




Article

Multi-Attribute Data-Driven Flight Departure Delay Prediction for Airport System Using Deep Learning Method

Yujie Yuan ¹ , Yantao Wang ¹  and Chun Sing Lai ^{2,*} 

¹ School of Air Traffic Management, Civil Aviation University of China, Tianjin 300300, China; yjyuan@cauc.edu.cn (Y.Y.); yt-wang@cauc.edu.cn (Y.W.)

² Department of Electronic and Electrical Engineering, Brunel University of London, London UB8 3PH, UK

* Correspondence: chunsing.lai@brunel.ac.uk

Abstract: Complex and diverse multi-attribute flight data can provide data-driven opportunities for airport flight delay prediction. However, it is a challenge to effectively and efficiently process multi-attribute flight data. This paper proposes a hybrid dynamic spatial-temporal long short-term memory network (LSTM) with 3D-directional multi-attribute features (3DF-DSCL) for departure flight delay prediction. The model is based on a 3D convolutional neural network (3D-CNN), graph convolutional network (GCN) and long short-term memory networks (LSTM) model. Firstly, the dataset divides the state and environment of departure flight delay into three situations, including the dynamic operation link, which integrates the trajectory system of aircraft movement in the terminal area, the network congestion link caused by aircraft multi-area movement in the air and ground, and other delay factors determined by the airport take-off and landing requirements. Multi-attribute data are divided into time series, spatial-temporal network and dynamic moving trajectory grid input variables. Among them, the spatial network and dynamic moving trajectory grid data are the inputs of GCN and 3D CNN models, which aim to extract spatial-temporal features. The time series input variables are fed into LSTM. These features are then integrated and fed into LSTM for flight delay prediction, where the flight delay of airport outbound flights is taken as the output variable. The case study shows that the proposed method can significantly improve the accuracy of flight prediction delay. The Mean Absolute Error (MAE) can reach 0.26, which is a 14.47% reduction compared with 2D CNN+GCN+LSTM.

Keywords: dynamic system; multi-attribute data; deep learning; airport delay prediction



Academic Editor: Judith Rosenow

Received: 17 December 2024

Revised: 27 February 2025

Accepted: 7 March 2025

Published: 17 March 2025

Citation: Yuan, Y.; Wang, Y.; Lai, C.S.

Multi-Attribute Data-Driven Flight Departure Delay Prediction for Airport System Using Deep Learning Method. *Aerospace* **2025**, *12*, 246.

<https://doi.org/10.3390/aerospace12030246>

aerospace12030246

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

licenses/by/4.0/).

1. Introduction

Multi-determinants can cause airport delays [1]. The key to improving the precision of airport delay prediction is to comprehensively take all factors into account [2]. Meanwhile, the increasingly widespread applications of large-scale datasets and sensor technologies have promoted a revolutionary information age for airport operation and management. The applicability of large-scale data enables the airport operation control (AOC) system to capture the potential features of complex dynamic systems from the perspective of big aviation data, such as aircraft trajectory, meteorology and network systems. Diversity of data provides more comprehensive and intelligent data support to accurately predict airport flight delays. However, the dynamic system data also bring several significant challenges to the modeling of flight delay prediction:

(1) Complexity and interdependence of data subjects: The AOC system includes the subsystems of humans, equipment and weather conditions. These factors may affect the

flight delay independently or cause a composite effect on the flight delay [1]. Because of the complexity and interdependency between subsystems, the extracted flight delay data will be affected by multi-factors, and some of them are difficult to be directly used in functional models, such as stress response of human body and abrupt change of weather.

(2) Diversity of data attribute: In AOC system, the research object of airport flight departure delay is the movement of aircraft, which leads to the increase in data dimensions, including spatial dimension (i.e., the state and position of aircraft operation), time series dimension (i.e., the process of aircraft operation, the corresponding weather, delay time, flight crews, passenger and other related characteristics) and non-time series dimension (i.e., the equipment state and operators of aircraft, etc.). The increase in dimensions degrades the accuracy and efficiency of the existing flight delay prediction models.

(3) Dependency of spatial-temporal dimensions: The dynamic movement of aircraft means that the state of the system is continuous, and the states of time and position are simultaneously correlated, thus there is a spatial-temporal dependency between the delayed state and the past operating state of the aircraft.

In response to the challenges, this paper proposes 3D-directional multi-attribute features (3DF-DSCL), a deep learning (DL) framework to predict the multi-attribute data-driven flight delay. The main contributions of this study are as follows:

(1) A new DL method is proposed based on multi-attribute dynamic data (including spatial dimension, temporal dimension and spatial-temporal dimension), which comprehensively considers various determinants of flight delay in airport operation system. The proposed DL method can effectively extract complex and interdependent multi-attribute data.

(2) Based on the features of graph convolutional network (GCN), long-term, short-term memory (LSTM) and 3D-convolutional neural network (CNN), the multi-attribute datasets are processed respectively by integrating the hybrid framework of 3DF-DSCL.

(3) By applying 3DF-DSCL model, the spatial-temporal relationship in each stage of aircraft operation delay is obtained.

The remainder of this paper is structured as follows: Section 2 provides a comprehensive review of the relevant literature. Section 3 formulates the multi-attribute data-based prediction problem and delineates the proposed methodology and framework. Section 4 presents the validation of the proposed approach through case studies, along with the determination of optimal training parameter values. Finally, Section 5 concludes the study and outlines potential avenues for future research.

2. Related Work

Classification and regression are two mainstream modeling methods for flight delay prediction. Regression methods include the automatic regression model, automatic regression moving average, automatic reverse integrated moving average (ARIMA) and Markov model [2–4]. In the classification model [5], a large amount of research has applied machine learning methods and achieved promising results [6]. For example, Khan et al. [7] used regression models for three major commercial airports in New York to predict flight delays. Highly skewed and scattered historical datasets make it difficult to accurately predict flight delay [8]. Chakrabarty et al. [9] compared four supervised machine learning algorithms, including random forest, support vector machine, gradient enhanced classifier and K algorithm, which are applied in American Airlines arrival delay prediction. In addition, the unbalanced distribution of the classes of delay [10], backpropagation (BP) neural network, support vector machine (SVM) and decision tree [11] have also attracted attention in flight delay prediction [12,13]. Gui et al. [14] used LSTM to forecast flight delays by collecting automatic dependent surveillance-broadcast (ADS-B) data and meteorological information.

The results showed that the LSTM model was an effective method for investigating time dependence in sequence data. However, the correlation and complexity of multi-factors may degrade the accuracy of these existing methods in flight delay prediction.

Mining and extracting multi-attribute, dynamic and high-dimensional data from large-scale datasets requires advanced methods to guarantee excellent performance, reliability and scalability. Deep learning (DL) is one of the most used methods for delay prediction. Through the hierarchical learning approach, it can extract a variety of complex and abstract data [15]. For example, recurrent neural network (RNN) [16] or its variants, LSTM [17,18], was used in time series data, two-dimensional convolutional neural network (2D CNN) was used to investigate spatial relations (images) [19], and spatial-temporal features selection of three-dimensional convolutional neural network (3D-CNN) [20] was used to extract dynamic and continuous movement of objects and spatial topological network relations of graph convolutional networks (GCNs) [21]. These applications can effectively provide state-of-the-art multi-attribute data processing methods for developing and integrating new DL architectures.

Recently, researchers applied machine learning technology to flight delay prediction [22,23]. Yu et al. [24] used a novel deep belief network method and support vector regression method to identify the micro-influencing factors in flight delay so that aviation authorities can explore the fundamental behavior and mechanism of flight delay. Margarida et al. [25] combined a dynamic Bayesian network to simulate flight delay. Some research combines the Autoregressive Integrated Moving Average (ARIMA) model [26], Alarm model [27], SVM [24] and BP [28,29] in flight delay prediction. The hybrid machine learning algorithm-based prediction model can achieve higher prediction accuracy than the traditional methods with mathematical statistics, but most of these models cannot deal with the complex spatial-temporal correlations among flights. Thus, the real-time prediction for multi-stage aircraft dynamic trajectory networks requires further exploration. Based on the convolution neural network, after RNN is introduced into the field of delay prediction [29], convolution operation can be used to better capture the spatial-temporal characteristics of aircraft taxiing [30], and deep-CNN was used to forecast flight delay by integrating flight data and meteorological data [15]. Although CNN has good feature learning ability, it can only be applied to Euclidean structure data. All non-Euclidean structure data must be transformed into a fixed format before being fed into the CNN model. In data transformation, some network topology data were disrupted or lost, so the relevant model (GCN) considering adjacency matrix and network topology has come into being [26,31]. However, GCN mostly uses shallow networks with one to four layers. When constructing a deep-seated graph network, the model performance is not as good as using a residual connection to construct a deep neural network (CNN) [32]. Therefore, high-order spatial features cannot be effectively extracted.

In order to overcome the shortcomings of using a single deep learning model and extract the spatial-temporal dependence of the network more effectively, various complex deep learning frameworks have been developed based on RNN, CNN, GCN, deep neural networks (DNNs) [33], graph temporal convolutional networks (TCNs) [34] and so on. For example, CNN and LSTM were combined for prediction [35,36]. Ai et al. [37] predicted the flight delay in the network structure by innovatively combining the summary layer and LSTM layer into an end-to-end deep learning architecture, which can capture the spatial and temporal features of delay variables at the same time. Yang et al. [38] proposed a hybrid spatial longitudinal long-short-term memory network (SCLN-TTF) for airway prediction index prediction. The model combined GCN and LSTM neural networks to extract the congestion of each operation state of aircraft effectively. Some studies embedded LSTM, Gaussian regression and other models under the framework of Auto encoder to

predict short-term delay [39,40]. Deep learning framework has better model performance than a single RNN, CNN and GCN model in most cases. However, some deep learning framework models cannot effectively deal with complex multi-attribute datasets. Therefore, it is necessary to consider the model complexity, model performance and application value comprehensively.

