Received: 7 October 2021 | Revised: 1 March 2022 | Accepted: 9 March 2022

IET Intelligent Transport Systems

DOI: 10.1049/itr2.12183

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Flight delay forecasting and analysis of direct and indirect factors

Fujun Wang[1,2] | Jun Bi[1,2] | Dongfan Xie[1] | Xiaomei Zhao[1]

[1]School of Traffic and Transportation, Beijing Jiaotong University, Beijing, China

[2]Key Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Beijing Jiaotong University, Beijing, China

Correspondence
Jun Bi, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China.
Email: jbi@bjtu.edu.cn

Funding information
National Natural Science Foundation of China, Grant/Award Number: 72171019; Key Laboratory of Airports Cluster Intelligent Operation under Grant, Grant/Award Number: KLAGIO20180801

Abstract

The accurate prediction of flight delays is of great significance to airports, airlines and passengers. This paper presents a causal flight delay prediction model developed for a single airport. A long short-term memory network of delay prediction with an attention mechanism (LSTM-AM) is established to predict flight delays and analyse their primary causes. In this model, the direct and indirect factors related to delays are comprehensively considered. LSTM-AM can focus on input data combined with the attention vector to capture the critical time points, which can make the prediction more accurate. The model's performance is verified by actual operational data of Beijing International Airport, one of the busiest airports in the world. Experimental results show that LSTM-AM has better prediction accuracy than baseline algorithms such as some machine learning methods and deep learning methods. The mean absolute error of LSTM-AM is about 8.15 min on the test dataset. The study found that using the predicted results of this paper to release delayed information in advance can effectively alleviate the nervousness of passengers. The critical time point captured by LSTM-AM combined with runway and apron flow control can reduce or eliminate delays of one flight.

1 | INTRODUCTION

Flight delays are a major problem in civil aviation. They incur direct and indirect costs, such as for maintenance at the gate, extra fees for crew, food service, and lodging. They also affect passenger satisfaction [1, 2]. An airport is a maintenance and transit hub where flight service begins and ends. Inaccurate forecasts of flight delays will result in losses to industries dependent on aviation and passengers, meanwhile, delays will harm the transportation network's service capacity and lead to delays in other airports [3]. The unpunctual arrival of aircraft (both earlier and later than expected) will take significant impacts on an airport's management, such as the reallocation of parking gates, runways, ferries, and scheduling of ground crew. So, the prediction and the analysis of flight delays are of great significance to airlines, passengers, and airports.

The punctuality of flights is an important indicator to assess airlines and airports. Predicting delays will help an airport to adjust resource allocations, quickly analyse the causes, and take measures to reduce or eliminate delays. This study stems from a cooperative project with Beijing International Airport (PEK), which is becoming a data-dominated and intelligent airport. PEK predicts flight delays through the combination of flight, air traffic control, weather, and environmental data. This study decomposes and locates the causes of delays in real-time to provide a basis for managers to make decisions.

Forecasts of flight delays fall broadly into two modes. In Mode 1, the model considers static data about airports, routes, and aircraft attributes in a period. In Mode 2, the model dynamically captures time-series data related to the airport, airline, arrivals, and departures. But an arrival or departure delay is not only related to the flight's attributes, the state of the airport at the time of arrival or departure, and the state of the flight's air route, as seen under Mode 1 in Figure 1. It is also associated with the airport's or air route history sequence (such as pre-arrival delays due to ground management), as seen under Mode 2 in Figure 1. We refer to factors in mode 1 as direct, and time-related factors in mode 2 as indirect. In forecast results, airport managers not only want to know how long the flight deviates from the schedule (delay time) but the main factors (primary time steps) causing the delay (cause focus in Figure 1). In the forecast results, delays caused by direct and indirect factors should be treated differently. Delays caused by direct factors (e.g., weather, holidays) can only be prepared in advance.

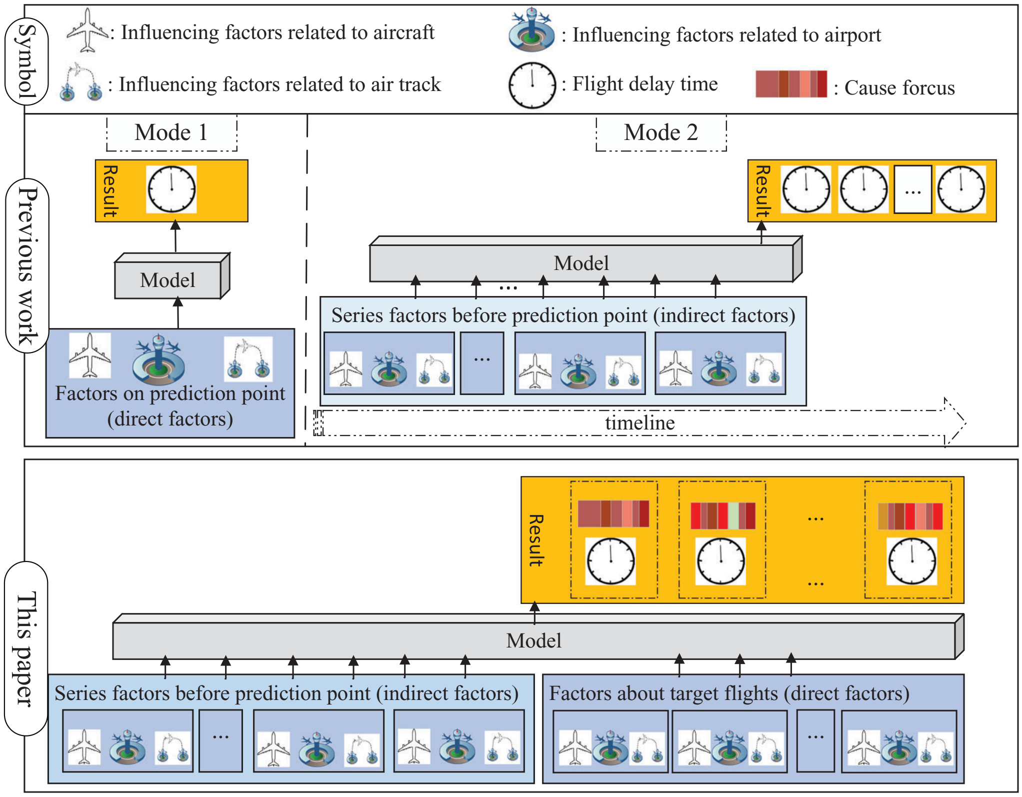

FIGURE 1 Factor and prediction mode comparison of flight delays

However, the management can avoid the delay caused by indirect factors with some measures, for example, to ensure the quick exit of the previous flights. Meanwhile, the influencing factors that managers can deal with over a period are limited, a model that can output the most critical factors (or time points) can save humans and equipment in the process of dealing with the delay. Therefore, finding critical time points or factors is of great significance for managers to alleviate or eliminate delays at the tactical level.

The challenges of this research include the following aspects. State-of-art studies classify and analyse the causes of delays from a statistical or macro perspective. First, these analyses can provide intuitive help to deal with uncertainty, but not specific countermeasures. Second, considering direct and indirect factors equally results in a huge dataset, which is challenging, or even impossible, to optimize. Finally, a suitable method to cut time segments will significantly impact the precision and training time. Current studies isometrically discretize time series, which is not conducive to determining the detailed causes of delays.

To solve the above problems, we established a long short-term memory network of delay prediction with an attention mechanism (LSTM-AM). At present, some airport only stores their detailed indirect factors, so it is difficult to obtain information of all airports in network. At the same time, to ensure a good prediction accuracy, our model needs to be updated according to the latest data. In order to facilitate the acquisition of data and ensure the speed of training, the model proposed in this study was used to predict the flight delay by the historical information of a single airport. The interaction of flights in the aviation network is not within the scope of this paper. The direct and indirect factors are considered in this deep learning network. The model's attention vector can directly show each time series' contribution to a delay, playing a guiding role in establishing countermeasures. Data related to flight delays contain many features, leading to a massive data dimension. Combining the one-hot encoding method and the statistical characteristics of input data, Pareto principle encoding is proposed to reduce the size of the input variables. We considered features that account for 80% of the attributes and ignore the rest, to reduce the data dimension and improve the speed of model training. We take the time of flight arrival or departure as the data collection point, which is intuitive and easy to use for airport operators.

The contributions of this paper can be summarized as follows: (1) The direct and indirect factors are taken into account in the delay prediction, which improved the prediction accuracy; (2) The direct and indirect influencing factors choose different networks, which reduces the data redundancy; (3) The attention mechanism is used in the model so that the primary time points of delay can be traced back and located; (4) We use the real data of PEK for one year to verify the effect of the model, and confirm that our model can meet the requirements of practical application.

The remainder of this paper is arranged as follows. Section 2 reviews the literature. Section 3 introduces and classifies delay-related factors. In Section 4, a flight delay prediction network is built using LSTM-AM, and some related methods are explained. Section 5 presents a case study using PEK data. Section 6 analyses prediction results and reasons for delays. Section 7 provides our conclusions and proposes future work.

2 | LITERATURE REVIEW

With increasingly tight flight schedules, the prediction of aviation resources is developing rapidly. The differences in the current research are mainly in the prediction methods and the input factors considered. Prediction methods are either based on statistics (Stats) or based on machine learning (ML) or deep learning (DL). The influencing factors considered are mainly divided into direct and indirect factors. As mentioned earlier, the direct influencing factors are those that have nothing to do with the time series, which will not be accumulated. However, the indirect factors are related to the time series, these factors will accumulate over time, and finally affect the delay of a flight.

Much literature addresses the statistical analysis. Tu et al. used a genetic algorithm to fit delay data and study long- and short-term flight departure trends [4]. The model included seasonal influences, daily trends, and random trends, enabling users to grasp general delay characteristics. Hsiao and Hansen considered the influence of arrival queues, passenger flow, weather and other factors on flight delays [5]. Through econometric analysis of the contribution rates of various factors to delays, the model explained 72–73% of the variation in the average delay. Hao et al. used econometric and simulation models to calculate and decompose delays, considering direct factors such as quarter-hourly data on throughput, demand, and arrival rates [6]. Rodriguez-Sanza et al. [7] used a Bayesian network and time-series features to model randomness and time variation of flight delays. However, the prediction results consisted of statistical guidance rather than a tactical operation.

ML and DL are developing rapidly. Rebollo and Balakrishnan divided flight delay data into temporal data (e.g., day of week, month of year, and time of day) and spatial data (e.g., delay state and type of delay) [8]. They used a random forest model to predict departure delays in the next two hours, with an average error of about 21 min on the test dataset. However, they only conducted a sensitivity analysis of influencing factors and not the primary factors (or time points) of delays. Manna et al. used the gradient boosted decision tree model (GBDT) to predict the delay of a flight with six directed factors [9]. Their model can also be used by passengers or airline agencies. Kim et al. placed delays at different levels for prediction, using a recurrent neural network (RNN) to consider time and other direct factors such as weather and visibility [10]. Based on a single airport, McCarthy et al. studied the delays of multiple airports with a long short-term memory (LSTM) algorithm, using the time series of the past 24 h to predict delays [11]. The method was shown to be accurate and robust for low-cost airlines in Europe. The analysis of causes on the whole network is helpful to gain an overall understanding of critical factors. Qiang uses the Random Forest (RF) algorithm to predict the delay of a single airport [12], and the method was validated by U.S. domestic flights. However, it cannot explain the delay for one flight.

Zhen and Bin et al. combine the RF and the maximal information coefficient to analyse the flight delay of PEK, but there is no analysis of factors of delay in detail [13].

