Expert Systems With Applications 205 (2022) 117662

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Flight delay prediction from spatial and temporal perspective

Qiang Li *, Ranzhe Jing

School of Information Management & Engineering, Shanghai University of Finance and Economics, Shanghai 200433, PR China

ARTICLE INFO

Keywords:
Flight delay prediction
Complex network theory
LSTM approach
ST-Random Forest

ABSTRACT

In this paper, we propose a novel prediction framework (ST-Random Forest) for flight delay prediction from temporal and spatial perspective. We first apply complex network theory to extract the spatial feature of the aviation network at edge-, node-, and network-level. Furthermore, considering the temporal correlation of weather condition and airport crowdedness on flight delays, we create a prediction framework based on LSTM units to extract the temporal property of crowdedness and weather condition. Finally, we use the factors (e.g., spatial, temporal and extrinsic) that affect flight delays as inputs and apply Random Forest as classifier to predict flight delays. We apply and test our approach in a case study at China domestic flights between Jun and Aug 2016; after evaluation, we find that the accuracy of our proposed model reaches 92.39%. For the on-time samples, approximately 86% are correct identified; for the delayed samples, the classification accuracy reaches 95%. The ST-Random Forest model contributes to aviation authorities and airport regulators by creating real-time monitoring and high accuracy prediction system to alleviate flight delays and providing insightful suggestions to develop effective air traffic control strategies.

1. Introduction

The rapid growth of the air traffic demand and the limited airspace have brought pressure on the air transportation system and increased the frequency of flight delays (Pyrgiotis, Malone, & Odoni, 2013). According to the report of Civil Aviation Administration of China (CAAC), about 40% commercial flights are delayed every day. Flight delays have become one of the major issues that passengers complain about (Variflight, 2018). Furthermore, according to the estimation by Cook and Tanner (2015), the total delay cost of the departure flights in Feb 2018 in Beijing Capital International airport was up to 55 million euros. Those negative impacts have motivated the development of an intelligent delay management system to reduce consequent flight delays.

The motivation for this work is to promote the existing delay prediction model by considering the open innovation variable (Baierle, Benitez, & Nara, 2020), such as the spatial feature of the aviation network and temporal correlation of airport crowdedness, to develop a high accuracy prediction model from the spatial and the temporal perspective. A high accuracy prediction model is important and commonly known as an essential tool to monitor flight delays in real-time and help the system operator in forecasting the future condition of flights (Thiagarajan, Srinivasan, Sharma, Sreekanthan, & Vijayaraghavan, 2017), which in turn benefits the reduce of passengers' dissatisfaction and economic losses of the airline by optimizing flight schedule (Shumsky, 1997).

Despite the advance of existing approaches in flight delay prediction, most factors that have been explored in existing studies for delay prediction are at the macro-level, such as carrier or airport type (Alonso & Loureiro, 2015; Khanmohammadi, Chou, Lewis, Elias, & ., 2014), temporal variables (Gui et al., 2020; Lambelho, Mitici, & Pickup, 2020) and seasonal effects (Rebollo & Balakrishnan, 2014; Tu, Ball, & Jank, 2008). However, some macro-level factors may not be directly associated with real-time flight delays unless broken down into more detailed subfactors (Yu, Guo, & Wang, 2019). Therefore, this study replaces the macro-level factor with more specific micro influential factors that directly affect flight delays at the operational level. In addition, from the spatial perspective, congestion of the airport (number of flights in the airport) and how the structure of the aviation network (topological property of aviation network) allows for responding to the congestion are the two primary determinants of flight delays (Li & Jing, 2021; Stefan & Christian, 2019). Rebollo and Balakrishnan (2014) also suggested that network effects of the aviation system were expected to be a good indicator to predict flight delays. Consequently, it is necessary to extract the spatial feature of the aviation network and embed them into the prediction model to improve the prediction accuracy. At last, flight delays are associated with past and current congestion and weather

* Corresponding author.
E-mail address: liqiang@163.sufe.edu.cn (Q. Li).

https://doi.org/10.1016/j.eswa.2022.117662
Received 24 October 2021; Received in revised form 8 May 2022; Accepted 27 May 2022
Available online 1 June 2022

Table 1
Summary of most relevant studies of flight delay prediction.

Reference	Methodology	Considered factors
Tu et al. (2008)	Expectation Maximization algorithm	Seasonal trend, daily propagation pattern and random residuals
Wu (2014)	Generalized Extreme Value (GEV) model	Number of flights, average flight delays
Khanmohammadi et al. (2014)	Adaptive network	Origin airport, departure delays, departure time, scheduled arrival time
Alonso and Loureiro (2015)	Unimodal model	Arrival delay (in minutes), origin and destination of the flight, weekday, hour, day and month of the flight, meteorological conditions, aircraft type, aircraft parking stand, ground operation time (in minutes) and take-off runway
Belcastro et al. (2016)	Random forest	Origin and destination airport, scheduled departure and arrival time, weather condition in origin and destination airports
Thiagarajan et al. (2017)	Gradient Boosting, random forest, extra-trees, AdaBoost	Origin and destination airport, quarter of year, month, time-of-day, day-of-week, scheduled departure and arrival time, weather condition at destination airport
Pamplona, Weigang, Barros, Shiguemori, and Alves (2018)	Neural Network	Airline, route type code, departure airport, arrival airport, scheduled day and hour of departure, scheduled day and hour of arrival
Yu et al. (2019)	Deep belief network	Time-of-day, day-of-week, month-of-year, number of passengers, aircraft capacity, air route situation, airline properties, boarding option, origin or pass-by flight, flight terminal, gap between check-in time, scheduled departure time, closing time of gate, closing time of cargo-hold door, and ready time of shuttles or jet bridge etc.
Lambelho et al. (2020)	LightGBM, multilayer perceptron and random forest	Airline, route type code, departure airport, arrival airport, scheduled day and hour of departure, scheduled day and hour of arrival
Guan et al., (2020)	Random forest and LSTM	Weather condition of departure/arrival airport, day of month, month, day of week, season, departure airport, scheduled time of departure, destination airport, scheduled time of arrival
This work	ST-Random forest	Spatial features of the aviation network, the temporal correlations of weather conditions and airport/aviation network crowdedness, delay of previous flights, scheduled turnaround time, average turnaround time, distance, scheduled fly time and average fly time

condition in the proximal area or the whole airspace (Mueller & Chatterji, 2002). For instance, congestion at a particular airport will increase flight delays in the current and propagate congestion to the next moment as time goes by (Wu, 2014). Consequently, considering the temporal correlation of relevant features is critical for improving the accuracy of the prediction model.

This study focuses on predicting flight delays from spatial and temporal perspectives and can be distinguished from the existing research in the following aspects. First, we build an aviation network which is defined as a graph with nodes and edges (Lordan, Sallan, Simo, & Gonzalez-Prieto, 2014), where nodes represent airports and edges represent flights (Du, Zhang, Zhang, Cao, & Zhang, 2018) to extract the spatial feature of the aviation network based on the complex network theory at node-, edge-, and network specific. Second, we propose a novel framework that considers the temporal effect of the recent and current congestion of the airport and air traffic system, especially temporal effect of the recent, current and future weather condition on flight delays. We divide the associated temporal factors into four categories: Airport crowdedness, aviation crowdedness, past weather conditions and future weather conditions. Four types of factors in each time fragment are then fed into four architectures based on LSTM units to model the temporal property. Using operational data obtained from China domestic flights, we conduct extensive sets of experiments to examine the performance of our proposed model. The prediction results indicate that our model noticeably improves the delay prediction accuracy from 0.76 to 0.92 compared with eight baselines constructed by macro-level factors and different classifiers.

The remainder of the paper is organized as follows: Section 2 summarizes the related literature on flight delay prediction. Section 3 describes the details of the data set being used in this work, followed by demonstrating of the proposed method ST-Random Forest. In Section 4, we report the results of the proposed model and discuss the results. Section 5 concludes the results and future discussions.

2. Literature review

Flight delays are a critical element of air traffic network efficiency and have been extensively investigated in literatures. According to our survey, the majority of the prior studies of flight delay prediction methodologies can be classified into three categories: The probability of flight delay estimation, or in other words, assessing the distribution of flight delays. Statistical modeling approach to estimate the main factor affecting flight delays and machine learning techniques to predict flight delays.

From the probability of flight delay estimation perspective, Mueller and Chatterji (2002) explored the characteristic of arrival and departure delays and concluded that departure delays were approximately fitted by Poisson distribution while arrival delays were better modeled by Normal distribution. Meanwhile, Tu et al. (2008) decomposed the observed flight delays into daily propagation patterns, seasonal trends, and random residuals to model the distribution of flight delays. Data from the Denver International Airport in Jan. 2000 was utilized for experiments, and results indicated that 90.34% of the validation data fell into a 90% confidence interval. Wu (2014) further proposed an optimal generalized extreme value (GEV) model to estimate the probabilistic distribution of arrival and departure delays in Beijing Capital International airport. Historical delay data in Aug. 2012 was utilized for the experiment, and results proved the good fitness of their proposed model.

