Journal of Hospitality and Tourism Management 49 (2021) 25–33

Contents lists available at ScienceDirect

Journal of Hospitality and Tourism Management

journal homepage: www.elsevier.com/locate/jhtm

# Using SARIMA–CNN–LSTM approach to forecast daily tourism demand☆

Kaijian He[a], Lei Ji[b], Chi Wai Don Wu[c,*], Kwok Fai Geoffrey Tso[d]

[a] College of Tourism, Hunan Normal University, Changsha, China
[b] Shanghai Kaiyu Information Technology Co., Ltd., Shanghai, China
[c] School of Tourism Management, Macao Institute of Tourism Studies, Macao SAR, China
[d] Department of Management Sciences, City University of Hong Kong, Kowloon, Hong Kong SAR, China

ARTICLE INFO

*Keywords:*
LSTM
Tourist arrivals
Deep neural network
Time-series model
CNN

ABSTRACT

Timely tourist demand forecasting is essential for the operation of the tourism industry; however, most studies focus on quarterly- or monthly-basis data, whose low-frequency nature makes it less informative than data at higher frequencies. In this article, we introduced a SARIMA–CNN–LSTM model to forecast tourist demand data at daily frequency, whose movement demonstrates the mixture of linear and nonlinear data features, difficult to model in the traditional framework. The SARIMA–CNN–LSTM model employs the SARIMA model and the deep neural network structure that combines the CNN and LSTM layers to capture linear and nonlinear data features. In the SARIMA–CNN–LSTM model structure, the SARIMA is used to capture the linear features. The Convolutional Neural Network (CNN) is used to capture the hierarchical data structure, while the Long Short Term Memory network (LSTM) is used to capture the long-term dependencies in the data. Our results confirmed that the SARIMA–CNN–LSTM model yields greater forecast accuracy than the individual models. The subtle nonlinear details in the residual are modeled better using the deep learning model. We found that the SARIMA–CNN–LSTM model can take advantage of the rich information in the high-frequency data better in the forecasting process.

## 1. Introduction

The tourism industry plays a key role in the global economy as it contributed 5% direct GDP and employed about 235 million people in 2018. Fluctuations of tourist arrivals have been a key factor for tourism service planning, due to the characteristics of services, namely, intangibility, heterogeneity, perishability, and inseparability. Thus, obtaining an accurate forecast for tourism arrivals is necessary for the efficiency of operating tourism-related companies.

In this paper, daily tourist arrival data are used to generate daily forecast as it is considered to be advantageous compared with lower frequency data at the monthly or quarterly levels (Divino & McAleer, 2010). Most studies use monthly and quarterly data to make travel demand forecasts (Jiao & Chen, 2019); however, accurate daily forecasts, compared with monthly forecasts and quarterly forecasts, is important as it provides information for decision on optimal daily operation, such as environmental and tourism tax determination, differential pricing strategies, and tourism packages formulation especially in periods of low demand. Accurate daily tourist arrival forecasting is helpful for planning and arranging tourist related business. With precise daily forecasts, government and tourism businesses can provide timely arrangements of manpower and increase capacity. Daily tourist forecasting is very important for the operations in the tourism industry.

Most researchers agreed that the fluctuation in tourist arrivals could be divided into three components, namely, trend, seasonal and irregular components. Daily tourist forecasting is more challenging as daily tourist arrivals are highly complex and mixed with both linear trend and nonlinear patterns. A wide range of statistical prediction models have been applied to tourist forecasting. We review the common prediction models. Time series models have long been used to predict tourist arrivals. The typical statistical methods include the naïve-1, autoregressive (AR), single exponential smoothing (ES) and moving average (MA) methods. Autoregressive integrated moving average (ARIMA) models and Seasonal autoregressive integrated moving average (SARIMA) models are widely used for tourism demand forecasting as they take into account the trends and/or seasonality components of the time

☆ The work described in this paper was supported by grants from the National Natural Science Foundation of China (NSFC No. 71671013), the Humanities and Social Sciences Youth Foundation of Ministry of Education of China (No. 16YJC790026), partial supported by the City University of Hong Kong funding (Project No. 11504320 and 11501019), and sponsored by a scholarship from the Macao Foundation.

* Corresponding author. School of Tourism Management, Macao Institute of Tourism Studies, Macao SAR, China.
*E-mail address:* donwu@ift.edu.mo (C.W.D. Wu).

https://doi.org/10.1016/j.jhtm.2021.08.022
Received 9 October 2020; Received in revised form 11 July 2021; Accepted 21 August 2021
Available online 4 September 2021

series. Most notably, the ARIMA model is the most commonly used model in tourism, according to Song, Qiu, and Park (2019). The ARIMA model considers current and lagged observations (AR), current and lagged random shocks (MA), degrees of integration (I). The expansion of ARIMA, Seasonal ARIMA (SARIMA) model considers seasonality adjustments (S) in addition to ARIMA. Both models have been shown to provide outstanding forecasts compared with other models for most cases. (Song et al., 2019). Econometric models such as the Autoregressive distributed lag model (ADLM) (Song, Wong, & Chon, 2003), the Bayesian Global Vector Autoregressive Model (Assaf, Li, Song, & Tsionas, 2019), the error correction model (ECM) (Kulendran & Wilson, 2000) and the vector autoregressive (VAR) modeling (Song & Witt, 2006) are also widely used. Comparing the performance of time series models and econometric models, time series models are considered to be generally better in the study conducted by Athanasopoulos, Hyndman, Song, and Wu (2011) study.

The Artificial intelligence (AI) models such as neural network and SVR have found successful applications in tourism demand forecasting (Chen & Wang, 2007; Cho, 2003; Kon & Turner, 2005). In computer terminology, AI refers to the human intelligence demonstrated by the artificially made objects. It may involve different types of human intelligence to solve complex practical problems, which range from the conscious self-awareness, the unconscious mind, the analytic, reasoning and logic capability, the learning, the planning, and the like. Machine learning is the subset of AI, which refers to the intelligence demonstrated by the computer with specially designed programs and algorithms. There exist different machine learning algorithms. The most recent prominent examples would be the deep learning model, where the intelligence is achieved by mimicking the human reasoning process in the program and algorithm design.

The main difference between the deep learning model and the traditional neural network model is the exponentially increasing level of complexity in the model structure, most notably in the depth of hidden layers in the neural network, compared to the traditional neural network. The deep learning model shares more resemblance to complex neuron structure in the human brain, which usually contains millions of connected neurons. The recent success of the deep learning model in various visual and data processing tasks demonstrates the importance of using this complicated neuron structure in achieving a higher level of nonlinear data processing power.