Ai and Pan et al. used a convolutional LSTM (Conv-LSTM) algorithm to analyse the flight delay distribution, considering indirect factors such as pre-order flight delay, route congestion and airport capacity [14]. Because of the large amount of data and wide range of prediction time window, the study used relatively few factors. Hao et al. make a multi-step prediction of airport delay with spatio-temporal data [15], and the model used historical data of multiple airports for training. The prediction accuracy of the model is high, but it is unable to analyse a single flight. Yu et al. combined a deep belief network and SVR to mine the influencing factors of a flight delay with better prediction results than benchmark methods such as k-nearest neighbours (KNN), support vector machine (SVM) and linear regression (LR) [16]. However, the paper used a statistical method to analyse critical factors and could not give a more intuitive, detailed, tactical description of each delay. The analysis of flight delay and discovery of influencing factors are becoming more and more detailed. Liu et al. used machine learning methods including SVM, LR and RF to analyse the impact of convective weather, local weather and airport traffic demand on ground delay [17]. Still, they could not guide specific work. Research on the prediction of delay is developing from two classifications to multi-classification. Gui and Liu combined several machine learning and deep learning methods with information on weather, flights, airports, and other direct influencing factors to classify and regress airport delays [18]. Jia and Honghai et al. combined KNN, RF, LG, Decision Tree and Gaussian Naïve Bayes with directed factors to predict flight delays of Boston Logan International Airport [19]. The results showed that the stochastic forest model could achieve better prediction accuracy than baseline algorithms, but factors affecting prediction results were not discussed. Khan et al. [20] integrates machine learning algorithms to study the delay and the duration of the delay, but they didn't consider indirect factors. Junfeng et al. built eight widely used modes including linear regression models, non-linear regression models, and tree-based ensemble models [21], and they found that if the feature set could capture the arrival characteristics, even simple linear regression models or algorithms could fulfil the prediction task. With the improvement of computing power, the DL algorithm based on multi-airport data has received more and more attention [22–24]. Multi-airport delay prediction needs to consider higher data dimensions, such as OD (origin-destination) data between airports, which will not be conducive to the updating of the model. At the same time, They are difficult to give a specific analysis for one specific flight.

Table 1 shows the influencing factors, research methods, and causes of delays in the literature ($\surd$ means that a factor or method is taken into account, and × indicates otherwise). The last column indicates whether the results of delays have been

TABLE 1 Influencing factors, methods, and cause analysis of current studies

	Influencing factors		Methods		
References	Direct	Indirect	Stats	ML/DL	Cause analysis
[8, 9, 12, 13, 19, 21, 25–27]	√	×	×	√	×
[4–7, 18, 21, 28–30]	√	×	√	×	√
[10, 11, 14, 15, 22, 23]	×	√	×	√	×
[16, 17, 20]	√	×	×	√	√

analyzed. It can be seen from the table that there is little literature considering both direct and indirect factors, and most of the analyses are based on statistical experience. There is no research on delay analysis for every flight.

We establish a deep learning network based on an attention mechanism and considering direct and time-related indirect factors. The network can dynamically capture the contribution rate of input data to flight delays for each prediction result. The results can provide a basis for tactical airport management and air traffic control.

3 | METHODOLOGY

3.1 | LSTM via attention-based approach for flight delay

LSTM is an important category of RNN that was proposed in 1997 [31]. RNN has seen good use in text, video, and audio processing, and is gradually being used in other fields. However, when the input sequence data is too long, it is difficult for the standard RNN to save the historical information far away from the decision point [32]. LSTM has a natural advantage in solving the long-term dependence of serial data [33]. In each LSTM unit, the input information is filtered while absorbing previous historical information to maintain the long-term memory of sequence data. LSTM can well maintain the internal relationships between sequences, which helps to preserve distant information. From the airport's point of view, it is hoped to predict the number of aircraft arriving next. Therefore, in our prediction of delay, the input of our model should be sequence data, and the output is also sequence data. However, structural limitations make it difficult for LSTM to make a serial-data prediction.

The seq2seq model was proposed by Cho et al. [34] with sequence data as input and output. Seq2seq is a deep learning technology with critical applications such as machine translation, document extraction, and man-machine dialogue, whose input and output characteristics are similar to delay prediction in this study. Seq2seq model has an encoder and decoder. The encoder's length is fixed, and the historical information used by it to predict results is entirely contained in the context vector, which contains limited information. For example, when using data of the past 50 flights to predict the delays of the next 10 flights at the airport, if we code the data of the past 50 flights

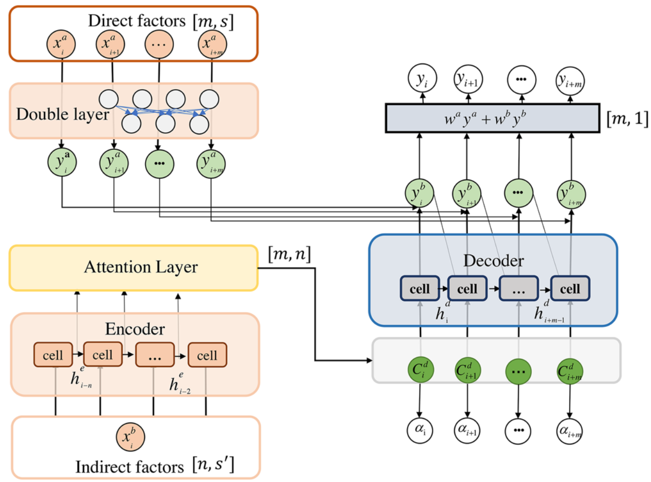

FIGURE 2 Structure of LSTM-AM

into a vector and then use the vector to predict the next 10 delays one by one, the result would be the same. Because they pay attention to the same parts of the vector. It is difficult to achieve the desired effect with just the seq2seq model.

Bahdanau and Cho et al. added an attention mechanism to make up for the limited detail of the seq2seq model [35]. In the decoder layer, seq2seq with an attention mechanism considers the context vector and the various effects of historical information. After adding the attention mechanism, data with input and output sequences can be modelled without distance on the input and output [36].

Our model is shown in Figure 2. We use the past n flights to predict the delays of $m+1$ flights in the future. In our model, y_i means the delay of the flight i which is the output of our model. The direct and indirect factors of the flight i are recorded as x_i^a and x_i^b. In Figure 2, x_i^a and x_i^b are all input of our model. Delays caused by x_i^a and x_i^b are y_i^a and y_i^b, respectively, as shown in Equations (1) and 2). What is stored in square brackets ($[*, *]$) is the dimension of the input and output, as well as the dimension of the attention matrix, and s and s' are respectively related to the length of direct and indirect factors after data processing (Section 3.2).

$$y_i^a = f^{dl}\left(x_i^a\right); \tag{1}$$

$$y_i^b,\ \alpha_i = f^{lstm}\left(x_i^b\right); \tag{2}$$

$$C_i^d = \sum_{j=1}^{n} \alpha_{ij} h_j^e; \tag{3}$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{n} \exp(e_{ik})}; \tag{4}$$

$$e_{ij} = f^w\left(h_{i-1}^d, h_j^e\right); \tag{5}$$

$$\alpha_i = \left[\alpha_{i1}, \alpha_{i2}, \alpha_{i3} \cdots \alpha_{in}\right]. \tag{6}$$

As mentioned earlier, the delay caused by direct factors (Equation 1) is different from that caused by indirect factors (Equation 2)). If the weather conditions are poor at the time of landing or the flight has a lower priority at the airport, the delay will be longer. But in this situation, the direct influencing factors will have a greater weight on the final delay (y_i). When a flight breaks down on the tarmac, or ground support capability conflicts with the target flight, the delay will also be longer, but the weight of indirect factors will be larger. Delays caused by x_i^a can not be intervened during operation, so our model does not pay more attention x_i^a on the final prediction result. In (Equation 1), we choose a double-layer feedforward neural network ($f^{dl}(\cdot)$) to learn the nonlinear influence of x_i^a factors on y_i. In Equation (2), $f^{lstm}(\cdot)$ represents the seq2seq model with attention mechanism. The decoder and encoder of $f^{lstm}(\cdot)$ are LSTM. In Equation (5), $f^{w}(\cdot)$ is a feedforward neural network jointly trained with all the other proposed system components (see [35]). In Equation 3), C_i^d is the context vector for each prediction. $\tilde{h}_j^e$ and $\tilde{h}_j^d$ represent the hidden layer output of the encoding layer and the decoding layer, respectively.

In Equation 4) α_{ij} contains various levels of attention to the input data in the decoding process. Our input is divided into two parts: x_i^a and x_i^b. The attention matrix α_i contains the importance of the model to all indirect factors of flight delays. Combined with Equation (6), it can be found that α_i learns the different attention of input data for every prediction result. Weights of non-important factors in the indirect data (x^b) are gradually reduced after training. Therefore, the vector of α_i can focus on critical influencing factors for a flight delay prediction result. In airport management, this vector can provide a basis to eliminate or slow down flight delays at the operational level.

We use a full connection layer to learn the relationship between these two factors (x_i^a and x_i^b), as shown in Equation 7).

$$y_i = w_i^{\mathrm{a}} \cdot y_i^{\mathrm{a}} + w_i^{\mathrm{b}} \cdot y_i^{b} \tag{7}$$

In Equation (7), w_i^{a} and w_i^{b} are weight vectors indicating the contributions of x_i^a and x_i^b to the final delay (y_i).

Our flight delay model can optimize parameters by backpropagation [37]. Because the actual delay will affect our tolerance to the predicted results, the loss function is chosen as

$$MAPE\left(\mathrm{y}_i, \mathrm{y}_i'\right) = \frac{\mathrm{y}_i - \mathrm{y}_i'}{\varepsilon + \mathrm{y}_i'}. \tag{8}$$

MAPE indicates the percentage of the predicted value (y_i) deviating from the actual value (y_i'), and ε is a very small number to prevent undefined losses caused by 0 in the denominator.

The optimization strategy used is Adam [38], an efficient random optimization method with less memory in the training process, which combines the advantages of AdaGrad [39] and RMSProp [40].

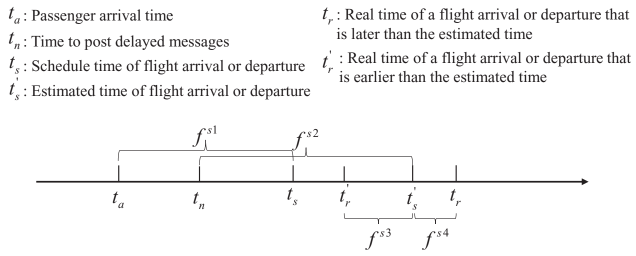

FIGURE 3 Stress changes in passengers waiting for a flight

3.2 | Passenger waiting stress

Accurately released delay information can reduce passenger anxiety and improve the travel experience, which is important to improve passengers' loyalty to an airline. The research of Osuna [41] and Such et al. [42] shows that customer waiting stress consists of actual waiting time and the gap between real and expected waiting time. At the same time, pressure accumulated by customers increases marginally over time. Therefore, the pressure on waiting passengers consists of the pressure of normal waiting (s1), pressure after being informed of a delay (s2), and pressure between the actual departure or arrival time and the estimated time (s3, s4). The stress density functions of the above three stages are expressed by $f^{s1}(t)$, $f^{s2}(t)$, ($f^{s3}(t)$, $f^{s4}(t)$) as shown in Figure 3.

For flight i, passenger stress ps_i can be expressed as

$$ps_i = \sum_{j=0}^{np^j} \left[\left(t_s^i - t_a^{ij}\right) \cdot \int_{t_a^{ij}}^{t_n^i} f^{s1}(t)\, dt + \left(t_r^i - t_n^i\right) \times \int_{t_n^i}^{t_s^{i'}} f^{s2}(t)\, dt + \left(t_r^i - t_s^{i'}\right) \cdot \int_{t_r^i}^{t_s^{i'}} f^{s*}(t)\, dt \right], \tag{9}$$

where t_r^i, t_s^i, t_n^i, and $t_s^{i'}$ are the actual take-off or landing time, planned take-off or landing time, notification time of delay information, and estimated take-off or arrival time, respectively, for flight i.

$$f^{s*} = \begin{cases} f^{s3}, \left(t_r^i - t_s'\right) \leq 0 \\ f^{s4}, \left(t_r - t_s'\right) > 0 \end{cases}. \tag{10}$$

If the flight set of airline j is al_j, all passenger stress can be defined as

$$APS_{al_j} = \sum_{i \in al_j} ps_i. \tag{11}$$

In the process of waiting, the pressure of normal waiting is the least, and pressure increases when a delay is announced. It is greatest when the delay time is not consistent with expectations, and passengers question the airline's credibility at this time. So, in general, $f^{s4}(t) > f^{s3}(t) > f^{s2}(t) > f^{s1}(t)$.

