Annals of Tourism Research 83 (2020) 102925

Contents lists available at ScienceDirect

Annals of Tourism Research

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

RESEARCH ARTICLE

Bayesian BILSTM approach for tourism demand forecasting

Anurag Kulshrestha*, Venkataraghavan Krishnaswamy, Mayank Sharma

Indian Institute of Management Kashipur, India

ARTICLE INFO

Associate editor: Haiyan Song

Keywords:
Tourism demand forecasting
BiLSTM
LSTM
SVR
RBFNN
Bayesian optimization

ABSTRACT

The tourism sector, with its perishable nature of products, requires precise estimation of demand. To this effect, we propose a deep learning methodology, namely Bayesian Bidirectional Long Short-Term Memory (BBiLSTM) network. BiLSTM is a deep learning model, and Bayesian optimization is utilized to optimize the hyperparameters of this model. Five experiments using the tourism demand data of Singapore are conducted to ascertain the validity and benchmark the proposed BBiLSTM model. The experimental findings suggest that the BBiLSTM model outperforms other competing models like Long Short-Term Memory (LSTM) network, Support Vector Regression (SVR), Radial Basis Function Neural Network (RBFNN) and Autoregressive Distributed Lag Model (ADLM). The study contributes to tourism literature by proposing a superior deep-learning method for demand forecasting.

Introduction

The tourism sector is an engine of economic growth, contributing 10.4% to worldwide gross domestic product (GDP) and a similar percentage of jobs on the planet in 2018 (WEF, 2019). A critical activity in the tourism sector is Tourism Demand (TD) forecasting. Organizations need precise forecasts to aid tactical and operational decisions such as pricing, staff, capacity, resources, and revenue management (Frechtling & Frechtling, 2001; Jiao & Chen, 2019; Wu, Song, & Shen, 2017). Governments require precise TD predictions for destination infrastructure planning, environmental quality control, and operational flexibility (Li, Wong, Song, & Witt, 2006).

TD forecasting is studied using both quantitative and qualitative methodologies. Qualitative methodologies such as the consensus and Delphi depend upon experience, intuition, and understanding of a specific destination market and usually suffer from poor adaptability (Witt & Witt, 1995). Unlike the qualitative approach, quantitative approaches are used widely and forecast tourist arrivals (demand) using historical tourist arrivals and other factors that affect tourism volume (Wu et al., 2017). Three categories of methods, namely time-series, econometric, and Artificial intelligence (AI) models, are prominent in quantitative approaches. Time-Series and econometric models are used in tourism forecasting and demand analysis (Law, Li, Fong, & Han, 2019). This study focusses on the third category, namely, AI models, which are advanced models for TD forecasting (Jiao & Chen, 2019; Song & Li, 2008).

The objective of time-series models and econometric models is to improve TD forecasting by incorporating relevant factors (Andrew, Cranage, & Lee, 1990; Frechtling & Frechtling, 2001). This category employs autoregressive moving average (ARMA) models and its variants, autoregressive distributed lag models (ADLMs or ARDL) (Gunter & Onder, 2015; Huang, Zhang, & Ding, 2017), Bayesian models, generalized dynamic factor models, and markov-switching models (Assaf, Li, Song, & Tsionas, 2019; Guizzardi & Stacchini, 2015; Li, Pan, Law, & Huang, 2017). Among the AI-based models, artificial neural networks (ANNs) provide a potential replacement to conventional linear statistical methods in terms of forecasting (Zhang, Patuwo, & Hu, 1998). With the

* Corresponding author at: Indian Institute of Management Kashipur, Uttarakhand, India.
E-mail address: anurag.fpm1802@iimkashipur.ac.in (A. Kulshrestha).

https://doi.org/10.1016/j.annals.2020.102925
Received 2 December 2019; Received in revised form 7 April 2020; Accepted 15 April 2020
0160-7383/

advancements in ANNs, deep learning methodologies, specifically the Recurrent Neural Network (RNN) architectures, are found more appropriate than Feed Forward Neural Network in dealing with time series complexities (Husken & Stagge, 2003).

Prior research has shown that LSTM, a type of RNN, and it's variant Bidirectional LSTM (BiLSTM) networks are efficient in handling long-term dependencies (Cheng, Ding, Zhou, & Ding, 2019; Law et al., 2019). Subsequently, BiLSTMs have been utilized in various contexts such as phoneme classification (Graves & Schmidhuber, 2005), natural language processing (Wollmer, Eyben, Graves, Schuller, & Rigoll, 2010), and electrocardiogram (ECG) signal classification (Yildirim, 2018). A critical step in ANNs is hyperparameter tuning, which identifies the parameters of the model that leads to the optimal solution. Experienced human domain experts have traditionally executed the hyperparameter identification process, but their conventional knowledge is being increasingly challenged by optimization methods such as Bayesian optimization. Herein, we utilize an optimization method to identify the relevant hyperparameters of the proposed model. We draw motivation from the application of deep learning methodologies as an alternate mechanism to forecast tourist demand. We establish the superiority of deep learning technologies, specifically that of BiLSTM, along with Bayesian optimization over specific AI-Models in predicting tourism demand in this work. The contributions of this article are as follows:

1. A novel deep learning model is proposed for tourism demand forecasting.
2. Bayesian optimization is employed to optimize the hyperparameters.
3. The effectiveness of the proposed model is validated via robustness analysis with multiple experiments.
4. The effect of multi-lagged variables on model performance is studied.

The rest of the article is organized as follows. In Literature review section, we review the literature on AI-based methods in tourism demand forecasting. In Material and methods section, we describe the materials and methods used in this study. In Data section, we describe the tourism dataset employed for the study. We present the results of benchmarking in Results section, followed by validation of the BBiLSTM method using various case studies as part of robustness analysis in Robustness analysis section. Conclusion and implications of this research work are presented in the final section.

Literature review

Economic variables related to origin and destination countries, such as tourist income level and tourism price, are predominant factors in TD forecasting due to their ability to quantify tourist's spending power (Jiao & Chen, 2019). Other variables include competing destinations prices (Gunter & Onder, 2015; Li, Song, & Witt, 2005), climate change (Moore, 2010), marketing expenditures (Lim, 1997), travel agencies (Icoz, Var, & Kozak, 1998), political stability (Saha & Yap, 2014), hotel's age (Hanly & Wade, 2007), terrorist attacks (Bonham, Edmonds, & Mak, 2006), search engine data from Google and Baidu (Dergiades, Mavragani, & Pan, 2018; Law et al., 2019; Yang, Pan, Evans, & Lv, 2015).

Among demand forecasting methods, AI-based methods have certain advantages over the time series and econometric methods. Time Series and econometric models require the probability distribution of the data and specification of the chosen model beforehand (Palmer, Montano, & Sese, 2006). Non-parametric AI-based methods do not require a priori model specification and data distribution (Hansen, McDonald, & Nelson, 1999). Besides, AI-based methods can capture the nonlinear relationships among time series attributes and exogenous variables (Jiao & Chen, 2019).

AI-based methods can be broadly categorized into five subtypes: Fuzzy time series, Support Vector Machines (SVMs), Rough Sets, Grey Theory, and ANNs (Song & Li, 2008). Fuzzy time series in TD forecasting is particularly useful in modeling short time-series data (Chen, Ying, & Pan, 2010; Hadavandi, Ghanbari, Shahanaghi, & Abbasian-Naghneh, 2011; Shahrabi, Hadavandi, & Asadi, 2013). SVMs are shown to perform better than multiple regression and time series models in predicting TD (Kon & Turner, 2005; Law et al., 2019). The Rough Set approach utilizes classical set theory to deal with imprecise or uncertain data (Au & Law, 2000; Goh, Law, & Mok, 2008). Wang (2004) applied the Grey Theory in predicting TD.

Artificial neural networks are non-parametric models, inspired by the learning process of the human brain. Uysal and El Roubi (1999) showed that ANN outperformed multiple regression in TD forecasting, leading to an increased research focus on the application of ANNs in TD forecasting (refer Fig. 1). Various ANNs architectures like Multilayer Perceptron (MLP) (Claveria, Monte, & Torra, 2015), Elman NN (Claveria et al., 2015), Radial Basis Function Neural Network (RBFNN) (Claveria et al., 2015) and Denoised neural networks (Silva, Hassani, Heravi, & Huang, 2019) are used in TD analysis. RBFNN has been shown to outperform Elman NN and MLP in forecasting TD to Catalonia, Spain (Claveria et al., 2015). Except for Law et al. (2019), who applied LSTM, a type of RNN for TD forecasting, Fig. 1 shows that the application of deep learning methodologies in TD forecasting is scant. Notably, the Singapore tourism dataset has been studied with numerous approaches, including Copula-ECM, ARIMA, NN, static regression, and ARFIMA, and by various researchers (Zhu, Lim, Xie, & Wu, 2018).

The models presented above, except RNNs, take into account each data point separately and perform forecasting at each time step (Wu, Yuan, Dong, Lin, & Liu, 2018). Unlike RNNs, such models are static and ignore historical information. RNN-based Deep Learning is employed in wide-ranging areas ranging from image classification (Krizhevsky, Sutskever, & Hinton, 2012; Yildirim, Uçar, & Baloglu, 2017), electricity price prediction (Cheng et al., 2019), financial markets (Fischer & Krauss, 2018), and forecasting short-term load (He, Zhou, Feng, Liu, & Yang, 2019). RNN overcomes the shortcoming of regular ANN but suffers from training issues in the presence of long-term dependencies, which may introduce instability in performance (Bengio, Simard, & Frasconi, 1994). A type of RNN, namely, LSTM networks, overcomes the shortcomings of RNN.

LSTM with attention mechanism is used in forecasting tourist arrivals to Macau with (Law et al., 2019). Bidirectional LSTM

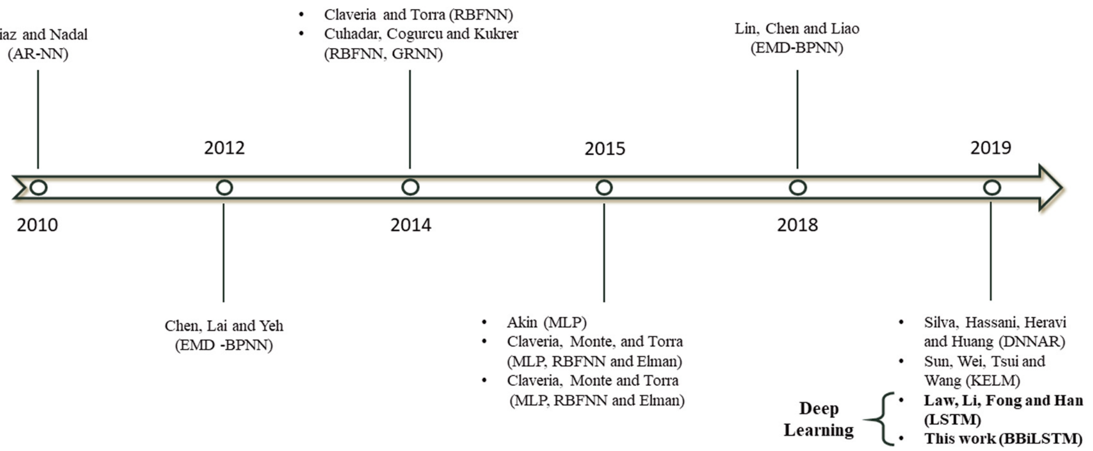

Fig. 1. Progression Neural Networks in tourism demand forecasting.
Note: AR-NN: Autoregressive neural network; ARIMA – AR integrated moving average, TFM – Transfer function model; ARIMAX - ARIMA with explanatory variables; BPNN – Backpropagation NN; GRNN – Generalized Regression NN; NNAR - NN Autoregression; DNNAR - Denoised NN Autoregression; EMD – Empirical Mode Decomposition; SVR – Support Vector Regression; LSSVR – Least Square SVR; LSTM – Long Short-Term Memory; SARIMA – Seasonal ARIMA; MLP – Multi-Layer Perceptron; RBFNN – Radial Basis Function NN; SETAR - Self-Exciting Threshold Autoregression, KELM – Kernel Extreme Learning Machine; BBiLSTM – Bayesian Bidirectional Long-short Memory.

network, a variant of LSTM, is especially apt for solving time series-based problems. BiLSTM, in contrast to LSTM, can utilize both backward and forward information. However, the research on the use of BiLSTM networks to TD prediction is scant. It is pertinent to note that research calls for the use of novel methods to predict TD (Song & Li, 2008). To this effect, this study utilizes BiLSTM along with Bayesian Optimization (BO) to predict quarterly tourist arrivals (demand) to Singapore by utilizing tourism price and income level as explanatory variables. Through this work, the study strives to improve the prediction accuracy of TD forecasts.