Although the deep tilt model has made significant progress in flight delay prediction, there are some key limitations in the existing deep learning architecture. First, multi-factors of dynamic datasets affect airport flight delays. The datasets have synchronization in spatial-temporal dimensions and have a 3D data structure. Secondly, in the airport, the operating paths of multiple aircraft will establish a multi-stage path network. The delay time series data on the whole network are mobile data with topological structure and delay propagation. Traditional CNN and GCN models can only process Euclidean structured data. Traditional LSTM and RNN structural models are inflexible, and most of them are suitable for multi-step-ahead flight delay prediction [22], while others are suitable for high-dimensional, temporarily ordered data with noise [1].

Based on the features of the above-mentioned DL methods, this paper proposes a DL architecture that combines 3D-CNN, GCN and LSTM. The proposed model is called 3DF-DSCL, which can effectively deal with complex datasets in dynamic systems. Multi-attribute datasets are the input variables, and these three methods are selected to extract features respectively. This architecture can deal with multiple data subjects: (1) 3D-CNN captures dynamic and continuous aircraft movement data features with spatial-temporal dependence; (2) GCN mines dynamic topological relations in airport aircraft operation network; (3) LSTM focuses on time series dimensions, which combines the weather and other related factors and further integrates time relations to understand the underlying effects of static factors.

3. Methodology

To predict departure delays for short-term flights, a hybrid dynamic spatial-temporal convolution along short-term memory network with 3D-directional multi-attribute features (3DF-DSCL) model is created. Figure 1 shows the architecture of the 3DF-DSCL approach. The model takes the processed multi-attribute data as the input and the delay time as the output. The architecture consists of four layers, including input, feature extraction, LSTM and output.

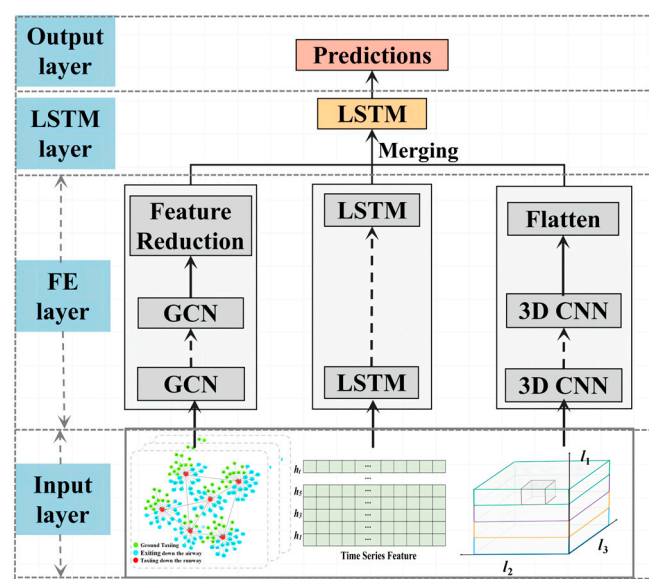


Figure 1. The architecture of the 3DF-DSCL model.

3.1. Input Data Processing

Previous studies have shown that there exist some connections between airport delays in aircraft operation flow, operation time, congestion and queuing, unfavorable weather, and other factors [41–43]. Among them, Roh et al. develop and assess winter climate hazard models to evaluate their impact on traffic volume. The study highlights how extreme weather conditions, such as snow, ice, and low temperatures, contribute to traffic delays, affecting both ground transportation and potentially air travel [41]. Sridhar et al. propose a short-term delay prediction model for the National Airspace System (NAS) using a Weather Impacted Traffic Index. This model considers adverse weather conditions, such as storms and turbulence, as key delay factors affecting air traffic operations [42]. Adacher et al. focus on solving air traffic congestion through rerouting algorithms. The study identifies congestion-related delays as a major challenge and explores optimization techniques to mitigate airspace bottlenecks and improve flight efficiency [43]. Based on these delay determinants, add the moving trajectory of the aircraft and the topology of the airport ground to better describe the aircraft operation state and taxi congestion variables. In our selected variables, the spatial feature data values are spatial variation, time static or relatively static data, including the spatial positions of the departure point, entry point and apron, and the moving distance between runway and taxiing record point. Time feature data are time-varying and spatial static data, including date, time, crew information and weather conditions. Spatial-temporal feature data refers to data that has changed in both time and space dimensions and has time-space dependence and correlation at the same time. Spatial-temporal data includes aircraft operation status data and the aircraft queuing congestion index within a fixed time interval.

3.1.1. Congestion Index Calculation

Congestion index (CI) refers to the degree of queuing and waiting of aircraft within the scope of airport operation control. Usually, the degree of airport congestion is estimated according to the ratio of the number of aircraft to the operating capacity in a fixed time interval. In order to consider different congestion states, it is widely used to calculate CI by combining the information on running flow, running distance and running time [44–47]. The airport CI is described as the cumulative running time divided by the running length in each state. The congestion index is calculated as $C_z = \sum_{n=1}^N \sum_{s=1}^S t_{z,s,n} / L_s$ of the airport during the z period [38].

L_i represents the actual length of the I section (km). S indicates the moving state of the aircraft in the air and on the ground during the z period. $t_{z,i,n}$ represents the taxiing or moving time of aircraft n in state i during z . N represents all aircraft operating during the flight f . The unit of C_z is in minutes per kilometer (min/km).

3.1.2. Aircraft Operational Stages Extraction

The aircraft operation in an airport can be divided into five stages. The length of time horizon between each stage may lead to the departure delays of the corresponding flight, and even the continuity between stages may aggravate the delay. Figure 2 shows the five stages of aircraft operation at the airport, which are closing the hatch stage, removing the chocks stage, holding-point stage, exiting clearance stage and departure stage. In data collection, the time point of each flight stage is recorded. All data processing results are shown in Table 1.

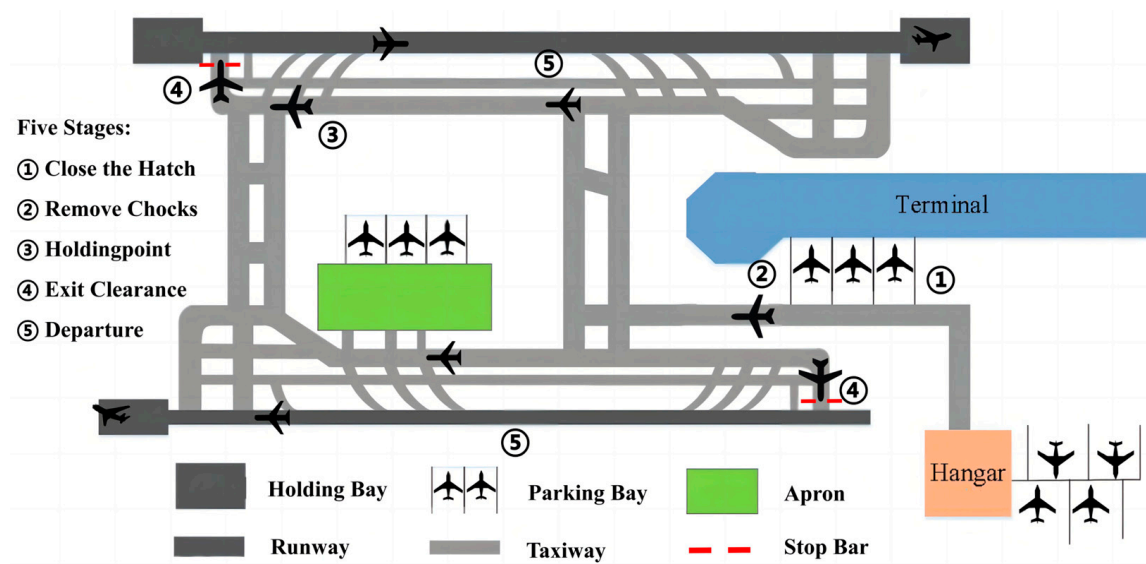


Figure 2. All stages of aircraft operation.

Table 1. Multi-attribution data processing.

Dimension	Name	Description
Temporal dataLa	Time	Time in a day (yyyy/mm/dd hh:mm:ss).
	Flight ID	The ID of each flight (e.g., CA975).
	Historic delay	Historic delay time of each flight (min).
	Weather	Dummy variable, general (or severe) unfavorable weather (mist, light rain, shower, rain, thunderstorm and snowfall) = 1; others = 0.
	Special situation	Dummy variable, Special situation (Political regulation, holidays, winter and summer vacations) = 1; others = 0.
	Aircrew ID	Dummy variable, crew scheduled = 1; others = 0.
	Aircraft ID	The type of aircraft (e.g., b737).
Spatial data	Longitude	Longitude of the sections and stages point.
	LatLatitude	Latitude of the sections and stages point.
	Moving sections length	Length for each moving sections (km), including taxiing, runway and airway.
	Congestion index	Congestion index for each moving sections (min/km).
Spatial-temporal data	Scheduled flow	Flow of each moving stages according to flight plan (aircraft/km).
	Aircraft operation status	Historic aircraft operation status of flight flow for each moving stages (aircraft/min).

3.2. Input Layer

The data from AOCC systems have many attributes, including static, dynamic, temporal, spatial and spatial-temporal data. These data are often difficult to process because of complexity, independence, and interactions [4,11]. The input layer processes and extracts the data into the standard data of each feature extraction layer.

3.2.1. Spatial Data

The airport structure layout and the CI of each section are the data with spatial features. The latitude and longitude coordinates of airport parking, taxiway, runway, and waterway are taken as points, and the aircraft moving congestion state is considered as connecting arcs; thus, the spatial characteristics can be effectively extracted, and the network topology can be formed as the basis of neural network. As shown in Figure 3, there are 3 runways, 15 apron areas, 87 taxiways and 59 airplanes in the airport structure layout. In order to construct the neural network input graph, the ground structure, runway distribution and air channel structure graph of the airport is transformed into the network structure graph as shown in the figure, and the topology graph is constructed according to the aircraft congestion state.