Following the statistical modeling approach, researchers also estimate the main factors affecting flight delays using regression model, correlation analysis and econometric model (Allan, Beesley, Evans, & Gaddy, 2001; Mofokeng & Marnewick, 2017). For example, Reynolds-Feighan and Button (1999) explored the impact of air traffic control factors, demand characteristics and environmental conditions on flight delays in European airports by using a correlation analysis approach. Results indicated that traffic levels had the most significant association with flight delays. Abdel-Aty, Lee, and Bai (2007) considered the features such as the maximum hourly flow rate, arrival demand, aircraft departures and the weather condition of the airport and designed a multinomial logistic regression approach to estimate the impact of those features on flight delays at Orlando International Airport during 2002–2003. Xiong and Hansen (2013) considered the features such as distance, pre-existing delays, characteristics of the destination airport and airline, potential delay savings, frequency, aircraft size, and applied a binary choice model to estimate the main reasons for the airline to cancel their flights. Results indicated that the influence of delays on flight cancellation utility was non-linear, and cancellation probability was impacted at an increasing rate between 15 min and 90 min, while when delays were above 90 min, cancellation probability was impacted by a decreasing rate.

Since flights operate in an interconnected air transportation network, each flight will be affected by the air traffic network effect (Du et al., 2018; Kafle & Zou, 2016). Consequently, Kwan and Hansen (2011) presented an econometric model to model flight delays at 10 airports with the worst on-time statistics in 2007 using airport congestion, total traffic, and route weather. Experimental results suggested that airport congestion had a major contributor to flight delays.

It is apparent that the modeling approaches mentioned in the study above can offer explanations of the underlying behavior of flight delays. Despite their success, these methods present shortcomings when predicting the potential flight delays in real-time because of their limitations in dealing with the high-volume and high-dimensional data and extracting nonlinear relationships (Wang, Brownlee, Woodward, Weiszer, & Chen, 2021; Yin, Hu, Ma, Yan, & Dan, 2018). Given the excellent performance of machine learning techniques in various fields, many tries have been applied to predict flight delays (Belcastro, Marozzo, Talia, & Trunfio, 2016). Balakrishna, Ganesan, and Sherry (2010) presented a nonparametric reinforcement learning model for taxi time prediction. Data from Tampa International Airport (TPA) from Jun 1, 2007 to Aug 25, 2007 was utilized for experiments, the scheduled and actual push back time, actual wheels off time, actual wheels on time and seasonal average taxi-out/taxi-in time were considered for the prediction model. Rebollo and Balakrishnan (2014) established a flight delay prediction model with a Random Forest algorithm to predict departure delays from 2 to 24 h in the future. Several features including time-of-day, day-of-week, month-of-year, delay state and type of delay day were identified, and a high prediction accuracy up to 88.7% when classifying delays as above or below 60 min. Khanmohammadi et al. (2014) proposed an adaptive network based on a fuzzy inference system to predict flight delays. Data from the John F. Kennedy International Airport (JFK) was utilized for the experiments, and the day-of-month, scheduled arrival time, origin airport and departure time were used as the input feature for prediction. The prediction results were further used as inputs of a fuzzy decision-making procedure to schedule aircraft landings.

Following a multifactor approach, Yu et al. (2019) designed a novel deep belief network model to predict flight delays at PEK International airport. Input features including air route situation, delay of previous flight, airline properties and air traffic control were used for experiments. Results of the proposed model demonstrated that 99.3% of the test errors were within 25 min. Recently, Lambelho et al. (2020) focused on lightGBM, multilayer perceptron (MLP), and random forest (RF) methods for flight delay prediction at London Heathrow Airport (LHR). The feature such as airline type, distance, time-of-day, day-of-week, month-of-year and number of seats of the aircraft assigned to a flight were chosen and the result indicated that using LightGBM showed the most efficient prediction. Other prediction models, Sridhar and Chen (2009), Klein, Craun, and Lee (2010), focused on the development of a Weather Impacted Traffic Index (WITI) for flight delay prediction.

As shown in Table 1, the majority of the prior studies for flight delay prediction mainly considered the macro-level factor. However, several features that may affect flight delays have not been comprehensively explored especially the spatial characteristics of the aviation network. Meanwhile, the temporal correlation of some features such as weather conditions, crowdedness of the airport has not been investigated. In theory, flight delays in an airport are affected by recent time intervals, both near and far. For instance, one aircraft delay in the current time might prevent using an airport gate by another. Consequently, considering the spatial characteristics of the aviation network and the temporal correlations of some features are essential for flight delay prediction. In this work, we apply a ST-Random Forest model to predict flight delays. By considering the characteristics of the aviation network from the spatial perspective, this approach can capture the network effect of flight delays in the air traffic system. Furthermore, due to the time series effect of airport crowdedness, aviation crowdedness and weather condition on flight delays, the influence of these features cannot be understood from the current level along. Therefore, we construct an LSTM model to extract the temporal and dynamic influence of these features. The ST-Random Forest model are compared with eight benchmark models and the result indicates that the prediction accuracy can be improved significantly by considering the spatial–temporal effect. Our proposed prediction model ST-Random Forest helps the airport managers in understanding the flight delays in real-time, and provides practical support to airport managers on flight scheduling optimization and decision-making.

Table 2
Information included in the Flight Records.

Type of data	Information
Airports	Airports describe the origin and destination of the flights
Tail number	Identify a unique aircraft
Scheduled time	Scheduled time describes the departure and arrival time of the flight in plan, i.e., 08:10 am
Actual time	Actual time describes the departure and arrival time of the flight in actual, i.e., 08:15 am
Departure date	Departure date of the flight and week, i.e., 2016/06/01, Sunday
Carrier ID	An identification number to identify a unique airline
Distance	Distance between airports (miles)
Weather condition	Temperature, humidity, precipitation, pressure, wind speed and visibility

3. Method

In this section, the data and the prediction framework (ST-Random Forest) for flight delay prediction are introduced for the subsequent research.

3.1. Data

This study collected the departure and arrival flights crossing the Chinese airspace between Jun and Aug 2016 from VariFlight (https://data.variflight.com), the largest online platform maintaining real-time flight status data in China. The data set contains 762,415 samples connecting 260 airports. The real data includes multitude of attributes, such as origin and destination airports, airlines, tail number, scheduled departure/arrival time and actual departure/arrival time and weather condition. The average delay of all flights during the study period is 40.47 min. In this study, the data set is randomly divided into three sub-datasets: 80% for model training, 10% for validation, and the remaining 10% for model testing. Table 2 illustrates the detailed information of each flight.

The distribution of flight delays of all domestic flights from Jun to Aug 2016 is depicted in Fig. 1. We noticed that only 11.97% of flights arrive earlier than the scheduled time, approximately 32.84% of flight delays are kept within 15 min, while 55.19% of flight delays are over 15 min. In this paper, 15 min is used as the threshold for determining flight delays (the 15 min threshold for defining delay has also historically been common to both Europe and the US (Jetzki, 2009; Cook, Tanner,

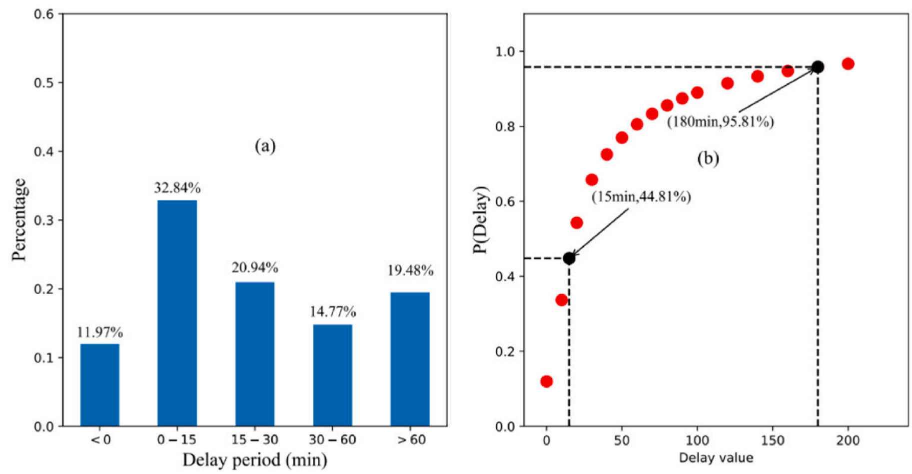

Fig. 1. Distribution of flight delays.

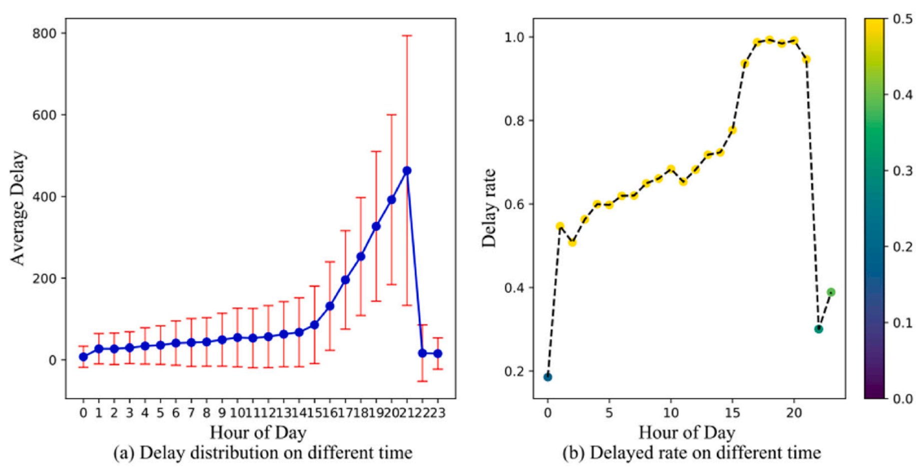

Fig. 2. The effect of temporal features on flight delays.