Compared with traditional AI models, deep learning has been widely applied in the field of prediction. As the deep learning technique provides a mechanism of feature engineering that extracts discriminative features with minimal domain knowledge and human effort, the application of deep learning approaches has received a great deal of attention from researchers and many studies have effectively used deep learning in handling complex data, for example, face verification (Taigman, Yang, Ranzato, & Wolf, 2014), breast cancer identification (Wang, Khosla, Gargeya, Irshad, & Beck, 2016), crude oil prices prediction (Chen, He, & Tso, 2017) and tourism prediction (Law, Li, Fong, & Han, 2019; Sun, Wei, Tsui, & Wang, 2019; Zhang, Li, Shi, & Law, 2020).

Although deep learning in tourism was developed, they lacked the large quantities of data and processing power needed to reach their full capabilities. Annual and quarterly tourism demands are widely used in current research, while limited research (less than 10%) uses monthly time series (Song & Li, 2008). There are even fewer studies focusing on daily arrivals in the field of tourism (Díaz & Mateu-Sbert, 2011; Divino & McAleer, 2010). In tourist arrival forecast, monthly data only have 12 observations per year; quarterly data only have four. Daily data generate at least 365 data points for each year. Many studies use deep learning for forecasts; however, these studies contain around 100 data points (Law et al., 2019) or only focus on a popular tourist spot (Lu et al., 2020). To our knowledge, there are not many studies using deep learning in tourism for large daily time-series datasets.

In the meantime, LSTM and CNN, two popular deep network architectures, have shown with great success on time series forecasting. LSTM networks is an extension of RNN, which is one of the most advanced deep learning architectures for sequence learning tasks, such as handwriting recognition and speech recognition (Hochreiter & Schmidhuber, 1997). LSTM has a recurrent learning unit inside the network with gates to capture the longer states from the beginning unit and the shorter states from the last unit. By having this feature, LSTM has been broadly used to solve time series forecasting problems, such as solar irradiance prediction (Qing & Niu, 2018), petroleum production (Sagheer & Kotb, 2019), rainfall-runoff (Hu et al., 2018; Kratzert, Klotz, Brenner, Schulz, & Herrnegger, 2018) and financial market (Fischer & Krauss, 2018).

Convolutional neural networks (CNN) have been widely used for various prediction tasks since it has a strong ability to capture local trend features and scale-invariant features when the nearby data points typically have a strong relationship with each other (LeCun, Bottou, Bengio, & Haffner, 1998). CNN has been applied to a humanoid robot to improve the human-robot interaction. CNN can also be applied to semantic segmentation. Regarding time series forecasting, by integrating the hidden features of LSTM and CNN, waterworks operational data are used to improve the accuracy and stability of the load forecast (Cao, Kim, Hwang, & Jung, 2018). CNN is also used for crude oil risk forecasting (Zou, Yu, Tso, & He, 2020).

Considering the complexity of tourist arrival time series, the introduction of deep learning to tourist arrival prediction is regarded as one of the most charming topics. After decades of development, deep learning has experienced phenomenal success in a wide range of challenging applications, for example, image captioning (LeCun, Bengio, & Hinton, 2015) and finance prediction (Bao, Yue, & Rao, 2017). With the application of a deep learning technique, the pattern can be analyzed adaptively as it has the ability to capture the highly nonlinear correlations, extract and model suitable data features at different network layers (Law et al., 2019). Despite the wide use of deep learning, this technique only started to draw attention in tourism in 2019 (Law et al., 2019; Sun et al., 2019; Zhang et al., 2020). With these successful applications, it is clear that deep learning models can perform extremely well in feature learning in tasks related to pattern recognition and prediction in a variety of application domains. Surprisingly, to our knowledge, there have been no previous attempts to deploy CNN or LSTM on a large dataset to assess its performance in tourism arrival prediction tasks on a region. In this paper, CNN and LSTM are applied to tourism arrivals from January 1, 2017 to October 31, 2019 with the intent to fill this gap. Hereby, we provide an in-depth guide on data preprocessing, as well as development, training, and deployment of both networks for tourist arrival time series prediction tasks. To the best of our knowledge, no previous work has compared deep learning techniques in large with the traditional SARIMA model.

However, the deep learning model may not apply solely for tourist arrival forecast as tourist arrival data is a mix of trend, seasonal and irregular components. Although in theory neural network is capable of modeling any nonlinear functions at arbitrary precision, in practice it can often be trapped in the local minimum problem as it overfits the data disrupted with noises. In this regard, a hybrid model serves as an important alternative. Hybrid methods have been proved to provide better prediction performance in the field outside of tourism. For example, different hybrid models are proposed in carbon futures price forecasting, including a hybrid model with the ARIMA model and the least squares support vector machine (LSSVM) (Zhu & Wei, 2013), a hybrid prediction with multi-output support vector regression (MSVR) and particle swarm optimization (PSO) (Sun et al., 2016), a hybrid forecasting model that combined the variational mode decomposition (VMD) and spiking neural networks (SNNs) (Zhang, Zhang, Xiong, & Su, 2017). The deep neural network models combining CNN with LSTM have been widely used in different disciplines, including energy consumption prediction, stock price forecasting, air quality forecasting and waterworks operation prediction (Cao et al., 2018; Huang & Kuo, 2018; Kim & Cho, 2019; Kim & Kim, 2019). However, tourism forecasting using hybrid models remains under discussion, Silva, Hassani, Heravi, and Huang (2019) is one of few studies focusing on the hybrid models.

Against this background, we introduced a new hybrid method in the field of tourism, more specifically we proposed a linear method combined with the nonlinear methods (SARIMA-LSTM-CNN) for tourist forecasts. The hybrid approach is proposed in this paper as it is widely used in different research to form a robust method which results in better forecasts. The main reason for better forecasts is that the mixed pattern is complex to deal with since nonlinear patterns cannot be dealt with using a linear model, and nonlinear patterns cannot be handled with the nonlinear model. Neither the linear model (e.g., ARIMA or SARIMA) nor nonlinear model (such as neutral network or LSTM) can sufficiently model and predict the tourist arrival. A novel hybrid approach that combines different linear and nonlinear models for better forecast results is proposed. In our model, SARIMA was used to capture the trend and seasonal components in tourist arrivals; then, high-level denoising features were fed into LSTM-CNN to examine the performance of the proposed model. We have conducted an extensive empirical study to compare deep learning techniques in large with the traditional SARIMA model.

To summarize, this paper contributes to the extant literature by applying deep learning in tourism for large daily time-series datasets. We deploy CNN and LSTM on a large dataset to assess its performance in tourism arrival prediction tasks on a region. We propose SARIMA-LSTM-CNN which are expected to capture trend, seasonal and random components.

The remainder of this paper is organized as follows. The following section explained the proposed SARIMA-LSTM-CNN method and presented the results of the experiment. Finally, the summary of the study, discussion, and implications are presented in the last section.