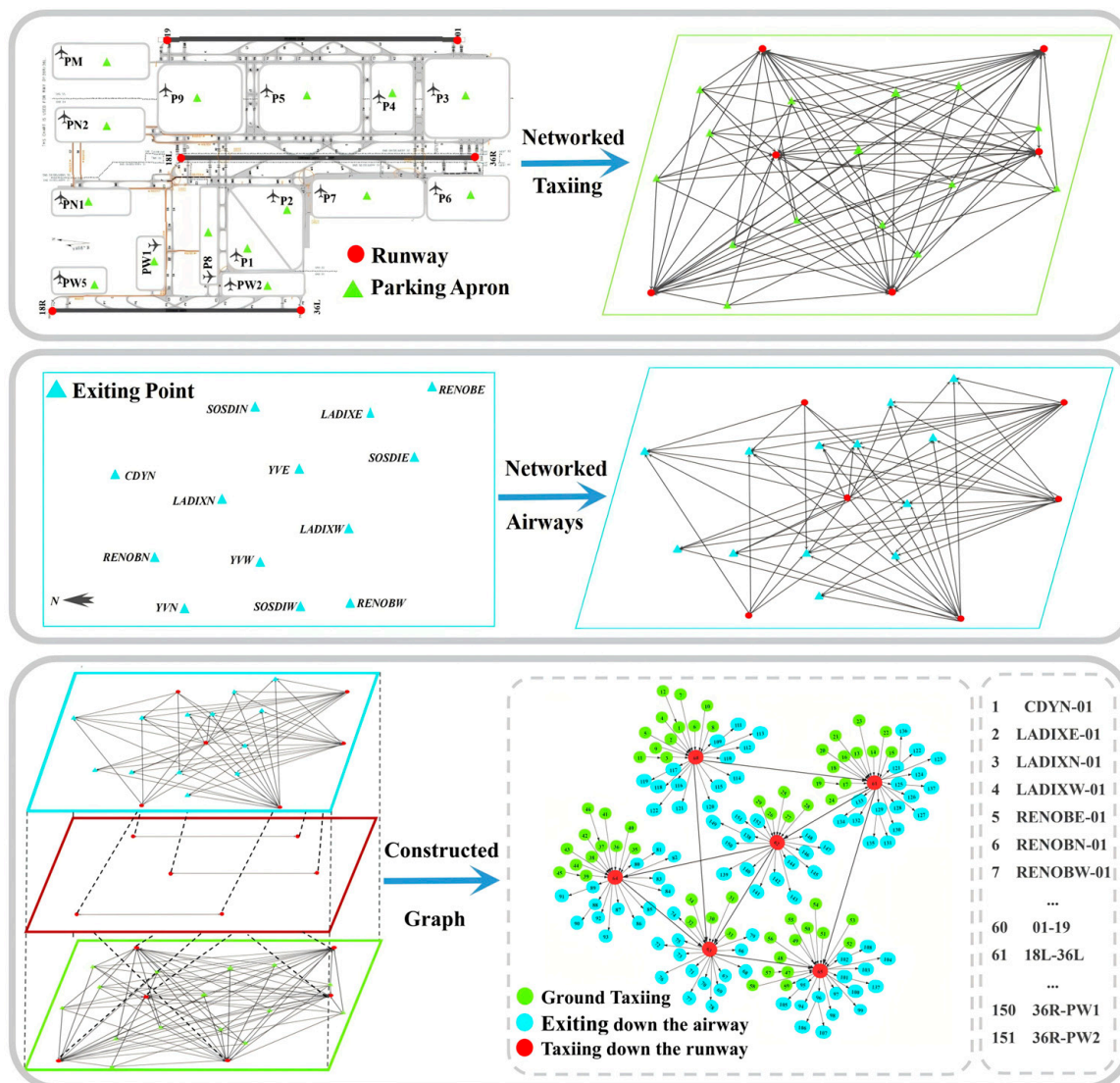


Figure 3. Spatial data construction.

When constructing the delay prediction influencing factors, the CI data are the spatial feature data of each road section, which describes the density of aircraft in the road section within a fixed time interval. The connection features of arcs in the network are adopted to reflect the dense situation. Therefore, when constructing CI features, the nodes in the topology network are defined as the connection arcs in the original network, and the connection lines between nodes are the direction of aircraft operation. Node 1 represents

section information between apron PW1 and runway 36R. The edge between nodes 1 and 65 corresponds to the holding point for Runway 36R. In the final topology graph, nodes 1–93 are the congestion of the ground channel, and nodes 94–152 are the congestion of the air channel.

Data definition is carried out based on the establishment of a network. Firstly, the topology network matrix airport network $G = (L, E)$ is constructed, where L is a set of air channel nodes, $L = \{l_1, l_2, \dots, l_M\}$, M is the number of nodes and E is aggregate of edges. Adjacency matrix A corresponds to the concatenate between air channels, $A \in R^{M \times M}$. The adjacency matrix only comprises the 0 and 1 elements. If there is no link between the two roads, the element is 0, and 1 indicates that there is a link. Secondly, the characteristic matrix $X^{M \times P}$ is defined. The CI information of each road section is taken as the attribute characteristics of network nodes, which is expressed as $X \in R^{M \times P}$. P is the number of attribute characteristics of nodes (the length of historical time series), and $X_i \in R^{M \times C_i}$ is the CI, i is used to represent each channel. Finally, learning mapping function f is based on network topology G and characteristic matrix X , and then the CI is calculated at the next time T as shown in Equation (1):

$$[X_{t+1}, X_{t+2}, \dots, X_{t+T}] = f(G; (X_{t-n}, \dots, X_{t-1}, X_t)), \quad (1)$$

where n and T represent the length of historical time series and the length of time series to be predicted, respectively.

3.2.2. Temporal Data

The flight delay influencing factors are transformed into a sequence form with time features (TFs). The quantitative features of classified variables in the data are transformed into digital features by one-hot coding, including crew information, airline information, aircraft information, special situations and some weather information. Figure 4 is the structure of TF, and the data are sorted and fed into the LSTM layer.

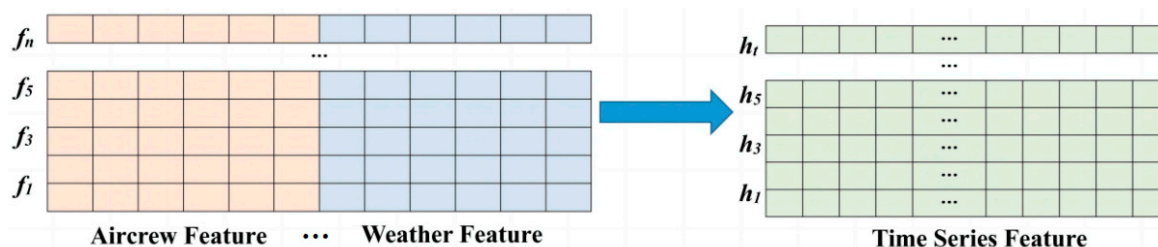


Figure 4. Temporal data construction.

Weather parameters are incorporated into the time series as binary (0–1) variables to quantify their impact on airport departure delays. Based on the relevant literature [48,49], the selection criteria for delay-inducing weather conditions are defined as follows:

If a severe thunderstorm is reported within a 50-m radius of the airport, the indicator variable is assigned a value of 1; otherwise, it is set to 0.

If moderate to severe thunderstorms occur along the flight route, the indicator variable is assigned a value of 1; otherwise, it is set to 0.

If the airport experiences heavy snowfall (24 h accumulation between 5.0 and 10 mm) or a blizzard (24 h accumulation exceeding 10 mm), the indicator variable is assigned a value of 1; otherwise, it is set to 0.

If the airport experiences intense precipitation (hourly rainfall exceeding 16 mm, cumulative precipitation exceeding 30 mm over 12 h, or exceeding 50 mm over 24 h), the indicator variable is assigned a value of 1; otherwise, it is set to 0.

If strong winds occur at the airport (wind force exceeding Level 4), the indicator variable is assigned a value of 1; otherwise, it is set to 0.

If fog or haze is present at the airport (relative humidity exceeding 80%), the indicator variable is assigned a value of 1; otherwise, it is set to 0.

If the cloud ceiling at the airport falls below the minimum decision height for the instrument landing system (10 meters), the indicator variable is assigned a value of 1; otherwise, it is set to 0.

If a sandstorm occurs at the airport (visibility less than 1 km), the indicator variable is assigned a value of 1; otherwise, it is set to 0.

3.2.3. Spatial-Temporal Data

The five stages of aircraft operation are firstly converted into a transportation grid. The dynamic aircraft movement states are added to the running network as the inputs of 3D-CNN. To illustrate, the operation phases on the left side of Figure 5 are transformed into a spatial-temporal image of the traffic network, where the horizontal axis represents the slip time interval and the vertical axis represents the spatial positions of the five phases. Each operating line represents the operating status of each aircraft. The right side of Figure 5 is the input layer of the constructed 3D-CNN with the stacked grid. The parameter l_1 is the length of the input quantity, representing the operating stage of the aircraft. l_2 (width of input volume) describes the number of the aircraft trajectories involved in operation in each grid, and l_3 (height of input quantity) is the number of aircraft states in each grid. The depth of the input cube is set to 1, which means the output is flight delay. The square motion trajectory of the aircraft is superimposed into a cube and applied to the input quantity to generate the output.