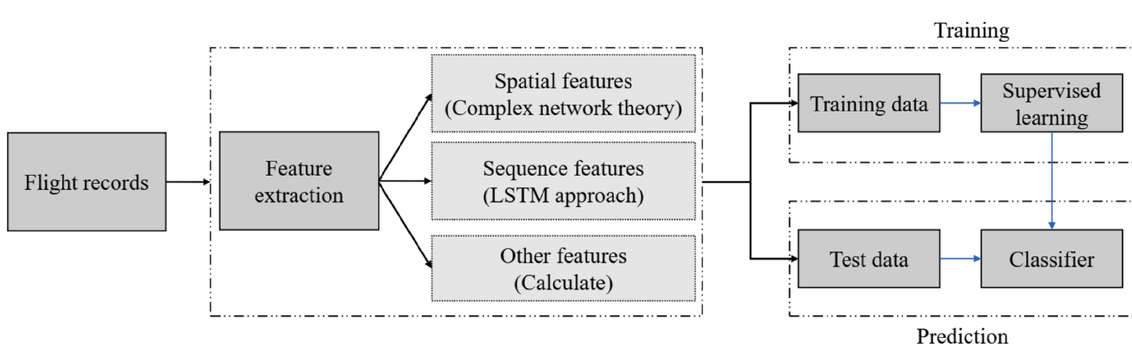

Fig. 3. Algorithm flow chart of the proposed model.

Cristóbal, & Zanin, 2012)).

The changes of flight delays within one day (averaged by all days) and the delay rate are shown in Fig. 2. From the result in Fig. 2, we discover that flight delays and delay rate (Proportion of delayed flights to total flights) reach their highest value at 9 pm. Another interesting finding described in Fig. 2 is that flight delays progressively increase first and then decrease as time goes by, indicating that flight delays tend to accumulate during the day and are mainly recovered during nighttime. Pyrgiotis et al. (2013) explained that this phenomenon was caused by delay propagation, and they found that delay propagation tended to "smoothen" daily airport demand and push more demands into late evening hours. The accumulated demand profiles further increased flight delays in late evening hours. So, flight delays have an empirical temporal correlation in time series; that is, flight delays of current time intervals are relevant to recent demand profiles. Therefore, in this study, we propose the ST-Random Forest model to capture the temporal property of demand profiles.

3.2. Prediction Model: ST-Random Forest

In this section, we introduce the details of our proposed model namely ST-Random Forest for flight delay prediction. The algorithm

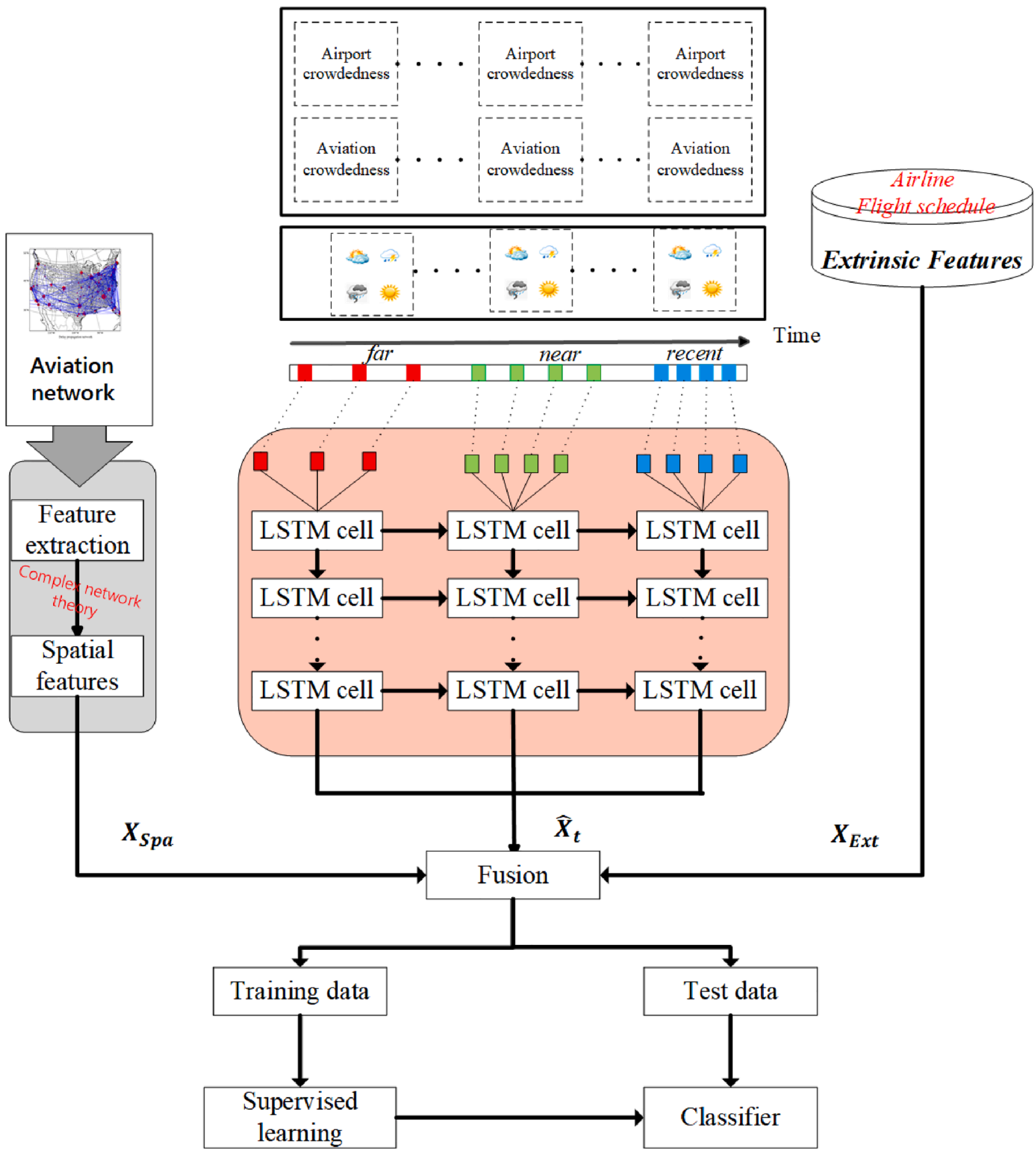

Fig. 4. The architecture of the ST-Random Forest.

flow chart of the proposed model is shown in Fig. 3 and the architecture of the ST-Random Forest is presented in Fig. 4.

As shown in Fig. 4, the ST-Random Forest is a two-stage approach that consists of three main components.

- Aviation network features extraction;
- Capturing the temporal correlation of features;
- Classifier: Random Forest.

In the first stage, we create an aviation network based on flight data and apply the complex network theory to extract the spatial feature of the aviation network, we also apply the LSTM model based on time series data of congestion and weather condition to extract the temporal effect. In the second stage, we use the spatial features and the temporal effect extracted in the first stage and the extrinsic features such as airline issues and flight schedule, and apply the Random Forest as classifier for flight delay prediction.

The detailed description of the three main components is as follows.

3.2.1. Aviation network features extraction

The aviation network is an extensive, complex system consisting of many airports and complex interactions. Consequently, the spatial characteristic of the aviation network cannot be illustrated from the information of the airport level alone. In the current research, an aviation network is widely defined as a graph with nodes and edges, where nodes represent airports and edges represent flights (Guimera, Mossa, Turtschi, & Amaral, 2005; Lordan et al., 2014). In this work, we model the aviation network as a weighted directed graph $G = (\mathrm{N}, \varepsilon)$, where N is

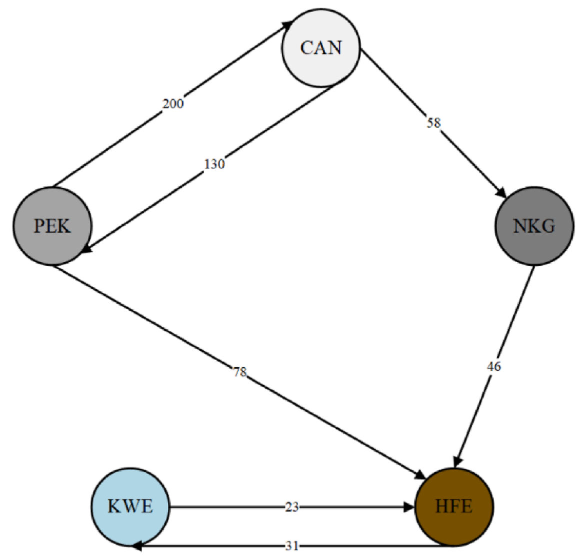

Fig. 5. A sample aviation network.

the set of airports and ε is the set of flights. We define $A = (a_{ij})_{N*N}$ as the adjacency matrix of the weighted directed aviation network, where $a_{ij} = w$ represents there exists w flights from airport i to j; otherwise, $a_{ij} = 0$ (Du et al., 2018). Fig. 5 shows a sample aviation network which contains five airports and seven directed flights. For edge *PEK*→*CAN*, $w = 200$ represents there exists 200 flights from airport *PEK* to airport *CAN*.