Material and methods

In this section, we introduce RNN followed by its advancements, namely, LSTM and Bidirectional LSTM, along with on Bayesian optimization.

Recurrent neural networks

RNN, first proposed by J.J. Hopfield, is a preferred architecture for sequential data (Hopfield, 1982). The traditional ANN incorporates an input layer, a single hidden layer, and an output layer, connected through weights. The major difference between RNN and traditional feed-forward ANN is that outputs are fed back as inputs in RNN. This connection between outputs and inputs serve as dynamic memory and make RNN more effective in learning temporal dependencies, thus making it better suited for time series data such as TD forecasting.

Long short-term memory network

RNNs are better suited than ANNs for sequential learning problems with short-term dependencies. However, they are not effective for handling problems with long-term dependencies, as the long training process may make the network unstable due to sudden large changes in training weights, also referred to as exploding gradients. LSTM, a type of RNN, introduced by Hochreiter and Schmidhuber (1997), uses the concept of memory cells to handle long-term dependencies. The distinguishing feature of LSTM networks as compared to RNN is the memory cells in the hidden layer(s). Memory cells selectively retain or forget relevant contextual information. Architecturally, each memory cell comprises an input gate ($\boldsymbol{i_t}$), forget gate ($\boldsymbol{f_t}$) and output gate ($\boldsymbol{o_t}$) for regulating the information flow as shown in Fig. 2. The input and forget gates decide the information that is to be added and discarded from the cell state, respectively. The output gate decides the parts of the cell state to be produced as output. The learning process of an LSTM is briefly described below (Fischer & Krauss, 2018).

At every time step *t*, the memory cells in the LSTM layer are updated in a multi-step process, described below, by vectorized equations with the following notation:

- The input vector at time step t is represented by $\boldsymbol{x_t}$.
- Bias vectors are indicated by $\boldsymbol{b_f}$, $\boldsymbol{b_{\tilde{c}}}$, $\boldsymbol{b_i}$, $\boldsymbol{b_o}$.
- The activation value vectors of the respective gates are represented by $\boldsymbol{f_t}$, $\boldsymbol{i_t}$, $\boldsymbol{o_t}$.
- The cell states and candidate values vectors are indicated by $\boldsymbol{c_t}$ and $\boldsymbol{\tilde{c}_t}$.

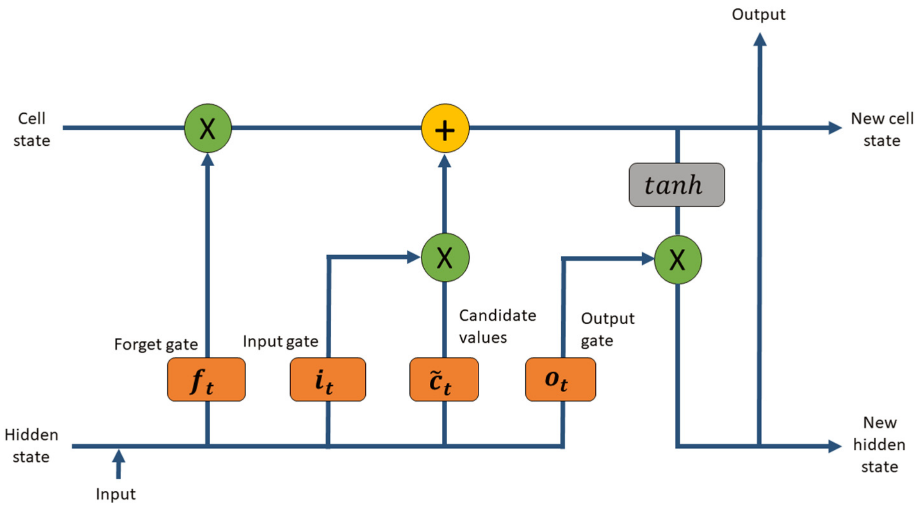

Fig. 2. Structure of LSTM network.

- $\boldsymbol{h_t}$ is the output vector of the LSTM layer.
- Weight matrices are denoted by $\boldsymbol{W_{f,\,x}}$, $\boldsymbol{W_{f,\,h}}$, $\boldsymbol{W_{\tilde{c},x}}$, $\boldsymbol{W_{\tilde{c},h}}$, $\boldsymbol{W_{i,\,x}}$, $\boldsymbol{W_{i,\,h}}$, $\boldsymbol{W_{o,\,x}}$ and $\boldsymbol{W_{o,\,h}}$.

The cell states $\boldsymbol{c_t}$ and outputs $\boldsymbol{h_t}$ at time step $\boldsymbol{t}$ during a forward pass are computed as follows:

First, the LSTM layer regulates the information that should be discarded its previous cell states $\boldsymbol{c_{t-1}}$. The current input $\boldsymbol{x_t}$, the outputs $\boldsymbol{h_{t-1}}$ of the memory cells at the previous time step $(t-1)$, and the bias terms $\boldsymbol{b_f}$ of the forget gates are utilized for computing the activation values $\boldsymbol{f_t}$ of the forgot gates at time step t through a sigmoid activation function $(\boldsymbol{\sigma})$.

$$\boldsymbol{f_t = \sigma(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f)} \tag{1}$$

In step two, the LSTM layer decides the new information that needs to be stored in the cell states $(\boldsymbol{c_t})$. This process involves two steps: (a) the activation values $(\boldsymbol{i_t})$ of the input gate at time step t are calculated and (b) new candidate values $\boldsymbol{\tilde{c}_t}$ are computed through a hyperbolic tangent function.

$$\boldsymbol{\tilde{c}_t = tanh(W_{\tilde{c},x}x_t + W_{\tilde{c},h}h_{t-1} + b_{\tilde{c}})} \tag{2}$$

$$\boldsymbol{i_t = \sigma(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i)} \tag{3}$$

In step three, we use the results in previous steps to calculate new call states $\boldsymbol{c_t}$ with $\circ$ depicting the Hadamard product:

$$\boldsymbol{c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t} \tag{4}$$

Lastly, the output h_t of memory cells of the LSTM layer is derived from the following equations:

$$\boldsymbol{o_t = \sigma(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o)} \tag{5}$$

$$\boldsymbol{h_t = o_t \circ tanh(c_t)} \tag{6}$$

Bidirectional LSTM

As described in Long short-term memory network section, while LSTM overcomes the limitations of RNN, it is still only able to process information from the past in a sequence. Whereas, BiLSTM utilizes future information as well. To achieve this, architecturally, BiLSTM contains forward and backward LSTM layers, as shown in Fig. 3. The inputs from the forward and backward layers are processed simultaneously by the output layer (Yildirim, 2018).

The output layer is updated in BiLSTM through the computation of the forward hidden sequence $\boldsymbol{\overrightarrow{h}_t}$, the backward hidden sequence $\boldsymbol{\overleftarrow{h}_t}$ and the output sequence $\boldsymbol{y_t}$ by iterating the backward layer from $t = T$ to 1 and the forward layer from $t = 1$ to T. Neural network update can be formulated as:

$$\boldsymbol{\overrightarrow{h}_t = H(W_1 x_t + W_2 \overrightarrow{h}_{t-1} + \overrightarrow{b})} \tag{7}$$

$$\boldsymbol{\overleftarrow{h}_t = H(W_3 x_t + W_5 \overleftarrow{h}_{t-1} + \overleftarrow{b})} \tag{8}$$

$$\boldsymbol{y_t = W_4 \overrightarrow{h}_t + W_6 \overleftarrow{h}_t + b_y} \tag{9}$$

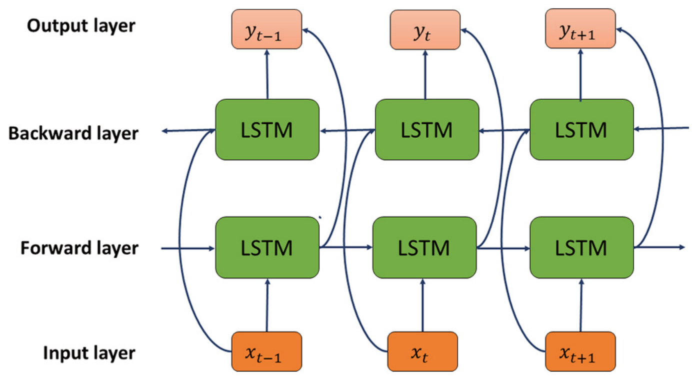

Fig. 3. Structure of BiLSTM neural network.
(Adapted from Yildirim, 2018.)

where $\overleftarrow{h_t}$, $\overrightarrow{h_t}$, y_t are the vectors for backward propagation, forward propagation and, an output layer, respectively. W's are the weight coefficients. $\overrightarrow{b}$, $\overleftarrow{b}$, b_y are the respective bias vectors (Cheng et al., 2019).

We use the adaptive moment (Adam) to optimize the learning rate of the BiLSTM network (Kingma & Ba, 2014). It has been shown to outperform SGD and RMSProp (Wu et al., 2018).

Benchmark models – LSTM, SVR, RBFNN, and ADLM

For benchmarking the BBiLSTM model, we choose LSTM, SVR, RBFNN, and Autoregressive Distributed Lag Model (ADLM) models. In the case of SVR, we optimize the parameters epsilon (ε) box constraint (C), and kernel scale (γ) (Alade, Rahman, & Saleh, 2019). RBFNN, a three-layered feed-forward neural network, employs a radial basis function (RBF), usually a Gaussian function, the spread of which should be appropriate so as not to cause overfitting or underfitting (Claveria et al., 2015). Hence, the spread of the Gaussian curve is the parameter of RBFNN that needs optimization. In the case of the ADLM model, the critical parameter is *lag*, which is selected through automatic lag selection based on the BIC criterion (Onder, Gunter, & Gindl, 2019).