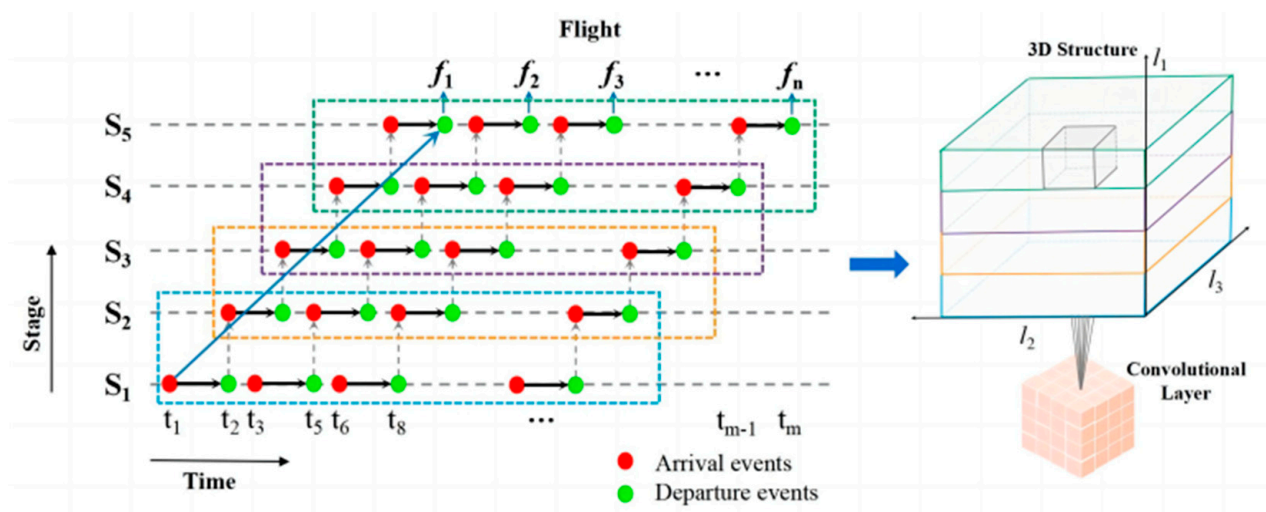


Figure 5. Dynamic spatial-temporal data construction.

3.3. Feature Extraction Layer

3.3.1. Spatial Features Extracted from GCN

The CI of road sections is spatially affected by the CI of adjacent road sections. In a fixed time interval, the CI is relatively stable, but from the perspective of overall flight operation range, the CI is affected by the congestion situation at the historical moments. Therefore, the feature extraction of the CI in the target road sections should consider its time and space characteristics. The traffic network based on the airport distribution structure can well reflect the congested road data source. Based on the CI of the window-time traffic topology network, the GCN neural network is used to extract the spatial features of the

actual congestion of the target road section in each window time and the actual congestion state of the target road section at the output window time is mapped accordingly. Given an adjacency matrix and a characteristic matrix, the GCN model creates a filter in the Fourier domain. The filter acts on the nodes in the graph, extracts the space features among nodes through its first-order neighborhood, and then builds a GCN model by superimposing multi-convolutional layers. It is described as follows [21]:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} \theta^{(l)}\right) \quad (2)$$

Among them, $\tilde{A} = A + I_M$ is the matrix adding self-connection, I_M is the $M \times M$ identity matrix, \tilde{D} is the degree matrix, $D_{ii} = \sum_j A_{ij}$, and $H^{(l)}$ is the output of the layer l , $H^{(0)}$ is X_t . $\theta^{(l)}$ includes the parameters of the layer, and $\sigma(\cdot)$ represents the sigmoid function of the nonlinear model.

3.3.2. Temporal-Spatial Features Extracted from 3D-CNN

A 3D convolution is carried out in the convolutional stage of CNN, and features are calculated from spatial and spatial-temporal dimensions. The 3D convolution is realized by convolving 3D kernels into a cube, which is formed by stacking a plurality of consecutive frames together. With this configuration, the feature map in the convolutional layer is connected to a plurality of adjacent frames in the upper layer so that the airport aircraft slip information is obtained. Formally, the value of (x, y, z) on the feature map j of the layer i is [47].

$$v_{ij}^{xyz} = \tanh\left(b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} \omega_{ijm}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right), \quad (3)$$

where \tanh is hyperbolic tangent function, b_{ij} is a deviation of the feature map, m indicates the feature map set in the layer $(i-1)$ connected with the current feature map, and ω_{ijm}^{pqr} is the value of (p, q, r) connected to the kernel of the m feature map in the previous layer. They are the length, width and height of the nucleus, respectively. CNN parameters, such as deviation b_{ij} and kernel weight ω_{ijm}^{pqr} , are usually learned using supervised or unsupervised methods [50–52].

3.3.3. Temporal Features Extracted from LSTM

This work extracts time features by using the LSTM algorithm, such as delay time, flight features, meteorological features and other time features [53]. For various modified versions of LSTM, it has been proved that the prediction accuracy is usually equivalent on condition that the LSTM includes forgetting gate and activation function [54]. In the literature, the most commonly used LSTM framework is the modified LSTM network, which achieves information retention and forgetting by setting more complex frameworks, such as input gate, forget gate, unit gate and output gate [55]. The structure of LSTM mainly includes the input layer, stack hidden layer and output layer. LSTM mainly uses a gate mechanism for a series of operations. The three mechanism gates are the input gate i , forgetting gate f and output gate o . The two basic units are cell state \tilde{C} and hidden unit h [55]. The input of neurons is the unit and meteorological information x_t of time t , the memory information of time $t-1$ is called hidden state h_{t-1} and cell state C_{t-1} , and the output is hidden state h_t and cell state C_t of time t . The calculation equations are the followings:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$c_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (6)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot c_t \quad (7)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

$$h_t = o_t \odot \tanh(C_t) \quad (9)$$

Equation (4) controls the degree of retention of historical information, which is completed by the Sigmoid function of the forgetting gate. Among them, W_f and b_f are the weight matrix and paranoia of f_t . Equation (5) controls the retention of the current input information, which is determined by the Sigmoid function of the input gate. Among them, W_i and b_i are the weight matrix and paranoia of i_t . Equation (6) is to establish a new candidate vector c_t using the Tanh function. W_c is Tanh's weight matrix, and b_c is the bias. The weights of C_{t-1} and c_t are f_t and i_t , respectively, and the two states are merged to update the current state, which is expressed as Equations (7)–(9) output the latest information, which is completed by the Sigmoid function of the output gate.

3.3.4. Fusion Technique

To integrate the input data from the three feature extraction layers, a fusion technique was implemented, wherein the output vectors of the GCN, 3D CNN and LSTM layers were concatenated side by side. If the output dimension of GCN is (N, G), the 3D CNN is (N, D), and the LSTM layer is (N, L), then the merged vector will be (N, G+D+L). N is the amount of training or testing data. The merged vector is then used as input and used as another LSTM layer to complete the prediction.

3.4. Loss Function and Evaluation Methods

The L2 regularization method is used to prevent over-fitting caused by insufficient samples. The loss function is calculated as follows:

$$Loss = \frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2 + \lambda \sum_k w_k^2, \quad (10)$$

where y_i and y'_i are the actual and the predicted value of the sample i , respectively; λ is the ratio of regularization in total loss; and w is trainable weight for regression models. n is the total prediction number. The MAE, RMSE and MAPE are calculated to evaluate the performance, as shown in Equations (11)–(13).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y'_i - y_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y'_i - y_i)^2} \quad (12)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y'_i - y_i}{y_i} \right| \quad (13)$$

4. Case Analysis Results

4.1. Data Collection

In this study, the flight data were collected from Beijing Capital International Airport (PEK) during 1 December 2017 to 30 November 2018, which can be classified as flight datasets and meteorological datasets. These data consist of a set of 305,643 flights. Each flight data point is recorded as: inbound and outbound date (UTC time), flight ID, aircraft type, apron area, runway information, historical time of aircraft taxiing record point, inbound and outbound point, inbound and outbound direction, and delay time. Airport Meteorological Information collected meteorological conditions in 2018, including airport

code, UTC time, average hourly temperature, wind direction, wind speed, precipitation, thunderstorms, fog, snowfall, smog and other weather phenomena. The variables of these weather phenomena in data processioning are set as dummy variables.

Figure 6 shows PEK airport delay information, which recorded the airport delay time at different timeslots from December 2017 to November 2018. For example, catastrophic delay period in the red area and typical strong delay propagation stage. The graph also describes the randomness and volatility of delay values. In the vertical time scale, the peak period of airport delays is concentrated from 10:00 to 17:00, and the delay peaked in June, July and August. The flight delays present the phenomenon of short-term propagation fluctuation and multi-frequency triggers. In the horizontal time scale, flight delays occur regularly, and the data points in red with high values of delays are presented in the figure with horizontal broken bars, which shows that airports often repeat delays every day in the same time intervals.

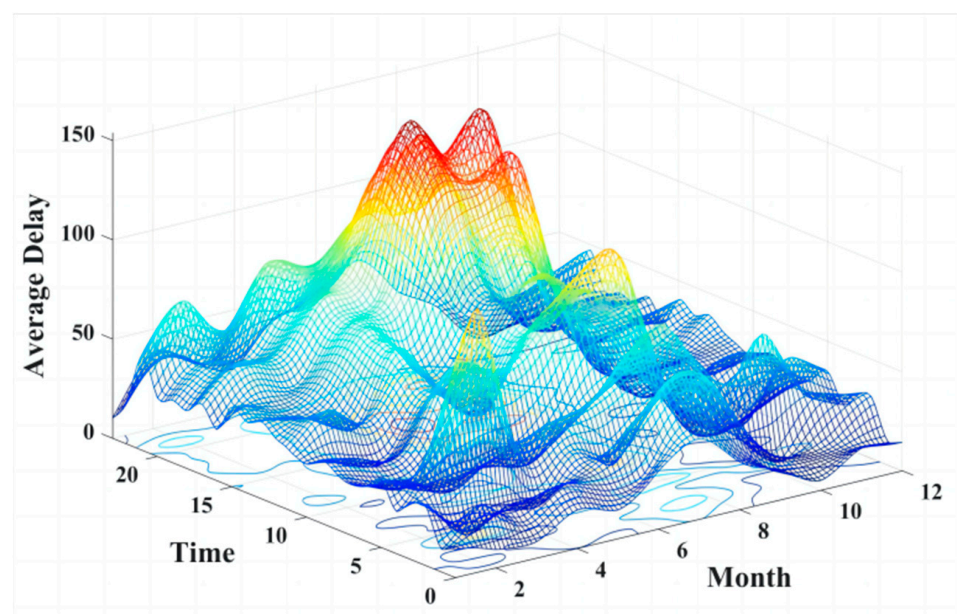


Figure 6. Total delay time at PEK Airport.

