study. Implementation of the proposed imaginary framework was empirically evaluated through the prediction of the monthly TAV in Macau, an autonomous region of China that is located across the Pearl River Delta from Hong Kong. The economy of this region mainly depends on its gaming and tourism sectors. So, it is important to have an accurate and timely prediction of the TAV to sustain the economy of the region. The TAV of Macau in this study was predicted using secondary indicators such as SII data due to the lack of reliable data sources and expertise.

4.2. Performance Evaluation. To proposed model was evaluated for TD prediction performance by comparison with some conventional prediction models that included the SVR, ANN, ARIMA, and ARIMAX models [18, 20]. The predictions with the conventional models relied on the use of the TAVit dates back 12 months as an estimate of y_{k+1} , with the name of $\hat{y}_{k+1} = y_{k+11}$. For the ANN and SVR models, the input data used were the data from the past 12 months $(x_l, y_l)_{l=k-11}^k$ for the prediction of \hat{y}_{k+1} . The ANN model was built with a sigmoid activation function and one hidden layer; it was trained using the backpropagation algorithm. Stationary series are achieved in the ARIMAX and ARIMA using the AR order of (p, d) times of difference; they are trained on the MA order of q using the tourist data from the past 12 months for the prediction of the next 12-month TAV series; this is gradually increased during the process of model validation. ARIMA differed from ARIMAX by using only the TAV for the prediction task while ARIMAX requires other external factors, x_{l-1}^T . Features that contribute less to the TAV were eliminated using the MIC because such a large number of features cannot be handled by the SVR and ANN. The walk-forward model validation was employed to mimic a real-world scenario where new TAVs are made available monthly for the prediction of the TAV for the following month. The

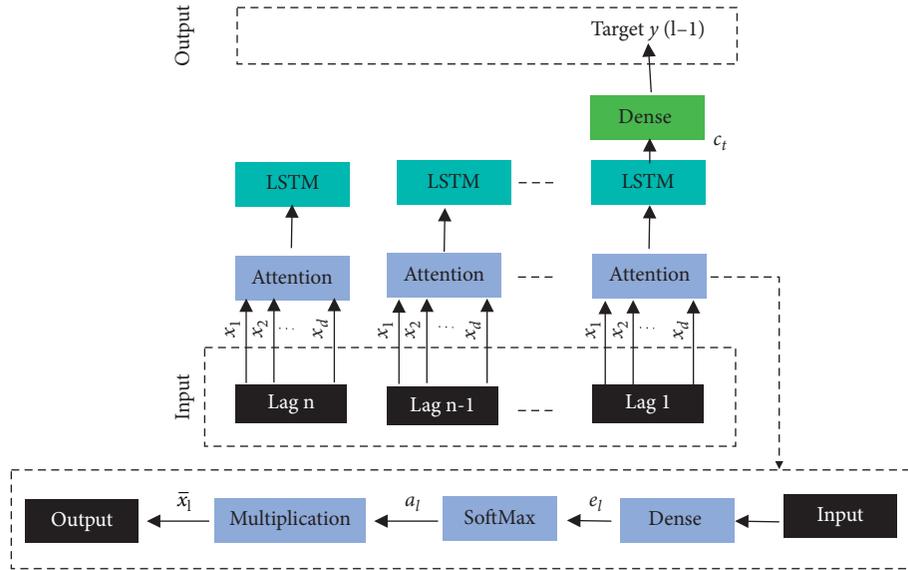


FIGURE 4: The deep network structure for tourism demand (TD).

prediction model is trained in each step and used to predict the TAV for the next month.

Three measures are used with this work of forecasting accuracy to the acquired predicted values (RMSE, MAE, and MAPE). They are defined as follows.

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \\ \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \\ \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i}. \end{aligned} \quad (14)$$

The best prediction model is determined as the model which achieved the lowest measurement values. The robustness of the developed model in this study was ensured by repeating the walk-forward validation process for the benchmarking models five times in the prediction of 12-month TAVs from 2016 to 2019. Note that the comparative models (SVR & ANN) have no expert-crafted features as their input features were automated from the MIC filtering process. Both models were further implemented using the features selected by the deep learning model; here, the models were recognized as SVR + F and ANN + F. Tables 1–3 summarize the global MAPE, MAE, and RMSE of the comparative models over a five-year period. The table shows that the performance of the deep network is better in terms of achieving low measurement errors for the 4 consecutive years.