In this study, we apply the complex network theory and its associated metrics and tools to extract the spatial feature of the aviation network at node-, edge- and network-level.

Degree of an airport i reflects the number of flights arrive to and depart from airport i, which is defined as $k_i = k_i^{in} + k_i^{out}$. Among them, $k_i^{in} = \sum_{j=1}^{N} a_{ji}$ is the number of flights arrive to airport i and $k_i^{out} = \sum_{j=1}^{N} a_{ij}$ is the number of departure flights. Therefore, k_i represents the crowdedness degree of airport i. Equivalently, the crowdedness degree of the aviation network is defined as $K = \sum_{i=1}^{N} k_i$, which reflects the number of flights in the aviation network.

In this study, we introduce two types of airport crowdedness ($B^s_{t-nh,t-(n-1)h}(i)$, $B^a_{t-nh,t-(n-1)h}(i)$), that is, the scheduled and actual crowdedness, which is quantified as the number of scheduled flights and actual flights in a unit time interval as shown in Eq. (1) and Eq. (2).

$$B^s_{t-nh,t-(n-1)h}(i) = k^s_{t-(n-1)h}(i) - k^s_{t-nh}(i) \tag{1}$$

$$B^a_{t-nh,t-(n-1)h}(i) = k^a_{t-(n-1)h}(i) - k^a_{t-nh}(i) \tag{2}$$

The aviation crowdedness in a unit time interval is described in Eq. (3) and Eq. (4), which quantified the number of scheduled flights or actual flights in the aviation network.

$$C^s_{t-nh,t-(n-1)h}(i) = K^s_{t-(n-1)h}(i) - K^s_{t-nh}(i) \tag{3}$$

$$C^a_{t-nh,t-(n-1)h}(i) = K^a_{t-(n-1)h}(i) - K^a_{t-nh}(i) \tag{4}$$

where t represents the departure time, $B^s_{t-nh,t-(n-1)h}(i)$ and $C^s_{t-nh,t-(n-1)h}(i)$ represent the number of flights in airport and aviation network during time interval $[t-nh, t-(n-1)h]$. In order to reflect the influence of the past and current crowdedness on individual airport, in this study, we set $n = \{1, 2, \cdots, 12\}$. For instance, $B_{t-2h,t-1h}(i)$ measures the number of flights in airport i between $t-2h$ and $t-1h$.

Density is the ratio of the number of edges in the network to the maximum possible number of edges in the network. This indicator is used to describe the density of the aviation network. In theory, a larger density of the aviation network should result in a higher probability of flight delays. The definition of density is shown in Eq. (5).

$$D_t = \frac{m}{n(n-1)}, t \in \{0, 1, 2, \cdots, 23\} \tag{5}$$

where m is the number of flights and n is the number of airports. D_t represents the density of the aviation network during $(t, t+1)$.

Edge betweenness centrality of a link e is defined as Eq. (6), which illustrates the sum of the fraction of all-pairs shortest paths that pass through e (Girvan & Newman, 2002).

$$g_t(e) = \sum_{s,t \in V} \frac{\sigma(s, t/e)}{\sigma(s, t)}, t \in \{0, 1, 2, \cdots, 23\} \tag{6}$$

where V is the set of nodes, $\sigma(s, t)$ is the number of shortest (s, t) *−paths*, and $\sigma(s, t/e)$ is the number of those paths passing through link e. Edge betweenness centrality quantifies the importance of flight route e in aviation network during $(t, t+1)$ by assessing how often it appears in shortest paths.

3.2.2. Capturing the temporal correlation of features

As mentioned in Introduction, the current flight delay can be affected by current and recent factors (e.g., Flight delays occurred in the past moments would not only increase the congestion of the airport, but also propagate the congestion to the next moment), which means some features exhibit strong temporal correlations on flight delays. In this study, we present a Recurrent Neural Networks (RNNs) to capture temporal near dependencies of relevant features. RNNs are called recurrent because they perform the same task for every element of a sequence, with the output depends on the previous computations, which means RNNs has a "memory" that can capture the information about what happened in the past (Ahmet, 2021; Graves, Mohamed, & Hinton, 2013; Shen, Lee, Liu, Chang, & Yang, 2021).

As shown in the middle corner of Fig. 4, we consider three temporal correlations on historical data, including airport crowdedness, aviation crowdedness, and weather condition. Airport crowdedness refers to sample data at a given time interval (e.g., an hour) from the recent number of flights. For instance, at a given time t, we sample the number of flights at airport i at time interval $[t-12h, t-11h]$, $\cdots$, $[t-1h, t]$. Similarly, the aviation crowdedness and weather condition indicate that sampling is performed at number of flights in the aviation network and meteorological data, respectively. For weather condition, we also consider the influence of future weather condition, that is, we sample meteorological data at a given time interval from the future meteorological data (e.g., considering the weather effect at $t+1, t+2, ..., t+n$ on current time t). In this work, we assume that we can use the forecasting weather at future time. Therefore, the temporal information of the airport crowdedness (I_a), aviation crowdedness (I_b) and weather condition (I_{wb}, I_{wf}) are defined as Eq.(7)-(10).

$$I_a = \left[B_{t-nh,t-(n-1)h}, B_{t-(n-1)h,t-(n-2)h}, \cdots, B_{t-h,t}\right] \tag{7}$$

$$I_b = \left[C_{t-nh,t-(n-1)h}, C_{t-(n-1)h,t-(n-2)h}, \cdots, C_{t-h,t}\right] \tag{8}$$

$$I_{wb} = \left[W_{t-nh}, W_{t-(n-1)h}, \cdots, W_t\right] \tag{9}$$

$$I_{wf} = \left[W_{t+nh}, W_{t+(n-1)h}, \cdots, W_t\right] \tag{10}$$

where n is the input length of $I_*(* \in \{a, b, wb, wf\})$. $I_a = [[B^s, B^a]] \in \mathbb{R}^{n\times 2}$, $I_b = [[C^s, C^a]] \in \mathbb{R}^{n\times 2}$ and $W = \left[\left[w^{temperature}, w^{wind\ speed}, \cdots, w^{visibility}\right]\right] \in \mathbb{R}^{n\times 6}$.

After creating the input tensor, we then fed them into four LSTM structures to model the aforementioned temporal properties: airport crowdedness, aviation crowdedness, and weather condition, respectively. The detailed information of this process is as follows.

3.2.2.1. Long Short-Term memory model (LSTM). The idea behind LSTM is to use sequential information (Evermann, Rehse, & Fettke, 2017). In

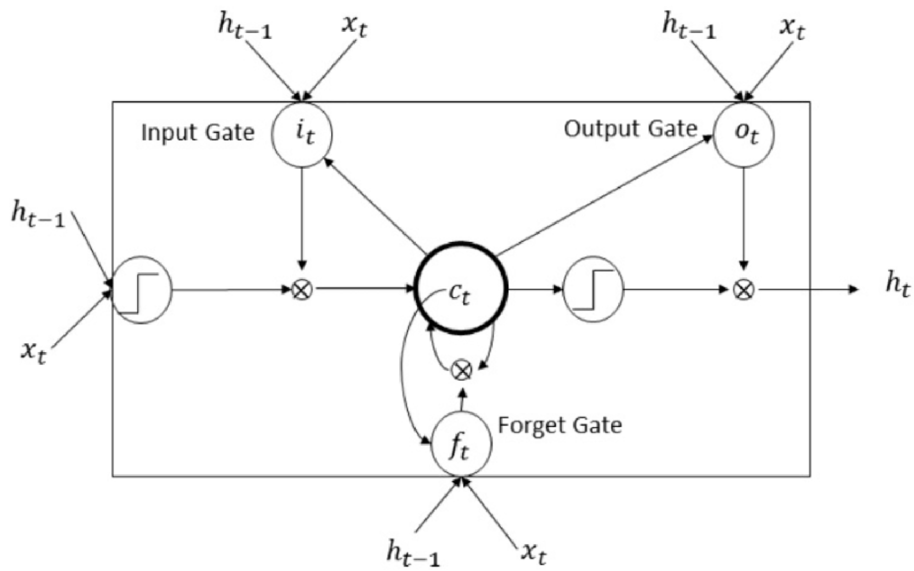

Fig. 6. Long Short-term Memory Cell (Graves et al., 2013).