Figure 7 presents the weather conditions at the PEK airport. Gale is defined as an average wind greater than or equal to 10 m/s; extreme gale is a gale with an average wind greater than or equal to 17 m/s. The strong winds in PEK mainly occur in January–February, April–May, September–October and December, and the extreme winds mainly occur in April–May in spring or December–February in winter. The frequency is the highest from afternoon to evening, the lowest from night to morning, and the peak occurs at 14:00–15:00. The extreme gale of Capital Airport is mainly caused by the northwest gale after the cold front with a wind direction of up to 330 degrees. The extreme gale at Capital Airport can be divided into three categories: spring gale, summer gale and winter gale. Among them, in spring, the surface water content is relatively small, the cold air passes quickly, and the strong wind is often accompanied by dusty weather. Summer gale is a strong convective weather associated with thunderstorms. In winter, the temperature is low, and the surface water content is large. The strong winds are often accompanied by rain, snow and freezing weather.

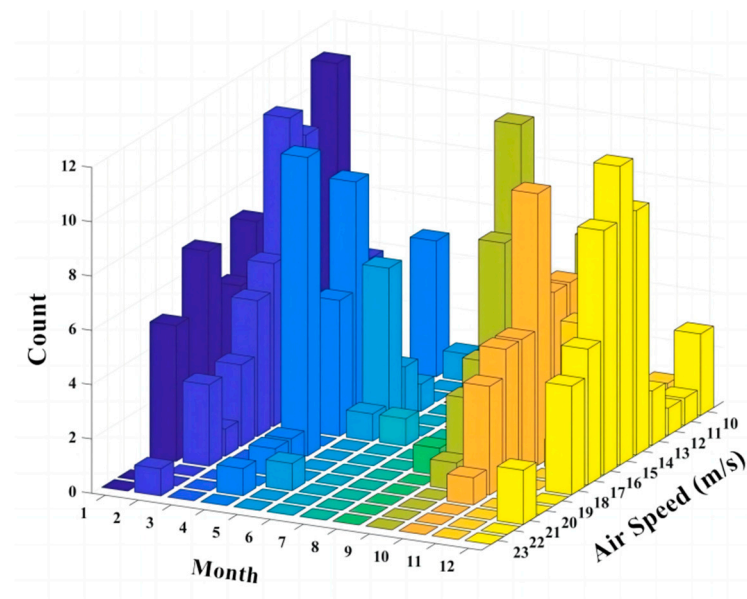


Figure 7. Airport weather conditions (statistics of times when the wind speed is greater than 10 m/s per month).

Figure 8 shows the time distribution of aircraft operation during various stages of airport flight delays. When there is a delay, the aircraft operates for a longer period of time in stages 1 and 5, which are the closing, the hatch stage and the departure stage. The average running time of these two stages is 10 min (with an operation time concentrated ranging between 3 and 13 min) and 16 min (with an operation time ranging between 13 and 25 min), which easily leads to a large delay. The remaining three stages, including the removing chocks stage, holding-point stage, and exiting clearance stage, operate for a relatively short period of time, averaging 3–7 min.

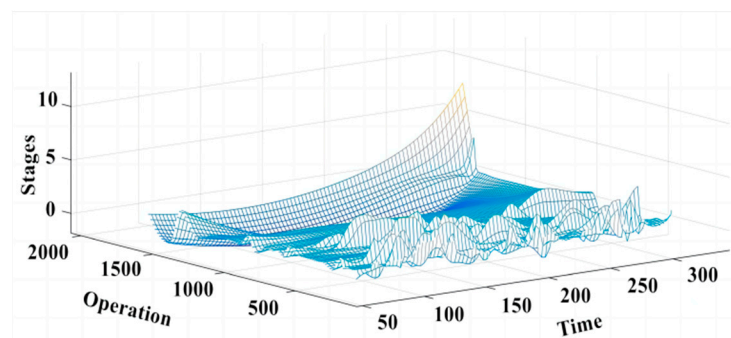


Figure 8. Aircraft operation time during various stages.

Based on the calculation formula of CI, the congestion index of each road section (edge) is obtained, as shown in Figure 9. The higher the CI value, the higher the aircraft density in the corresponding road section. The highest CI value is 11.74 min/km. The higher CI values are mostly concentrated in the section from airport apron P1 to P6 to runway 18L/R (moving routes ≤ 50). The taxiing distance of the relevant section is relatively short, but the density of aircraft on the section is high, often exhibiting peak aggregation characteristics throughout the day. The CI of the remaining sections (moving routes > 50) is lower, which is characterized by longer taxiing distances or lower aircraft density on the sections. The CI of the entire road section is concentrated between 1–3 min/km.

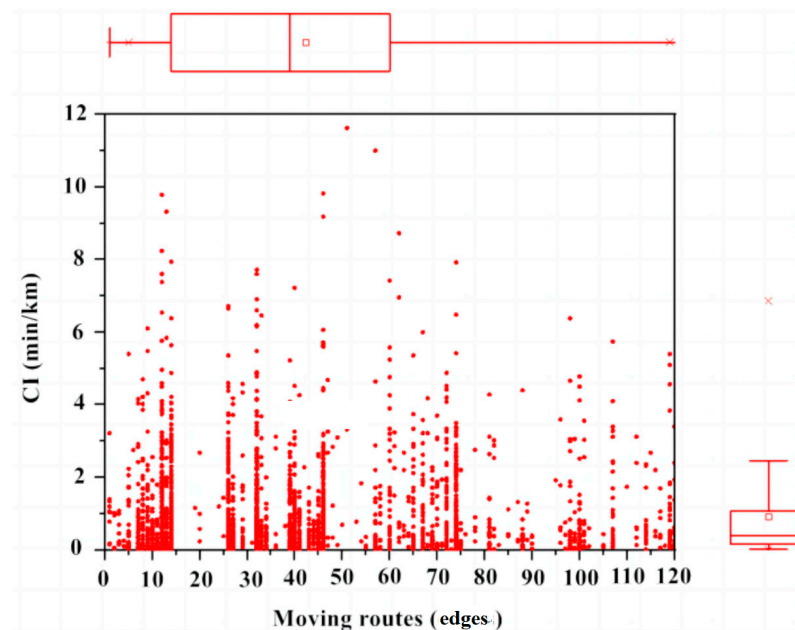


Figure 9. Airport CI index.

4.2. Model Training and Testing

The convergence of the DL model largely depends on the selection of the optimizer. A superior optimizer can result in faster convergence and lower errors [48]. The results of the model using the Nadam [49] optimizer with five widely used optimizers, namely Adam [49], Adamax [56], SGD [57], Adagrad [58] and Adadelata [58], were compared in this study. Firstly, batch_size was set to 32. MSE as used as the loss function. The Loss function using l2 regularization is shown in Equation (10). Epochs was set to 200. TEST sums five different optimizers under different learning raftes (LR). Figure 10 shows the lossy convergence rates using the SGD, Adagrad and Adadelata optimizers are rather slow compared with the other three optimizers. For Adam, Adamax and Nadam, their convergence speeds and training performance are similar. By comparing Nadam and Adam, TEST's performance at different learning rates was determined to decide on an appropriate value. The figure shows that the performance of these two optimizers is similar under different learning rates. Thus, the Nadam optimizer was used in the following experiments.

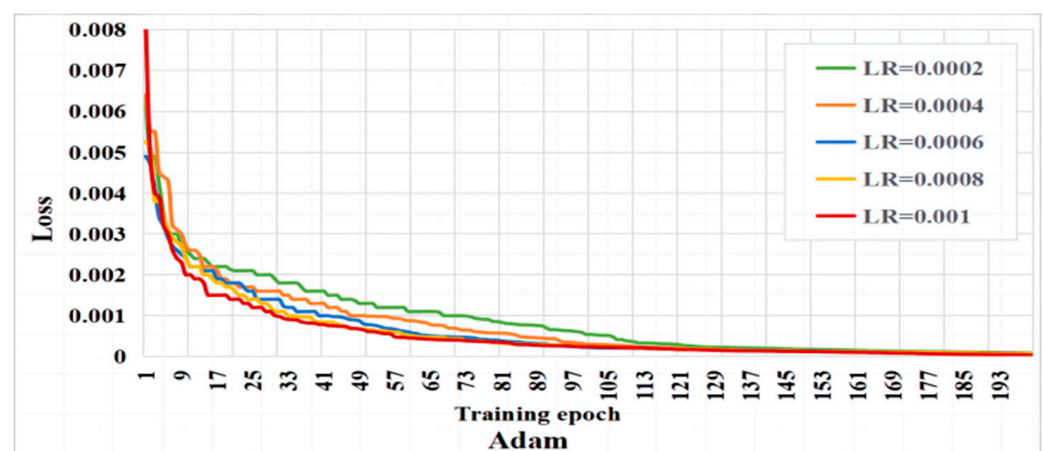


Figure 10. Cont.