TABLE 2 Details of direct and indirect factors

Factors	Elements	Meanings	Examples
Direct (x_i^a)	W_{st_i}	The weather at the time of arrival or departure of flight i	Weather condition, wind speed, wind direction, wind force
	B_{st_i}	Airport congestion at the time of arrival or departure of flight i	Number of passengers, number of flights
	FA_i	Attributes of flight i	Aircraft size, airline properties
	AR_i	Congestion and weather conditions on the route of flight i	The number of flights in the past interval on the rout of flight i
	PD_i	Periodic data of flight i	Month, day of the month, day of the week, season, holiday
Indirect (x_i^b)	PB_{st_i}	Previous airport of flight i	Time-series of airport congestion
	PFA_i	Previous flights of flight i	Time-series of flight delays

Scenario s2 indicates the psychological change process of a passenger after a delay notification. The function is not only affected by the moment but by the interval between the notification time and the expected time. Therefore, in this study, $f^{s2}(t)$ is related to $t_s' - t_n$ as

$$f^{s2}\ (t) = f^{s2}\ \left(t, t_s' - t_n\right)\ :\ \frac{1}{t_s' - t_n} f^{s2}\ (t)\ . \tag{12}$$

Combining Equations (9–12), the stress of the passenger is related to the time of the notification and the accuracy of the expected time. In this paper, Equation (9) is used to measure the impact of delay information released by the forecast results in Section 3.1 on passenger stress.

4 | FACTORS INFLUENCING DELAY

As described in Section 2, we divided the influencing factors into direct (x_i^a in Figure 2) and indirect (x_i^b in Figure 2) factors according to whether the influencing factors have a cumulative effect on delay. The details of the input data in the model of Section 3.1 are shown in Table 2. In Table 2, st_i means the schedule time of aircraft i, and each element is calculated in detail in Sections 4.1 and 4.2.

4.1 | Direct influencing factors

4.1.1 | Weather

According to the Civil Aviation Administration of China (CAAC), 47.46% of delays are related to bad weather [18]. Weather indicators include weather conditions, wind speed, wind direction, wind force, and air humidity at arrival and departure airports. Of course, weather along the route also impacts flight delays, but there is no comprehensive database of such information. It is challenging to obtain worldwide airport weather data. So, we use weather conditions of PEK to characterize the impact of weather on flights:

$$W_t = \left[wca_t, wsa_t, wda_t, wfa_t, aha_t\right], \tag{13}$$

where wca_t, wsa_t, wda_t, wfa_t, and aha_t denote weather condition, wind speed, wind direction, wind force, and temperature at the arrival airport, respectively, at time t.

4.1.2 | Congestion degree in airport

Airports are busy at different times of the day. The more planes arrive or leave per unit time, the greater the pressure on airport ground services. For departing or arriving flights, if the airport is too crowded for a while before arrival, the air traffic control or tower will delay a flight's arrival or departure, resulting in a delay.

We use the number of scheduled flights (Equation (14)), the actual number of flights (Equation (15)), and the expected number of passengers (Equation (16)) per unit time before this flight's arrival or departure [16] to describe airport congestion (B_t),

$$B_{\mathrm{t}}^s = f_{t+\Delta t}^s - f_t^s, \tag{14}$$

$$B_{\mathrm{t}}^a = f_{t+\Delta t}^a - f_t^a, \tag{15}$$

$$B_{\mathrm{t}}^p = f_{t+\Delta t}^p - f_t^p, \tag{16}$$

$$B_t = \left[B_t^s, B_t^a, B_t^p\right], \tag{17}$$

where f_t^s, f_t^a, and f_t^p are the number of cumulative planned flights, the cumulative actual number of flights, and the cumulative number of planned passengers, respectively, at time t, and Δt is the time interval. Based on discussions with PEK staff and consulting the relevant literature [16], Δt is set to half an hour. B_{t}^s, B_{t}^a, and B_{t}^p are respectively the planned number of flights, the actual number of flights, and the actual number of passengers in the airport at time t, and B_t is the congestion degree in the airport at time t.

4.1.3 | Flight attributes

Flights at PEK have different priorities in scheduling and resource allocation based on airlines and aircraft sizes.

Number of passengers, aircraft size, international and domestic attributes

The attributes of passengers, aircraft size, and domestic attributes are similar. To ensure airport operational efficiency, improve passenger satisfaction, and reduce security risks, airport parking gate and ground support resources generally prioritize flights of larger aircraft or with more passengers. For example, PEK has special parking gates for larger aircraft, which will reduce ground service time.

$$FA_i^1 = \left[np_i, as_i, id_i\right], \tag{18}$$

where np_i is the number of passengers on flight i, and as_i is the aircraft size of flight i. np_i and as_i can indicate the priority of the flight being served within the airport, which is associated with both departure and arrival delays. id_i is a 0–1 dummy variable depending on whether flight i is international.

Airline properties

Airline capacity scales differ by airport, that is, the number of routes and aircrafts owned by different airlines at this airport is different, which affects their scheduling strategies. Larger airlines usually choose certain airports as bases, where they will be more flexible in controlling take-off and landing times. Airlines not based at an airport will pay more passing fees if they spend more time at a station; hence they exercise more precise control over these times. We characterize an airline's attributes according to its name and whether the airport is a base.

$$FA_i^2 = \left[al_i, ba_i\right], \tag{19}$$

where al_i is the name of the airline to which the flight i belongs, which is one-hot coded; and ba_i is a dummy variable depending on whether the airport is the base for i.

Runway, terminal, scheduled time

The use of runways and terminals differ during the day, and the lack of parking spaces in runways or terminals can cause flight delays. These two factors represent airport congestion at the tactical level. Combined with the target flight planning time, they can dynamically characterize the changes of runway and terminal resources in the time dimension, and their impact on delay.

$$FA_i^3 = \left[rw_i, te_i, st_i\right], \tag{20}$$

where rw_i and te_i are the runway and terminal, respectively, where flight i lands, and st_i is the flight's scheduled time.

Historical delay and delay of previous flight

The historical delay of the target flight can somewhat characterize the statistical influence of other factors that are difficult to quantify, such as weather and guaranteed resources of the route. A previous flight delay will directly affect the flight schedule and ground support plan of the follow-up flight during the airport's regular operation. We characterize the historical flight delay by the historical average delay and standard deviation.

$$FA_i^4 = \left[E\left(h\left(y_i'\right)\right), \sigma\left(h\left(y_i'\right)\right), y_{i-1}'\right], \tag{21}$$

where y_i' is the delay time of flight i; $h(y_i')$ is the set of delays for flights that have the same flight number with flight i in the dataset ; $E(*)$ and $\sigma(*)$ are average and standard deviation operations, respectively; y_{i-j}' is the j-th flight's delay before flight i in the flight schedule.

$$FA_i = \left[FA_i^1, FA_i^2, FA_i^3, FA_i^4\right], \tag{22}$$

where FA_i represents all factors directly related to the scheduled arrival time and attributes of flight i.

4.1.4 | Air route

Congestion and weather conditions on the route also significantly impact flight delays [16, 18]. However, at present, there is a lack of databases for weather and congestion conditions for each route segment. We use the airline to approximately replace the air routes of flight i, considering the speed of data processing and scale of input. Each air line is uniquely marked by departure or arrival airport code. The air traffic control department will control the flow of specific routes, and when the density of aircraft on a route is too high, it will delay take-offs and arrivals. Therefore, historical delay data on a route can represent the periodic characteristics of bad weather, emergencies on the route, and the relationship between flight delay and route flow. We characterize routes by the average and standard deviation of flow for a while before take-off or landing,

$$AR_i = \left[E\left(\mathrm{h}\left(ar_i\right)\right), \sigma\left(\mathrm{h}\left(ar_i\right)\right)\right], \tag{23}$$

where ar_i denotes the air route of flight i; $\mathrm{h}(ar_i)$ is the set of delays for flights on ar_i; $E(*)$ and $\sigma(*)$ are average and standard deviation operations, respectively. Statistical indicators $E(\mathrm{h}(ar_i))$ and $\sigma(\mathrm{h}(ar_i))$ can reflect the congestion between different airports.

4.1.5 | Periodic data

The degree of business for a particular airport or route is generally affected by the season or month in the long term, and by holidays and weekends in the short term. We consider the month, day of the month, day of the week, season, and holidays.

$$PD_i = \left[tm_i, td_i, tw_i, ts_i, th_i\right], \tag{24}$$

where tm_i,td_i,tw_i,ts_i, and th_i denote the month, day of the month, day of the week, season, and holiday, respectively, at the scheduled time of flight i, and

$$x_i^a = \left[W_{st_i}, B_{st_i}, FA_i, AR_i, PD_i\right] \tag{25}$$

sums the above information to represent all direct factors related to the delay of flight i. In Equation (25), st_i is scheduled time of flight i.

4.2 | Indirect influencing factors

Indirect factors refer to the impact of continuous arrival and departure flights in a short period, as well as past constant state changes at airports on a flight.

In dealing with indirect factors, we abandon the concept of the "equal time interval". Time-series processing is not to discretize the time and extract the statistical data of flights or airports at equal time intervals, but to take the time of departure or arrival at the airport as the collection node of historical data. Because the airport's resources or schedule will be adjusted only if previous flight changes are different than expected.

4.2.1 | Previous airport

The influence of pre-ordered flights should be an ordered time series. If the data collection range of pre-ordered flights for i is n, the impact of pre-ordered flights can be expressed as

$$P\,B_{st_i} = \left[B_{st_{i-n}}, B_{st_{i-(n+1)}}, B_{st_{i-(n+2)}} \cdots B_{st_{i-2}}, B_{st_{i-1}}\right], \tag{26}$$

where st_{i-j} is the j-th flight's schedule time before flight i; B_* represents the direct influencing factors with flight i in Equation (17); PB_{st_i} is the short-term effect of the previous airport status on flight i.

4.2.2 | Previous flights

An arriving flight will occupy various security resources, and a departing flight will release such resources, so the impact of pre-ordered flights can be expressed as

$$\begin{aligned} PFA_i = \big[&\left[ad_{i-n}, FA_{i-n}\right], \left[ad_{i-(n+1)}, FA_{i-(n+1)}\right], \\ &\cdots, \left[ad_{i-2}, FA_{i-2}\right], \left[ad_{i-1}, FA_{i-1}\right]\big] \end{aligned} \tag{27}$$

where ad_{i-j} is a dummy variable depending on whether the j-th flight before flight i is arriving or departing; FA_* is direct influencing factors in Equation (22), and PFA_i indicates the influence factors of previous flights to i in a short period.

Combining Equations (26) and (27), we obtain the indirect influence factors of flight i as

$$x_i^b = \left[PB_{st_i}, PFA_i\right]. \tag{28}$$

5 | CASE STUDY

Affected by the COVID-19, the number of flights at PEK has been greatly reduced since 2020, and the data during the epidemic cannot explain the normal situation, so we use the data of 2019 to replace the operation of the airport after it returns to normal, which can better show the universality of our method. The number of passengers at PEK reached 100 million in 2019, ranking second globally [43]. Its punctuality rate was 70.83%, much lower than Tokyo's Haneda International Airport (85.25%), which is of the same magnitude (about 80 million passengers) [44]. To train and verify our model's prediction performance and its explanation of delays, we collected 606,139 PEK flight records, including arrivals and departures from 1 January 2019 to 1 January 2020.