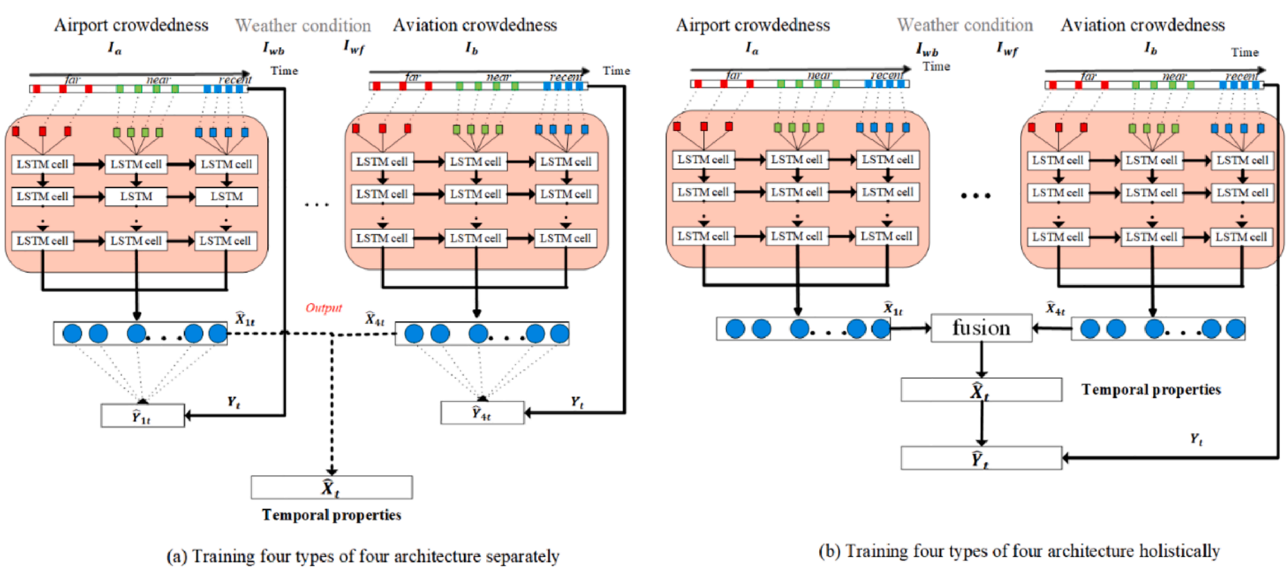

Fig. 7. The architecture for extracting the temporal properties of crowdedness and weather condition.

general, flight departure status at time t depends not only on the current weather condition and crowdedness, but also on the past condition. Given an input sequence: $X = (x_{t-nh}, x_{t-(n-1)h}, \cdots, x_t)$ which represents the crowdedness or weather condition of the airport before n hours.

In this work, we use the long-short term memory (LSTM) architecture to capture the temporal correction. LSTM is a special Recurrent Neural Network (RNN) proposed by Hochreiter and Schmidhuber (1997) as a remedy to the vanishing gradient problem of traditional RNN. The processing framework of LSTM is shown in Fig. 6.

The LSTM consists of four parts: Input gate (i), forget gate (f), output gate (o) and cell activation vectors (c), all of which are the same size as the hidden vector h. The value of the input gate and the input unit of all memory cells in LSTM layer are given in Eq. (11) - (15).

$$i_t = \varnothing(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \tag{11}$$

$$f_t = \varnothing\left(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \tag{12}$$

$$c_t = f_t c_{t-1} + i_t tanh(W_{xc}x_t + W_{hc}h_{t-1} + W_{ci}c_{t-1} + b_c) \tag{13}$$

$$o_t = \varnothing(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \tag{14}$$

$$h_t = o_t \tanh(c_t) \tag{15}$$

In this paper, $\varnothing$ represents the *ReLu* function and the calculation formulas of *ReLu* function and *tanh* function are shown in Eq. (16) and Eq. (17).

$$R\ eLu(x) = max\{0, x\} \tag{16}$$

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{17}$$

As mentioned before, we first divide associated temporal features into four categories: Airport crowdedness, aviation crowdedness, past weather condition, and future weather condition. To extract the temporal property of crowdedness and weather condition, we create two different architectures, as shown in Fig. 7. In the first way, as shown in Fig. 7 (a), the four types of features in each time fragment are fed into four architectures separately to model the temporal property. The four architectures share the same network structure with a recurrent neural network followed by a LSTM unit and then connected with a dense layer. In the training step, the structure described in Fig. 7 (a) trains four architectures separately, that is, for each architecture, we input the temporal information of the airport crowdedness (I_a), aviation crowdedness (I_b) and weather condition (I_{wb}, I_{wf}) into four architectures respectively. The outputs of the four architectures are defined as $\widehat{Y}_{i,t}$. We learn the parameter in the four architectures by minimizing the mean squared error between the predicted flight delay ($\widehat{Y}_{i,t}$) and the true flight delay (Y_t).

In the prediction step, the first way (as shown in Fig. 7 (a)) inputs the airport crowdedness (I_a), aviation crowdedness (I_b) and weather condition (I_{wb}, I_{wf}) into four architectures and obtains the $\widehat{X}_{1,t}$, $\widehat{X}_{2,t}$, $\widehat{X}_{3,t}$, $\widehat{X}_{4,t}$ respectively. Then $\widehat{X}_{1,t}$, $\widehat{X}_{2,t}$, $\widehat{X}_{3,t}$, $\widehat{X}_{4,t}$ is fused as $\widehat{X}_t$ which is further integrated with the spatial feature (X_{Spa}) and external feature (X_{Ext}) as shown in Fig. 4 to create the input of the classification model. The second way (as shown in Fig. 7 (b)) inputs $[I_a, I_b, I_{wb}, I_{wf}]$ into the multi-input model and then obtains the output of the fused layer $\widehat{X}_t$. This result is further integrated with the spatial feature (X_{Spa}) and external feature (X_{Ext}) to predict flight delays.

3.2.3. Classifier: Random Forest

As stated above, we create three types of features (spatial features, extrinsic features and temporal properties) for flight delay prediction. In this section, we introduce the classifier, Random Forest algorithm. The Random Forest model (RF) is an ensemble classifier that uses decision trees as sub-classifiers. For each decision tree, only a subset of features are selected from the full attribute set. In the training step, RF first draws a bootstrap sample Z^* from the training data, and then grows a decision tree T_b by using the bootstrapped data. To increase the ensemble diversity, only a fraction of the total number of features are selected and split at each node by minimizing the Gini criteria to build the tree. Once the decision tree is constructed, the ensemble classifier uses the majority voting results of the sub-classifiers to predict whether the flight will delay (Breiman, 2001).

4. Experimental observations

In this section, we introduce the detailed experimental setting, including the inputting feature, benchmark model, and demonstrate the advantage of the ST-Random Forest model in prediction performance.

4.1. Experimental settings: Potential explanatory variables

The present work predicts flight delays with spatial features, temporal properties and extrinsic features. To start with, we now summarize the factors used in this study for flight delays prediction.

Definition 1. (***Spatial features***): *Density of the air traffic system and importance of the route.*

Definition 2. (***Temporal properties***): *The temporal properties contain four categories: Airport crowdedness, aviation crowdedness, past weather conditions and future weather conditions, which are extracted by using the architectures as described in* Fig. 7.

Definition 3. (***Extrinsic features***):

1) Delay of previous flights. Due to the effects of delay propagation, the delay of an earlier flight operated by the same aircraft can affect the subsequent flight. In this paper, by tracking the tail number of aircraft, we extract the delay of previous flights from the collected data as an essential factor for flight delays.
2) Airline issues. Different airlines may operate differently, the airline operational capability can be measured by the turnaround time, which stands for the time spent by an aircraft on ground from arrival to departure from the gate (Fleurquin, Ramasco, & Eguiluz, 2013). In this paper, we consider two indicators to represent airline factors: Scheduled turnaround time and average turnaround time. Average turnaround time is the average actual turnaround time for each combination of airline, airport and departure time (to the hour).
3) Distance between origin and destination airport.
4) Scheduled fly time (the difference between the scheduled departure time and arrival time), and average fly time for each combination of flight segment, airline and departure time (to the hour).

Table 3
Parameter settings in LSTM.

Parameters	Value
No of layers	{1,2,3}
No of neurons	[64,32,32]
Time step	[1,2,3,…,12]
Epochs	50
Optimizer	Adam
Batch size	850
Activation function	Sigmoid
Learning rate	0.001

Definition 4. (***Flight delays***): *The magnitude of flight delay is defined as the difference between actual and scheduled departure time. Given a flight F and a threshold Th, F is a Delayed Flight when the magnitude of flight delays over Th. Otherwise, F is an on-time flight. In this work, we set Th as* 15 *min.*

Problem 1. (***Flight delay prediction***): *Given the observation, the temporal property* $\widehat{X}_t$, *the spatial feature* (X_{Spa}) *and external features* (X_{Ext}), *predict whether the flight will delay.*

4.2. Model structure selection

In this work, we design two different architectures to extract the temporal property of crowdedness and weather condition. In the first way, we train four architectures separately to model the temporal property. In the second way, we build a multi-input model that trains four architectures holistically to extract the temporal effect. Keras tuner is employed to identify the optimal hyper parameters and the final parameters of LSTMs is presented in Table 3.

We explore the performance of the two architectures to determine the appropriate structure of the prediction model. In this work, we apply the accuracy, precision, recall scores and Receiver Operating Characteristic curves (ROCs) to measure the performance of predictor. The Receiver Operating Characteristic curves (ROCs) are an effective technique for assessing the performance in binary classification, which illustrates the model's diagnostic ability by plotting the true positive rate against the false positive rate relative to a certain threshold. The prediction performance for different structure is shown in Fig. 8. Results indicate, creating a multi-input model and training four architectures holistically will improve the accuracy, precision and recall from 0.88, 0.72 and 0.78 to 0.89, 0.75 and 0.81. Therefore, we concatenate four types of four architectures into a single multi-input neural network in this work instead of training four types architectures separately.