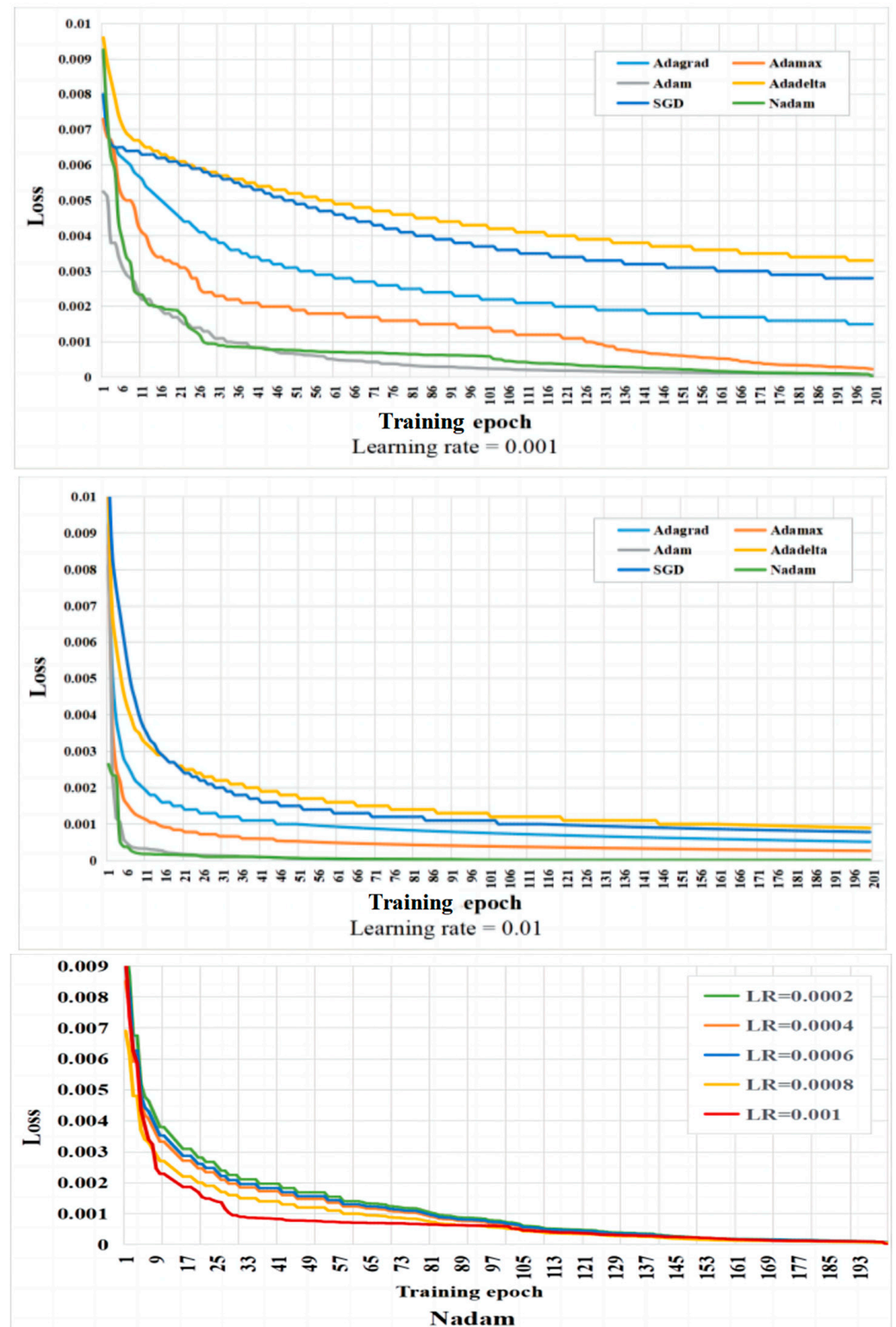


Figure 10. Comparison of results by different optimizers.

4.3. Flight Delay Prediction Based on the 3DF-DSCL Method

In this case, epochs = 200 was specified, and the batch_size value was specified according to the change of the loss function and the situation in the training process, generally ensuring that it was less than 20% of the number of samples. In order to verify the performance of the proposed 3DF-DSCL model, a set of commonly used methods were also applied to test cases as a benchmark. Based on the literature and our data features,

using selected benchmark models can effectively compare the performance. All benchmark methods use the same input variables for training and test to ensure the comparability of models. Table 2 summarizes MAE, MAPE and RMSE for each method. According to the analysis, the proposed 3DF-DSCL method achieves the most stable and satisfactory performance in terms of MAE, MAPE and RMSE among all the methods of GCN+LSTM, 2D CNN+LSTM, 3D CNN+LSTM, LSTM, ARIMA, GCN and CNN. The performance of 2D CNN+GCN+LSTM, GCN+LSTM and 3D CNN+LSTM is far superior to the other methods because they integrate LSTM and GCN (CNN) structures and consider spatial and temporal features at the same time. In addition, 3DF-DSCL has more benefits than 2D CNN+GCN+LSTM because it is more accurate in extracting dynamic data.

Table 2. The MAE, MAPE and RMSE of different methods.

Model	MAE (min)	RMSE (min)	MAPE (*)
LSTM	0.430	0.728	0.297
ARIMA	0.681	0.927	0.265
CNN	0.581	0.860	0.584
GCN	0.584	0.847	0.568
GCN+LSTM	0.403	0.728	0.296
2D CNN+LSTM	0.418	0.598	0.332
3D CNN+LSTM	0.341	0.551	0.322
2D CNN+GCN+LSTM	0.304	0.454	0.163
3DF-DSCL	0.260	0.363	0.053

*: Between 0 and 1, it is represented as a decimal.

Finally, the robustness of the proposed model is verified from the perspective of data scales and data dimensions by using different sizes and dimensions of datasets in flight delay prediction. The size of datasets is distinguished by monthly, quarterly, semi-yearly, and annual historical data, respectively. Figure 11 shows the prediction results between 3DF-DSCL and 2D CNN+GCN+LSTM models under different data scales and data dimensions. In the specified prediction scenarios, all the MAE values of the proposed method are less than 0.6, which proves the model's robustness to different data scales and dimensions. Moreover, it can be observed that the prediction accuracy of large-scale datasets is relatively higher than that of small-size datasets, which indicates that the proposed model is more applicable in the scenarios of large-scale datasets. Overall, the proposed 3DF-DSCL has the least MAPE, RMSE and MAE compared with the 2D CNN+GCN+LSTM model in all test cases.

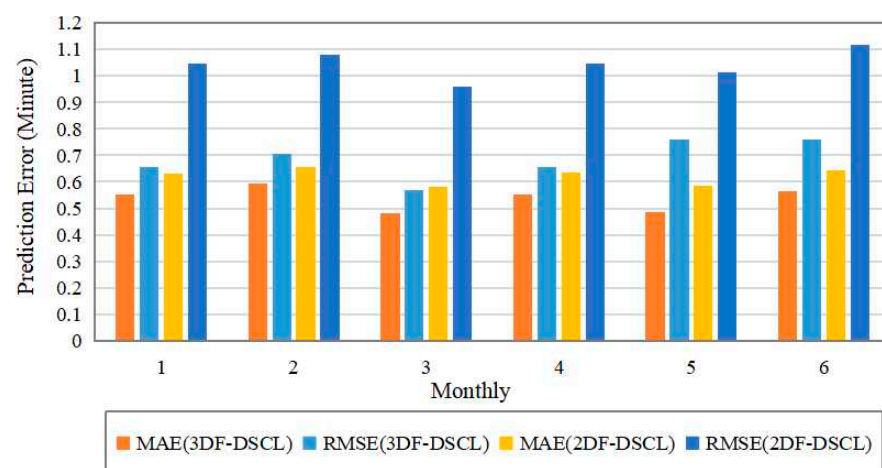


Figure 11. Cont.