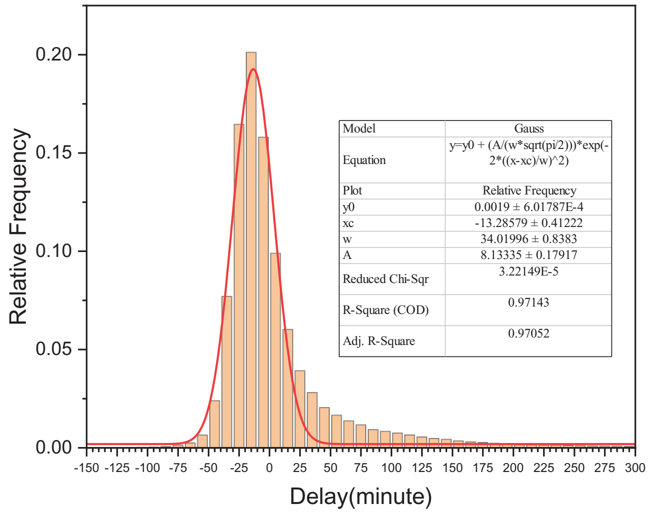

FIGURE 4 Flight delay frequency statistics and Gaussian fitting distribution. The bar chart is the frequency distribution of the delay, and the curve represents the Gaussian fitting

Because we do not analyse flight cancellations, we eliminated them, along with some empty data. The final dataset involved 96 airlines and 593,666 flight records. The dataset contained all the attributes mentioned in Section 3.1.

5.1 | Data description

Data of flight delays can be positive or negative, where a negative value indicates an arrival earlier than scheduled, which significantly impacts the airport's equipment and staffing. Therefore, negative values of delays are also within the scope of our study. As shown in Figure 4, flight delay approximately obeyed a Gaussian distribution, and the R-squared of Gaussian fitting was 0.97143, indicating an excellent fitting effect.

As shown in Figure 5, the impact of weather on flights is measured by four indicators: wind power, wind direction, weather conditions, and temperature. The wind has six levels, from small to large, where zero indicates no wind. The wind has eight directions, where 1 to 8 denote northeast, east, southeast, south, southwest, west, northwest, and north, respectively, and zero indicates no continuous wind. A total of 11 weather conditions are involved in the historical data at PEK. The weather code and its specific meanings are shown in Table 3.

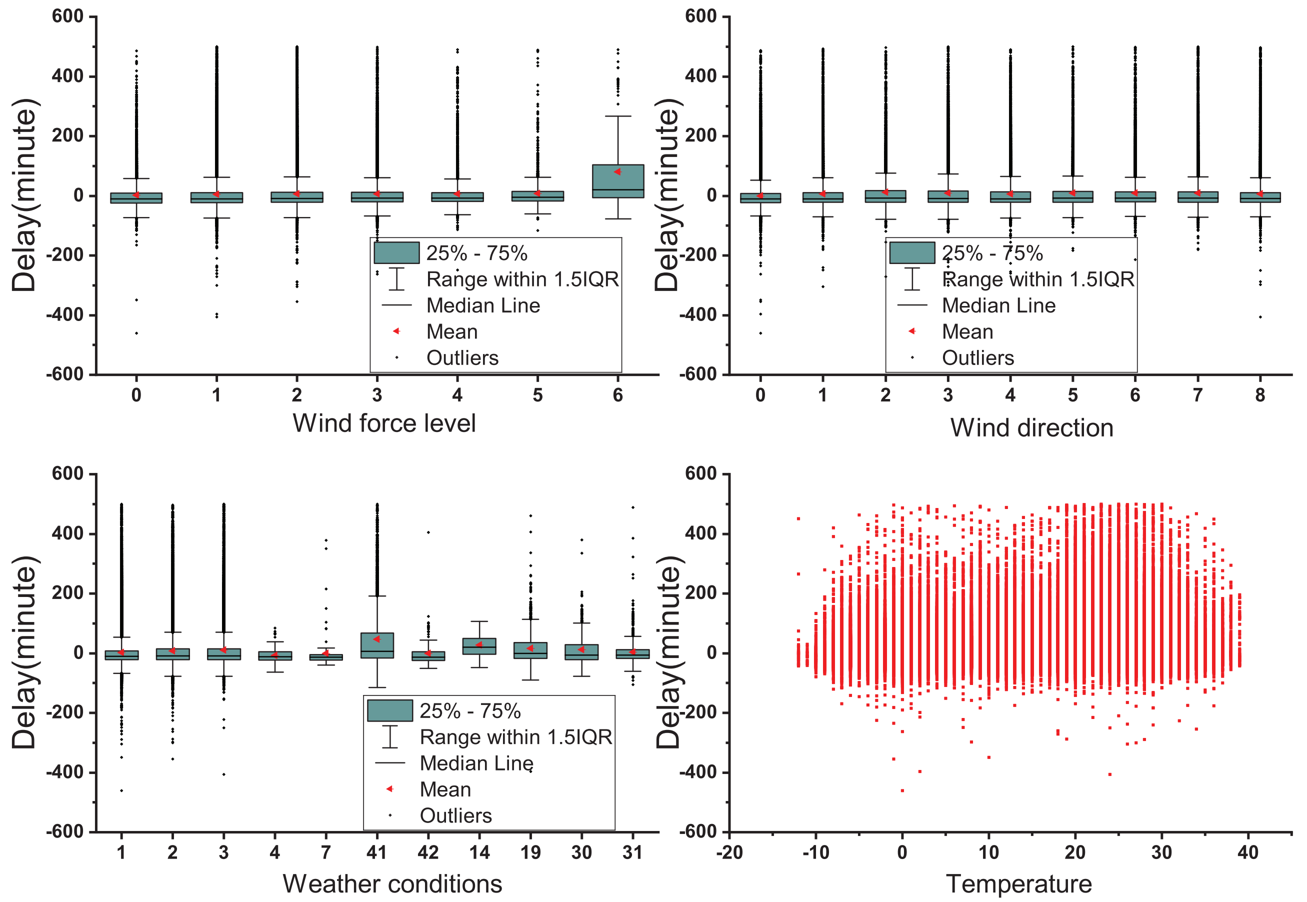

FIGURE 5 Influencing factors about the weather

TABLE 3 Weather conditions and corresponding codes

Code	Weather condition
1	Sunny
2	Overcast
3	Cloudy
4	Shower
7	Sleet
41	Rain
42	Snow
14	Snow shower
19	Fog
30	Floating dust
31	Blowing sand

Statistically, weather conditions have a significant impact on flight delays, whose average time is higher in rainy or snowy weather. The average flight delay shows a Gaussian distribution with an average of zero. When the wind is strong (such as level 6), arrivals will be delayed for safety reasons. When the temperature is between 5 and 20 °C, the distribution of flight delays is more concentrated. However, the absolute value of delay is higher in cold and hot conditions. Temperature also must be considered. For example, lower temperatures during rainfall or snowfall can freeze the runway, causing delays. Our double-layer feedforward network can capture these inherent nonlinear factors.

In Figure 6, the abscissas of airline and air route, respectively, represent airline and airport codes assigned by the International Air Transportation Association (IATA). There are significant differences in the distribution of flight delays between airlines, aircraft types, and routes. Our study involves 94 airlines and 5715 aircraft of 90 types. As shown in Figure 6 (Aircraft type), the larger the aircraft the less the average delay (A380). There are 478 routes related to PEK, and only some are listed in Figure 6. Obvious differences exist in the distribution of delays among routes.

For Figure 7, 1–7 denote Monday to Sunday, respectively, in the Week chart; 1–12 denote January through December, respectively, in the Month chart; 0 means 00:00–01:00 and 23 means 23:00–00:00 in the Hour chart; and 1,2,3,4 respectively denote January to March, March to June, June to September, and September to December in the Season chart.

The statistical characteristics (mean, median) of periodic data vary little, as shown in Figure 7. However, the distribution of the total number of flights in each cycle is uneven, and the distribution of outliers varies greatly by statistical period. For

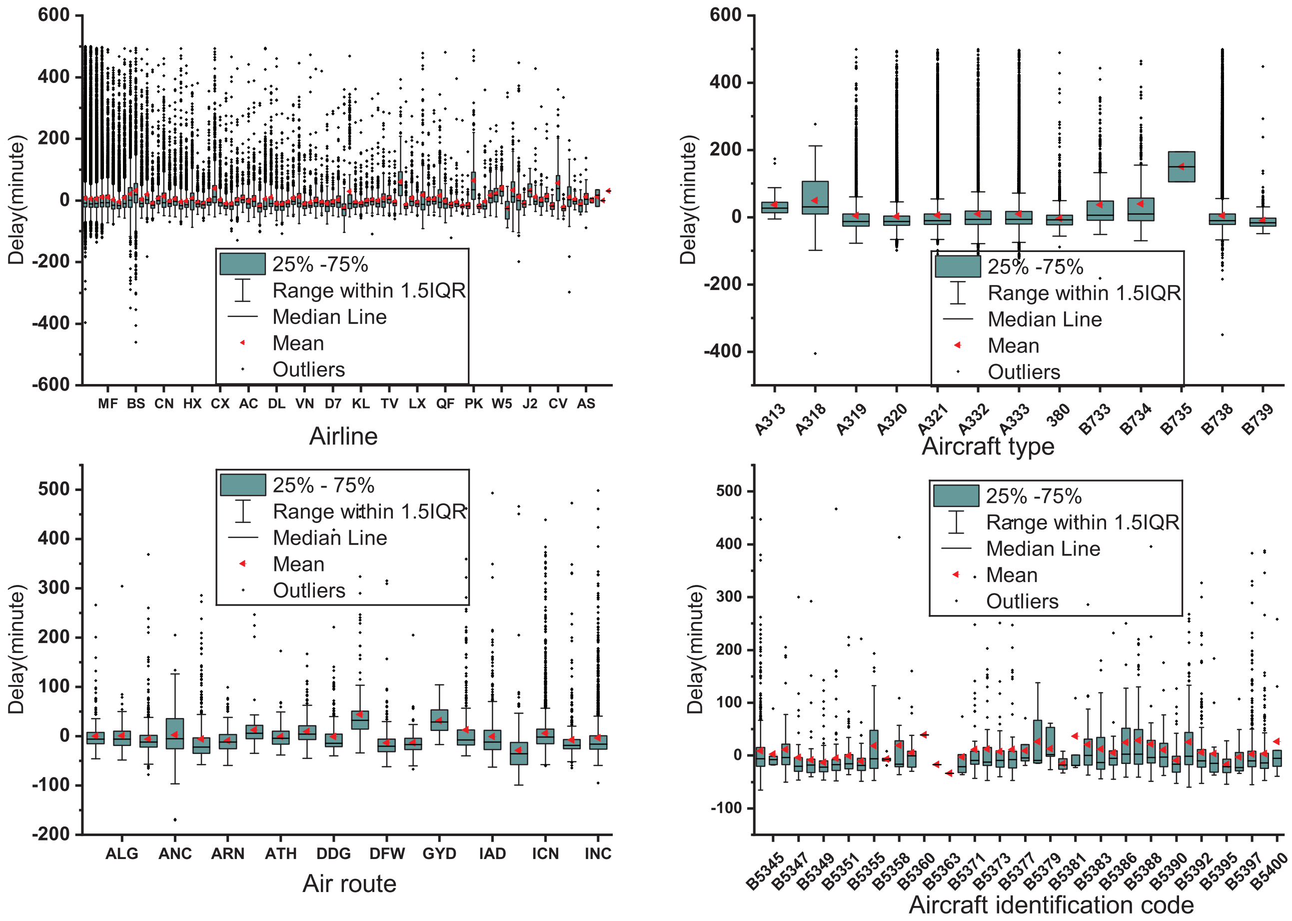

FIGURE 6 Influencing factors about flight attributes

example, from the "Hour" statistics, we can see that arrivals are greater during the peak hours from 7 AM to 12:00 PM, while in the off-peak period, the numbers of flights later and earlier than the scheduled time are the same. Beijing has a temperate monsoon climate, and the weather changes frequently between May and August, so the average flight delay is greater in those months.