## 2. Methodology and data

We follow the classical linear and nonlinear data feature fusion strategies in the literature to separate the time series into different main components for further modeling. The main data features in the tourist demand data such as trend, seasonality, autocorrelation, cyclical, irregular behaviors can be broadly classified into linearly deterministic component and nonlinear stochastic component. Many methods have been developed, such as Wold decomposition for the stationary data in the econometric theory. The linear and nonlinear fusion modeling approach has gained significant attention in the tourist and economic forecasting literature. Although ARIMA and SARIMA model has been recognized as one of the most widely used models for the linear data component in the tourist daily tourism demand forecasting, there is not much consensus on the optimal nonlinear modeling methodology. It poses a difficult research problem in the field. CNN-LSTM model is one of the new developments in the deep learning research field that could potentially contribute to the nonlinear modeling in the general feature level fusion modeling framework. Work in this paper provided the new empirical evidence on the effectiveness and the merit of the linear and nonlinear data feature fusion modeling framework, with the proposition of the new proposed SARIMA–CNN–LSTM models.

The proposed hybrid SARIMA–CNN–LSTM model consisted of the linear SARIMA model and nonlinear hybrid CNN-LSTM model. In the hybrid forecasting model, the linear SARIMA model is used to capture the linear features of the time series dataset while the nonlinear hybrid CNN-LSTM model is proposed to extract the nonlinear spatio-temporal features of the time series. The forecasting process of the hybrid SARIMA–CNN–LSTM model can be divided into three stages. The general procedure is illustrated in Fig. 1 and the detailed data processing steps in CNN-LSTM model is illustrated in Fig. 2.

In the first phase, the linear SARIMA model is used to model the linear trend and seasonal component of the time series. The linear SARIMA model can only extract the linear data feature in the time series, and the nonlinear patterns are retained in the residuals of SARIMA model, the hybrid CNN-LSTM model is applied to recognize the spatial

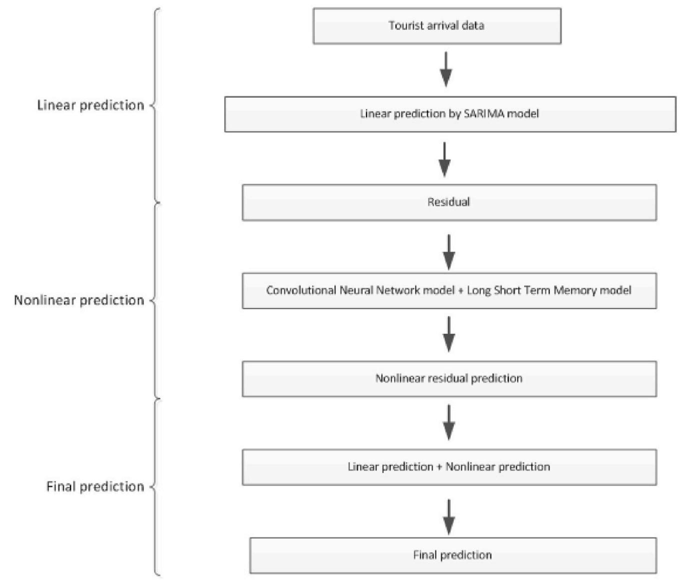


**Fig. 1.** Flow chart for SARIMA–CNN–LSTM model.

relationship and temporal dependencies in the time series, and then utilize these specific nonlinear data patterns to model the residuals sequence of the SARIMA model.

In the linear forecasting process, the rolling windows of SARIMA model is set as *m*, which are less than the length of raw time series. Each time the rolling window of the SARIMA model scroll forward one step to fit the observations and make predictions for the raw time series, until obtaining all predicted values, the difference between observed values and predicted values are used to obtain the residuals sequence of the SARIMA model. The calculation formula of the residuals is as follows:

$$y_R = y_O - y_S$$

Where $y_O$represents the real value corresponding to the predicted value, $y_S$represents the predicted value of the SARIMA model, and$y_R$represents the residuals of the SARIMA model.

In the second phase, the residuals sequence of the SARIMA model is modeled using CNN-the LSTM model. The CNN-LSTM model consists of convolutional neural network (CNN) and long short-term memory network (LSTM), as illustrated in Fig. 2.

The convolutional neural network is a popular deep learning model, due to the fact that it is successfully used in the field of image classification research and time series forecasting research. The convolutional neural network has the ability to capture the nonlinear spatial interaction relationship features in the observations (LeCun et al., 1998). There are six different components in the CNN model, which are the input layer, convolutional layer, pooling layer, flatten layer, fully connected layer and output layer (LeCun et al., 1998). In the convolutional layer, the filters (or kernels) capture nonlinear spatial features from the fixed-length sub-sample of the overall time series by making convolution operations with the local observations, which can derive the nonlinear spatial location relationship between the adjacent observed values within the fixed-length sub-sample. Then, these extracted nonlinear spatial features (also named feature maps) will be fed into the next layer called the pooling layer. The role of the pooling layer is that it is used to reduce the complexity of the feature maps generated by the previous convolutional layer and draw the most important feature information from these feature maps by a pooling operation. In the pooling layer, there are different pooling functions to deal with the input feature maps such as the max pooling function and average poling function. Following the pooling layer is the flatten layer, which is used to transform the multi-dimensional feature maps received from the previous pooling layer into a one-dimension array to meet the data processing

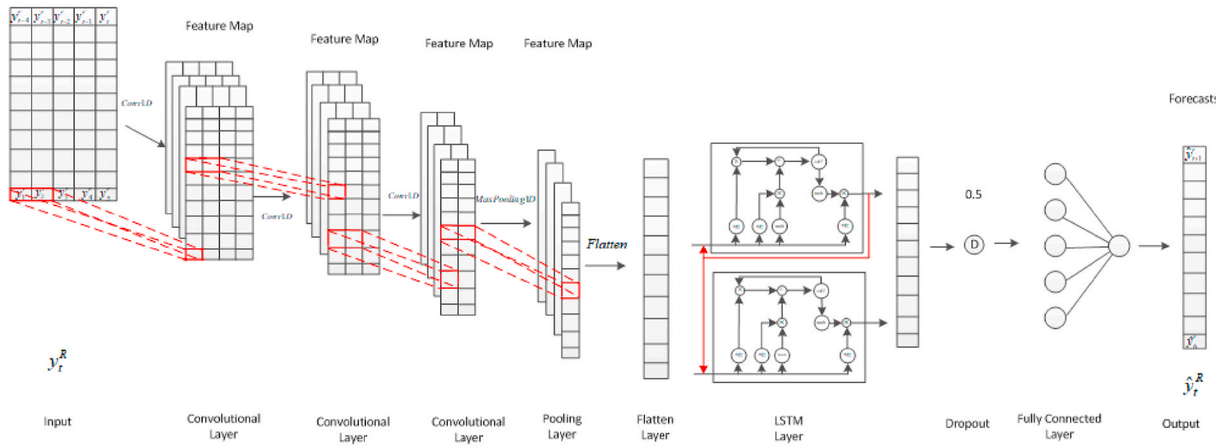


**Fig. 2.** Flow diagram for CNN-LSTM model.

requirements of the next full connected layer. Lastly, the full connected layer makes predictions of the future values according to these extracted spatial correlation patterns (Zou et al., 2020).