4.3. Variable selection

The data set contains three categories potential factors that contributes to flight delays. For the temporal property, we consider the influence of past crowdedness or weather condition on current condition. For instance, we consider the influence of crowdedness of airport i during $\{[t-12h, t-11h], [t-11h, t-10h], \ldots, [t-1h, t]\}$, which contains 12 observations. However, if we consider a large observation may also contains some redundant information, that is, crowdedness of airport during $[t-11h, t-12h]$ may have insignificant impact on current flight

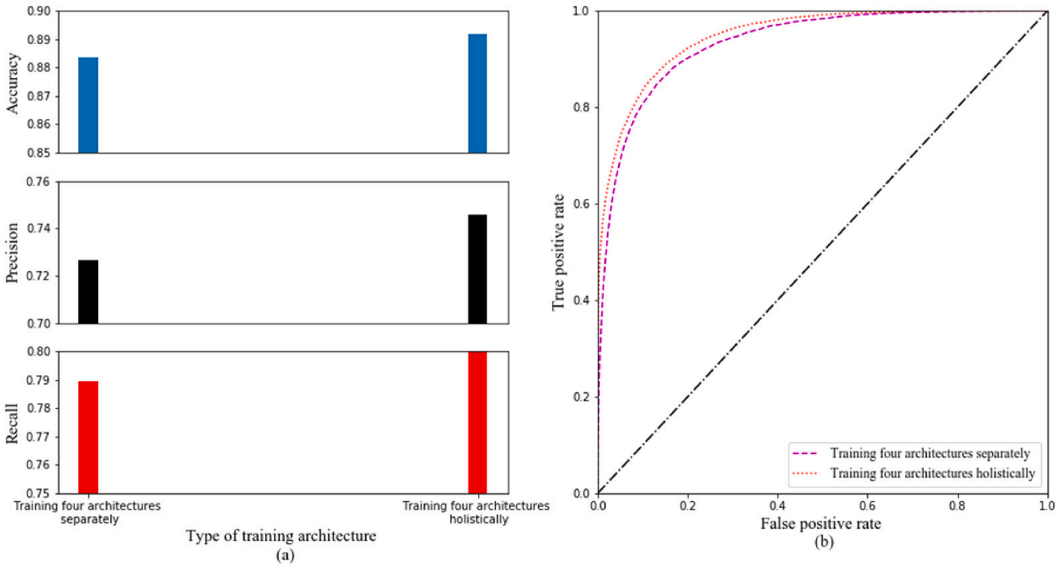

Fig. 8. Prediction performance *vs* model structure.

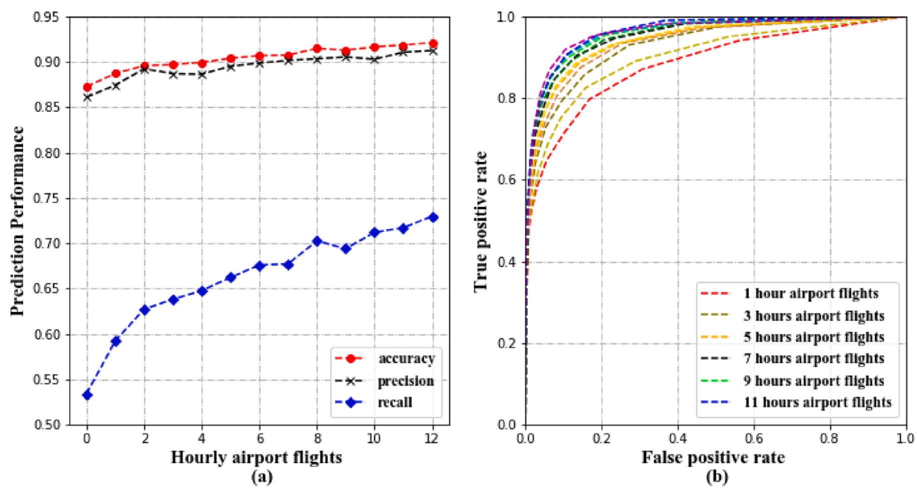

Fig. 9. Prediction performance *vs* number of airport crowdedness observations in airports.

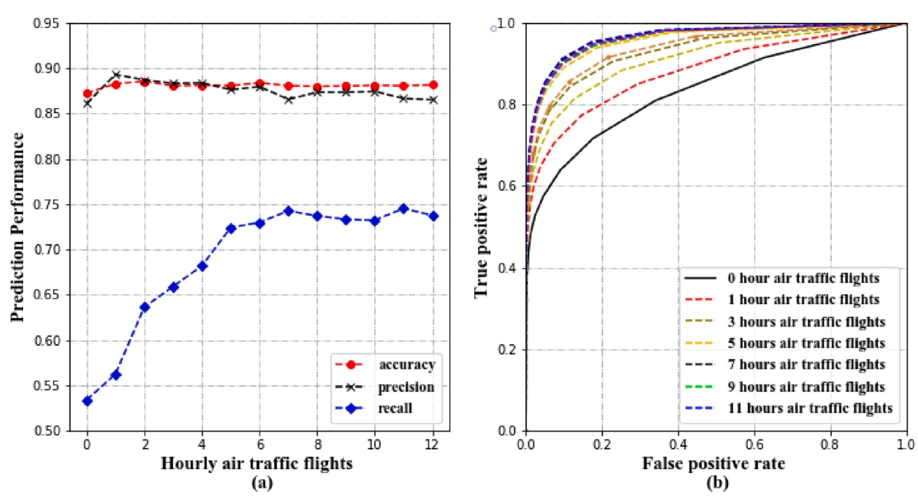

Fig. 10. Prediction performance *vs* number of aviation crowdedness observations.

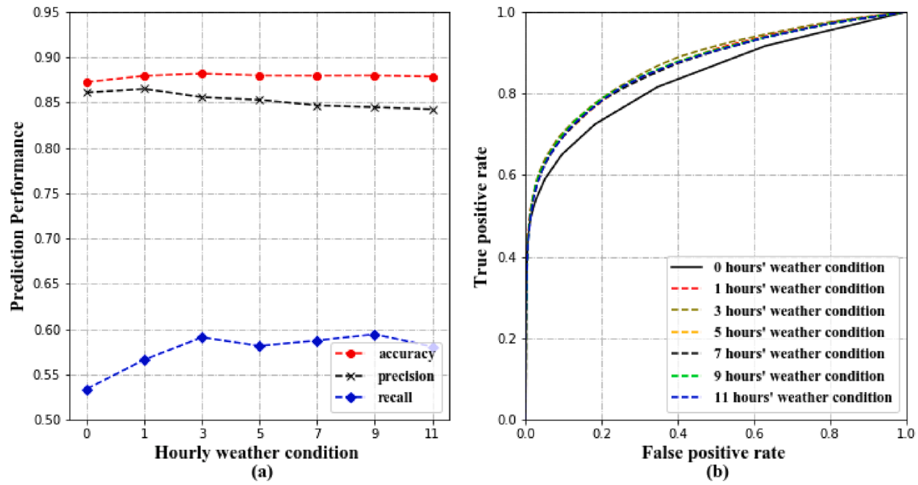

Fig. 11. Prediction performance *vs* number of previous weather observations in airports.

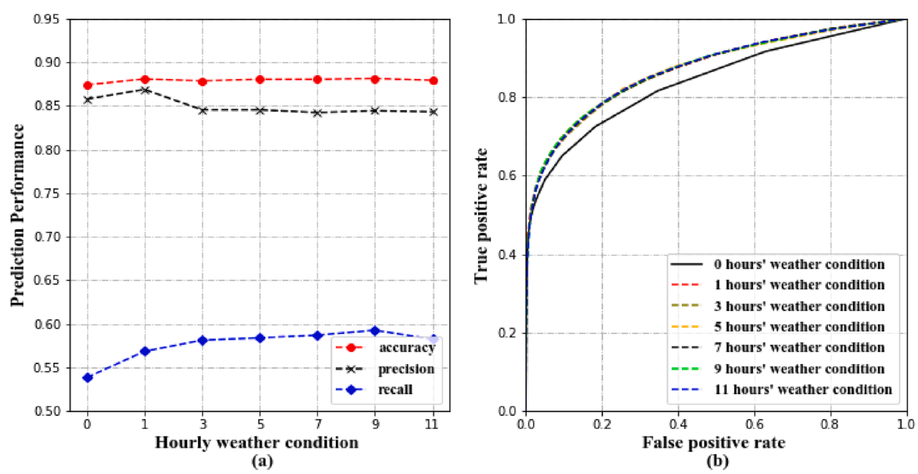

Fig. 12. Prediction performance *vs* number of future weather observations in airports.

delays. Therefore, in this section, we conduct the following experiment to determine the appropriate time intervals for prediction model.

- Number of observations for airport crowdedness;
- Number of observations for aviation crowdedness;
- Number of observations for weather condition.

The first experiment tries to identify the length of observations that has to be considered for the airport crowdedness. Fig. 9 shows the prediction performance obtained for length of observations of airport crowdedness varying from 0 to 12 h, and zero length of observations for aviation crowdedness and weather condition. As illustrated in Fig. 9, the prediction performance improved with increasing the length of observations. Therefore, 12 airport crowdedness observations in airports are considered in the proposed model.

The second experiment tries to determine the length of observations for aviation crowdedness. Fig. 10 shows the prediction performance obtained for length of observations of aviation crowdedness varying from 0 to 12 h, and zero length of observations for airport crowdedness and weather condition. As plotted in Fig. 10, the prediction performance indicates that using more than six observations does not significantly increase of the prediction performance. Therefore, six aviation crowdedness observations are considered in this work.