The Kruskal-Wallis (KW) test and Median test on the factors mentioned above are listed in Table 4. The *p*-values of each factor are less than 0.05, so there are significant differences in the data set distribution divided by the selected features on the confidence interval of 95%.

This part makes a statistical analysis of the attributes in the dataset. For some computationally available features, such as airport and airline congestion, existing studies have confirmed the effectiveness of these attributes in flight delay [45, 46], about which we will not give a more detailed description in this study.

5.2 | Data preprocessing

To verify the model, original data is randomly disrupted. And data were divided into 80% for model training, 10% for model verification, and 10% for testing. The data were divided into continuous and discrete parts, as shown in Table 5. The continuous data were normalized to eliminate the influence of dimension on training results, improve convergence speed, and avoid over-fitting. Attribute data, such as weather and periodic data, were discrete, could not be directly input into the model. These attributes had to be coded. We used a Pareto encoding method based on conventional one-hot coding, which could effectively reduce the data dimension to ensure the model's accuracy.

5.2.1 | Normalization

It is well known that normalizing can eliminate differences in data dimensions, reduce the influence of outliers on results, and improve training speed. Because our sample dimension is not as high as those of videos and pictures, the sharp increase of error with batch normalization (BN) [47] can be avoided. Considering the speed of optimization, we selected BN to normalize the data. The process of BN is shown in Algorithm 1.

Algorithm 1: Batch normalization

Input: Continuous data x^c over a mini-batch; $B = \{x_1^c, x_2^c, x_3^c \ldots x_{bs}^c\}$; Learning parameters: γ, βOutput: y_i^c

$$\mu_B \leftarrow \frac{1}{bs} \sum_{i=1}^{mb} x_i^c \tag{29}$$

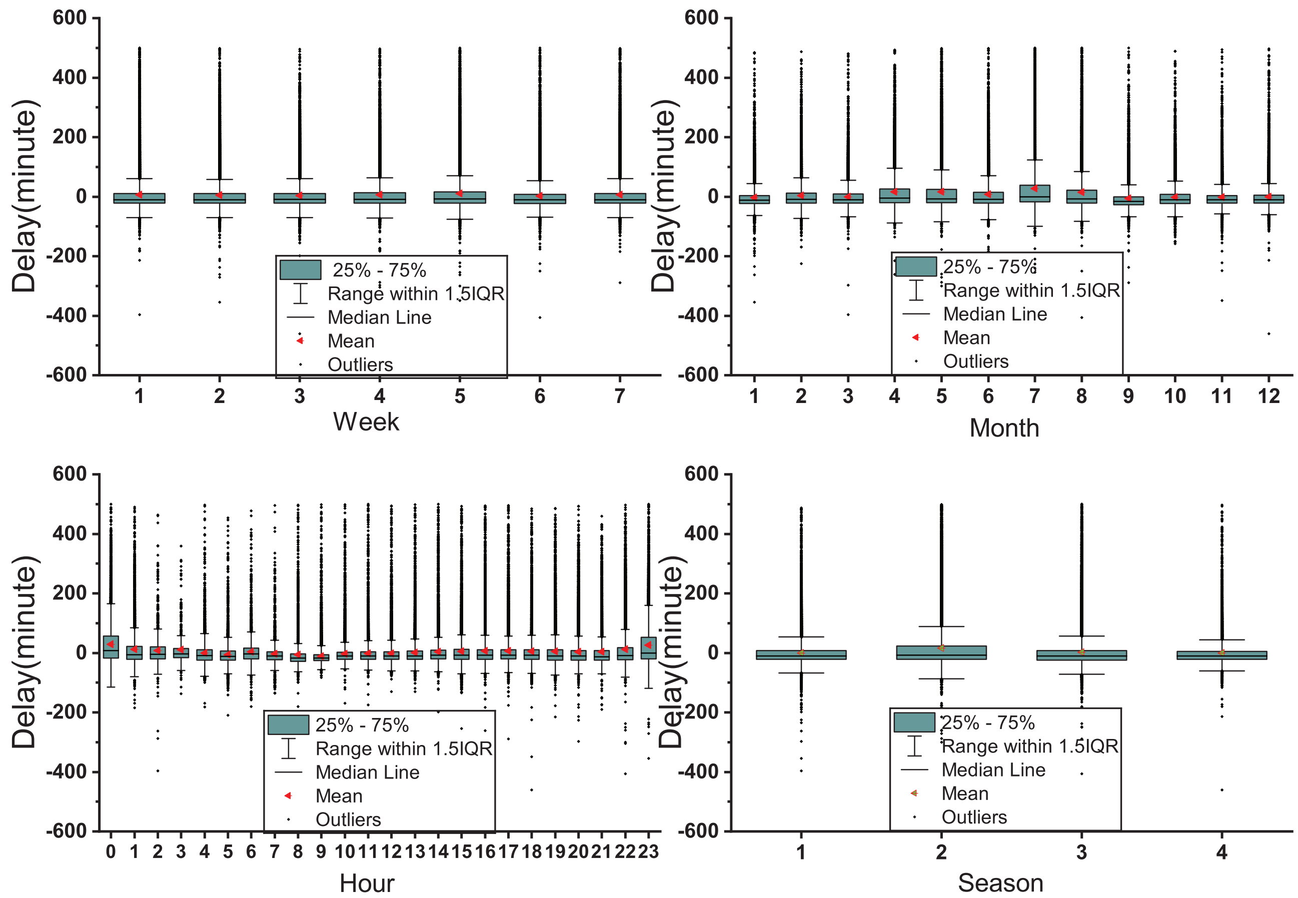

FIGURE 7 Temporal variation in flight delays

TABLE 4 Kruskal-Wallis and Median test for factors of weather, flight attributes and temporal variables

Category	Factors	KW H	KW p	Median H	Median p
Weather	Wind force level	570	0.0	450	0.0
	Wind direction	816	0.0	524	0.0
	Weather conditions	1847	0.0	1072	0.0046
	Temperature	1209	0.0014	648	0.0
Flight attributes	Airline	8488	0.0	7281	0.0
	Aircraft type	5070	0.0	4214	0.0
	Air route	27110	0.0	19696	0.0
	Aircraft identification code	2505	0.0	2084	0.0021
Temporal variation	Week	598	0.0	499	0.0
	Month	721	0.0	606	0.0
	Hour	9768	0.0	6693	0.0
	Season	436	0.0006	389	0.0

Note: KW and Median are Kruskal-Wallis and Median test, respectively; H and p represent statistics and p-value of the test, respectively.

TABLE 5 Continuous and discrete data of factors influencing delays

Data type	Factors
Continuous	Historical delay, crowdedness degree of airport, delay of previous flight, scheduled time, condition of air route
Discrete	Weather, aircraft identification code, aircraft size, airline properties, flight terminal, taxiway

$$\delta_B^2 \leftarrow \frac{1}{bs}\sum_{i=1}^{mb}(x_i-\mu_B)^2 \tag{30}$$

$$\widehat{x_i^c} \leftarrow \frac{x_i^c-\mu_B}{\sqrt{\delta_B^2+\varepsilon}} \tag{31}$$

$$y_i^c = \gamma\hat{x}_i^c+\beta = BN_{\gamma,\beta}\left(x_i^c\right) \tag{32}$$

The sample batch size in Algorithm 1 is bs. Equations (29–32) limit the input data from 0 to 1, and the transformed sample follows a Gaussian distribution. Equation (32) is the most crucial step. Through the combination of γ and β, the expression ability of the normalized data model is improved so that it is not

TABLE 6 Some attributes with a large dimension

Attribute	Size
Aircraft code	5715
Aircraft type	90
Airline	94
Air route	478

limited by a Gaussian distribution. ε has mean 0 and variance 1, and is used to avoid zeros in the denominator.

The attributes of limited categories such as weather conditions were transformed to digit code by mapping; for example, 0 represents sunny days and 1 represents rainy days. However, these methods will change samples' internal relationships. For example, if we express sunny, cloudy, and rainy days as 0, 1, and 2, we can find that rainy days > cloudy days > sunny days, but this relationship is meaningless and it will increase the difficulty of network training.

5.2.2 | Encoding

In our study, weather, flight number, aircraft size, airline properties, flight terminal, and taxiway are all categorical data that can be handled by one-hot encoding.

Some attributes have many features. In Table 6, we show only four attributes, but to store just these features requires a vector of size 6377 (5715+90+94+478), so coding directly with one-hot will increase the dimension of the dataset, increasing the difficulty of training and storage. Attributes with many features require special treatment.

In the late 1800s, Vilfredo Pareto observed that 80 percent of the land in Italy was owned by 20 percent of the population [48]. The Pareto principle, also called the 80/20 rule, states that roughly 80% of consequences come from 20% of causes for many outcomes.

Through data analysis, we found that the Pareto principle applies to airport flight data, including aircraft identity code, aircraft type, airline properties, and air route, for example, the studied dataset involves 94 airlines, but the top seven account for 81.15% of all flights, as shown in Figure 8.

We propose Pareto principle encoding. Table 7 shows the process of encoding Airline to reduce storage space by 94% compared to one-hot encoding. We only care about the features of the top 80% and ignore the rest 20%.

6 | RESULTS AND DISCUSSION

6.1 | Analysis of prediction result

An epoch consists of input of the dataset, running the model once, and optimizing its parameters by backpropagation. A total of 100 epochs were set, within which we found that the model could consistently converge. We use historical data related to the past 50 flights at the airport to predict the delays of the next 10 flights. Each data point represents an aircraft arriving or departing. In each epoch, the data are read in batches of 65, that is, $bs = 65$.

Figure 9 shows the optimization process. After 30 epochs, the mean absolute error (MAE) of the model was about 8.15 min. Figure 10 shows the observed and predicted delays of some flights arriving sequentially (about 400 flights). It can be seen from Figure 10 that the predicted values are close to the observed values, which shows the prediction effect of the LSTM-AM. Table 8 is a statistical description of the prediction results of 58,600 samples from the test dataset. The absolute values of predicted and observed values within 30 min accounted for 92.91% of all forecast samples. Those with absolute values within 20 min accounted for 88.43%, and those within 15 min for 83.9%. To verify the model's stability, we randomly divided the test set into 100 parts by bagging, and predicted the 100 datasets. The results show that their maximum MAE deviation was less than 4.5 min. Therefore, the stability of the model can meet the requirements of practical application.

To further evaluate the performance of the proposed algorithm, we selected several algorithms suitable for large-scale dataset prediction for comparison, including random forest regression tree (RFRT), k-nearest neighbours (KNN), LSTM, and Deep Belief Network (DBN).We compare the performance of each algorithm through the following five indicators, CPU running time (Time), average-absolute-error (MAE), mean-square-error (MSE), root-mean-square-error (RMSE), and variance (VAR) in the predictions. In this study, for the setting of super parameters and the training process of RFRT, we referred [18]. Imitating the treatment of the KNN algorithm in [49], to improve the prediction accuracy, the number of neighbours in this study is different under different weather conditions. The design of the LSTM structure in this study is similar to that of [11]. In the process of predicting delay by DBN, we referred [16].