Bayesian optimization

Hyperparameters are the parameters that need to be initialized before training a model. BiLSTM has several hyperparameters which control the training process and directly affect the predictive performance of the model. The hyperparameters considered for Bayesian Optimization (BO) are the number of neurons (NN), learning rate, L2 regularization, and the probability of dropout layer in the BBiLSTM model.

The first hyperparameter is *NN*. If the *NN* is too small, the model may not be able to generalize well, and significantly large *NN* may impact the computational efficacy (Guler, Ubeyli, & Guler, 2005). The second hyperparameter is the *learning rate*. The learning rate controls the rate at which a model adapts to the problem. A large learning rate can lead to a model that converges too quickly to a suboptimal solution, while a smaller learning rate can lead to a model that may never converge in a given time. The third and fourth hyperparameters are *dropout* (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) and *L2 regularization* (Lago, De Ridder, & De Schutter, 2018), which are used to prevent overfitting of the model. Dropout is the probability with which some neurons at each iteration are chosen and prevented from being trained. L2 regularization is exploited to optimize the weight parameters of the BBiLSTM model, thereby avoiding overfitting (Zhou, Chang, Chang, Kao, & Wang, 2019).

Bayesian Optimization (BO) is a method for identifying hyperparameters (Law & Shawe-Taylor, 2017). Traditionally, human domain experts with prior user experience identify the optimum set of hyperparameters. This approach often yields suboptimal results (Bergstra & Bengio, 2012). Another alternative is the grid search or random search. However, these approaches become complicated, with a large number of dimensions and are time-consuming (Cornejo-Bueno, Garrido-Merchán, Hernández-Lobato, & Salcedo-Sanz, 2018). Moreover, BO has outperformed grid search, and random search in the tuning of hyperparameters in various applications (Bergstra & Bengio, 2012; Cornejo-Bueno et al., 2018) and outperformed human domain experts in some cases (Eggensperger et al., 2013). It is capable of efficiently finding a global optimum for unknown and complex functions, thereby making it suitable for the automatic tuning process of various algorithms. BO can also obtain satisfactory results with fewer iterations in comparison with traditional optimization algorithms (He et al., 2019).

Bayesian Optimization tries to approximate the posterior distribution of the black-box function or objective function $f(X)$ to be optimized using Bayes theorem.

$$\boldsymbol{X}^{*} = \boldsymbol{arg}_{X \in U} \max \boldsymbol{f}(\boldsymbol{X}) \tag{10}$$

where X^{*} is the set of optimal hyperparameters.

Fig. 4 shows the procedure of Bayesian optimization. The objective function depicted in Fig. 4 is for visualization of BO procedure.

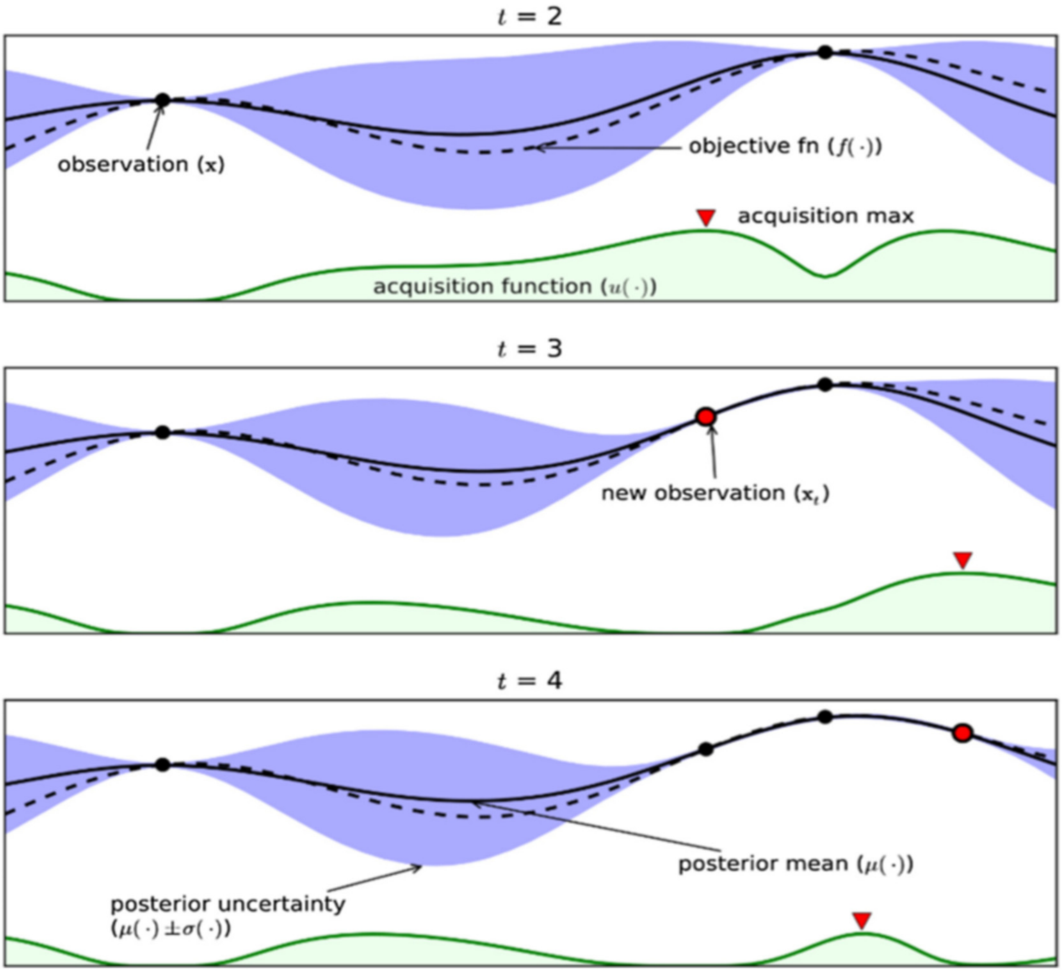

Fig. 4. Bayesian optimization procedure.
(Reproduced from Shahriari, Swersky, Wang, Adams, & De Freitas, 2016).

In the first step, the algorithm tries to fit a surrogate function (dark line) over the objective function (dotted line) by randomly selecting a few data points. In the second step, the surrogate model is updated through a Gaussian process (GP) to form the posterior distribution over the objective function as the GP model is robust, analytically traceable and accurate (Martinez-Cantin, 2017). Next, the posterior distribution is utilized to construct an acquisition function (bottom line), which is used to find out where to sample next in the hyperparameter space. The acquisition function balances the compromise between exploitation and exploration. In the exploitation phase, sampling is performed in the regions where the surrogate model most likely expects to find the global solution. Exploration means sampling in the areas where prediction ambiguity is high. Exploration and exploitation occur in the third and fourth steps, respectively.

This exploration and exploitation process and updates of the posterior GP continue till a predefined stopping criterion is reached. In this work, this criterion was set to 200 iterations. The aim is to maximize the acquisition function in finding the next sampling point. Following Cheng et al., 2019, Expected improvement (EI), a widely used acquisition function is employed in this study. EI is the improvement associated with the candidate solution over the current best solution.

In the current study, we take the first 56 data points (60%) for training the models. The next 19 data points (20%) are used as the validation set to select the optimal set of hyperparameters, and the last 19 data points (20%) that are not utilized at any step, is implemented as out-of-sample data to compare the performance of the models (He et al., 2019). The data is then transformed into the format required by supervised learning techniques. Additionally, TD is a dynamic process, as tourists decide on the choice of destination country beforehand. Behavioral patterns such as habit formation, the expectation of tourists, and the effect of "word of mouth" are captured by lagged dependent variables. Meanwhile, the dynamic effects of numerous factors that affect TD are incorporated by including lags of explanatory variables (Lim, 1997). Moreover, the inclusion of the lagged dependent variable of tourism demand has been found to be one of the critical factors that affect TD (Song & Witt, 2006). Hence, we consider the lag of dependent and explanatory variables in the proposed model. The steps adopted in this study are shown in Fig. 5.

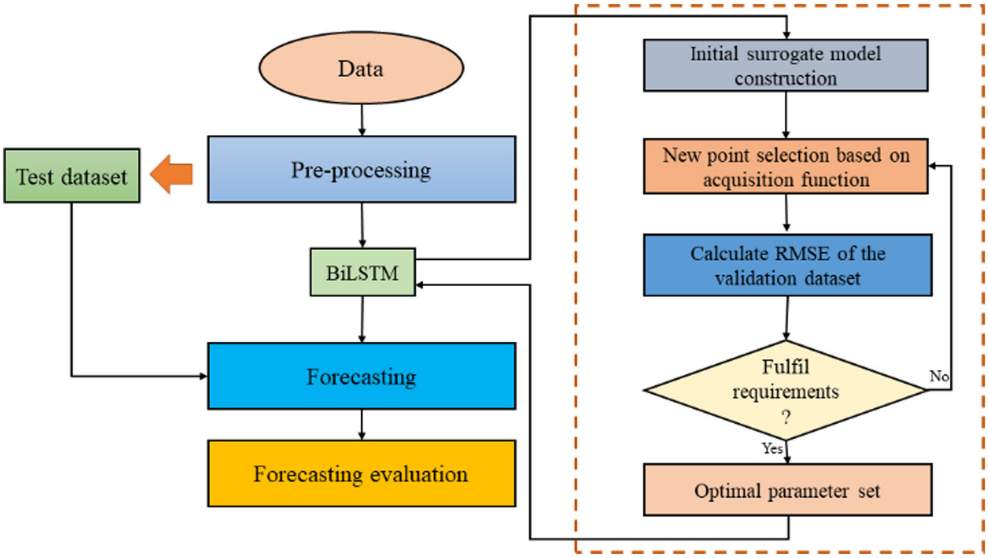

Fig. 5. Flowchart of the proposed BBiLSTM model.

Table 1
Descriptive statistics for tourist arrivals to Singapore.

Variable	Mean	Median	Std dev.	IQR	SW (p)
Ln*tourist arrivals*					
Australia	12.0400	12.0800	0.4158	0.7474	< 0.01
France	10.1640	10.0520	0.3943	0.7320	< 0.01
Germany	10.8120	10.7210	0.3054	0.4216	< 0.01
Netherlands	9.8130	9.8140	0.1449	0.1806	< 0.01
New Zealand	10.1410	10.2120	0.2361	0.2988	< 0.01
Ln*income*					
Australia	10.470	10.523	0.2718	0.4653	< 0.01
France	10.36	10.41	0.2352	0.3501	< 0.01
Germany	10.47	10.49	0.2598	0.4417	< 0.01
Netherlands	10.57	10.66	0.2602	0.3824	< 0.01
New Zealand	10.211	10.240	0.2646	0.4540	< 0.01
Ln*price*					
Australia	−0.0326	−0.0073	0.1103	0.1836	< 0.01
France	0.0296	0.0196	0.0196	0.0766	< 0.01
Germany	0.0173	0.0160	0.0581	0.0769	> 0.01
Netherlands	0.0352	0.0196	0.0902	0.0997	< 0.01
New Zealand	0.0732	0.0695	0.0957	0.1265	> 0.01

Note: SW denotes Shapiro-Wilk Test. Cut-off p-value is 0.05. IQR – Interquartile range.