The long short-term memory (LSTM) network is a popular artificial intelligence neural network presented by Hochreiter and Schmodedhuber in 1997 (Hochreiter & Schmidhuber, 1997), which is a special improved recurrent neural network model. In recent years, it is widely used in the field of time series forecasting research due to its ability to model nonlinear long-term temporal dependencies patterns in time series. In order to solve the gradient vanishing and gradient explosion problems of the RNN model in long-term prediction research, the LSTM model is introduced to improve the modeling ability of the RNN model (Law et al., 2019). Unlike the traditional RNN model, the hidden layer of the LSTM model consists of a memory block, which has a two-layer loop structure. The two-layer loop structure of memory block in the LSTM hidden layer control the memorization and process of the long-term temporal features in the time series. The memory block contains three adaptive multiplicative gates and one or more memory cells; these units jointly mange the input and flow of information and the output and status of the memory block (Chang, Zhang, & Chen, 2019). The memory cell is a self-recurrent unit that is a main component in the memory block, which is also an important part of the internal circulation of the memory block. Memory cell recurrently processes the historical observations and update the cell state when the new information is fed into the memory block. The memory cell is then capable of storing the temporal data features (Cao, Li, & Li, 2019). The three types of adaptive multiplicative gates are respectively named the input gate, the forget gate and the output gate. The role of input gate is that it is applied to determine which new information can be fed into the memory cell to update the cell status. The forget gate is used to decide which information in the memory cell should be given up and which information can be retained to update the status of cell units along with new input information. The function of the output gate is that it is used to filter out the information in the memory cell and output the most desired information to make prediction for the future values.

In the nonlinear prediction of the hybrid CNN-LSTM model, the residuals sequence of the SARIMA model is divided into a training set and testing set, which are fed into the CNN-LSTM model. The training set is used to train the parameters of the hybrid CNN-LSTM model while the testing set is used to verify the predictive performance of the trained model in the out-sample data. The convolutional layer of the hybrid CNN-LSTM model recognizes the spatial relationship structures between several residual data at different adjacent time points by sliding the filters over the residual data. The LSTM layer of hybrid CNN-LSTM model is used to model the long-term temporal relationship between several feature maps extracted by the convolutional layer. At last, the fully connected layer of the hybrid CNN-LSTM model make predictions for residuals sequence according to the extracted spatio-temporal correlation features.

Finally, the predicted values of the hybrid SARIMA–CNN–LSTM model is calculated as the sum of the predicted values of SARIMA model and hybrid CNN-LSTM model. The calculation equation of the final predicted value of the hybrid SARIMA–CNN–LSTM model is presented as follows:

$$y_F = y_S + y_D$$

Where $y_F$ represents the final predicted value of hybrid SARIMA–CNN–LSTM model, $y_S$and $y_D$ respectively denote the linear predicted value of the SARIMA model and nonlinear residuals predicted value of the hybrid CNN-LSTM model.

To evaluate the performance of the proposed model, we have used the real tourist arrival data to conduct the empirical study. In this paper, we have chosen Macao, a Chinese special administration region as the research subject to conduct the empirical studies. Macao is one of the most well-known tourist attractions in the world, with active tourist inflows from around the world daily. It is representative of the tourist destinations in the world. In this paper, we have collected tourist arrival data from 6 major countries and regions that Macao has active tourist inflow from January 1, 2017 to October 31, 2019. The constructed data set, as a result, include 1033 observations for each origin.

To facilitate the empirical studies, we have divided the dataset into three parts, i.e. training set, validation set and test set. The training set is used to estimate the model parameters for the seasonal models. The validation data set is used to estimate the deep learning model specification and the parameters. The test set is used to evaluate the performance of the proposed SARIMA–CNN–LSTM model against the benchmark models, including SARIMA, CNN and LSTM models. The test set includes 204 observations (i.e. 20% of the total dataset).

We have conducted the exhaustive search for the optimal deep learning model specification and parameters using the grid search method and the validation data set based on the forecasting performance measured by the Mean Squared Error (MSE) for the candidate model specification and the parameters. The one with the minimum mean squared area is selected as the optimal model specification of parameters.

In the meantime, the tradeoff between accuracy and time complexity may not be the major concern compared to the model accuracy as the difference between the computational time for each individual model is marginal and insignificant, after the model specification has been determined in the model training period. For example, for China dataset, a single one-step-ahead prediction takes 0.5407 s for SARIMA and 1.3294 s for SARIMA–CNN–LSTM model. Provided that the difference in forecasting accuracy is statistically significant and the accuracy improvement is attributed to the more accurate modeling of nonlinear components using the Deep Learning (DL) model, the tradeoff between accuracy and time complexity in the hybrid SARIMA–CNN–LSTM model is justifiable.

## 3. Results

We firstly plotted the original daily tourist arrivals in Fig. 3 and calculated the descriptive statistics in Table 1. Fig. 3 shows the original observations of daily tourist arrivals in 6 countries and regions.

It can be seen from Fig. 3 that the daily tourist arrivals demonstrate significant volatility. There are seasonal and cyclical patterns, as well as random fluctuations due to nonlinear dynamics. Table 1 lists the descriptive statistics such as four moments for the daily tourist arrivals data in 6 countries and regions.

The descriptive statistics in Table 1 show significant kurtosis values for all 6 countries or regions. Among them, Hong Kong has the largest kurtosis value while Singapore has the smallest one. There is significant volatility as measured by standard deviation values. The distributions of tourist arrivals in 6 regions clearly deviate from the standard normal distribution.

We then consider individual models such as SARIMA, CNN and LSTM. The network structure for LSTM is set as follows: LSTM layer - Dropout layer - Dense layer. The network structure for CNN is set as follows: Conv layer - Pooling layer - Dense layer - Dropout layer - Dense layer. The network structure for SARIMA–CNN–LSTM is set as follows: Conv layer - Flatten layer - Pooling layer - LSTM layer - Dropout layer - Dense layer - Dropout layer - Dense layer. CNN and LSTM models may have different network structures and parameters. For example, the CNN model may have a different number of layers. There are also several hyper-parameters, such as the number of filters in the CNN model, the number of hidden neurons in the LSTM model, the learning rate, and the training epochs.

Based on the minimization of the Akaike information criteria, the SARIMA(1,0,0)x(7,7) model is selected, where the autoregression lag term is 1, seasonal integration is 7 and the seasonal moving averaging lag term is 7.

For CNN, LSTM, and SARMA–CNN–LSTM models, we used the Greedy Search method to search exhaustively for the optimal network structure that would produce the lowest level of in-sample forecast errors, measured using the Mean Squared Error (MSE). A range of values for the deep learning model parameters are used to construct the candidate deep learning models during the search process. The lower and the upper bound for these values are listed in the parenthesis as follows. CNN layers: (1,3), Neurons in CNN layers (1,10), neurons in LSTM layers (1,20). The kernel size is set to $4 \times 4$, $3 \times 3$, $2 \times 2$ for the one layer, two layers, and three-layer structure for both CNN and SARIMA–CNN–LSTM models. The dropout ratio is set to 0.5. The activation function is relu. The learning rate is 0.001. The training algorithm is Adam. The pooling layer size is set to 2. It is worth noting that the optimal network structure for the deep learning model changes for different datasets. For example, the optimal model network structure contains 1 CNN layers and 1 LSTM layers for China dataset while the optimal model network structure contains 2 CNN layers and 1 LSTM layers for Korea dataset.