The third experiment evaluates the predictor performance by varying the length of weather observation, and zero observations are considered for airport and aviation crowdedness. Fig. 11 shows the weather condition observations in the airport from the departure time back to 12 h before. As shown in Fig. 11, we note that using more than three observations does not significantly increase the prediction performance. Therefore, three previous weather observations are considered in this work. Fig. 12 plots the weather observations in airport from the departure time 12 h later. Results indicate that one observation is the best choice.

Finally, the observations for airport crowdedness are 12, for aviation crowdedness observations are six, the number of previous and future weather observations in airports are considered as three and one, respectively.

4.4. Baselines

In this paper, we predict flight delays by using the spatial factor, temporal properties and extrinsic features. As a comparison, we future check the prediction result with macro-level factors and the influence of different classifiers on prediction results. According to our survey, the macro-level factors used in previous studies, including the origin and destination airports, airlines, day-of-month, day-of-week, scheduled departure and arrival time, which are modeled as a 0–1 dummy variable. Delay of previous flights and distance, weather condition of the airport at departure time. In this work, we compare our ST-Random Forest with the following eight baselines:

Table 4
Final Random Forest description.

Parameters	Value
No of estimators	150
Max depth of decision tree	10
Max feature of decision tree	auto
random forest type	classification

- **KNN-Macro**: Considering macro-level factors as inputs and using k-nearest neighbors (KNN) as a classifier to predict flight delays;
- **RF-Macro**: Considering macro-level factors as inputs and using random forest (RF) as a classifier to predict flight delays;
- **LR-Macro**: Considering macro-level factors as inputs and using linear regression (LR) as a classifier to predict flight delays;
- **NN-Macro**: Considering macro-level factors as inputs and using neural network (NN) as a classifier to predict flight delays;
- **ST-KNN**: Considering spatial features, temporal properties and extrinsic features as inputs and using k-nearest neighbors (KNN) as a classifier to predict flight delays;
- **ST-LR**: Considering spatial features, temporal properties and extrinsic features as inputs and using linear regression (LR) as a classifier to predict flight delays;
- **ST-NN**: ST-KNN, ST-LR and ST-RF are two-stage approaches that use the structure as shown in Fig. 4 to extract the spatial and temporal features, and then uses KNN, Logistic Regression and Random Forest as classifier to predict flight delays. ST-NN is an end-to-end model that uses a dense layer followed by usual *Sigmoid* classifier to predict flight delays.
- **S-T-Random Forest**: In this study, we design a Recurrent Neural Networks (RNNs) to capture the temporal near dependencies of airport crowdedness, aviation crowdedness and weather condition. As a comparison, this baseline indicates that we directly use the airport congestion, aviation crowdedness and weather condition as input features to predict flight delays. That is, we consider the crowdedness of airport during $[t-nh, t-(n-1)h]$ as input features for classifier.

4.5. Performance comparison

In this work, we use the scikit-learn (Pedregosa, Varoquaux, & Gramfort, 2011) to construct the Random Forest model. A grid-search strategy was used to adjust the training parameters and the final parameters are listed in Table 4.

Detailed experimental results of our proposed model are plotted in Fig. 13, where Fig. 13 (a) shows the ROC curve of prediction results, Fig. 13 (b) illustrates the confusion matrix of our proposed model.

The result in Fig. 13 shows that the accuracy of our proposed model reaches 92.39%. For the on-time samples, approximately 86% are correct identified; for the delayed samples, the classification accuracy reaches 95%. An area under the ROC curve (AUC score) reaches to 0.89, implying a high prediction performance of our proposed model.

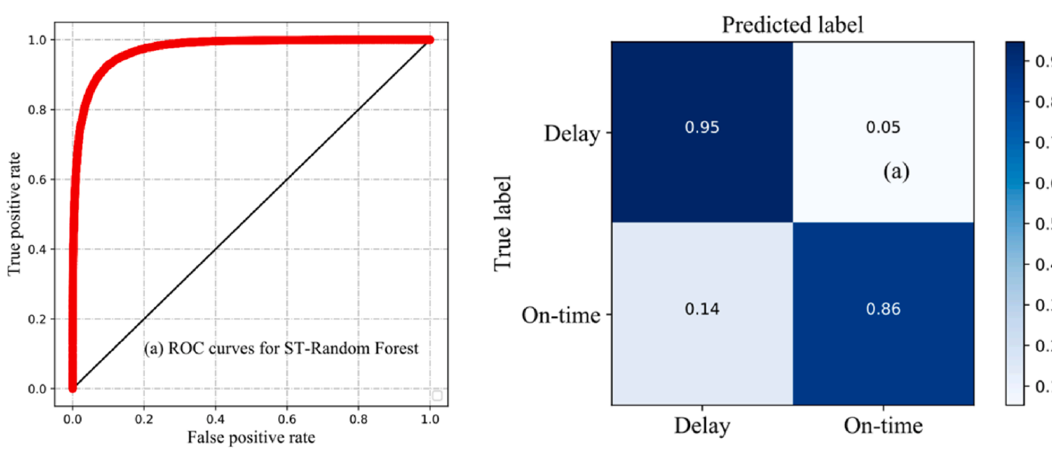

Fig. 13. Prediction performance of our proposed model.

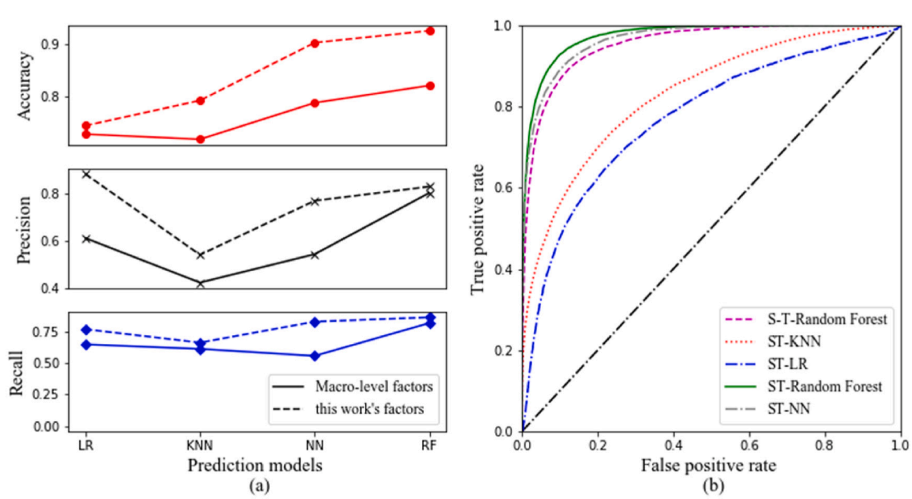

Fig. 14. Comparison of prediction performance with different models and factors.

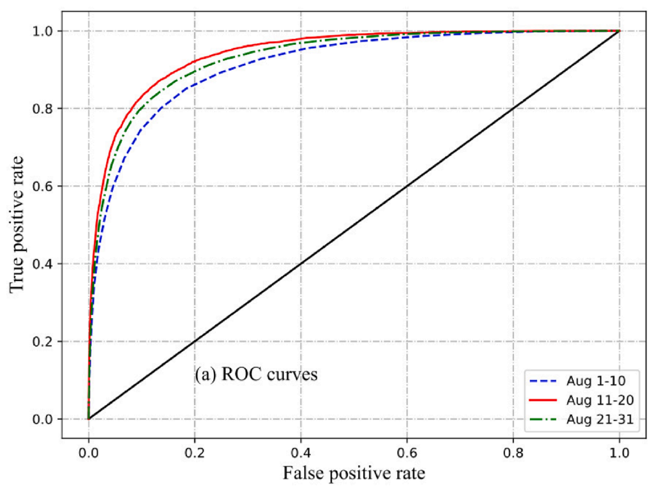

Fig. 15. Prediction performance of ST-Random Forest in different data sets.

The prediction of our model and other baselines is plotted in Fig. 14. Fig. 14 (a) compares the prediction results with different inputs while Fig. 14 (b) compares the prediction results with different classifiers.

In Fig. 14 (a), we consider the input features as macro-level factors and the features extracted in this work (spatial features, temporal properties and extrinsic features) to predict flight delays. We observe that all of these four models that uses the features we extracted are better than the baselines. Compared to the baselines, the ST- has relatively 13.0% up to 29.5% higher accuracy. In Fig. 14 (b), we consider the input features as spatial features, temporal properties and extrinsic features, and use different classifiers (e.g., RF, KNN, LR) to predict flight delays. Results also imply that our proposed model outperforms the baselines.

Comparing the prediction results of ST-NN and ST- Random Forest, we observe that uses Random Forest as a classifier instead of a dense layer followed by the usual Sigmoid classifier improves the accuracy, precision and recall from 0.89, 0.77 and 0.81 to 0.92, 0.83 and 0.85 which indicates the effectiveness of our proposed model. Furthermore, as shown in Fig. 14(b), the ST-Random Forest shows better performance than the S-T-Random Forest, which indicates the existence of temporal correlations between features.

To test the robustness of our proposed model, the prediction for different dates (Aug 1 to 10, Aug 11 to 20, and Aug 21 to 31,) was also observed in Fig. 15. From the results in Fig. 15, we found a small gap in different data sets, implying that ST-Random Forest has a good robustness for different data sets regardless of their date.