Performance indices

We use four widely used performance measures to evaluate the proposed model. The measures are root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the ratio of RMSE (RRMSE). RRMSE has been employed as a predictive performance evaluation metric in TD forecasting (Silva et al., 2019).

$$\mathbf{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_t - x_t^{\wedge})^2}$$

$$\mathbf{MAE} = \frac{1}{N}\sum\nolimits_{t=1}^{N} |x_t - x_t^{\wedge}|$$

$$\mathbf{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x_t - x_t^{\wedge}}{x_t}\right|$$

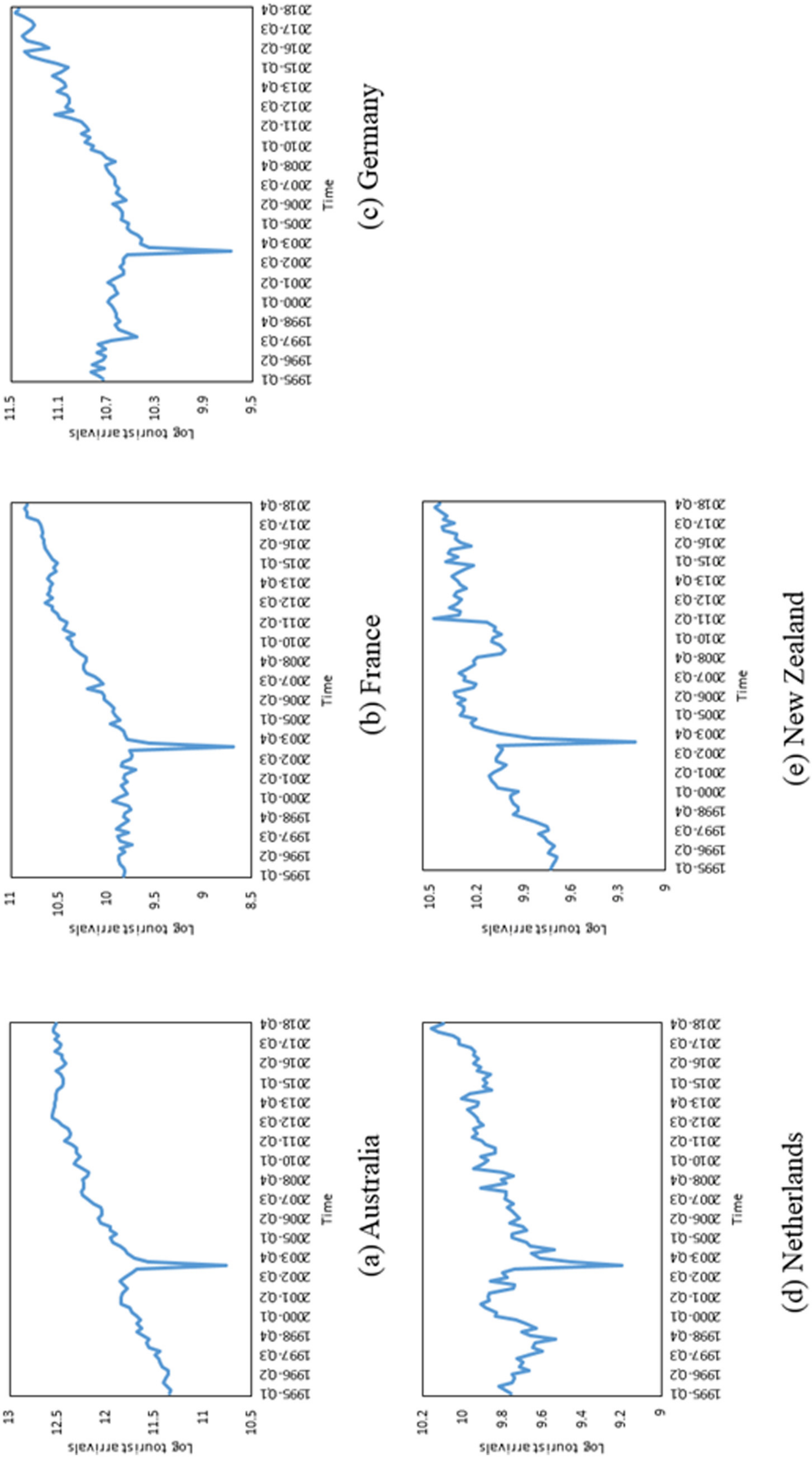

Fig. 6. Tourism demand to Singapore from five source countries.

Table 2
ADF and KPSS test p-values for the dataset.

Variables	I(0)		I(1)	
	ADF	KPSS	ADF	KPSS
Ln*tourist arrivals*				
Australia	0.3726	< 0.01	< 0.01	> 0.1
France	0.7538	< 0.01	< 0.01	> 0.1
Germany	0.8015	< 0.01	< 0.01	> 0.1
Netherlands	0.2639	< 0.01	< 0.01	> 0.1
New Zealand	0.0817	< 0.01	< 0.01	> 0.1
ln*income*				
Australia	0.9639	< 0.01	< 0.01	> 0.1
France	0.5072	< 0.01	0.0360	> 0.1
Germany	0.0140	< 0.01	< 0.01	> 0.1
Netherlands	0.5073	< 0.01	0.0242	0.0933
New Zealand	0.5089	< 0.01	< 0.01	> 0.1
ln*price*				
Australia	0.5387	< 0.01	< 0.01	> 0.1
France	0.01822	< 0.01	0.0254	> 0.1
Germany	0.1298	< 0.01	0.0263	> 0.1
Netherlands	0.03495	< 0.01	0.0587	> 0.1
New Zealand	0.2714	0.0386	0.01	> 0.1

Note: In ADF Test, p-value < 0.05 indicates that data is stationary. In KPSS Test, p-value > 0.05 indicates that data is stationary.

Table 3
Forecasting performance of BBiLSTM and benchmark models.

Source country Hyperparameter	Model	RMSE	MAE	MAPE	RRMSE $\frac{\text{BBiLSTM}}{\text{LSTM}}$	RRMSE $\frac{\text{BBiLSTM}}{\text{SVR}}$	RRMSE $\frac{\text{BBiLSTM}}{\text{RBFNN}}$	RRMSE $\frac{\text{BBiLSTM}}{\text{ADLM}}$
Australia	**BBiLSTM**	**0.0025**	**0.0022**	**0.0002**	0.0671	0.0595	0.0056	0.0585
NN – 172	LSTM	0.0372	0.0326	0.0026				
L2 – 0.0021	SVR	0.0420	0.0363	0.0029				
DP – 0.0163	RBFNN	0.4497	0.3946	0.0316				
LR - 0.0020	ADLM	0.0428	0.0340	0.0027				
France	**BBiLSTM**	**0.0206**	**0.0068**	**0.0006**	0.4993	0.5334	0.1361	0.4218
NN – 42	LSTM	0.0413	0.0284	0.0027				
L2 – 0.0003	SVR	0.0386	0.0254	0.0024				
DP – 0.0202	RBFNN	0.1514	0.1023	0.0096				
LR – 0.0196	ADLM	0.0488	0.0343	0.0032				
Germany	**BBiLSTM**	**0.0204**	**0.0090**	**0.0008**	0.2607	0.1923	0.1524	0.2338
NN – 90	LSTM	0.0783	0.0635	0.0056				
L2 – 0.0110	SVR	0.1061	0.0861	0.0076				
DP – 0.0889	RBFNN	0.1339	0.1158	0.0102				
LR – 0.0108	ADLM	0.0873	0.0655	0.0058				
Netherlands	**BBiLSTM**	**0.0275**	**0.0086**	**0.0009**	0.4636	0.6274	0.4326	0.3178
NN – 138	LSTM	0.0593	0.0541	0.0054				
L2 – 0.0940	SVR	0.0438	0.0352	0.0035				
DP – 0.0111	RBFNN	0.0636	0.0519	0.0052				
LR – 0.0056	ADLM	0.0865	0.0596	0.0060				
New Zealand	**BBiLSTM**	**0.0094**	**0.0078**	**0.0007**	0.1072	0.1364	0.1024	0.1033
NN – 65	LSTM	0.0877	0.0764	0.0074				
L2 – 0.0091	SVR	0.0689	0.0564	0.0054				
DP – 0.1134	RBFNN	0.0918	0.0793	0.0077				
LR – 0.0415	ADLM	0.0910	0.0777	0.0075				

Note: NN is the number of neurons in the BiLSTM layer, L2 denotes L2 regularization, DP depicts the probability of dropout layer, and LR represents learning rate. Values in bold indicate best performance among models.

$$\mathbf{RRMSE_i} = \frac{\boldsymbol{RMSE\_BBiLSTM_i}}{\boldsymbol{RMSE\_Model_i}}$$

where N is the number of samples, $\hat{x_t}$ is the forecasted tourism value and x_t is the actual tourism value. The five source countries are represented by *i,* and *Model* represents the four benchmark models. As an example, when the RRMSE ratio of BBiLSTM to LSTM is < 1, then BBiLSTM forecast performs better than LSTM forecast by 1- $\frac{\text{BBiLSTM}}{\text{LSTM}}$ percent and vice versa.

Table 4
Optimal lag corresponding to information criteria.

Source country	AIC	BIC	HIC
Australia	3	1	1
France	1	1	1
Germany	1	1	1
Netherlands	1	1	1
New Zealand	1	1	1

Note: AIC - Akaike information criterion; BIC- Bayesian Information Criterion; HIC - Hannan–Quinn information criterion. The optimal lag was chosen based on BIC value.

Data

In this section, we describe the data source and the stationarity test applied to the tourism variables. As indicated earlier, Singapore tourism dataset has been studied with numerous approaches, including Copula-ECM, ARIMA, NN, static regression, and ARFIMA (Zhu et al., 2018)

Data source and description

Singapore was chosen as the location of this empirical study. Singapore is ranked 17th in global travel and tourism (WEF, 2019). Approximately 18.5 million visitors arrived in Singapore, contributing $26.9 billion as tourism receipts in 2018. In this study, we estimate tourist arrivals to Singapore from five major source countries, namely Australia, France, Germany, Netherlands, and New Zealand over 24 years (1995Q1 – 2018Q4). Other countries were not selected due to data availability issues. The definitions of variables that are utilized in this study are largely drawn from Song, Witt, and Li (2009) and Gunter and Onder (2015).

The variable *International TD* measuring the number of tourist arrivals to Singapore is defined as the natural logarithm of quarterly tourist arrivals to Singapore from five major source countries (France, Germany, Australia, Netherlands, and New Zealand). It is represented by $q_{i,\ t} = ln\,(Q_{i,\ t})$ where $i\in$ (FR, DE, AU, NL, NZ) at time $t=$ 1995Q1, …,2018Q4, thereby resulting in 96 observations per source market. The tourist arrivals dataset used in the study is obtained from Singapore the Tourism Board (STB, 2019).

The *relative price* variable measures the living cost in Singapore compared to the price level in the respective market, adjusted by the corresponding exchange rates against the US dollar. The relative price variable is denoted by $PR_{i,\ t}$ where

$$\boldsymbol{PR_{i,t} = ln\,\frac{(CPI_{Des,t}/ER_{Des,t})}{(CPI_{i,t}/ER_{i,t})}}$$

where $CPI_{Des,\ t}$ and $CPI_{i,\ t}$ are the quarterly consumer price index, $ER_{Des,\ t}$ and $ER_{i,\ t}$ are the consumer price index and exchange rates for destination and origin respectively against the US dollar. The CPI and exchange rate data are obtained from the International Monetary Fund (IMF, 2019).