The number of learnable parameters is large, which is usually in terms of thousands. This is due to the CNN-LSTM model and layers involved. For example, for the CNN-LSTM model with 2 CNN layers and 1 LSTM layers for Korea dataset, the number of learnable parameters in CNN-LSTM model is 2098. The number of learnable parameters is calculated as the sum of the number of parameters in different layers.

As for the overfitting issue, the traditional neural network model is known to overfit the data when its exponentially increasing number of parameters quickly exceeds the number of observations in the data set of relatively small size. This would result in a lower level of

**Table 1**
Descriptive statistics for daily tourist arrivals to Macao.

| | Mean | Standard Deviation | Skewness | Kurtosis |
|---|---|---|---|---|
| China | 68977.5286 | 18175.1731 | 1.4110 | 6.9042 |
| Hong Kong | 18026.0784 | 5090.9640 | 1.7774 | 8.7747 |
| Taiwan | 2925.1723 | 600.3478 | 0.7770 | 5.8203 |
| Philippines | 930.5024 | 386.6689 | 1.6881 | 7.4589 |
| Korea | 2265.2459 | 729.5240 | 0.8791 | 4.1566 |
| Singapore | 357.7483 | 146.3025 | 0.7427 | 3.4064 |

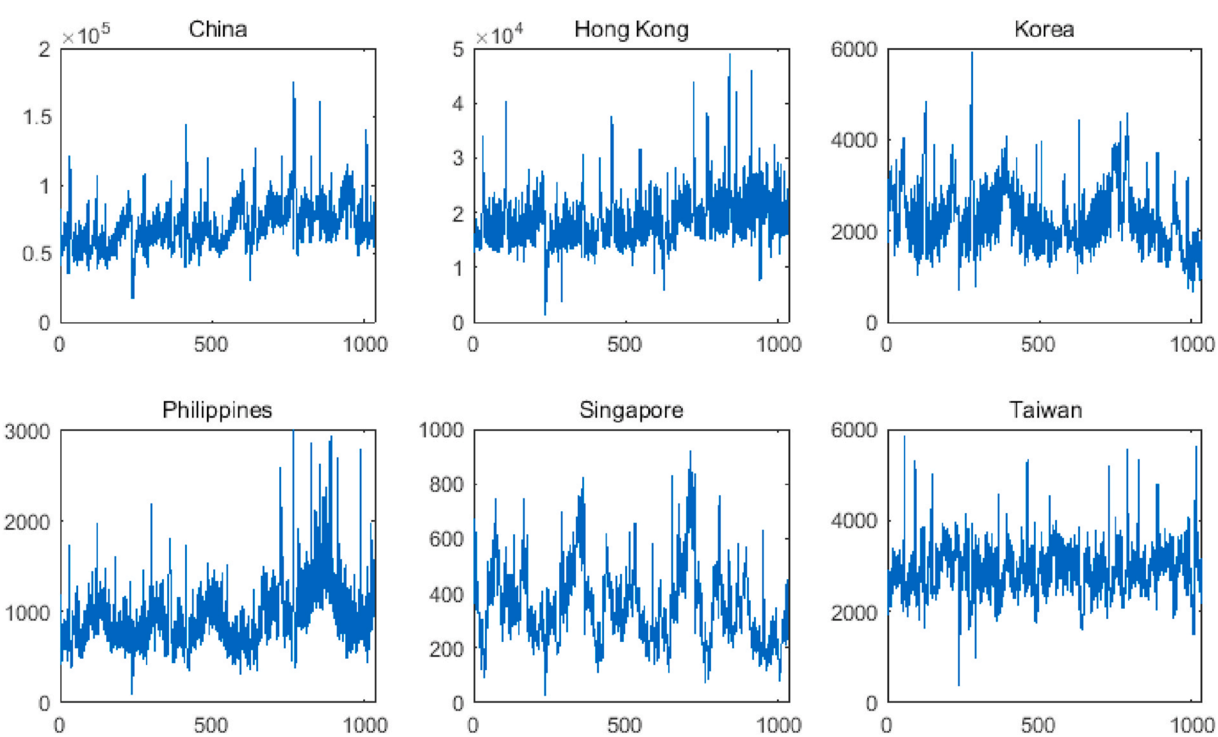


**Fig. 3.** Time series plots of daily tourist arrivals.

generalizability when the model is applied to forecast the out-of-sample data. This is less of an issue for the deep learning model. Recent theoretical studies suggest that the deep learning models have better generalizability with a lower number of parameters to estimate compared to the new network model (Zhang, Bengio, Hardt, Recht, & Vinyals, 2017). The increasing level of network depth and the increasing number of neurons do not significantly affect the forecasting accuracy in the out-of-sample forecast. Table 2 summarizes and compares the results as generated by the LSTM model, CNN model, SARIMA model.

Notably, the performance of CNN and LSTM (both are deep learning methods) in Table 2 are unanimously worse than pure linear SARIMA in all 6 countries and regions. RMSE and MAPE for CNN and LSTM models are higher than that of SARIMA model. As highlighted by Hassani, Silva, Antonakakis, Filis, and Gupta (2017), relying on the RMSE alone for determining the best forecasting model is not statistically efficient. Therefore, the modified Diebold-Mariano (DM) test (Harvey, Leybourne, & Newbold, 1997) and the Kolmogorov-Smirnov Predictive Accuracy (KSPA) test (Hassani & Silva, 2015) have been applied for comparing the predictive accuracy between SARIMA, CNN and LSTM. Test results using DM tests show that their forecasting performance are different, which align with the result of RMSE and MAPE. Overall, the null hypothesis for DM test is rejected at 0.1 cutoff value. This indicates that the performance gap is statistically significant. The null hypothesis for KSPA test is rejected except for three cases (LSTM in China, Korea and Singapore). This indicates that the error distribution from both SARIMA and CNN, LSTM is statistically different, which provides statistical justifications to the superior forecasting performance of SARIMA model against CNN and LSTM models.

Therefore, we propose SARIMA–CNN–LSTM model. Table 3 summarizes and compares the results as generated by the SARIMA model and SARIMA–CNN–LSTM model.