Table 4 shows the result of the one-tailed l-test for the MAPE, MAE, and RMSE of the proposed and comparative

TABLE 1: The MAPE comparison for last 4-year prediction (global).

	DLM	SVR + F	ANN + F	ANN	SVR	ARIMA	NALVE
2016	0.925	7.441	5.474	7.656	8.718	5.31	7.177
2017	1.133	4.899	7.312	5.191	5.717	6.31	4.386
2018	2.434	4.564	4.981	30.051	4.298	6.93	2.958
2019	1.772	4.697	5.898	12.228	8.434	6.50	5.963

models in this study at $\alpha = 0.05$. The results showed that the null hypothesis is rejected as it suggests the equality of the mean values of the proposed DLM and the comparative model. This is supported by the lower MAPE, MAE, and RMSE values of the DLM compared to the comparative models. The SVR and ANN with no feature selection performed comparatively with the conventional method. However, the proposed DLM which required no feature selection process performed better in terms of prediction than the other models. Note that the performance of SVR and ANN was improved by incorporating the DLM-selected features into their framework; the MAPE of SVR reduced to 5.086% from 6.482% while that of ANN reduced to 5.922% from 14.319%. Hence, the feature selection capability of the proposed approach is validated.

Although all DLM, ARIMA, and ARIMAX do not require preselected features, the ability of DLM to autoselect relevant features from the raw SII data improved its performance more than that of the ARIMA model. The comparison was made using only the Baidu keywords and Macau tourist arrival from Mainland China. Tables 5–7 show the MAPE, MAE, and RMSE of the comparative models within the studied five-year period. These tables also show similar improvements in performance; for instance, there were decreases in the MAPE. The results of

TABLE 2: The MAE comparison for last 4-year prediction (global).

	DLM	SVR + F	ANN + F	ANN	SVR	ARIMA	NALVE
2016	22,257	201,484	154,196	222,701	210,318	118,139	191,189
2017	29,479	122,697	178,699	128,183	148,698	141,279	108,102
2018	64,598	114,743	122,549	607,331	113,690	182,658	77,967
2019	42,556	128,299	163,971	295,977	217,234	192,388	165,904

TABLE 3: The RMSE comparison for last 4-year prediction (global).

	DLM	SVR + F	ANN + F	ANN	SVR	ARIMA	NALVE
2016	27,102	269,194	211,419	279,121	276,138	162,255	208,665
2017	40,030	158,672	219,891	166,446	211,729	188,802	143,833
2018	88,811	154,531	134,545	632,044	135,048	234,939	95,596
2019	51,613	167,243	190,895	352,530	216,289	213,798	202,697

TABLE 4: The l -test results (global).

Comparison	MAPE		MAE		RMSE	
	p value	l -interval	p value	l -interval	p value	l -interval
DLM vs. SVR + F	0.0044	-5.22, ∞	0.0055	-143,255, ∞	0.0046	-194,018, ∞
DLM vs. ANN + F	0.0006	-5.48, ∞	0.0007	-143,121, ∞	0.0024	-188,382, ∞
DLM vs. ANN	0.0112	-24.27, ∞	0.0098	-501,812, ∞	0.006	-514,013, ∞
DLM vs. ARIMA	0.00003	-6.02, ∞	0.0002	-174,193, ∞	0.0003	-153,644, ∞
DLM vs. Naive	0.0042	-5.58, ∞	0.0097	-152,931, ∞	0.0094	-168,319, ∞

TABLE 5: The MAPE comparison for last 4-year prediction at China.

	DLM	SVR + F	ANN + F	ANN	SVR	ARIMA	NALVE
2016	1.846	6.86	9.45	12.62	14.1	8.39	12.11
2017	1.533	7.18	7.89	9.88	8.26	12.88	6.36
2018	1.658	4.59	5.18	6.78	6.39	8.31	3.21
2019	1.548	4.45	6.31	5.93	8.42	8.98	8.84

TABLE 6: The MAE comparison for last 4-year prediction at China.

	DLM	SVR + F	ANN + F	ANN	SVR	ARIMA	NALVE
2016	32,175	120,539	175,722	233,449	263,029	164,588	218,354
2017	27,316	118,491	136,835	160,881	136,998	203,611	104,771
2018	31,564	77,343	95,732	108,242	110,638	136,998	54,763
2019	28,944	75,383	109,944	109,874	165,568	156,299	168,568

TABLE 7: The RMSE comparison for last 4-year prediction at China.