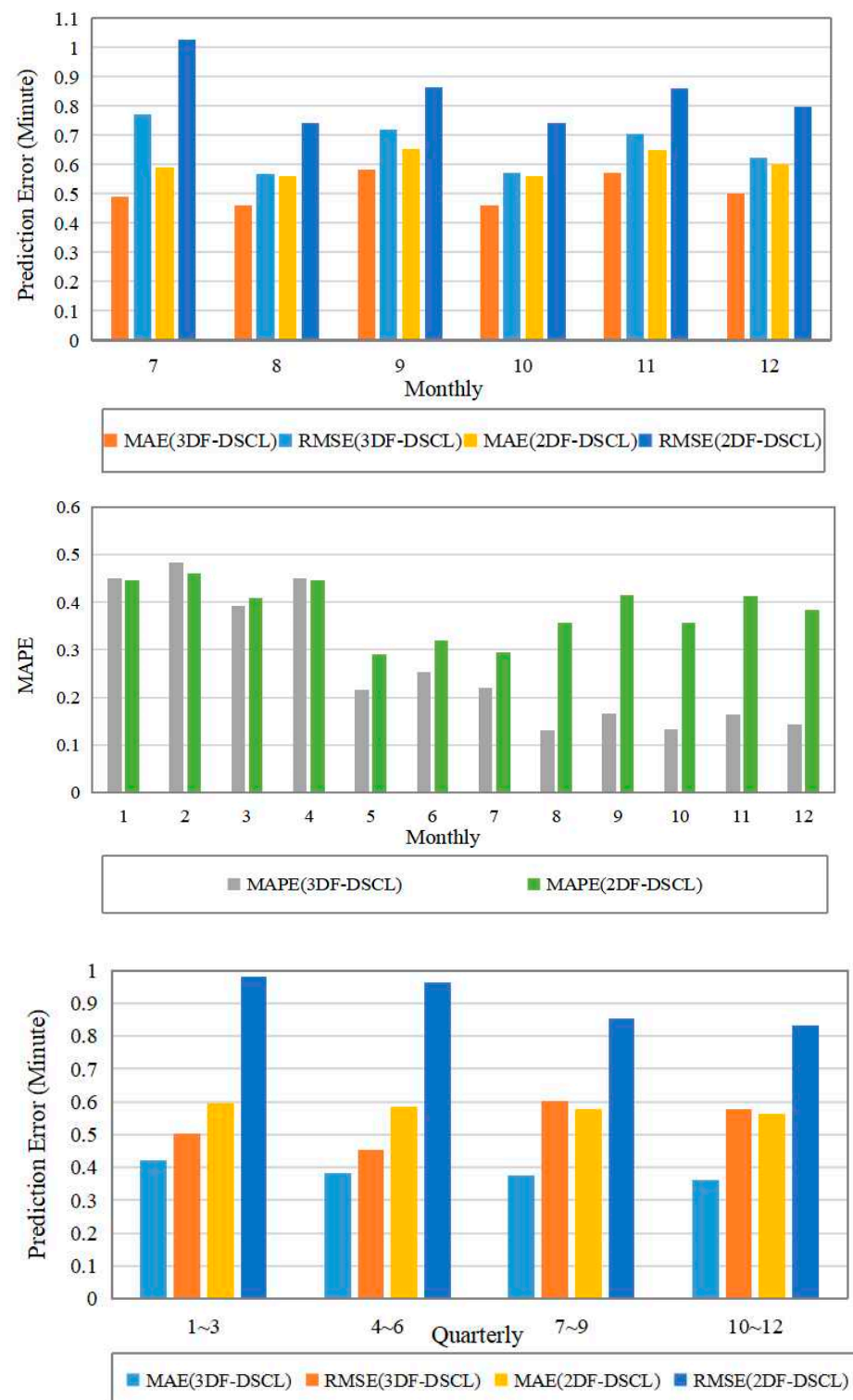


Figure 11. Cont.

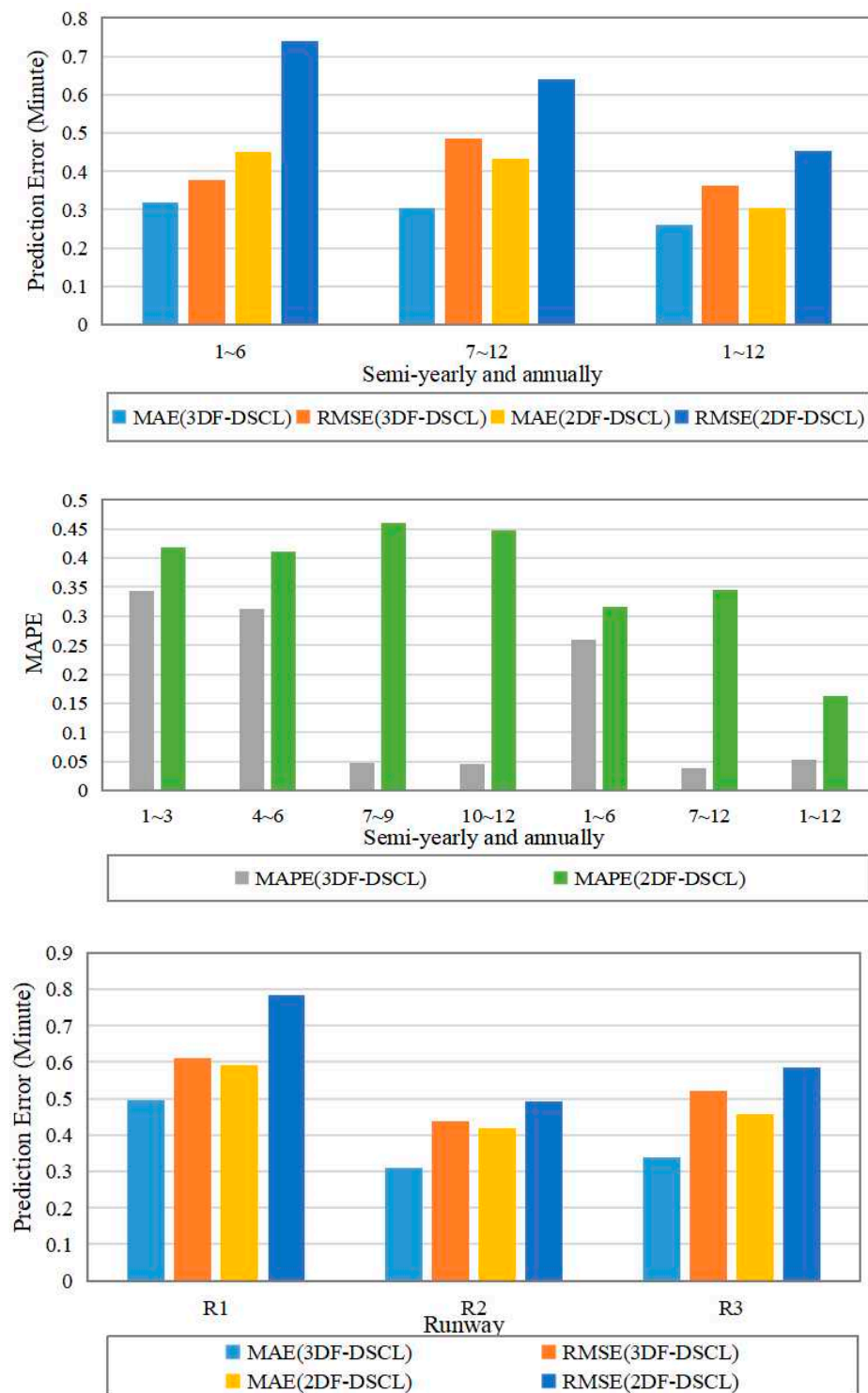


Figure 11. Cont.

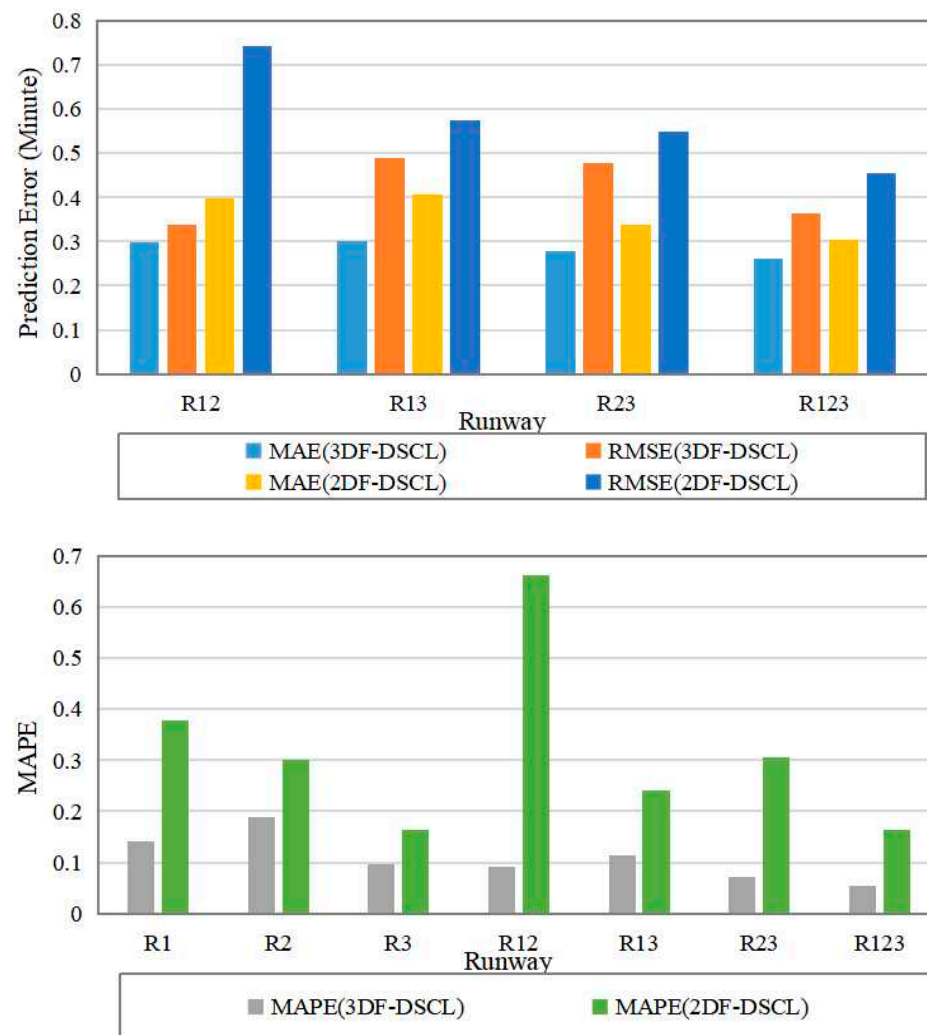


Figure 11. The robustness of the 3DF-DSCL model and 2D CNN+GCN+LSTM model in different data sizes and airport zones.

5. Conclusions

In this study, a deep learning network was developed to model the non-temporal, temporal and extemporization features of complex airport systems. Specifically, in order to solve the complex multi-attribute datasets caused by moving objects, the proposed 3DF-DSCL model merges three different model architectures with self-dominance neural network features, named 3D CNN, GCN and LSTM. The 3D CNN layer is the input of the stacked grid/volume, which aims to capture the continuous dynamic spatial-temporal dependence. The LSTM layer, which processes temporal data as a sequence, is developed to extract time dependence. The GCN, which imports topology network features, is used to identify the influence factors of spatial-temporal correlation. The proposed model utilizes three-dimensional convolution to extract features from spatial, temporal, and spatiotemporal dimensions, enabling the capture of motion trajectory information embedded across multiple adjacent frames. The model generates multiple channel representations, which are subsequently integrated into a unified feature representation. To further enhance its performance, feature regularization is applied, and the model's recognition capabilities are strengthened by incorporating multiple complementary models. The proposed approach is deployed in a real-world airport environment to identify aircraft operational behaviors. Its effectiveness is then systematically evaluated through a comparative analysis against baseline methods to assess its advantages. In the case analysis, the model can effectively

and accurately predict airport delays. The MAE reaches 0.26, which is at least a 14.47% reduction as compared to 2D CNN+GCN+LSTM. Future research still needs to determine the impact of multivariable delay. For example, the airport dynamic weather map data obtained by weather radar, wind profile radar and meteorological satellites are used as data input for delay prediction. Secondly, this paper takes single airport data as a case study. The multi-airport prediction model can also be used in further study of the model's robustness and applicability.