Notably, the RMSE and MAPE from the SARIMA–CNN–LSTM model in Table 3 are lower than those of SARIMA model. These are the lowest results compared with those of other benchmarking models such as CNN and LSTM models. To statistically evaluate the performance of SARIMA–CNN–LSTM model and SARIMA model, Clark West test of predictive accuracy has also been introduced and used (Clark & West, 2007). The reason is as follows. Compare SARIMA with SARIMA–CNN–LSTM models, SARIMA can be viewed as the parsimonious model, nested within the larger and more complex SARIMA–CNN–LSTM model, which is constructed based on SARIMA model. In this case, DM and KSPA tests are biased while CW test provides adjusted test statistics, taking into account the noises introduced into forecasts in the larger model. As suggested by CW test, the superior performance achieved by SARIMA–CNN–LSTM model is statistically significant. The null hypothesis for CW test is rejected as p value is less than 0.1 cutoff value.

**Table 2**
Predictive accuracy for SARIMA, CNN and LSTM model.

| | SARIMA | CNN | LSTM |
|---|---|---|---|
| China | | | |
| RMSE | 11546.2366 | 15386.0376 | 13639.5221 |
| MAPE | 0.0999 | 0.1355 | 0.1195 |
| $P_{DM}$ | N/A | 0.0936 | 0.0054 |
| $P_{KSPA}$ | N/A | 0.0003 | 0.967 |
| HK | | | |
| RMSE | 3768.0701 | 5205.9553 | 6612.45548 |
| MAPE | 0.1087 | 0.1611 | 0.1970 |
| $P_{DM}$ | N/A | 0 | 0 |
| $P_{KSPA}$ | N/A | 0.0002 | 0 |
| Taiwan | | | |
| RMSE | 387.9463 | 482.3959 | 520.0313 |
| MAPE | 0.0972 | 0.1172 | 0.1217 |
| $P_{DM}$ | N/A | 0 | 0 |
| $P_{KSPA}$ | N/A | 0.0001 | 0 |
| Philippines | | | |
| RMSE | 311.7108 | 417.4435 | 404.4726 |
| MAPE | 0.1766 | 0.2537 | 0.2417 |
| $P_{DM}$ | N/A | 0 | 0 |
| $P_{KSPA}$ | N/A | 0.0012 | 0.0429 |
| Korea | | | |
| RMSE | 306.0057 | 373.8844 | 381.7019 |
| MAPE | 0.1311 | 0.1611 | 0.1723 |
| $P_{DM}$ | N/A | 0 | 0 |
| $P_{KSPA}$ | N/A | 0.0049 | 0.4051 |
| Singapore | | | |
| RMSE | 62.8027 | 75.2052 | 68.9796 |
| MAPE | 0.1896 | 0.1949 | 0.2040 |
| $P_{DM}$ | N/A | 0.0182 | 0.06 |
| $P_{KSPA}$ | N/A | 0.0069 | 0.0728 |

**Table 3**
Predictive accuracy for SARIMA and SARIMA–CNN–LSTM model.

| | SARIMA | SARIMA–CNN–LSTM |
|---|---|---|
| China | | |
| RMSE | 11546.2366 | 11247.4200 |
| MAPE | 0.0999 | 0.0986 |
| $P_{CW}$ | N/A | 0.0424 |
| HK | | |
| RMSE | 3768.0701 | 3724.0853 |
| MAPE | 0.1087 | 0.1087 |
| $P_{CW}$ | N/A | 0.0633 |
| Taiwan | | |
| RMSE | 387.9463 | 375.9924 |
| MAPE | 0.0972 | 0.0964 |
| $P_{CW}$ | N/A | 0.0207 |
| Philippines | | |
| RMSE | 311.7108 | 309.1444 |
| MAPE | 0.1766 | 0.1742 |
| $P_{CW}$ | N/A | 0.0897 |
| Korea | | |
| RMSE | 306.0057 | 301.0451 |
| MAPE | 0.1311 | 0.1280 |
| $P_{CW}$ | N/A | 0.0231 |
| Singapore | | |
| RMSE | 62.8027 | 62.5311 |
| MAPE | 0.1896 | 0.1866 |
| $P_{CW}$ | N/A | 0.0897 |

This research study well demonstrates the accuracy of the SARIMA–CNN–LSTM model applied, as they can minimize the error levels. According to the comparisons and results above, the SARIMA–CNN–LSTM model can improve the rate of error and enhance the accuracy rate. Therefore, the model is proposed and recommended in this study. In particular, SARIMA–CNN–LSTM shows a lower error than any single model and has the highest prediction performance.

We have also plotted the forecasts from the proposed model against the actual values for the tourist arrival data in each origin in Fig. 4.

Since the main difference between the proposed SARIMA–CNN–LSTM model and the benchmark SARIMA model is the introduction of the deep learning model to capture the nonlinear dynamics in the residual, we may argue that the modeling of nonlinear dynamics in the residual result in the forecasts when the predictive accuracy improves. This is another desired feature for the tourist arrival forecasts.

The proposed hybrid model has achieved performance improvement in terms of lower levels of MSE for all 6 countries or regions. The performance improvement is statistically significant for all 6 countries or regions, as indicated by the P-values of a set of statistical tests, including the Clark West test and Diebold Mariano test. Therefore, performance improvement of the proposed SARIMA–CNN–LSTM model is both statistically significant and robust (i.e., consistently better). Notably, the performance of the deep learning model is very sensitive to the size of the dataset and the computational power, as is evidenced in empirical applications in diverse disciplines. Given the constrained data set size and computational power in this paper, the proposed model has good potential for further performance improvement when more data are available at a higher frequency, and more powerful computational resources are available. As the data become available at a higher frequency, the impact of transient and nonlinear data features would