Lastly, the *tourist income* variable is measured as real GDP per capita of the source market and is depicted by $y_{i,\ t} = ln\,(Y_{i,\ t})$. The data was obtained from OECD (2019). All the datasets employed in the study were log-transformed before estimation.

In this study, transportation cost, despite being an explanatory variable for TD, has been precluded. There is no accepted definition, and lack of reliable data for various countries, and multicollinearity may also be introduced when transportation cost is included with the tourism price variable (Song et al., 2009).

Table 1 shows descriptive statistics for the dataset. Most of the variables related to each source country did not follow a normal distribution, as suggested by Shapiro-Wilk (SW) test (p-value < 0.01). Thereby indicating that non-parametric methods are more suitable for such a dataset (Hansen et al., 1999). Hence, the median should be regarded as a measure of central tendency and Interquartile range (IQR) as a measure of variation in the data. From Table 1, we can see that Australia shows the highest variation and mean in the dataset. Fig. 6 shows the progression of TD to Singapore from five source countries. The entire dataset consisted of 96 data points from 1995Q1 to 2018Q4.

Stationarity

The tourism variables were tested for stationarity. ADF (Dickey & Fuller, 1979) and KPSS (Kwiatkowski, Phillips, Schmidt, & Shin, 1992) tests were carried out to find out if the time series is stationary or not. The null hypothesis of the ADF test is: "unit root is present in time series." The null hypothesis of the KPSS test is "the time series is stationary." Based on both KPSS and ADF test results in Table 2, the time series is deemed non-stationary. In order to attain a stationary time series, the first difference of the series was taken (Lim, Chang, & McAleer, 2009).

Results

The empirical results from the proposed forecasting methodology are described in Table 3. The results shown in the tables depict the performance of the models on the test dataset at *lag1*. The optimal number of lags was selected based on the Bayesian information

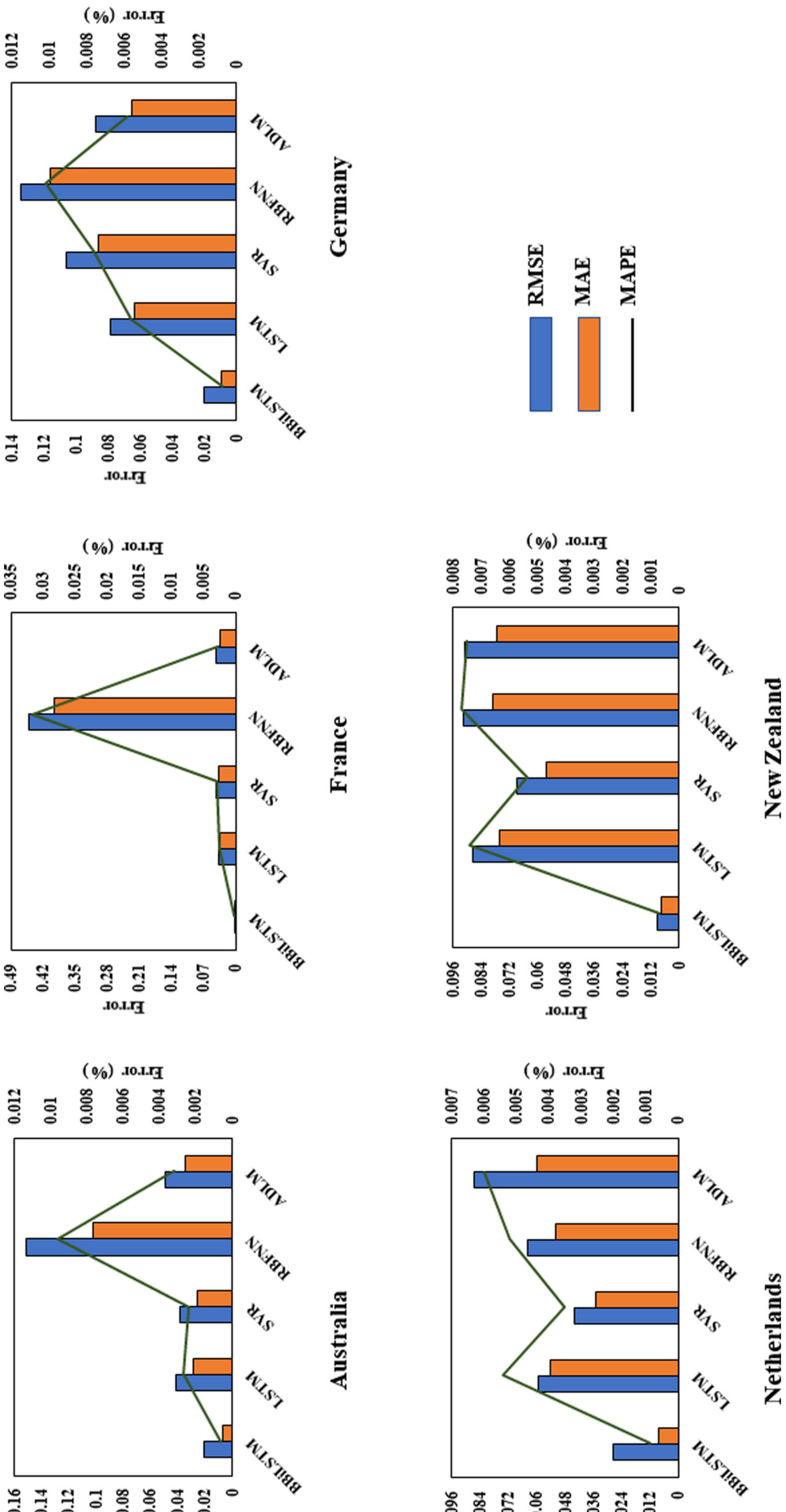

Fig. 7. Forecasting results for BBiLSTM and benchmark models at lag1.

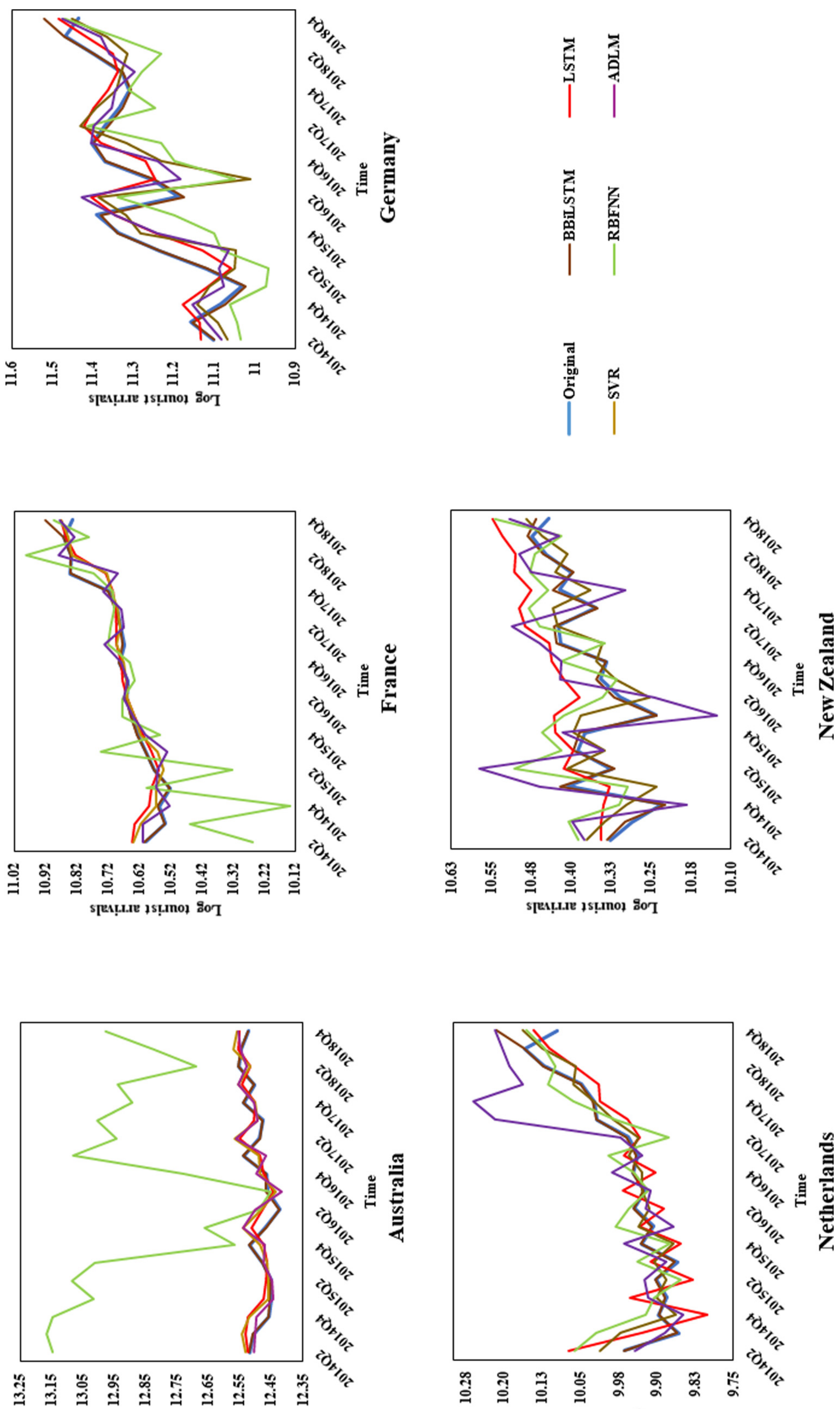

Fig. 8. Out of sample forecasts for tourist arrivals to Singapore.

Table 5
Effect of increasing lags on BBiLSTM model performance.

Source country	Lag1		Lag2		Lag3		Lag4	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
Australia	0.0025	0.0002	0.0133	0.0005	0.0093	0.0006	0.0144	0.0007
France	0.0206	0.0006	0.0122	0.0008	0.0199	0.0012	0.0075	0.0005
Germany	0.0204	0.0008	0.0164	0.0005	0.0163	0.0006	0.0190	0.0010
Netherlands	0.0275	0.0009	0.0222	0.0015	0.0206	0.0012	0.0073	0.0005
New Zealand	0.0094	0.0007	0.0113	0.0007	0.0064	0.0005	0.0153	0.0009

Table 6
Impact of number of neurons on the model performance.