5. Conclusion

Flight delays are an important problem in the aviation industry and incur large economic losses. Therefore, a high accuracy delay prediction system is indispensable for airports to monitor flight delays in real-time and improve the on-time performance.

The main goal of this work is to develop a high-accuracy prediction model with practical data. In this study, we extract a novel set of influential factors by using complex network theory and LSM approach, and employed a random forest method to predict fight delays, where complex network theory is utilized to extract the spatial feature of the aviation network at edge-, node-, and network-level, LSTM cell is embedded in the prediction model to extract the temporal near dependencies of crowdedness and weather condition. The proposed method has proven to be highly accurate for prediction compared with the previous research model.

The ST-Random Forest model can be embedded within flight information systems of airports or airlines to monitor and predict potential flight delays accurately and help the system operator understand the current and future status of flights in advance. Monitoring flight delays is essential for both airport operations and airlines. From the perspective of airports, identifying the potential flight delay in advance could help them carry out schedule optimization, such as flight scheduling, runway scheduling, and stand allocation to alleviate flight delays and improve real-time decision support. The presented research could help airlines arrange their operation plans and reduce passengers' dissatisfaction and economic loss.

The study using China domestic flight data between Jun and Aug 2016 for flight delay prediction can be extended further. First, considering the international flight information and proposing a flight delay prediction model for both domestic and international flights are also interesting. Second, the paramount consideration of this study is to capture the temporal correlation between crowdedness and weather condition. However, there also exists spatial correlation among crowdedness. For instance, congestion on a particular airport will also propagate to its neighboring airports. Therefore, developing a prediction model that considers the spatial correlation is a potential area for further research.

CRediT authorship contribution statement

Qiang Li: Conceptualization, Methodology, Software, Writing – review & editing. **Ranzhe Jing:** Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abdel-Aty, M., Lee, C., & Bai, Y. (2007). Detecting periodic patterns of arrival delay. *The Journal of Air Transport Management, 13*(6), 355–361.
Ahmet, K. (2021). Multi-step influenza outbreak forecasting using deep LSTM network and genetic algorithm. *Expert Systems With Applications, 180*, Article 115153.
Allan, S. S., Beesley, J. A., Evans, J. E., & Gaddy, S. G. (2001). Analysis of delay causality at Newark International Airport. *Proceedings of the 4th USA/Europe Air Traffic Management R&D Seminar*.
Alonso, H., & Loureiro, A. (2015). Predicting flight departure delay at Porto Airport: A preliminary study. In *2015 7th International Joint Conference on Computational Intelligence (IJCCI), Lisbon, Portugal* (pp. 93–98).
Baierle, I. C., Benitez, G. B., Nara, E. O. B., et al. (2020). Influence of open innovation variables on the competitive edge of small and medium enterprises. *Journal of Open Innovation: Technology, Market, and Complexity, 6*(4), 179.
Balakrishna, P., Ganesan, R., & Sherry, L. (2010). Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: A case-study of Tampa bay departures. *Transportation Research Part C: Emerging Technologies, 18*(6), 950–962.
Belcastro, L., Marozzo, F., Talia, D., & Trunfio, P. (2016). Using scalable data mining for predicting flight delays. *ACM Transactions on Intelligent Systems and Technology, 8*(1), 1–20.
Breiman, L. (2001). Random Forests. *Machine Learning, 45*(1), 5–32.
Cook, A. J., Tanner, G., Cristóbal, S., & Zanin, M., (2012). Passenger-oriented enhanced metrics. Second SESAR Innovation Days.
Cook, A., & Tanner, G. (2015). *European Airline delay cost reference values*. University of Westminster.
Du, W. B., Zhang, M. Y., Zhang, Y., Cao, X. B., & Zhang, J. (2018). Delay causality network in air transport systems. *Transportation Research Part E: Logistics, 118*, 466–476.
Evermann, J., Rehse, J. R., & Fettke, P. (2017). Predicting process behavior using deep learning. *Decision Support Systems, 100*, 129–140.
Fleurquin, P., Ramasco, J. J., & Eguiluz, V. M. (2013). Systemic delay propagation in the US airport network. *Scientific Reports, 3*(1), 1159-1159.
Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America, 99*(12), 7821–7826.
Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on* (pp. 6645–6649).
Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., & Zhao, D. (2020). Flight delay prediction based on aviation big data and machine learning. *IEEE Transactions on Vehicular Technology, 69*(1), 140–150.

Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences, 102*(22), 7794–7799.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Jetzki, M. (2009). *The propagation of air transport delays in Europe*. RWTH Aachen University, Airport and Air Transportation Research. Master's thesis.

Kafle, N., & Zou, B. (2016). Modeling flight delay propagation: A new analytical econometric approach. *Transportation Research Part B, 93*, 520–542.

Khanmohammadi, K. S., Chou, C., Lewis, H. W., & Elias, D. (2014). A systems approach for scheduling aircraft landings in JFK airport. In *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Beijing China* (pp. 1578–1585).

Klein, A., Craun, C., & Lee, R. S. (2010). Airport delay prediction using weather impacted traffic index (WITI) model. *Digital Avionics Systems Conference (DASC)*.

Kwan, I., & Hansen, M. (2011). *U.S. Flight Delay in the 2000s: An Econometric Analysis*. Transportation Research Board Meeting.

Lambelho, M., Mitici, M., & Pickup, S. (2020). Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions. *The Journal of Air Transport Management, 82*. https://doi.org/10.1016/j.jairtraman.2019.101737

Li, Q., & Jing, R. (2021). Generation and prediction of flight delays in air transport. *IET Intelligent Transport Systems*, 1–14. https://doi.org/10.1049/itr2.12057

Lordan, O., Sallan, J. M., Simo, P., & Gonzalez-Prieto, D. (2014). Robustness of the air transport network. *Transportation Research Part E: Logistics and Transportation Review, 68*, 155–163.

Mofokeng, T. J., & Marnewick, A. (2017). Factors contributing to delays regarding aircraft during A-check maintenance. In *IEEE Technology & Engineering Management Conference (TEMSCON) 2017 Santa Clara CA, USA* (pp. 185–190).

Mueller, E., & Chatterji, G. (2002). Analysis of aircraft arrival and departure delay characteristics. *AIAA's Aircraft Technology, Integration, and Operations*.

Pamplona, D. A., Weigang, L., Barros, A. D., Shiguemori, E. H., & Alves, C. (2018). Supervised neural network with multilevel input layers for predicting of air traffic delays. In *Proc. Int. Jt. Conf Neural Networks* (pp. 1–6).

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in {P}ython. *JMLR, 12*, 2825–2830.

Pyrgiotis, N., Malone, K. M., & Odoni, A. (2013). Modelling delay propagation within an airport network. *Transportation Research Part C: Emerging Technologies, 27*(2), 60–75.

Rebollo, J. J., & Balakrishnan, H. (2014). Characterization and prediction of air traffic delays. *Transportation Research Part C: Emerging Technologies, 44*, 231–241.

Reynolds-Feighan, A. J., & Button, K. J. (1999). An assessment of the capacity and congestion levels at European airports. *The Journal of Air Transport Management, 5* (3), 113–134.

Shen, M. L., Lee, C. F., Liu, H. H., Chang, P. Y., & Yang, C. H. (2021). Effective multinational trade forecasting using LSTM recurrent neural network. *Expert Systems with Applications, 182*, Article 115199.

Shumsky, R. A. (1997). Real-time forecasts of aircraft departure queues. *Air Traffic Control Quart., 5*(4), 281–308.

Sridhar, B., & Chen, N. Y. (2009). Short-term national airspace system delay prediction using weather impacted traffic index. *Journal of Guidance, Control, and Dynamics, 32* (2), 657–662.

Stefan, B., & Christian, U. (2019). Bad weather and flight delays: The impact of sudden and slow onset weather events. *Economics of Transportation, 18*, 10–26.

Thiagarajan, B., Srinivasan, L., Sharma, A. V., Sreekanthan, D., & Vijayaraghavan, V. (2017). A machine learning approach for prediction of on-time performance of flights. In *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC), St. Petersburg FL, USA* (pp. 1–6).

Tu, Y., Ball, M. O., & Jank, W. (2008). Estimating flight departure delay distributions statistical approach with long-term trend and short-term pattern. *Journal of American Statistical Association, 103*(481), 112–125.

Variflight. (2018). On-time performance report of global airports. Retrieved from https://data.variflight.com/reports. Accessed: August 1, 2020.

Wang, X., Brownlee, A., Woodward, J. R., Weiszer, M., & Chen, J. (2021). Aircraft taxi time prediction: Feature importance and their implications. *Transportation Research. Part C: Emerging, 124*(1), Article 102892.

Wu, Q. (2014). A stochastic characterization-based data mining implementation for airport arrival and departure delay data. *Applied Mechanics and Materials., 668*, 1037–1040.

Xiong, J., & Hansen, M. (2013). Modelling airline flight cancellation decisions. *Transportation Research Part E, 56*, 64–80.

Yin, J., Hu, Y., Ma, Y., Yan, X., & Dan, C. (2018). Machine Learning Techniques for Taxi-out Time Prediction with a Macroscopic Network Topology. In *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC), London, UK* (pp. 1–8).

Yu, B., Guo, Z., & Wang, H. (2019). Flight delay prediction for commercial air transport: A deep learning approach. *Transportation Research Part E*, 203–221.