The parameters of all baseline algorithms are selected by the grid search method. All results are shown in Figure 11. RFRT and KNN are machine learning methods, and model training is fast; KNN only considers the distance between samples, so model training is the fastest. These algorithms can only input the direct influence factor (x^a) of samples, so their prediction accuracy is not as good as LSTM-AM. LSTM and DBN are typical deep learning algorithms. LSTM can only absorb sequence data (x^b) in the training process, so its accuracy is not high. Since DBN can process large-dimensional data, its prediction results have accuracy similar to that of the proposed algorithm, but the training time is longer due to a large number of internal parameters. The results of LSTM also show that in processing indirect influencing factors (sequence data), our proposed method of adding an attention mechanism can effectively improve the accuracy of prediction. As can be seen from the last column, none of the models can strike a good balance between bias and variance (bias-variance dilemma). Although the prediction accuracy of KNN is not high, the variance of the predicted value is small. The prediction variance of DBN, RFRT and LSTM is about 4120, 3500 and 3680 respectively, while the variance of

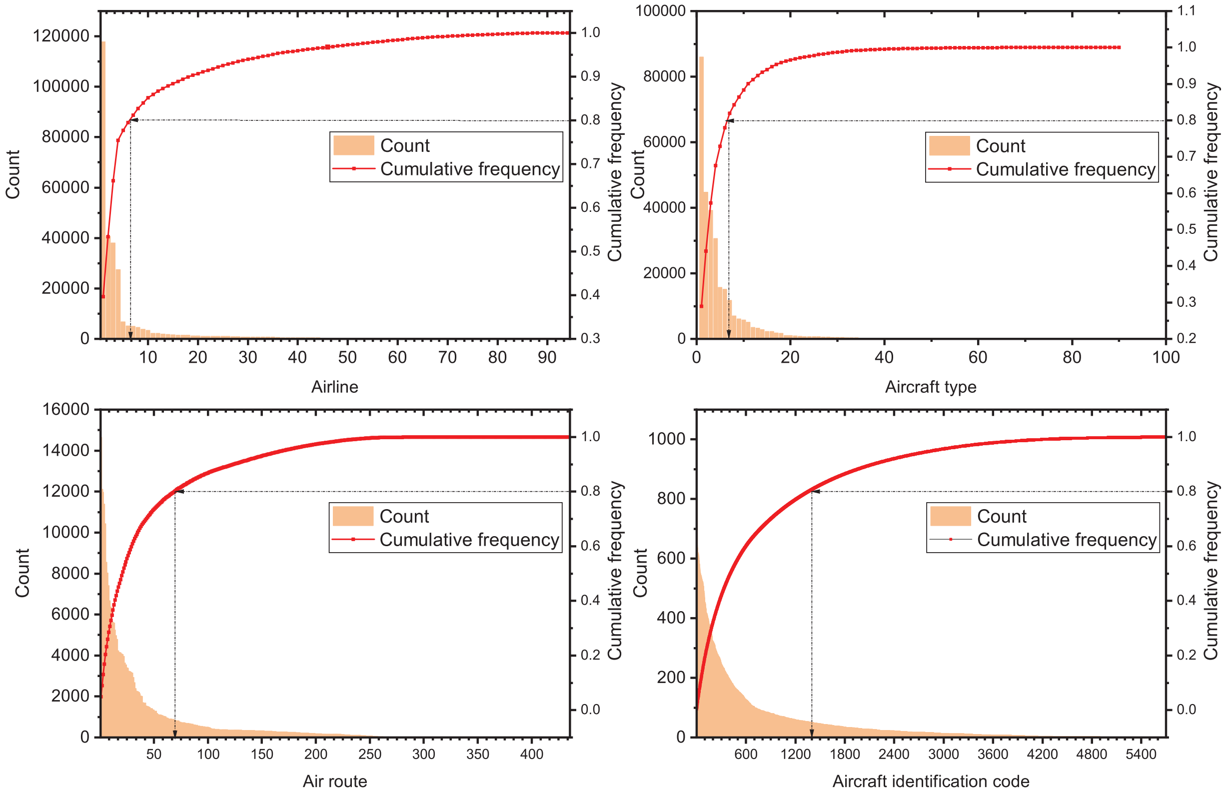

FIGURE 8 Bar chart and cumulative frequency curve of flight data grouped by different types of aircraft properties.

TABLE 7 An example of Pareto principle encoding

Airlines	Proportion	CF	Index	One-hot	Pareto principle
CA	39.68%	39.68%	1	[1,0,0,…,0]	[1,0,0,0,0,0,0,0]
CZ	13.70%	53.38%	2	[0,1,0,…,0]	[0,1,0,0,0,0,0,0]
MU	12.80%	66.18%	3	[0,0,1,…,0]	[0,0,1,0,0,0,0,0]
HU	9.28%	75.46%	4	[0,0,0,…,0]	[0,0,0,1,0,0,0,0]
MF	2.27%	77.43%	5	[0,0,0,…,0]	[0,0,0,0,1,0,0,0]
3U	1.75%	79.48%	6	[0,0,0,…,0]	[0,0,0,0,0,1,0,0]
SC	1.67%	81.15%	7	[0,0,0,…,0]	[0,0,0,0,0,0,1,0]
ZH	1.55%	82.70%	8	[0,0,0,…,0]	[0,0,0,0,0,0,0,1]
…	…	…	…	[0,0,0,…,0]	[0,0,0,0,0,0,0,1]
R3	0.0003%	100%	94	[0,0,0,…,1]	[0,0,0,0,0,0,0,1]

Note: CF means the cumulative frequency.

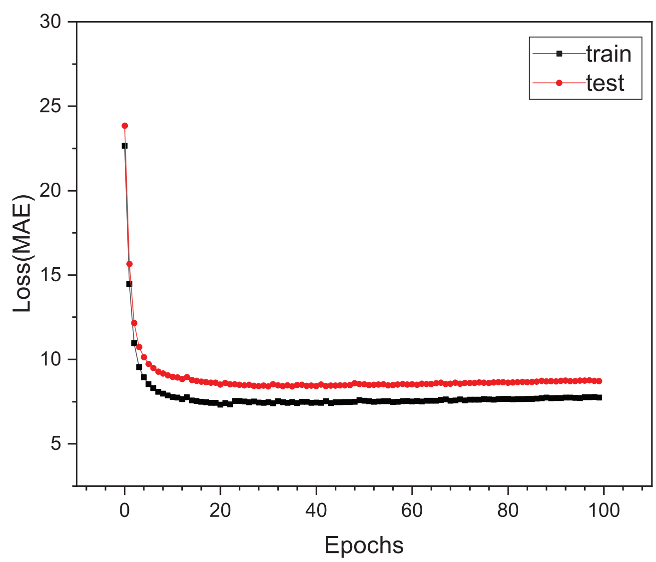

FIGURE 9 Model loss with the attention-based approach

the method used in this paper is about 2240. In contrast, our model can better balance deviation and variance. At the same time, the training time of the proposed method is 58 min so that the model can be scrolled and updated online with one-week or one-day new data.

6.2 | Analysis of critical time point

A more accurate forecast of delays can more comprehensively inform airport management and release news in advance to ease passengers' anxiety. However, to more accurately predict delay times is not enough in many scenarios. Managers want to know the factors causing delays, which can eliminate uncertainty in advance. Because without determining the real cause, hasty measures are likely to have a poor effect.

From Equation (7), we can know that the total delay y_i has two parts: $y_i^1 \cdot w_i^1$ and $y_i^2 \cdot w_i^2$. The first is related to the direct influence factor x^a. If this part makes a greater contribution to

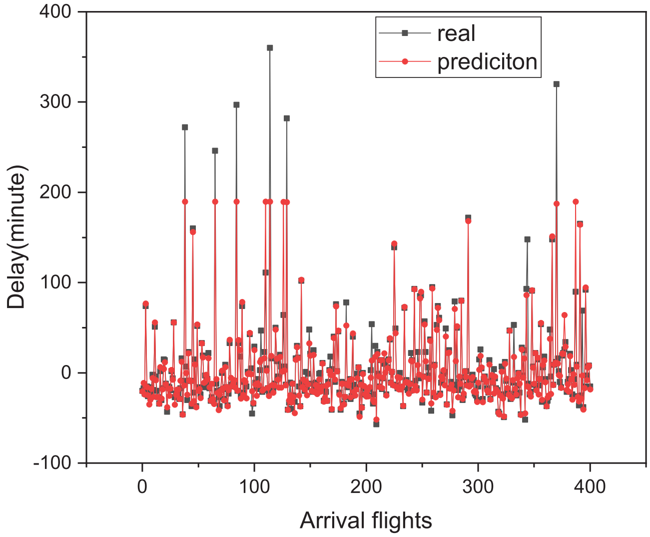

FIGURE 10 Part of real and predicted delays on test dataset

TABLE 8 Statistics between actual and predicted delays

Delay (min)	Proportion
30	92.91%
20	88.43%
15	83.90%
10	73.91%

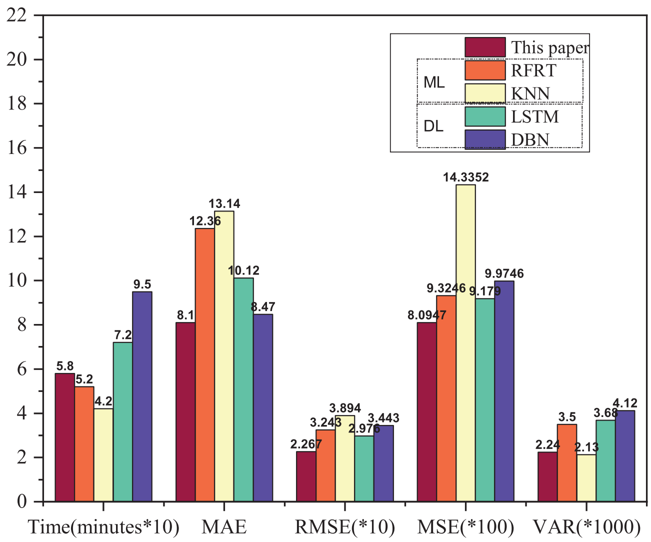

FIGURE 11 Comparison of proposed and baseline methods

the delay, then the flight has fewer resources or lower priority at the airport, which is difficult to eliminate at the operational level. The second term is related to the indirect factor x^b, and if this has a more significant impact on the delay, then it is due to the previous flight or poor management optimization at the airport. If we can determine that indirect factors play a major role in flight delays, then we can track the most critical factors in historical operations through the vector α^i.

Table 9 shows the prediction and breakdown of delays for 10 flights from 01:15 PM to 01:25 PM on 4 July 2019. It can be seen from Table 8 that CA flights have fewer delays, and the impact caused by direct factors is more significant than that caused by indirect factors ($y_i^1 \cdot w_i^1 > y_i^2 \cdot w_i^2$). This is mainly because Air China has a higher priority for use of the Capital Airport's resources, as the largest base airline. The delays predicted for CA4126, MU2105, CA1510, MU5107, ZH9103, CA1828 and HU7606 are less than 20 min, which can be easily absorbed and eliminated by managers. However, the delay predictions of CZ6138 and CA934 are 128.73 and 41.39 min, respectively, significantly affecting airport management and passenger satisfaction, so special attention and cause determination are required. In the delays of CZ6138 and CA934, indirect influence factors are much greater than direct influence factors ($y_i^2 \cdot w_i^2 >> y_i^1 \cdot w_i^1$). Therefore, these two delays can be explained and traced back to historical data.

In Figure 12, the attention matrix shows which parts of indirect factors pay more attention to the result. The horizontal axis shows the flight index of 10 flights in Table 9, and the vertical axis shows the position of the input sequence. The length of our model input sequence is 50, and the larger the ordinate, the closer to the prediction point. It can also be seen intuitively that the model gives more weight to input segments closer to the prediction point, which shows that the value of information will depreciate over time. Newer information has a more significant impact on the result.

From the green box in the picture, it can be seen that CZ6138 was greatly affected by the 14th and 15th previous flights and the airport resources at that moment. In contrast, CA934 was greatly influenced by the fourth and fifth previous flights arriving and/or departing, and the airport's situation at that point. When airport operators or air traffic controllers deploy resources to eliminate these delays, they can focus on adjusting resources at the point of 14th, 15th, fourth, and fifth flights arriving and/or departing, reducing the decision-making cost.

So, our model can output the time points that are closely related to the predicted results (the arrival or departure time of the previous flight). This result is of great significance to guide the ground management, just as shown in Section 6.3.

6.3 | Impact of prediction result on airline and airport management

Arrival and departure delays can be predicted by LSTM-AM. Airport managers can prepare for delays and optimize the decision process according to the result of this paper.