	DLM	SVR + F	ANN + F	ANN	SVR	ARIMA	NALVE
2016	41,159	140,911	231,644	282,367	302,988	205,286	242,181
2017	29,873	151,186	168,883	182,485	171,184	230,795	142,591
2018	35,701	127,488	117,526	128,468	132,871	171,182	69,855
2019	31,454	113,514	134,952	134,222	205,288	190,473	193,677

TABLE 8: The l -test results at China.

Comparison	MAPE		MAE		RMSE	
	p value	l -interval	p value	l -interval	p value	l -interval
DLM vs. SVR + F	0.0007	-6.24, ∞	0.0018	-119,022, ∞	0.0004	-150,781, ∞
DLM vs. ANN + F	0.0015	-9.08, ∞	0.0019	-171,432, ∞	0.0015	-217,779, ∞
DLM vs. ANN	0.0148	-18.5, ∞	0.0121	-298,133, ∞	0.0074	-322,915, ∞
DLM vs. SVR	0.0019	-11.24, ∞	0.0033	-206,089, ∞	0.0022	-239,332, ∞
DLM vs. ARIMA	0.0015	-9.81, ∞	0.0006	-161,044, ∞	0.0003	-185,229, ∞
DLM vs. Naive	0.0068	-10.22, ∞	0.0093	-182,015, ∞	0.0059	-202,661, ∞

the l-test for the proposed DLM and the comparative models are shown in Table 8; the results further supported better performance of the proposed DLM compared to the benchmark models.

5. Conclusion

The tourism sector requires accurate and timely demand predictions for making informed sector-related decisions. Studies have previously focused on the time-series, econometrics, and AI models for this task in the previous years; however, the performance of these conventional models is dependent on the goodness of the selected features. The feature selection process and their lag order determination are domain specific and demand significant human effort. This study proposed a deep learning-based approach to the selection of the relevant features for better performance of the predictive models. The evaluations showed better performance of the proposed deep network framework compared to the conventional methods possibly due to two reasons; firstly, the ability of the deep network model is to mimic the natural biological system. The successive network layers extract low-level features from the initial input layer for the subsequent abstraction of the high-level features that will capture the semantic relationships between features in the succeeding layers. Secondly, the LSTM has an attention mechanism that automatically selects the relevant features at each time step.

This study made two important contributions to TD prediction; the first one is the development of a systemic conceptual framework of TD prediction and the validation of its TD prediction capability. The proposed model in this work considered all the available TD prediction factors and required no human intervention in terms of feature selection. The second contribution is the use of the attention score for the interpretation is the trained deep network architecture. This provides the practitioners a new way of updating their TD prediction based on a set of relevant indicators at different time steps. The outcome of this work also suggests the ability of the proposed DLM to select a set of relevant features and determine their suitable lag orders. With these contributions, this work can be extended in two directions: firstly, incorporation of different types of indicators, such as Blogs and Tweets into the TD prediction task; which can address the issue of non-availability of training data as the DLM allows direct usage of these media data; secondly, feature sets with suitable lag orders can serve as inputs to other TD prediction models. The combination of DLM and the conventional prediction models can enable further theory development.

Future work can increase the features extracted and make the deep neural network more flexible in terms of the number nodes and layers. Prediction for a large variable number is required, and training should be made for the huge dataset.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the National Natural Science Foundation of China, grant number 71701141, and Soft Science Research Project of Shanxi of China, grant number 2017041007-3, for funding and supporting this work.