Source market	Constant parameters	NN	RMSE	MAE	MAPE
Australia	L2 – 0.0021	25	0.0492	0.0453	0.0036
	LR – 0.0020	75	0.0079	0.0074	0.0006
	DP – 0.0163	125	0.0048	0.0043	0.0003
		200	0.0040	0.0038	0.0003
		172	**0.0025**	**0.0022**	**0.0002**
France	L2 – 0.0003	25	0.0210	0.0073	0.0007
	LR – 0.0196	75	0.0204*	0.0093	0.0009
	DP – 0.0202	125	0.0211	0.0098	0.0009
		200	0.0216	0.0067	0.0006
		42	**0.0206**	**0.0068**	**0.0006**
Germany	L2 – 0.0110	25	0.0233	0.0178	0.0016
	LR – 0.0108	75	0.0209	0.0148	0.0012
	DP – 0.0889	125	0.0218	0.0150	0.0009
		200	0.0224	0.0153	0.0014
		90	**0.0204**	**0.0090**	**0.0008**
Netherlands	L2 – 0.0940	25	0.0372	0.0182	0.0018
	LR – 0.0056	75	0.0272*	0.0101	0.0010
	DP – 0.0111	125	0.0277	0.0165	0.0015
		200	0.0267*	0.0119	0.0012
		138	**0.0275**	**0.0086**	**0.0009**
New Zealand	L2 – 0.0091	25	0.0189	0.0162	0.0016
	LR – 0.0415	75	0.0103	0.0089	0.0009
	DP – 0.1134	125	0.0317	0.0221	0.0018
		200	0.0616	0.0539	0.0052
		65	**0.0094**	**0.0078**	**0.0007**

Note: In Tables 6–9, NN is the number of neurons in the BiLSTM layer, L2 denotes L2 regularization, Dropout depicts the probability of dropout layer, and LR represents learning rate. Values in bold show the optimal NN as selected by BO and corresponding performance. In three cases (indicated by *), the performance is superior to BO based hyperparameter.

criterion (BIC). From Table 4, we can see that *lag1* is the optimal lag for most countries. In order to assess the prediction performance genuinely, we used the out-of-sample data (Sagheer & Kotb, 2019).

First, based on the RMSE criterion, we can observe from Fig. 7 that the BBiLSTM outperforms all other forecasting models, namely LSTM, SVR, RBFNN, and ADLM. The MAPE and MAE criterion also strengthens the suggestion as MAPE and MAE values of the BBiLSTM model are comparatively lesser than the other four benchmark models.

Next, we examine the forecasting results at the level of source country from Table 3 and Fig. 8. In the case of Australian tourist arrivals, BBiLSTM predictions outperform the LSTM, SVR, RBFNN and ADLM forecasts based on RMSE, MAPE, and MAE criteria. As evident from the RRMSE values, the BBiLSTM forecasts are found to be 93.3%, 94.1%, 99.4%, and 94.2% better than LSTM, SVR, RBFNN, and ADLM forecasts, respectively.

For France tourist arrivals, we observe that the BBiLSTM forecasts perform better than the LSTM, SVR, RBFNN, and ADLM forecasts. As evident from the RRMSE values, the BBiLSTM forecasts are found to be 50.1%, 46.7%, 86.4%, and 57.8% better than LSTM, SVR, RBFNN, and ADLM forecasts, respectively.

In the case of German tourist arrivals, the BBiLSTM forecasts outperform the LSTM, SVR, RBFNN, and ADLM forecasts. Also, the BBiLSTM forecasts are found to be 74.0%, 80.8%, 84.8%, and 76.6% better than the LSTM, SVR, RBFNN, and ADLM forecasts, respectively.

BBiLSTM performs better than the LSTM, SVR, RBFNN, and ADLM in forecasting tourist arrivals from the Netherlands also. BBiLSTM forecasts are found to be 53.6%, 37.3%, 56.7% and 68.2% better than the LSTM, SVR, RBFNN and ADLM forecasts, respectively.

In the case of New Zealand tourist arrivals, the BBiLSTM forecasts outperform the LSTM, SVR, RBFNN and ADLM forecasts. BBiLSTM forecasts are found to be 89.3%, 86.4%, 89.8% and 89.7% better than the LSTM, SVR, RBFNN and ADLM forecasts,

Table 7
Impact of L2 regularization on the performance of the model.

Source market	Constant parameters	L2	RMSE	MAE	MAPE
Australia	NN – 172	0.01	0.0083	0.0075	0.0006
	LR – 0.0020	0.0001	0.0073	0.0068	0.0005
	DP – 0.0163	0.1	0.0619	0.0537	0.0043
		0.005	0.0026	0.0023	0.0002
		0.0021	**0.0025**	**0.0022**	**0.0002**
France	NN – 42	0.01	0.0192*	0.0110	0.0010
	LR – 0.0196	0.0001	0.0221	0.0096	0.0009
	DP – 0.0202	0.1	0.0179*	0.0102	0.0010
		0.005	0.0203*	0.0095	0.0009
		0.0003	**0.0206**	**0.0068**	**0.0006**
Germany	NN – 90	0.01	0.0251	0.0184	0.0016
	LR – 0.0108	0.0001	0.0220	0.0131	0.0012
	DP – 0.0889	0.1	0.0892	0.0722	0.0064
		0.005	0.0207	0.0109	0.0098
		0.0110	**0.0204**	**0.0090**	**0.0008**
Netherlands	NN – 138	0.01	0.0312	0.0096	0.0010
	LR – 0.0056	0.0001	0.0262*	0.0089	0.0009
	DP – 0.0111	0.1	0.0442	0.0365	0.0037
		0.005	0.0300	0.0092	0.0009
		0.0940	**0.0275**	**0.0086**	**0.0009**
New Zealand	NN – 65	0.01	0.0101	0.0095	0.0009
	LR – 0.0415	0.0001	0.0106	0.0097	0.0009
	DP – 0.1134	0.1	0.0674	0.0554	0.0053
		0.005	0.0167	0.0151	0.0015
		0.0091	**0.0094**	**0.0078**	**0.0007**

Note: Values in bold show the optimal L2 as selected by BO and corresponding performance. In four cases (indicated by *), the performance is superior to BO based hyperparameter.

Table 8
Impact of dropout on the model performance.

Source market	Constant parameters	DP	RMSE	MAE	MAPE
Australia	NN – 172	0.001	0.0048	0.0035	0.0003
	L2 – 0.0021	0.1	0.0088	0.0081	0.0007
	LR – 0.0020	0.05	0.0056	0.0048	0.0004
		0.5	0.0064	0.0048	0.0004
		0.0163	**0.0025**	**0.0022**	**0.0002**
France	NN – 42	0.001	0.0208	0.0083	0.0008
	L2 – 0.0003	0.1	0.0195*	0.0077	0.0007
	LR – 0.0196	0.05	0.0213	0.0090	0.0008
		0.5	0.0160*	0.0103	0.0010
		0.0202	**0.0206**	**0.0068**	**0.0006**
Germany	NN – 90	0.001	0.0228	0.0083	0.0007
	L2 – 0.0110	0.1	0.0233	0.0138	0.0012
	LR – 0.0108	0.05	0.0205	0.0105	0.0009
		0.5	0.0159*	0.0112	0.0010
		0.0889	**0.0204**	**0.0090**	**0.0008**
Netherlands	NN – 138	0.001	0.0286	0.0146	0.0015
	L2 – 0.0940	0.1	0.0263*	0.0101	0.0010
	LR – 0.0056	0.05	0.0261*	0.0104	0.0010
		0.5	0.0244*	0.0162	0.0016
		0.0111	**0.0275**	**0.0086**	**0.0009**
New Zealand	NN – 65	0.001	0.0107	0.0099	0.0009
	L2 – 0.0091	0.1	0.0126	0.0118	0.0010
	LR – 0.0415	0.05	0.0169	0.0164	0.0016
		0.5	0.0274	0.0235	0.0023
		0.1134	**0.0094**	**0.0078**	**0.0007**

Note: Values in bold show the optimal dropout as selected by BO and corresponding performance. In six cases (indicated by *), the performance is superior to BO based hyperparameter.

respectively.

From the above results, we may conclude that the BBiLSTM, with its two-way learning process, has better generalization capabilities than other competing models. All other competing models learn in a unidirectional way, thereby somewhat limiting their learning capability.

Table 9
Impact of learning rate on the model performance.

Source market	Constant parameters	LR	RMSE	MAE	MAPE
Australia	NN – 172	0.001	0.0393	0.0383	0.0031
	L2 – 0.0021	0.5	3.6911	3.6211	0.2899
	DP – 0.0163	0.06	0.0429	0.0410	0.0033
		0.15	0.2237	0.1607	0.0129
		0.0020	**0.0025**	**0.0022**	**0.0002**
France	NN – 42	0.001	0.0449	0.0402	0.0038
	L2 – 0.0003	0.5	0.1453	0.1369	0.0128
	DP – 0.0202	0.06	0.0353	0.0309	0.0029
		0.15	0.0250	0.0174	0.0016
		0.0196	**0.0206**	**0.0068**	**0.0006**
Germany	NN – 90	0.001	0.0962	0.0785	0.0069
	L2 – 0.0110	0.5	2.4059	2.3455	0.2083
	DP – 0.0889	0.06	0.0620	0.0509	0.0045
		0.15	0.0607	0.0505	0.0045
		0.0108	**0.0204**	**0.0090**	**0.0008**
Netherlands	NN – 138	0.001	0.0307	0.0282	0.0028
	L2 – 0.0940	0.5	7.1045	6.9523	0.6985
	DP – 0.0111	0.06	0.0423	0.0376	0.0038
		0.15	0.9360	0.9189	0.0921
		0.0056	**0.0275**	**0.0086**	**0.0009**
New Zealand	NN – 65	0.001	0.0726	0.0622	0.0060
	L2 – 0.0091	0.5	0.1064	0.0911	0.0088
	DP – 0.1134	0.06	0.0136	0.0113	0.0011
		0.15	0.0462	0.0369	0.0036
		0.0415	**0.0094**	**0.0078**	**0.0007**

Note: Values in bold show the optimal LR as selected by BO and the corresponding performance.

Robustness analysis

The forecasting performance of the BBiLSTM model under different conditions is studied in this section. Case study 1 examines the effect of increasing lags on forecasting performance. Case study 2 examines the effect of (a) number of neurons (b) L2 regularization, (c) probability of dropout layer, and (d) learning rate on the performance of the BBiLSTM model. Case study 3 utilizes multi-step ahead forecasting. Case study 4 validates the performance of the proposed BBiLSTM model in deseasonalized data. Case study 5 compares the performance of the BBiLSTM model using Diebold Mariano test.

Case study experiment 1 - effect of increasing lags

In this case study, we analyze the effect of increasing lags on model performance. Forecasts are generated for lag2 to lag4. We can see from Table 5, that upon increasing the number of lags, the BBiLSTM model performance differs according to the source country. Albeit, the effect of increasing the lag is negligible in terms of MAPE. In the case of France, Germany, and the Netherlands, the RMSE value at lag1 is greater than RMSE values at lag2, lag3, and lag4. This is in contrast to the case of Australia, where RMSE at lag1 is the lowest. In the case of New Zealand, the RMSE at lag1 is lower than lag2 and lag4, but not lag3. The findings also suggest that tourists start searching about the choice of destination, at least three months (lag1) before the actual travel by searching for information about prices, exchange rates, and new destinations that might fit within their budget. The planning period may be higher for tourists from European countries, as indicated by the lower RMSE values corresponding to higher lags.

Case study experiment 2 - impact of hyperparameters

In this case study, we analyze the impact of various parameter values on the performance of the BBiLSTM model. The optimal set of hyperparameters for each source country, as found by BO and their performance in terms of RMSE, MAE, and MAPE, are shown in Table 3. The parameters studied are the number of neurons (NN) (Table 6), L2 regularization (L2) (Table 7), dropout probability (Table 8), and learning rate (LR) (Table 9).