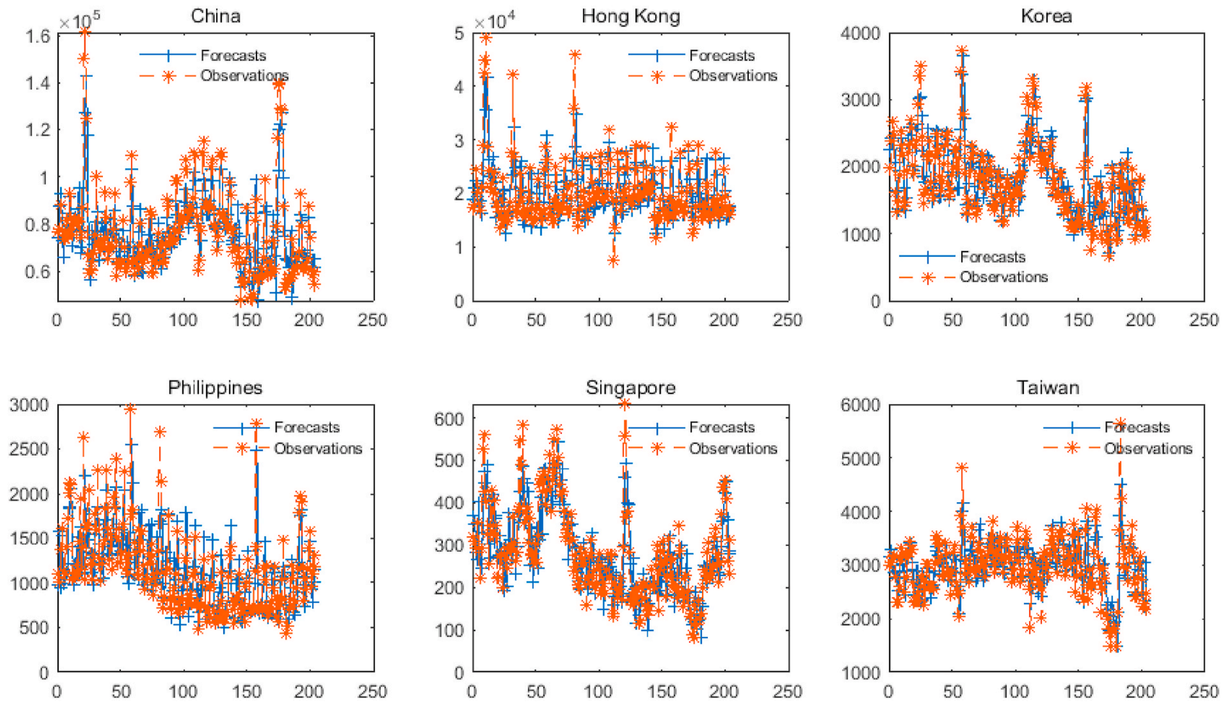


**Fig. 4.** Forecasts from SARIMA–CNN–LSTM model.

become larger, which require powerful nonlinear models such as deep learning models.

There are many tourist demand forecasting models developed over the years, such as the econometric or time-series approach or the artificial intelligence-based approach. Our model is constructed based on the SARIMA model and deep learning models, which are two sub-models for the proposed hybrid model, which takes completely different approaches. SARIMA model is the most widely used econometric model, while CNN and LSTM models are the most recent development in the AI-based approach. Comparing our model with typical models in these two approaches can be conducted in different dimensions such as accuracy, interpretability, and application.

In terms of accuracy, our model demonstrates more superior performance than SARIMA, CNN, and LSTM models. Since these three models are the most widely accepted models in their individual modeling approach, our model provides state-of-the-art performance in modeling the mixture of linear and nonlinear data features.

For modeling the linear data feature, the SARIMA model provides the best modeling accuracy, as indicated by the literature review. In our empirical results, the SARIMA model also provides superior performance than other models. There are also multivariate regression types models that consider the exogenous influencing factor for the tourist arrival data. These models are parametric in nature that usually relied on strict assumptions for the residual distribution. Empirical data of higher frequency and nonlinear in nature may violate the stationarity and normality assumptions of the model. They are usually applied to tourist data of lower frequency. For modeling the nonlinear data feature, the Deep Learning model (DL) is known to provide better performance than the traditional neural network model in AI literature. Some recent studies have provided promising evidence in the tourism field. The main difficulty with the applications of the DL model in the tourism field is the widely concerned interpretability issue due to its black-box nature. However, the interpretability issue does not prevent DL to be one of the most useful tool to model the nonlinear part, where interpretability is less of an issue in the traditional time series modeling framework.

In terms of interpretability, our model provides better interpretability than AI models alone. The main linear data component is modeled using the SARIMA model, which captures long-term main trends in the data, which can be interpreted in the traditional time series modeling framework. The subtle nonlinear details in the residual are modeled using the CNN-LSTM model, which contributes mainly to the modeling of unexplainable nonlinear dynamics such as the transient and shorter-term influencing factors.

In terms of applications, our model is more suitable for modeling complex data characteristics as the data size grows exponentially in the era of big data. Although the traditional SARIMA model performs well over the longer time horizon when data are at a lower frequency and transient data features are smoothed out, its forecasting accuracy decreases over the shorter time horizon as those transient and nonlinear data features dominate at higher data frequency. The complexity of the hybrid model is designed to capture those subtle details in the nonstationary nonlinear data movement at a higher frequency. As the data moves to higher intradaily frequency in the big data research, our model would offer more flexibility and higher accuracy in modeling the nonlinear and nonstationary data features prevalent in the high-frequency data.

## 4. Discussion and conclusion

In this paper, deep learning methods including LSTM and CNN have been tested for forecasting Macao daily tourist arrivals. The results show that the performance of CNN or LSTM is worse than the traditional SARIMA model. A new model has been proposed based on SARIMA, LSTM and CNN. SARIMA is used to extract the linear component of tourist arrivals, CNN and LSTM are used to capture the nonlinear component of the tourist arrival. Most notably, the Convolutional Neural Network (CNN) is used to capture the hierarchical data structure, while the Long Short Term Memory network (LSTM) is used to capture the long-term dependencies in the data.

Our result provides future evidence for the use of deep learning and hybrid models which has limited discussion in tourist forecasting. Our results show that neither the linear model (SARIMA) nor nonlinear model (CNN or LSTM) can sufficiently model and predict the tourist arrivals. A novel hybrid approach SARIMA-LSTM-CNN that combines different linear and nonlinear models provides better forecast results using Macao daily data. Our findings will intensify interest in using new hybrid models for univariate tourist forecasting and stimulate subsequent research.

This study has several limitations. The first limitation is that we are only limited to one-step-ahead forecasts. The deep learning models in the literature have mostly been applied to short term forecasting, especially with one step ahead forecasting in tourism and economics field (Lago, Ridder, & Schutter, 2018; Shehhi & Karathanasopoulos, 2020; Wu, Ji, He, & Tso, 2021). Deep learning models are used to capture the short-term nonlinear dynamics that the mainstream econometrics and time series models have difficulty in dealing with. The current deep learning methodology has advantages of finding the optimal hyper parameters for the deep learning model that can better capture the nonlinear dynamics in data. But when it comes to the multiple steps ahead long-term forecasts, where the noises and nonlinear dynamics are smoothed out, the econometric and time series models prevail as their model specification and parameters fit the data better when specific assumptions are made about the data behavior. Although the deep learning model can also be applied in this case, it would require the re-design of the deep learning network structure and careful calibration of the hyper parameters for the deep learning model, which remains one of the most difficult problems in the area of deep learning research. The second limitation is that only Macao data are used. Research based on longer time series data and data from other destinations are possible directions for future research.

## References


Assaf, A. G., Li, G., Song, H., & Tsionas, M. G. (2019). Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model. *Journal of Travel Research, 58*(3), Article 0047287518759226. https://doi.org/10.1177/0047287518759226.

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting, 27*(3), 822–844. https://doi.org/10.1016/j.ijforecast.2010.04.009.

Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS One, 12*(7), Article e0180944. https://doi.org/10.1371/journal.pone.0180944.

Cao, K., Kim, H., Hwang, C., & Jung, H. (2018). CNN-LSTM coupled model for prediction of waterworks operation data. *Journal of Information Processing Systems, 14*(6).

Cao, J., Li, Z., & Li, J. (2019). Financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and Its Applications, 519*, 127–139.

Chang, Z., Zhang, Y., & Chen, W. (2019). Electricity price prediction based on hybrid model of adam optimized LSTM neural network and wavelet transform. *Energy, 187*, 115804.