References

- [1] A. A. Peprah and A. F. Adekoya, "Entrepreneurship and economic growth in developing countries: evidence from Africa," *Business Strategy & Development*, vol. 3, no. 3, pp. 388–394, 2020.
- [2] H. Li, M. Hu, and G. Li, "Forecasting tourism demand with multisource big data," *Annals of Tourism Research*, vol. 83, p. 102912, 2020.
- [3] Y. Yao and Y. Cao, "A neural network enhanced hidden Markov model for tourism demand forecasting," *Applied Soft Computing*, vol. 94, p. 106465, 2020.
- [4] V. Otrachshenko and F. Bosello, "Fishing for answers? impacts of marine ecosystem quality on coastal tourism demand," *Tourism Economics*, vol. 23, no. 5, pp. 963–980, 2017.
- [5] J. Yu, J. Seo, and S. S. Hyun, "Attributes of medical tourism destination brands: case study of the Korean medical tourism market," *Journal of Travel & Tourism Marketing*, vol. 38, no. 1, pp. 107–121, 2021.
- [6] X. Fu, J. Ridderstaat, and H. Jia, "Are all tourism markets equal? linkages between market-based tourism demand, quality of life, and economic development in Hong Kong," *Tourism Management*, vol. 77, p. 104015, 2020.
- [7] K. Anurag, K. Venkataraghavan, and M. Sharma, "Bayesian BILSTM approach for tourism demand forecasting," *Annals of Tourism Research*, vol. 83, 2020.
- [8] Y. Zhang, G. Li, B. Muskat, and R. Law, "Tourism demand forecasting: a decomposed deep learning approach," *Journal of Travel Research*, vol. 60, no. 5, pp. 981–997, 2021.
- [9] L. Wen, C. Liu, H. Song, and H. Liu, "Forecasting tourism demand with an improved mixed data sampling model," *Journal of Travel Research*, vol. 60, no. 2, pp. 336–353, 2021.
- [10] B. Ayaz, H. Ullah, M. K. Khan, and S. Jan, "The effect of google search volume index on the stock market excess returns. Evidence from listed firms in Pakistan stock exchange," *Review of Education, Administration & LAW*, vol. 4, no. 1, pp. 23–35, 2021.
- [11] C. F. Tang and E. Lau, "Exploring the impact of political stability and climate change on inbound tourism demand: evidence from dynamic panel data analysis," *Regional Science Perspectives on Tourism and Hospitality*, pp. 403–419, 2021.
- [12] J. Rosselló-Nadal and J. He, "Tourist arrivals versus tourist expenditures in modelling tourism demand," *Tourism Economics*, vol. 26, no. 8, pp. 1311–1326, 2020.
- [13] X.-H. Wang, Y. Zhang, X.-Y. Sun, Y.-L. Wang, and C.-H. Du, "Multi-objective feature selection based on artificial bee colony: an acceleration approach with variable sample size," *Applied Soft Computing*, vol. 88, p. 106041, 2020.
- [14] M. Spruit, T. Dedding, and D. Vijlbrief, "Self-service data science for healthcare professionals: a data preparation approach," in *Proceedings of the 13th International Joint*