6.3.1 | Influence of prediction result release on passenger waiting stress

Minimizing delays caused by lack of service capacity and providing a comfortable and convenient environment for passengers

TABLE 9 Result decomposition of a prediction result

Flight number	Flight index	Schedule time	Delay	Prediction	$y_i^a \cdot w_i^a$	$y_i^b \cdot w_i^b$
CA4126	1	2019/7/4 13:15	−21	−17.76	−17.7211	−0.03885
MU2105	2	2019/7/4 13:15	8	5.34	3.566366	1.773634
CA1510	3	2019/7/4 13:20	5	7.28	6.392766	0.887234
CZ6138	4	2019/7/4 13:20	133	128	10	118
MU5107	5	2019/7/4 13:20	−14	−17.19	−9.75326	−7.43674
ZH9103	6	2019/7/4 13:20	−16	−14.58	−13.5127	−1.06729
CA1828	7	2019/7/4 13:25	1	9.74	6.058745	3.681255
CA934	8	2019/7/4 13:25	50	41	8	32
HU7606	9	2019/7/4 13:25	−20	−17.37	−0.47585	−16.8942
TV9815	10	2019/7/4 13:25	−4	1.02	0.47	0.55

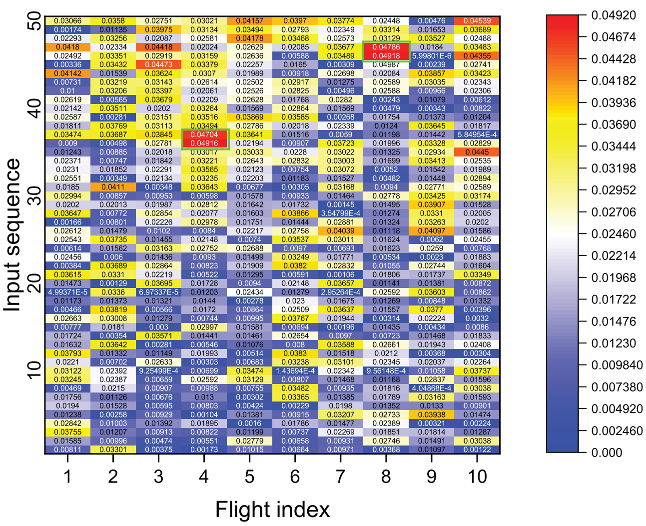

FIGURE 12 Attention matrix (α) of predicted results in Table 8

are two important functions of the airport. With the development of the aviation industry, passengers have higher and higher requirements for travel quality, in which the waiting stress at the airport is an important influencing factor. According to CAAC regulations, when latency occurs, the airline must inform passengers at least 30 min in advance. To verify that using the forecast results of this paper to notify passengers in advance can better alleviate the stress of passengers, we have done the following experiments.

We set up numerical experiments to compare our delay prediction results to CAAC regulations at reducing passenger stress. Since the dataset included no release time of delay information, we assumed a Gaussian distribution for advance time of delay notification, with mean μ and variance σ^2,

$$(t_s - t_n) \approx N\left(\mu, \sigma^2\right). \tag{43}$$

For the convenience of expression, μ_1 and σ_1^2 represent the mean and variance, respectively, of the current model (by CAAC regulations), and μ_2 and σ_2^2 are those of this study. The advance

TABLE 10 Scenario setting with constant stress function

Scenario	$f^{s1}(t)$	$f^{s2}(t)$	$f^{s3}(t)$	$f^{s4}(t)$
1	1	1.3	1.4	1.5
2	1	1.3	1.5	1.6
3	1	1.3	1.6	1.8
4	1	0.8	1	1.2
5	1	0.8	1.2	1.3
6	0.8	0.8	1.4	1.5
7	0.8	1.2	1.3	1.5
8	0.8	1.2	1.5	1.6
9	1.2	1.2	1.6	1.8
10	1.2	1.4	1.6	1.8

waiting time of passengers was set to a Gaussian distribution with mean μ' and variance σ'^2,

$$(t_s - t_a) \approx N\left(\mu', \sigma'^2\right). \tag{44}$$

According to CAAC regulations and our settings in Section 5 (the input sequence has 50-time steps), μ_1 and μ_2 were set to 60 minutes and 100 min, respectively. Airlines have different check-in time regulations, so μ' was set to the recommended early check-in time of each airline. Ten scenarios were set for cases of a constant pressure function and one following a Gaussian, as shown in Tables 10 and 11, respectively.

The test dataset was used to evaluate the stress values of passengers in the scenarios in Tables 10 and 11, with results as shown in Figures 13 and 14, respectively.

As can be seen from the table, the results fluctuated due to the arrival of passengers and the different airline check-in times. In general, the higher the mean of the pressure function, the greater the total passenger pressure (scenarios 1 and 3). However, it also can be seen that the prediction results of this paper can effectively reduce the stress of passengers through advance notification. When the stress function was constant, the airline could reduce the pressure value of passengers by about 35.56%

TABLE 11 Scenario setting with stress function following a Gaussian distribution

Scenario	$f^{s1}(t)$	$f^{s2}(t)$	$f^{s3}(t)$	$f^{s4}(t)$
1	[1,1]	[1.3,1]	[1.4,1]	[1.5,1]
2	[1,0.8]	[1.3,1.2]	[1.5,1]	[1.6,1]
3	[1,0.8]	[1.3,0.8]	[1.6,1]	[1.8,1]
4	[1,1.2]	[0.8,1]	[1,1]	[1.2,1]
5	[1,1.2]	[0.8,1.2]	[1.2,1]	[1.3,1]
6	[0.8,0.8]	[0.8,0.8]	[1.4,1.2]	[1.5,0.8]
7	[0.8,0.8]	[1.2,1]	[1.3,1]	[1.5,1.2]
8	[0.8,1.2]	[1.2,1.2]	[1.5,1.2]	[1.6,1.2]
9	[1.2,1]	[1.2,0.8]	[1.6,1.2]	[1.8,0.8]
10	[1.2,1.2]	[1.4,1]	[1.6,0.8]	[1.8,1.2]

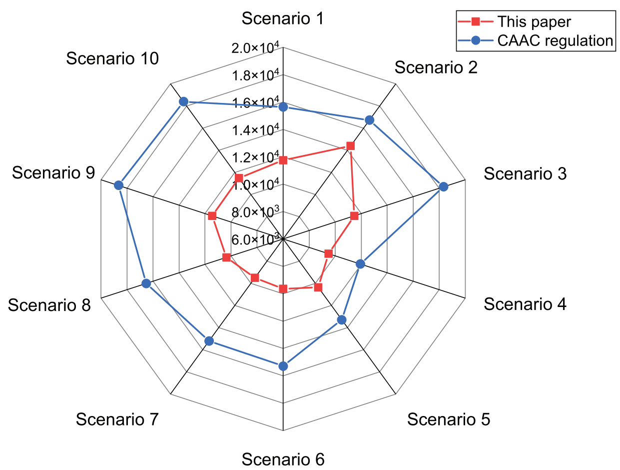

FIGURE 13 Comparison of stress values when stress function is constant

on average by announcing delay prediction results, and with a Gaussian pressure function, by 30.65% on average. Therefore, our prediction results can effectively guide passenger management.

6.3.2 | Eliminating delays by airport management with attention matrix

The relationship between inputs and outputs obtained through attentional mechanisms cannot be directly applied to airport or airline management. Because higher correlations in attention vectors do not indicate more significant effects. To prove that the method proposed in this paper can meet the needs of actual airport management, we set up the following experiment.

Attention matrix (α) contains the contribution of past time steps to the delay of the target flight, as shown in Figure 12. Many factors in past time steps are directly related to the management of the airport, such as the allocation of the airport apron, runway, and the control of the parking apron or runway congestion. The airport can reduce the delay time of the target flight by adjusting ground support resources. However, limited

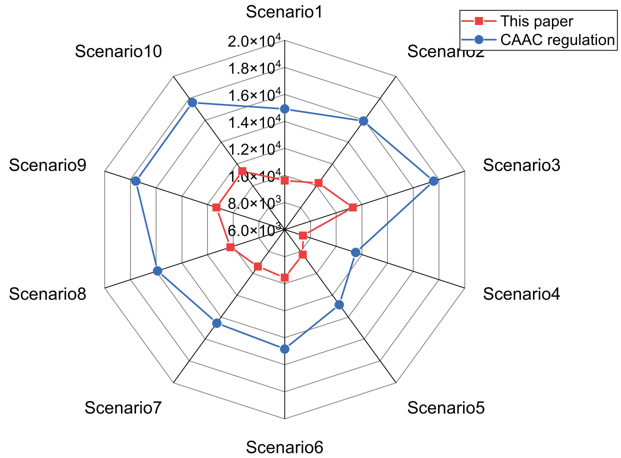

FIGURE 14 Comparison of stress values when stress function follows Gaussian distribution

manpower and equipment make it impossible to monitor and control the negative factors that cause flight delays. Through the analysis of α in each prediction result, we can grasp the important time step that causes the delay, to effectively eliminate or reduce it.

It is assumed in this study that measures taken by the airport to deal with delayed flights include the following aspects:

a. release the parking space occupancy of the target flight in advance (M−1);
b. reduce the parking density on the apron at important time steps for target flights (M−2);
c. reduce the runway density at critical time steps for target flights (M−3).

To simulate the implementation of the above three measures on the test set, the input data of the test set should be processed as follows:

(i) to implement M−1 in a hypothetical scenario, we adjust the parking space in the input data from occupied to idle;
(ii) to implement M−2 and M−3 in a hypothetical scenario, runway and apron density within the critical time step is changed to 60% of the original density.

Data from the test dataset were used to calculate the drop percentage of a predicted delay before and after the proposed measures were taken, with results as shown in Figure 15, where NIT is the number of top important time steps handled by the airport (NIT $\leq$ 50).

It can be seen from Figure 15 that, as more critical time steps are considered, the length of the delay becomes less. The three measures (M-1, M-2, and M-3) proposed by the airport in each time step can reduce the delay by up to about 80%. We note that actions taken within the first 10 time steps reduced delays by about 60%, and later important time steps had less impact on delays. It can also be seen from the experimental results that the correlation obtained through the attention mechanism is not directly linear with the relief of delay. A larger correlation

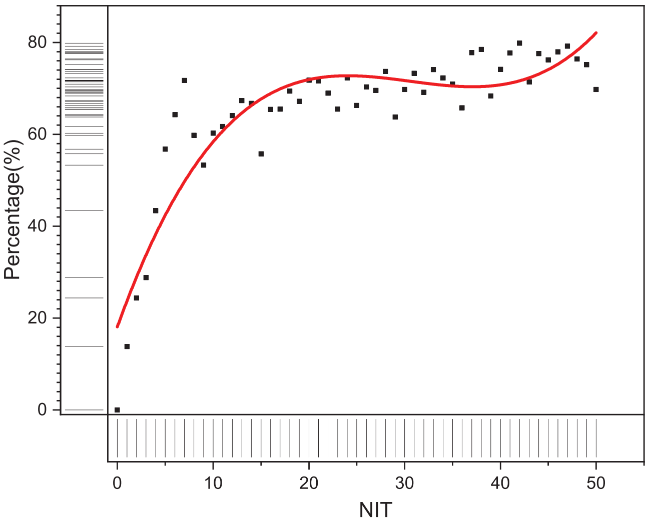

FIGURE 15 Drop percentage of predicted delay before and after with different NIT

coefficient does not guarantee a greater impact on delays. Therefore, considering all the important factors captured by the model in management cannot alleviate the congestion to the maximum extent. In this dataset, the airport management process should focus on the top 10 important time steps to minimize delays with the minimum cost of manpower and equipment.

7 | CONCLUSIONS AND FUTURE RESEARCH

With the rapid development of the air transport industry, in countries with less aviation airspace and high population density, such as China, flight delays increasingly affect the improvement of the air transport efficiency and travel experience. Among the policies and measures to alleviate flight delays, accurate prediction plays a critical role. The prediction of flight delays can provide a decision-making reference for airline crew configuration and ground guarantee resource allocation. The result will help ease passenger tension and anxiety, and provide a reference for passenger time management.