From Tables 6–9, we can see that different hyperparameters have a considerable effect on forecasting performance. Further, we compare the performance of the hyperparameters as identified by BO against various randomly selected hyperparameters in terms of RMSE, MAE, and MAPE. From Tables 6–9, we observe that the BO identified hyperparameters (shown in bold) show better performance than randomly selected hyperparameters in 67 (84%) out of the 80 cases (4 randomly selected hyperparameters × 4 hyperparameters × 5 countries) in terms of RMSE. The thirteen cases (16%) where the randomly selected hyperparameters fared better is restricted to mostly two source countries, France and the Netherlands, and in one case, Germany (indicated by a * mark in Tables 6–9). Considering those thirteen cases, none of the randomly selected parameters had MAE and MAPE values below BO identified parameters. Thus, we may conclude that BO identified hyperparameters are robust and show superior performance in

Table 10
Multi step ahead forecasting with BBiLSTM and benchmark models.

Source market Model	$h = 2$			$h = 3$			$h = 4$		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
Australia									
BBiLSTM	**0.0129**	**0.0114**	**0.0009**	**0.0188**	**0.0164**	**0.0013**	**0.0206**	**0.0178**	**0.0014**
LSTM	0.0728	0.0621	0.0058	0.0913	0.0847	0.0082	0.1026	0.0986	0.0091
SVR	0.0935	0.0876	0.0082	0.1127	0.1035	0.0098	0.1862	0.1631	0.0158
RBFNN	0.9567	0.9131	0.0894	1.2758	0.9968	0.0932	1.9536	1.6725	0.9994
ADLM	0.0921	0.0852	0.0083	0.1011	0.0981	0.0089	0.1793	0.1624	0.0153
France									
BBiLSTM	**0.0569**	**0.0326**	**0.0038**	**0.0815**	**0.0656**	**0.0063**	**0.0935**	**0.0890**	**0.0084**
LSTM	0.0719	0.0624	0.0060	0.0962	0.0767	0.0071	0.0989	0.0904	0.0089
SVR	0.0783	0.0721	0.0071	0.1038	0.0957	0.0082	0.1257	0.1011	0.0126
RBFNN	0.3589	0.3183	0.0298	0.6423	0.6108	0.0596	0.8582	0.7381	0.0729
ADLM	0.0876	0.0794	0.0073	0.1488	0.1246	0.0110	0.1765	0.1532	0.0131
Germany									
BBiLSTM	**0.0683**	**0.0551**	**0.0005**	**0.0728**	**0.0456**	**0.0041**	**0.0824**	**0.0607**	**0.0055**
LSTM	0.0953	0.0914	0.0089	0.1183	0.1019	0.0098	0.2136	0.1954	0.0176
SVR	0.1331	0.1153	0.0142	0.2280	0.1987	0.0188	0.2741	0.2593	0.0238
RBFNN	0.1989	0.1826	0.0179	0.2672	0.2231	0.0252	0.3114	0.2979	0.0283
ADLM	0.1029	0.0987	0.0092	0.1821	0.1747	0.0169	0.2530	0.2249	0.0216
Netherlands									
BBiLSTM	**0.0833**	**0.0690**	**0.0067**	**0.1112**	**0.0937**	**0.0095**	**0.1355**	**0.1214**	**0.0133**
LSTM	0.0989	0.0861	0.0082	0.1386	0.1156	0.0106	0.1924	0.1858	0.0179
0.0SVR	0.0922	0.0899	0.0084	0.1297	0.1034	0.0968	0.1835	0.1611	0.0143
RBFNN	0.1053	0.0963	0.0093	0.1681	0.1539	0.0144	0.2684	0.2431	0.0216
ADLM	0.1242	0.1117	0.0108	0.1694	0.1597	0.0150	0.2790	0.2528	0.0250
New Zealand									
BBiLSTM	**0.0320**	**0.0217**	**0.0022**	**0.0384**	**0.0273**	**0.0026**	**0.0967**	**0.0734**	**0.0071**
LSTM	0.1057	0.0979	0.0090	0.1267	0.1125	0.0131	0.2163	0.2014	0.0283
SVR	0.0941	0.0899	0.0083	0.1183	0.1024	0.0115	0.1616	0.1539	0.0148
RBFNN	0.1697	0.1385	0.0125	0.2134	0.1922	0.0183	0.2998	0.2712	0.0237
ADLM	0.1644	0.1288	0.0116	0.2582	0.2416	0.0234	0.3412	0.3187	0.0365

Note: Values in bold indicate best performance among models.

general. The implication is that a practitioner may confidently deploy BO for hyperparameter identification.

Case study experiment 3 - multi-step ahead forecasting

To further ascertain the robustness of our proposed model, multi-step ahead forecasting was performed with $h = 2$,3 and 4 steps. From Table 10, we can see that based on RMSE, MAE, and MAPE criteria, BBiLSTM outperforms all the other benchmark models at $h = 2$,3 and 4 steps ahead. Thus, implying that our model is robust upon increasing the forecasting horizon.

Case study experiment 4 – impact of data pre-processing

As described earlier, the dataset contains quarterly tourist arrivals. In our context, the seasonal difference is the difference between the quarterly natural logarithm of tourist arrivals in a given quarter and the corresponding quarter of the preceding year. We employ a seasonal differencing filter $D^4 = 1 - L_4$ where D^4 and L_4 are the seasonal difference and backshift operators, respectively (Dickey & Zhang, 2010). Upon seasonal differencing, the variables were transformed into year-on-year growth. The seasonally differenced dataset was found to be stationary, based on both ADF and KPSS tests. After seasonal differencing, 92 observations per variable and per source country were available for modeling and was split in 60:20:20 ratio for model building, validation, and benchmarking. Based on BIC values, lag1 was selected as the optimal lag. From Table 11, we observe that BBiLSTM outperforms LSTM, SVR, RBFNN and ADLM models in terms of RMSE, MAPE, and MAE. The performance of BBiLSTM on deseasonalized data is consistent with results in Table 3.

Case study experiment 5 - comparison of forecasting accuracy with Diebold Mariano test

To confirm the forecasting performance of the BBiLSTM model, Diebold and Mariano (DM) (1995) test is utilized. The null

Table 11
Comparison of performance of models upon seasonal differencing.

Source market / Hyperparameter	Model	RMSE	MAPE	MAE
Australia	**BBiLSTM**	**0.0495**	**0.0033**	**0.0412**
NN – 20	LSTM	0.0928	0.0061	0.0758
L2 – 3.2983E – 06	SVR	0.1477	0.0096	0.1198
DP – 0.0121	RBFNN	0.2022	0.0132	0.1649
LR – 0.1724	ADLM	0.1602	0.0105	0.1313
France	**BBiLSTM**	**0.0304**	**0.0024**	**0.0256**
NN – 199	LSTM	0.1653	0.0118	0.1258
L2 – 6.8216E – 06	SVR	0.1281	0.0093	0.0990
DP – 0.0104	RBFNN	0.2197	0.0166	0.1768
LR – 0.0130	ADLM	0.5859	0.0407	0.4335
Germany	**BBiLSTM**	**0.1074**	**0.0071**	**0.0801**
NN – 24	LSTM	0.6367	0.0463	0.5223
L2 – 9.6043E – 06	SVR	0.2505	0.0177	0.1994
DP – 0.0815	RBFNN	0.8256	0.0398	0.4513
LR – 0.0320	ADLM	0.6939	0.0563	0.6359
Netherlands	**BBiLSTM**	**0.0596**	**0.0022**	**0.0221**
NN – 200	LSTM	0.1223	0.0099	0.0988
L2 – 8.1997E – 04	SVR	0.1419	0.0109	0.1088
DP – 0.0205	RBFNN	0.2385	0.0200	0.1989
LR – 0.0161	ADLM	0.2642	0.0227	0.2262
New Zealand	**BBiLSTM**	**0.0382**	**0.0028**	**0.0295**
NN – 61	LSTM	0.1390	0.0110	0.1144
L2 – 0.0024	SVR	0.3834	0.0299s	0.3101
DP – 0.0103	RBFNN	0.4524	0.0364	0.3768
LR – 0.0411	ADLM	0.4295	0.0344	0.3562

Note: Values in bold indicate best performance among models.

Table 12
DM stat for comparison of predictive accuracy of BBiLSTM model.

Source	Benchmark model	DM stat
Australia	LSTM	− 4.4391*
	SVR	− 4.0798*
	RBFNN	− 5.4975*
	ADLM	− 3.7511*
France	LSTM	− 1.4306
	SVR	− 1.2406
	RBFNN	− 2.0813
	ADLM	− 1.5109
Germany	LSTM	− 2.3723*
	SVR	− 2.9058*
	RBFNN	− 4.562*
	ADLM	− 2.1824*
Netherlands	LSTM	− 2.5808*
	SVR	− 1.2756
	RBFNN	− 2.0832
	ADLM	− 1.9194
New Zealand	LSTM	− 3.8222*
	SVR	− 2.9014*
	RBFNN	− 3.8883*
	ADLM	− 2.6500*

Note: * indicates the p-value is < 0.05.

hypothesis of the DM test is "equal predictive accuracy of both the models." The negative sign of the DM stat in Table 12 indicates that the competing model has higher prediction errors. From Table 12, it is also evident that BBiLSTM model outperforms the LSTM, SVR, RBFNN, and ADLM models consistently. For most of the observations, the DM statistic values were statistically significant.

Conclusion

This study contributes to the hospitality literature by applying a deep learning approach to forecast tourism demand. The BBiLSTM model utilized income and relative price variables to predict tourism demand to Singapore from five visitor countries and was benchmarked against LSTM, SVR, and RBFNN models. The experimental findings suggest that the BBiLSTM model outperforms all four benchmark models. The findings were validated statistically through the DM test, thus implying that it may be worthwhile to consider BBiLSTM as a tourism demand forecasting model. The findings also showed that the hyperparameters of the model significantly impact the model's performance. Experiments also establish the superiority of Bayesian Optimization in identifying the hyperparameters in an informed manner, leading to superior model performance. We also found that the BBiLSTM approach takes only a few seconds to make predictions on the test data set. Hence, practitioners need not be concerned about the time taken to execute deep learning methods on data sets comparable to the one used in this article. While the model has its advantages, BBiLSTM, like any other AI model, is a black-box model and would be preferable where precision is more important than explanation.

Our work implies that practitioners may identify hyperparameters using the BO approach instead of manual, random, or grid search methods. BBiLSTM is especially suitable for data that might exhibit long-term dependencies such as in tourism demand. Practitioners may follow the steps in Fig. 5 in implementing a BBiLSTM model. The initial steps would entail data collection, preprocessing activities such as data transformation, seasonality, and stationarity checks, and splitting the data into training, validation, and testing. The next step is to construct the BBiLSTM model with suitable predictors using the training data. The hyperparameters of the model are then identified by BO using validation data. Finally, the performance is assessed using the testing data.