Chen, Y., He, K., & Tso, G. K. F. (2017). Forecasting crude oil prices: A deep learning based model. *Procedia Computer Science, 122*, 300–307. https://doi.org/10.1016/j.procs.2017.11.373.

Chen, K. Y., & Wang, C. H. (2007). Support vector regression with genetic algorithms in forecasting tourism demand. *Tourism Management, 28*(1), 215–226. https://doi.org/10.1016/j.tourman.2005.12.018.

Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management, 24*(3), 323–330. https://doi.org/10.1016/S0261-5177(02)00068-7.

Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics, 138*(1), 291–311. https://doi.org/10.1016/j.jeconom.2006.05.023.

Díaz, M.á., & Mateu-Sbert, J. (2011). Forecasting daily air arrivals in Mallorca island using nearest neighbour methods. *Tourism Economics, 17*(1), 191–208. https://doi.org/10.5367/te.2011.0022.

Divino, J. A., & McAleer, M. (2010). Modelling and forecasting daily international mass tourism to Peru. *Tourism Management, 31*(6), 846–854. https://doi.org/10.1016/j.tourman.2009.09.002.

Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research, 270*(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting, 13*(2), 281–291.

Hassani, H., & Silva, E. S. (2015). A Kolmogorov-Smirnov based test for comparing the predictive accuracy of two sets of forecasts. *Econometrics, 3*(3), 590–609.

Hassani, H., Silva, E. S., Antonakakis, N., Filis, G., & Gupta, R. (2017). Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research, 63*, 112–127.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Huang, C. J., & Kuo, P. H. (2018). A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. *Sensors, 18*(7). https://doi.org/10.3390/s18072220.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water, 10*(11), 1543.

Jiao, E. X., & Chen, J. L. (2019). Tourism forecasting: A review of methodological developments over the last decade. *Tourism Economics, 25*(3), 469–492. https://doi.org/10.1177/1354816618812588.

Kim, T.-Y., & Cho, S.-B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy, 182*, 72–81. https://doi.org/10.1016/j.energy.2019.05.230.

Kim, T., & Kim, H. Y. (2019). Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. *PloS One, 14*(2), Article e0212320. https://doi.org/10.1371/journal.pone.0212320.

Kon, S. C., & Turner, L. W. (2005). Neural network forecasting of tourism demand. *Tourism Economics, 11*(3), 301–328. https://doi.org/10.5367/000000005774353006.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks. *Hydrology and Earth System Sciences, 22*(11), 6005–6022. https://doi.org/10.5194/hess-22-6005-2018.

Kulendran, N., & Wilson, K. (2000). Modelling business travel. *Tourism Economics, 6*(1), 47–59. https://doi.org/10.5367/000000000101297460.

Lago, J., Ridder, F. D., & Schutter, B. D. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy, 221*, 386–405. https://doi.org/10.1016/j.apenergy.2018.02.069.

Law, R., Li, G., Fong, D. K. C., & Han, X. (2019). Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research, 75*, 410–423. https://doi.org/10.1016/j.annals.2019.01.014.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324. https://doi.org/10.1109/5.726791.

Lu, W., Rui, H., Liang, C., Jiang, L., Zhao, S., & Li, K. (2020). A method based on GA-CNN-LSTM for daily tourist flow prediction at scenic spots. *Entropy, 22*(3). https://doi.org/10.3390/e22030261.

Qing, X., & Niu, Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy, 148*, 461–468. https://doi.org/10.1016/j.energy.2018.01.177.

Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing, 323*, 203–213. https://doi.org/10.1016/j.neucom.2018.09.082.

Shehhi, M. A., & Karathanasopoulos, A. (2020). Forecasting hotel room prices in selected GCC cities using deep learning. *Journal of Hospitality and Tourism Management, 42*, 40–50. https://doi.org/10.1016/j.jhtm.2019.11.003.

Silva, E. S., Hassani, H., Heravi, S., & Huang, X. (2019). Forecasting tourism demand with denoised neural networks. *Annals of Tourism Research, 74*, 134–154. https://doi.org/10.1016/j.annals.2018.11.006.

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting—a review of recent research. *Tourism Management, 29*(2), 203–220. https://doi.org/10.1016/j.tourman.2007.07.016.

Song, H., Qiu, R. T. R., & Park, J. (2019). A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research, 75*, 338–362. https://doi.org/10.1016/j.annals.2018.12.001.

Song, H., & Witt, S. F. (2006). Forecasting international tourist flows to Macau. *Tourism Management, 27*(2), 214–224. https://doi.org/10.1016/j.tourman.2004.09.004.

Song, H., Wong, K. K. F., & Chon, K. K. S. (2003). Modelling and forecasting the demand for Hong Kong tourism. *International Journal of Hospitality Management, 22*(4), 435–451. https://doi.org/10.1016/S0278-4319(03)00047-1.

Sun, G., Chen, T., Wei, Z., Sun, Y., Zang, H., & Chen, S. (2016). A carbon price forecasting model based on variational mode decomposition and spiking neural networks. *Energies, 9*(1). https://doi.org/10.3390/en9010054.

Sun, S., Wei, Y., Tsui, K.-L., & Wang, S. (2019). Forecasting tourist arrivals with machine learning and internet search index. *Tourism Management, 70*, 1–10. https://doi.org/10.1016/j.tourman.2018.07.010.

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition*.

Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). *Deep learning for identifying metastatic breast cancer*. arXiv preprint arXiv:1606.05718.

Wu, D., Ji, L., He, K. J., & Tso, K. F. (2021). Forecasting tourst daily arrivals with a hybrid SARIMA-LSTM approach. *Journal of Hospitality & Tourism Research, 45*, 52–67. https://doi.org/10.1177/1096348020934046.

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2017b). Understanding deep learning requires rethinking generalization. In *Proceedings of the 5th international conference on learning representation (ICLR 2017)*. April 25-26, 2017. TOULON: ICLR, 2017.

Zhang, B., Li, N., Shi, F., & Law, R. (2020). A deep learning approach for daily tourist flow forecasting with consumer search data. *Asia Pacific Journal of Tourism Research, 25*(3), 323–339. https://doi.org/10.1080/10941665.2019.1709876.

Zhang, L., Zhang, J., Xiong, T., & Su, C. (2017a). Interval forecasting of carbon futures prices using a novel hybrid approach with exogenous variables. *Discrete Dynamics in Nature and Society*, 1–12. https://doi.org/10.1155/2017/5730295, 2017.

Zhu, B., & Wei, Y. (2013). Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. *Omega, 41*(3), 517–524. https://doi.org/10.1016/j.omega.2012.06.005.

Zou, Y., Yu, L., Tso, G. K. F., & He, K. (2020). Risk forecasting in the crude oil market: A multiscale convolutional neural network approach. *Physica A: Statistical Mechanics and Its Applications*, 541. https://doi.org/10.1016/j.physa.2019.123360.