To the extent possible, we considered all influencing factors related to flight delays. These are of two types: one is directly related to flight attributes and arrival times, which are not subject to direct intervention, and the other consists of indirect influencing factors associated with the pre-order flight and airport state. The indirect influencing factors allow human intervention before flight delay occurs. Such a data classification method reduces the burden of model training and enriches the expression of datasets. The attention model adopts a seq2seq structure. Adding an attention mechanism layer can intuitively show the degree of attention to input data, which is of great significance for the early mining and analysis of delay. We used PEK 2019 annual data for model training. Pareto encoding accelerated training and eliminated the influence of redundant information. Results showed that Pareto encoding could greatly reduce data storage compared to one-hot encoding. Experimental results show that the method in this paper has better prediction accuracy than baseline algorithms (RFRT, KNN, LSTM, DBNs). The mean absolute error (MAE) of LSTM-AM is about 8.15 min on a test dataset. The results of the model can clearly output the influence of direct and indirect factors on the final results. At the same time, the influence of indirect factors can be tracked by attention vector, which is very important for the fine management

Our findings focus on the following: (1) The prediction results of the LSTM-AM model considering the direct and indirect influencing factors are better than the current mainstream machine learning and deep learning algorithms mentioned in the paper. (2) Many influencing factors related to airport delay have Pareto properties. The dimension of input data can be reduced by using this property. (3) The delay can be classified by direct factors and indirect factors considered in prediction separately. In airport management, to save resources, more manpower and equipment can be used to mitigate or eliminate delays caused by indirect factors. (4)Accurate forecasts (LSTM-AM) of flight delays combining attention mechanisms can help relieve the waiting pressure of passengers.

There are still deficiencies in this study. When mining the cause of the delay through the attention vector, it can only be located at previous arrival or departure times. The roles of certain factors cannot be accurately analysed. At the same time, this paper only uses the historical data of a single airport, without considering the interaction between airports in the airport cluster, as well as the historical data of multi-airport terminal airspace, which is a future research direction.

CONFLICT OF INTEREST

All authors declare that they do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

FUNDING INFORMATION

This work was supported by NSFC (72171019) and the Key Laboratory of Airports Cluster Intelligent Operation under Grant no. KLAGIO20180801.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Fujun Wang https://orcid.org/0000-0001-7861-3526

REFERENCES

1. Zheng, Z., Wei, W., Zou, B., Hu, M.: How late does your flight depart? A quantile regression approach for a Chinese case study. Sustainability-Basel 12(1055324), 1–16 (2020)
2. Wang, Y., Zheng, H., Wu, F., Chen, J., Hansen, M.: A comparative study on flight delay networks of the USA and China. J. Adv. Transp. 2020, 1–11 (2020)
3. Asif, M.T., Dauwels, J., Goh, C.Y., Oran, A., Fathi, E., Xu, M., et al.: Spatiotemporal patterns in large-scale traffic speed prediction. IEEE Trans. Intell. Transp. 15(2), 794–804 (2014)

4. Tu, Y., Ball, M.O., Jank, W.S.: Estimating flight departure delay distributions - A statistical approach with long-term trend and short-term pattern. J. Am. Stat. Assoc. 103(481), 112–125 (2008)
5. Hsiao, C., Hansen, M.: Econometric analysis of US airline flight delays with time-of-day effects. Transp. Res. Rec. 1951, 104–112 (2006)
6. Hao, L., Hansen, M., Zhang, Y., Post, J.: New York: Two ways of estimating the delay impact of New York airports. Transp. Res. Part E Logist. Transp. Rev. 70:245–260 (2014)
7. Rodriguez-Sanza, A., Gomez Comendador, F., Arnaldo Valdes, R., Perez-Castan, J., Barragan Montes, R., Camara Serrano, S.: Assessment of airport arrival congestion and delay: Prediction and reliability. Transp. Res. Part C Emerging Technol. 98, 255–283 (2019)
8. Rebollo, J.J., Balakrishnan, H.: Characterization and prediction of air traffic delays. Transp. Res. Part C Emerging Technol. 44, 231–241 (2014)
9. Manna, S., Biswas, S., Kundu, R., Rakshit, S., Gupta, P., Barman, S.: A statistical approach to predict flight delay using gradient boosted decision tree. In: 2017 International Conference on Computational Intelligence in Data Science (ICCIDS). Chennai, Tamil Nadu (2017)
10. Kim, Y.J., Choi, S., Briceno, S., Mavris, D., (eds.): A Deep Learning Approach to Flight Delay Prediction. IEEE, Sacramento, CA (2016)
11. McCarthy, N., Karzand, M., Lecue, F.: Amsterdam to dublin eventually delayed? LSTM and transfer learning for predicting delays of low cost airlines. Proc. AAAI Conf. Artif. Intell. 33(01), 9541–9546 (2019)
12. Li, Q., Jing, R.: Generation and prediction of flight delays in air transport. IET Intell. Transp. Syst. 15(6), 740–753 (2021)
13. Guo, Z., Yu, B., Hao, M., Wang, W., Jiang, Y., Zong, F.: A novel hybrid method for flight departure delay prediction using Random Forest Regression and Maximal Information Coefficient. Aerosp. Sci. Technol. 116(106822), 1–10 (2021)
14. Ai, Y., Pan, W., Yang, C., Wu, D., Tang, J.: A deep learning approach to predict the spatial and temporal distribution of flight delay in network. J. Intell. Fuzzy Syst. 37(5), 6029–6037 (2019)
15. Zhang, H., Song, C., Zhang, J., Wang, H., Guo, J.: A multi-step airport delay prediction model based on spatial-temporal correlation and auxiliary features. IET Intell. Transp. Syst. 15(7), 916–928 (2021)
16. Yu, B., Guo, Z., Asian, S., Wang, H., Chen, G.: Flight delay prediction for commercial air transport: a deep learning approach. Transp. Res. E 125, 203–221 (2019)
17. Liu, Y., Liu, Y., Hansen, M., Pozdnukhov, A., Zhang, D.: Using machine learning to analyze air traffic management actions: Ground delay program case study. Transp. Res. Part E Logist. Transp. Rev. 131, 80–95 (2019)
18. Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., Zhao, D.: Flight delay prediction based on aviation big data and machine learning. IEEE Trans. Veh. Technol. 69(1), 140–150 (2020)
19. Yi, J., Zhang, H., Liu, H., Zhong, G., Li, G.: Flight delay classification prediction based on stacking algorithm. J. Adv. Transp. 2021(4292778), 1–10 (2021)
20. Khan, W.A., Ma, H., Chung, S., Wen, X.: Hierarchical integrated machine learning model for predicting flight departure delays and duration in series. Transp. Res. Part C Emerging Technol. 129, 103225 (2021)
21. Zhang, J., Peng, Z., Yang, C., Wang, B.: Data-driven flight time prediction for arrival aircraft within the terminal area. IET Intell. Transp. Syst. 16(2), 263–275 (2021)
22. Wang, C., Hu, M., Yang, L., Zhao, Z.: Prediction of air traffic delays: An agent-based model introducing refined parameter estimation methods. PLoS One 16(e02497544), 1–22 (2021)
23. Bao, J., Yang, Z., Zeng, W.: Graph to sequence learning with attention mechanism for network-wide multi-step-ahead flight delay prediction. Transp. Res. Part C Emerging Technol. 130(103323), 1–18 (2021)
24. Cai, K., Li, Y., Fang, Y., Zhu, Y.: A deep learning approach for flight delay prediction through time-evolving graphs. IEEE Trans. Intell. Transp. 162(1), 1–11 (2021)
25. Song, C., Guo, J., Zhuang, J.: Analyzing passengers' emotions following flight delays- a 2011–2019 case study on SKYTRAX comments. J. Air Transp. Manage. 89(101903), 1–7 (2020)
26. Schultz, M., Reitmann, S.: Machine learning approach to predict aircraft boarding. Transp. Res. Part C Emerging Technol. 98, 391–408 (2019)
27. Khan, W.A., Ma, H., Chung, S., Wen, X.: Hierarchical integrated machine learning model for predicting flight departure delays and duration in series. Transp. Res. Part C Emerging Technol. 129(103225) (2021)
28. Balakrishna, P., Ganesan, R., Sherry, L.: Accuracy of reinforcement learning algorithms for predicting aircraft taxi-out times: a case-study of Tampa Bay departures. Transp. Res. Part C Emerging Technol. 18(6SI), 950–962 (2010)
29. Liu, Y., Hansen, M., Gupta, G., Malik, W., Jung, Y.: Predictability impacts of airport surface automation. Transport. Res. Part C Emerging Technol. 44, 128–145 (2014)
30. Lordan, O., Sallan, J.M., Valenzuela-Arroyo, M.: Forecasting of taxi times: The case of Barcelona-El Prat airport. J. Air Transp. Manage. 56(SIB), 118–122 (2016)
31. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. 9(8), 1735–1780 (1997)
32. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Networks 5(2), 157–166 (1994)
33. Yang, Y., Sarkis, R., El Atrache, R., Loddenkemper, T., Meisel, C.: Video-based detection of generalized tonic-clonic seizures using deep learning. IEEE J. Biomed. Health 25(8), 2997–3008 (2021)
34. Cho, K., Van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., et al. (eds.): Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. Association for Computational Linguistics (ACL), Doha, Qatar (2014)
35. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. Comput. Sci. 1409(0473), 1–15 (2014)
36. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., et al. (eds.): Attention is All You Need. Neural Information Processing Systems Foundation, Long Beach, CA (2017)
37. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature 323(6088), 533–536 (1986)
38. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
39. Duchi, J., Hazan, E., Singer, Y.: Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn Res. 12(7), 2121–2159 (2011)
40. Hinton, G., Krizhevsky, A., Jaitly, N., Tieleman, T., Tang, Y. (eds.): Does the Brain do Inverse Graphics. 2, 1–46. Brain and Cognitive Sciences Fall Colloquium (2012)
41. Osuna, E.E.: The psychological cost of waiting. J. Math. Psychol. 29(1), 82–105 (1985)
42. Suck, R., Holling, H.: Stress caused by waiting: a theoretical evaluation of a mathematical model. J. Math. Psychol. 41(3), 280–286 (1997)
43. ACI reveals top 20 airports for passenger traffic, cargo, and aircraft movements - ACI World. (2021) [cited 2021 July]. Available from:https://aci.aero/2020/05/19/aci-reveals-top-20-airports-for-passenger-traffic-cargo-and-aircraft-movements/
44. Variflight. On-time performance report of global airports in 2019. (2021) [cited 2021 July]. Available from: https://data.variflight.com/reports
45. Wang, X., Brownlee, A., Woodward, J.R., Weiszer, M., Mahfouf, M., Chen, J.: Aircraft taxi time prediction: Feature importance and their implications. Transp. Res. Part C Emerging Technol. 124(102892), 1–23 (2021)
46. Wang, C., Wang, X.: Airport congestion delays and airline networks. Transp. Res. Part E Logist. Transp. Rev. 122, 328–349 (2019)
47. Ioffe, S., Szegedy, C. (eds.).: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning. Lille, France (2015)
48. Pelosi, G., Selleri, S.: To celigny, in the footprints of vilfredo pareto's optimum. IEEE Antennas Propag. Mag. 56(3), 249–254 (2014)
49. Sun, B., Cheng, W., Goswami, P., Bai, G.: Short-term traffic forecasting using self-adjusting k-nearest neighbours. IET Intell. Transp. Syst. 12(1), 41–48 (2018)

How to cite this article: Wang, F., Bi, J., Xie, D., Zhao, X.: Flight delay forecasting and analysis of direct and indirect factors. IET Intell. Transp. Syst. 16, 890–907 (2022). https://doi.org/10.1049/itr2.12183