Precise tourism demand forecasting can help practitioners effectively allocate and utilize resources and frame policies. Accuracy improvements in tourism demand forecasting translate to superior operational planning for the business involved in the tourism industry, for example, hotels and tour operators. The city administration, transport-hub, and transportation managers can use the BBiLSTM model to prepare for peak demand seasons more effectively and efficiently. Imprecise demand prediction may result in wastage of resources in the case of overestimation of demand and unpreparedness and loss of opportunity in the case of underestimation of tourist demand.

Our results show that the BBiLSTM model has consistently performed better across different source countries and different forecasting horizons. We also found that for these five source countries, tourist decides regarding the choice of destination county, at least three months before the actual travel. This, in turn, will aid the minimization of cost by businesses and the planning of foreign exchange currency as per the forecasted demand. This study utilized Singapore as a single test case, and quarterly data with only two explanatory variables were used for explaining tourism demand. For better generalization capability and robustness, more destination countries, yearly or monthly data, and more explanatory variables like marketing expenditure, and the number of hotels in the destination country can be tested in future work.

Declaration of competing interest

None.

References

Alade, I. O., Rahman, M. A. A., & Saleh, T. A. (2019). Predicting the specific heat capacity of alumina/ethylene glycol nanofluids using support vector regression model optimized with Bayesian algorithm. *Solar Energy, 183*, 74–82.
Andrew, W. P., Cranage, D. A., & Lee, C. K. (1990). Forecasting hotel occupancy rates with time series models: An empirical analysis. *Hospitality Research Journal, 14*(2), 173–182.
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), 383–397.
Au, N., & Law, R. (2000). The application of rough sets to sightseeing expenditures. *Journal of Travel Research, 39*(1), 70–77.
Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks, 5*(2), 157–166.
Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research, 13*, 281–305.
Bonham, C., Edmonds, C., & Mak, J. (2006). The impact of 9/11 and other terrible global events on tourism in the United States and Hawaii. *Journal of Travel Research, 45*(1), 99–110.
Chen, M. S., Ying, L. C., & Pan, M. C. (2010). Forecasting tourist arrivals by using the adaptive network-based fuzzy inference system. *Expert Systems with Applications, 37*(2), 1185–1191.
Cheng, H., Ding, X., Zhou, W., & Ding, R. (2019). A hybrid electricity price forecasting model with Bayesian Optimization for German energy exchange. *International Journal of Electrical Power & Energy Systems, 110*, 653–666.
Claveria, O., Monte, E., & Torra, S. (2015). Tourism demand forecasting with neural network models: Different ways of treating information. *International Journal of Tourism Research, 17*(5), 492–500.
Cornejo-Bueno, L., Garrido-Merchán, E. C., Hernández-Lobato, D., & Salcedo-Sanz, S. (2018). Bayesian optimization of a hybrid system for robust ocean wave features prediction. *Neurocomputing, 275*, 818–828.
Dergiades, T., Mavragani, E., & Pan, B. (2018). Google Trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management, 66*, 108–120.
Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of American Statistical Association, 74*(366), 427–431.
Dickey, D. A., & Zhang, Y. (2010). Seasonal unit root tests in long periodicity cases. *Journal of the Korean Statistical Society, 39*(3), 271–279.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*(3), 253–263.
Eggensperger, K., Feurer, M., Hutter, F., Bergstra, J., Snoek, J., Hoos, H., & Leyton-Brown, K. (2013). Towards an empirical foundation for assessing Bayesian optimization of hyperparameters. *NIPS workshop on Bayesian optimization in theory and practice*.
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.
Frechtling, D. C., & Frechtling, D. C. (2001). *Forecasting tourism demand: Methods and strategies.* Oxford, Boston: Butterworth-Heinemann.
Goh, C., Law, R., & Mok, H. M. (2008). Analyzing and forecasting tourism demand: A rough sets approach. *Journal of Travel Research, 46*(3), 327–338.
Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks, 18*(5–6), 602–610.
Guizzardi, A., & Stacchini, A. (2015). Real-time forecasting regional tourism with business sentiment surveys. *Tourism Management, 47*, 213–223.
Guler, N. F., Ubeyli, E. D., & Guler, I. (2005). Recurrent neural networks employing Lyapunov exponents for EEG signals classification. *Expert Systems with Applications, 29*(3), 506–514.
Gunter, U., & Onder, I. (2015). Forecasting international city tourism demand for Paris: Accuracy of uni and multivariate models employing monthly data. *Tourism Management, 46*, 123–135.
Hadavandi, E., Ghanbari, A., Shahanaghi, K., & Abbasian-Naghneh, S. (2011). Tourist arrival forecasting by evolutionary fuzzy systems. *Tourism Management, 32*(5),

1196–1203.
Hanly, P., & Wade, G. (2007). Research note: Modelling tourism demand–an econometric analysis of North American tourist expenditure in Ireland, 1985–2004. *Tourism Economics, 13*(2), 319–327.
Hansen, J. V., McDonald, J. B., & Nelson, R. D. (1999). Time series prediction with genetic-algorithm designed neural networks: An empirical comparison with modern statistical models. *Computational Intelligence, 15*, 171–184.
He, F., Zhou, J., Feng, Z., Liu, G., & Yang, Y. (2019). A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Applied Energy, 237*, 103–116.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.
Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, 79*(8), 2554–2558.
Huang, X., Zhang, L., & Ding, Y. (2017). The Baidu index: Uses in predicting tourism flows – A case study of the forbidden city. *Tourism Management, 58*, 301–306.
Husken, M., & Stagge, P. (2003). Recurrent neural networks for time series classification. *Neurocomputing, 50*, 223–235.
Icoz, O., Var, T., & Kozak, M. (1998). Tourism demand in Turkey. *Annals of Tourism Research, 25*(1), 236–240.
IMF (2019). *Access to macroeconomics and financial data.* Retrieved December 15, 2019, fromInternational Monetary Fund Websitehttps://data.imf.org/.
Jiao, E. X., & Chen, J. L. (2019). Tourism forecasting: A review of methodological developments over the last decade. *Tourism Economics, 25*(3), 469–492.
Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations*.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* (pp. 1097–1105). .
Kon, S. C., & Turner, L. W. (2005). Neural network forecasting of tourism demand. *Tourism Economics, 11*(3), 301–328.
Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics, 54*(1/3), 159–178.
Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy, 221*, 386–405.
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.
Law, T., & Shawe-Taylor, J. (2017). Practical Bayesian support vector regression for financial time series prediction and market condition change detection. *Quantitative Finance, 17*(9), 1403–1416.
Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research, 44*(1), 82–99.
Li, G., Wong, K. K. F., Song, H., & Witt, S. F. (2006). Tourism demand forecasting: A time varying parameter error correction model. *Journal of Travel Research, 45*(2), 175–185.
Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, 57–66.
Lim, C. (1997). Review of international tourism demand models. *Annals of Tourism Research, 24*(4), 835–849.
Lim, C., Chang, C., & McAleer, M. (2009). Forecasting h(m)otel guest nights in New Zealand. *International Journal of Hospitality Management, 28*(2), 228–235.
Martinez-Cantin, R. (2017). Bayesian optimization with adaptive kernels for robot control. *2017 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3350–3356). .
Moore, W. R. (2010). The impact of climate change on Caribbean tourism demand. *Current Issues in Tourism, 13*(5), 495–505.
OECD (2019). *Tourism statistics.* Retrieved December 15, 2019, fromOrganisation for Economic Co-operation and Development Websitehttps://www.oecd.org/.
Onder, I., Gunter, U., & Gindl, S. (2019). Utilizing Facebook statistics in tourism demand modeling and destination marketing. *Journal of Travel Research, 59*(2), 195–208.
Palmer, A., Montano, J. J., & Sese, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management, 27*(5), 781–790.
Sagheer, A., & Kotb, M. (2019). Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing, 323*, 203–213.
Saha, S., & Yap, G. (2014). The moderation effects of political instability and terrorism on tourism development: A cross-country panel analysis. *Journal of Travel Research, 53*(4), 509–521.
Shahrabi, J., Hadavandi, E., & Asadi, S. (2013). Developing a hybrid intelligent model for forecasting problems: Case study of tourism demand time series. *Knowledge-Based Systems, 43*, 112–122.
Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2016). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE, 104*(1), 148–175.
Silva, E. S., Hassani, H., Heravi, S., & Huang, X. (2019). Forecasting tourism demand with denoised neural networks. *Annals of Tourism Research, 74*, 134–154.
Song, H., & Li, G. (2008). Tourism demand modelling and forecasting: A review of recent research. *Tourism Management, 29*, 203–220.
Song, H., & Witt, S. F. (2006). Forecasting international tourist flows to Macau. *Tourism Management, 27*(2), 214–224.
Song, H., Witt, S. F., & Li, G. (2009). *The advanced econometrics of tourism demand.* New York, London: Routledge.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*, 1929–1958.
STB (2019). *Quarterly tourism performance report.* Retrieved December 15, 2019, fromSingapore Tourism Board Websitehttps://www.stb.gov.sg/content/stb/en/statistics-and-market-insights/tourism-statistics/quarterly-tourism-performance-report.html.
Uysal, M., & El Roubi, M. S. (1999). Artificial neural networks versus multiple regression in tourism demand analysis. *Journal of Travel Research, 38*(2), 111–118.
Wang, C. H. (2004). Predicting tourism demand using fuzzy time series and hybrid grey theory. *Tourism Management, 25*(3), 367–374.
WEF (2019). The travel & tourism competitiveness report 2019. Retrieved December 15, 2019, from World Economic Forum Website http://www3.weforum.org/docs/WEF_TTCR_2019.pdf.
Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting, 11*(3), 447–475.
Wollmer, M., Eyben, F., Graves, A., Schuller, B., & Rigoll, G. (2010). Bidirectional LSTM networks for context-sensitive keyword detection in a cognitive virtual agent framework. *Cognitive Computation, 2*(3), 180–190.
Wu, D. C., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management, 29*(1), 507–529.
Wu, Y., Yuan, M., Dong, S., Lin, L., & Liu, Y. (2018). Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing, 275*, 167–179.
Yang, X., Pan, B., Evans, J. A., & Lv, B. (2015). Forecasting Chinese tourist volume with search engine data. *Tourism Management, 46*, 386–397.
Yildirim, O. (2018). A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. *Computers in Biology and Medicine, 96*, 189–202.
Yildirim, O., Uçar, A., & Baloglu, U. B. (2017). *Recognition of real-world texture images under challenging conditions with deep learning.* ASYU52.
Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35–62.
Zhou, Y., Chang, F. J., Chang, L. C., Kao, I. F., & Wang, Y. S. (2019). Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *Journal of Cleaner Production, 209*, 134–145.
Zhu, L., Lim, C., Xie, W., & Wu, Y. (2018). Modelling tourist flow association for tourism demand forecasting. *Current Issues in Tourism, 21*(8), 902–916.

Anurag Kulshrestha is a PhD student in Information Technology & Systems area at Indian Institute of Management Kashipur, India. His research interests include tourism demand forecasting approaches and AI based models. E-mail: anurag.kulshrestha23@gmail.com and anurag.fpm1802@iimkashipur.ac.in.

Venkataraghavan Krishnaswamy is an Assistant Professor of Information Technology & Systems at Indian Institute of Management Kashipur, India.

Mayank Sharma is an Assistant Professor of Information Technology & Systems at Indian Institute of Management Kashipur, India.