- Conference on Biomedical Engineering Systems and Technologies*, pp. 724–734, Valletta, Malta, February 2020.
- [15] K. Takahashi, “Comparing the determinants of tourism demand in Singapore and French Polynesia: applying the tourism demand model to panel data analysis,” *Tourism Analysis*, vol. 25, no. 1, pp. 175–181, 2020.
 - [16] B. Zhang, X. Huang, N. Li, and R. Law, “A novel hybrid model for tourist volume forecasting incorporating search engine data,” *Asia Pacific Journal of Tourism Research*, vol. 22, no. 3, pp. 245–254, 2017.
 - [17] A. De Rienzo, L. R. Chirieac, Y. P. Hung et al., “Large-scale analysis of BAP1 expression reveals novel associations with clinical and molecular features of malignant pleural mesothelioma,” *The Journal of Pathology*, vol. 253, no. 1, pp. 68–79, 2021.
 - [18] T. Dergiades, E. Mavragani, and B. Pan, “Google Trends and tourists’ arrivals: emerging biases and proposed corrections,” *Tourism Management*, vol. 66, pp. 108–120, 2018.
 - [19] S. Pouyanfar, Y. Yang, S.-C. Chen et al., “Multimedia big data analytics,” *ACM Computing Surveys*, vol. 51, no. 1, pp. 1–34, 2018.
 - [20] R. Law, G. Li, D. K. C. Fong, and X. Han, “Tourism demand forecasting: a deep learning approach,” *Annals of Tourism Research*, vol. 75, pp. 410–423, 2019.
 - [21] N. Kumar and R. R. Kumar, “Relationship between ICT and international tourism demand: a study of major tourist destinations,” *Tourism Economics*, vol. 26, no. 6, pp. 908–925, 2020.
 - [22] K. Zhang, Y. Chen, and C. Li, “Discovering the tourists’ behaviors and perceptions in a tourism destination by analyzing photos’ visual content with a computer deep learning model: the case of Beijing,” *Tourism Management*, vol. 75, pp. 595–608, 2019.
 - [23] Y. Li and H. Cao, “Prediction for tourism flow based on LSTM neural network,” *Procedia Computer Science*, vol. 129, pp. 277–283, 2018.
 - [24] Y. Ma, Z. Xiang, Q. Du, and W. Fan, “Effects of user-provided photos on hotel review helpfulness: an analytical approach with deep leaning,” *International Journal of Hospitality Management*, vol. 71, pp. 120–131, 2018.
 - [25] M. W. Islam, L. Ruhanen, and B. W. Ritchie, “Adaptive co-management: a novel approach to tourism destination governance?” *Journal of Hospitality and Tourism Management*, vol. 37, pp. 97–106, 2018.
 - [26] E. E. Cust, A. J. Sweeting, K. Ball, and S. Robertson, “Machine and deep learning for sport-specific movement recognition: a systematic review of model development and performance,” *Journal of Sports Sciences*, vol. 37, no. 5, pp. 568–600, 2019.
 - [27] K. Z. Hilal, A. O. Akyuz, M. Uysal et al., “An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain,” *Complexity*, vol. 2019, Article ID 9067367, , 2019.
 - [28] X. Yang, B. Pan, J. A. Evans, and B. Lv, “Forecasting Chinese tourist volume with search engine data,” *Tourism Management*, vol. 46, pp. 386–397, 2015.
 - [29] X. Li, B. Pan, R. Law, and X. Huang, “Forecasting tourism demand with composite search index,” *Tourism Management*, vol. 59, pp. 57–66, 2017.
 - [30] P. F. Bangwayo-Skeete and R. W. Skeete, R. W. Skeete, Can Google data improve the forecasting performance of tourist arrivals? mixed-data sampling approach,” *Tourism Management*, vol. 46, pp. 454–464, 2015.
 - [31] J. M. Stokes, K. Yang, K. Swanson et al., “A deep learning approach to antibiotic discovery,” *Cell*, vol. 180, no. 4, pp. 688–702, 2020.
 - [32] S. L. Oh, Y. Hagiwara, U. Raghavendra et al., “A deep learning approach for Parkinson’s disease diagnosis from EEG signals,” *Neural Computing and Applications*, vol. 32, pp. 10927–10933, 2018.
 - [33] E. S. Silva, H. Hassani, S. Heravi, and X. Huang, “Forecasting tourism demand with denoised neural networks,” *Annals of Tourism Research*, vol. 74, pp. 134–154, 2019.
 - [34] A. S. M. Arif and J. T. Du, “Understanding collaborative tourism information searching to support online travel planning,” *Online Information Review*, vol. 43, no. 1, 2019.
 - [35] N. Chakrabarty, “A regression approach to distribution and trend analysis of quarterly foreign tourist arrivals in India,” *Journal of Soft Computing Paradigm*, vol. 2, no. 1, pp. 57–82, 2020.
 - [36] X. Dong, Z. Yu, W. Cao, Y. Shi, and Q. Ma, “A survey on ensemble learning,” *Frontiers of Computer Science*, vol. 14, no. 2, pp. 241–258, 2020.
 - [37] S. P. Adam, S.-A. N. Alexandropoulos, P. M. Pardalos, and M. N. Vrahatis, “No free lunch theorem: a review,” *Approximation and Optimization*, vol. 145, pp. 57–82, 2019.
 - [38] J. Chen, L. Liu, J. Pei, and M. Deng, “An ensemble risk assessment model for urban rainstorm disasters based on random forest and deep belief nets: a case study of Nanjing, China,” *Natural Hazards*, vol. 2021, pp. 1–22, 2021.
 - [39] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
 - [40] W. Bao, J. Yue, and Y. Rao, “A deep learning framework for financial time series using stacked autoencoders and long-short term memory,” *PloS One*, vol. 12, no. 7, p. e0180944, 2017.
 - [41] K. Cho, B. Van Merriënboer, C. Gulcehre et al., “Learning phrase representations using RNN encoder-decoder for statistical machine translation,” arXiv preprint arXiv:1406.1078, 2019.
 - [42] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, and M. Shi, “A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework,” *Energy Conversion and Management*, vol. 212, p. 112766, 2020.