Author Contributions: Conceptualization, Y.Y., Y.W. and C.S.L.; Methodology, Y.Y. and C.S.L.; Resources, Y.W.; Writing—original draft, Y.Y.; Writing—review & editing, Y.W. and C.S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Yujie Yuan grant number 3122024QD18 and by basic scientific research of central universities.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lall, A. Delays in the New York City metroplex. *Transp. Res. Part A Policy Pract.* **2017**, *114*, 10–21. [\[CrossRef\]](#)
2. Anguita, J.G.M.; Díaz Olariaga, O. Prediction of Departure Flight Delays through the Use of Predictive Tools Based on Machine Learning/Deep Learning Algorithms. *Aeronaut. J.* **2024**, *128*, 23.
3. Tu, Y.; Ball, M.; Jank, W. Estimating flight departure delay distributions, a statistical approach with long-term trend and short-term pattern. *J. Am. Stat. Assoc.* **2008**, *103*, 112–125. [\[CrossRef\]](#)
4. Fu, T.-C. A review on time series data mining. *Eng. Appl. Artif. Intell.* **2011**, *24*, 164–181. [\[CrossRef\]](#)
5. Belcastro, L.; Marozzo, F.; Talia, D.; Trunfio, P. Using scalable data mining for predicting flight delays. *ACM Trans. Intell. Syst. Technol.* **2016**, *8*, 1–22. [\[CrossRef\]](#)
6. Hao, L.; Hansen, M.; Zhang, Y.; Post, J. New York, New York: Two ways of estimating the delay impact of New York airports. *Transp. Res. Part E Logist. Transp. Rev.* **2014**, *70*, 245–260. [\[CrossRef\]](#)
7. Khan, W.; Ma, H.-L.; Chung, S.-H.; Wen, X. Hierarchical integrated machine learning model for predicting flight departure delays and duration in series. *Transp. Res. Part C Emerg. Technol.* **2021**, *129*, 103225. [\[CrossRef\]](#)
8. Liu, F.; Sun, J.; Liu, M.; Yang, J.; Gui, G. Generalized flight delay prediction method using gradient boosting decision tree. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020. [\[CrossRef\]](#)
9. Chakrabarty, N.; Kundu, T.; Dandapat, S.; Sarkar, A.; Kole, D.K. Flight arrival delay prediction using gradient boosting classifier. In *Emerging Technologies in Data Mining and Information Security*; Springer: Singapore, 2018; pp. 651–659. [\[CrossRef\]](#)
10. Moreira, L.; Dantas, C.; Oliveira, L.; Soares, J.; Ogasawara, E. On evaluating data preprocessing methods for machine learning models for flight delays. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8–13 July 2018. [\[CrossRef\]](#)
11. 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 Proceedings of the IEEE International Conference on Computational Intelligence and Data Science (ICCIDS), Chennai, India, 2–3 June 2017; pp. 1–5.
12. 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 Emerg. Technol.* **2021**, *130*, 103323. [\[CrossRef\]](#)
13. Huang, P.; Wen, C.; Fu, L.; Peng, Q.; Tang, Y. A deep learning approach for multi-attribute data: A study of train delay prediction in railway systems. *Inf. Sci.* **2019**, *216*, 234–253. [\[CrossRef\]](#)
14. 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.* **2019**, *69*, 140–150. [\[CrossRef\]](#)
15. Li, Z.; Chen, H.; Ge, J.; Ning, K. An airport scene delay prediction method based on LSTM: 14th International Conference on Advanced Data Mining and Applications, ADMA 2018, Nanjing, China, 16–18 November 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 174–184. [\[CrossRef\]](#)
16. Kim, Y.J.; Choi, S.; Briceno, S.; Mavris, D. A deep learning approach to flight delay prediction. In Proceedings of the 2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, CA, USA, 25–29 September 2016. [\[CrossRef\]](#)

17. Silvestre, J.; Martinez-Prieto, M.A.; Bregon, A.; Alvarez-Esteban, P.C. A Deep Learning-Based Approach for Predicting In-Flight Estimated Time of Arrival. *J. Supercomput.* **2024**, *80*, 80. [\[CrossRef\]](#)
18. Zhang, Z.; Li, Y.; Dong, H. Multiple-feature-based vehicle supply–demand difference prediction method for social transportation. *IEEE Trans. Comput. Soc. Syst.* **2020**, *7*, 1095–1103. [\[CrossRef\]](#)
19. Sanaei, R.; Pinto, B.; Gollnick, V. Toward ATM resiliency: A deep CNN to predict number of delayed flights and ATFM delay. *Aerospace* **2021**, *8*, 28. [\[CrossRef\]](#)
20. Liu, H.; Lin, Y.; Chen, Z.; Guo, D.; Zhang, J.; Jing, H. Research on the air traffic flow prediction using a deep learning approach. *IEEE Access* **2019**, *7*, 148019–148030. [\[CrossRef\]](#)
21. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Trans. Intell. Transp. Syst.* **2019**, *22*, 3848–3858. [\[CrossRef\]](#)
22. Alfarhood, M.; Alotaibi, R.; Abdulrahim, B.; Einieh, A.; Almousa, M.; Alkhanifer, A. Predicting Flight Delays with Machine Learning: A Case Study from Saudi Arabian Airlines. *Int. J. Aerosp. Eng.* **2024**, *2024*, 12. [\[CrossRef\]](#)
23. Carvalho, L.; Sternberg, A.; Maia Goncalves, L.; Beatriz Cruz, A.; Soares, J.A.; Brandão, D.; Ogasawara, E. On the relevance of data science for flight delay research: A systematic review. *Transp. Rev.* **2021**, *41*, 499–528. [\[CrossRef\]](#)
24. Yu, B.; Guo, Z.; Asian, S.; Wang, H.; Chen, G. Flight delay prediction for commercial air transport: A deep learning approach. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *125*, 203–221. [\[CrossRef\]](#)
25. Sousa, M.; Carvalho, A.M. Polynomial-time algorithm for learning optimal BFS-consistent dynamic Bayesian networks. *Entropy* **2018**, *20*, 274. [\[CrossRef\]](#)
26. Güvercin, M.; Ferhatosmanoglu, N.; Gedik, B. Forecasting flight delays using clustered models based on airport networks. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 3179–3189. [\[CrossRef\]](#)
27. Zonglei, L.; Jiandong, W.; Guansheng, Z. A new method to alarm large scale of flight delays based on machine learning. In Proceedings of the 2008 International Symposium on Knowledge Acquisition and Modeling, Wuhan, China, 21–22 December 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 589–592.
28. Wu, Z.J.; Tian, S.; Ma, L. A 4D trajectory prediction model based on the BP neural network. *J. Intell. Syst.* **2020**, *29*, 1545–1557. [\[CrossRef\]](#)
29. Zhang, K.; Chen, B. Phased Flight Trajectory Prediction with Deep Learning. *arXiv* **2022**, arXiv:2203.09033.
30. Qu, J.; Zhao, T.; Ye, M.; Li, J.; Liu, C. Flight delay prediction using deep convolutional neural network based on fusion of meteorological data. *Neural Process. Lett.* **2020**, *52*, 1461–1484. [\[CrossRef\]](#)
31. Cai, K.; Li, Y.; Fang, Y.P.; Zhu, Y. A deep learning approach for flight delay prediction through time-evolving graphs. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 11397–11407. [\[CrossRef\]](#)
32. Li, G.; Xiong, C.; Thabet, A.; Ghanem, B. DeeperGCN: All you need to train deeper GCNs. *arXiv* **2020**, arXiv:2006.07739.
33. Sun, C.; Li, S.; Cao, D.; Wang, F.-Y.; Khajepour, A. Tabular learning-based traffic event prediction for intelligent social transportation systems. *IEEE Trans. Comput. Soc. Syst.* **2023**, *10*, 1199–1210. [\[CrossRef\]](#)
34. Kong, X.; Huang, Z.; Shen, G.; Lin, H.; Lv, M. Urban overtourism detection based on graph temporal convolutional networks. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 442–454. [\[CrossRef\]](#)
35. Ma, L.; Tian, S. A hybrid CNN-LSTM model for aircraft 4D trajectory prediction. *IEEE Access* **2020**, *8*, 134668–134680. [\[CrossRef\]](#)
36. Khan, S.; Alarabi, L.; Basalamah, S. Toward smart lockdown: A novel approach for COVID-19 hotspots prediction using a deep hybrid neural network. *Computers* **2020**, *9*, 99. [\[CrossRef\]](#)
37. 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.* **2019**, *37*, 6029–6037. [\[CrossRef\]](#)
38. Yang, Z.; Tang, R.; Zeng, W.; Lu, J.; Zhang, Z. Short-term prediction of airway congestion index using machine learning methods. *Transp. Res. Part C Emerg. Technol.* **2021**, *125*, 103040. [\[CrossRef\]](#)
39. Chen, M.; Zeng, W.; Xu, Z.; Li, J. Delay prediction based on deep stacked autoencoder networks. In Proceedings of the Asia-Pacific Conference on Intelligent Medical 2018 & International Conference on Transportation and Traffic Engineering 2018, Beijing, China, 21–23 December 2018; pp. 238–242.
40. Yi, J.; Zhang, H.; Liu, H.; Zhong, G.; Li, G. Flight delay classification prediction based on stacking algorithm. *J. Adv. Transp.* **2021**, *2021*, 4292778. [\[CrossRef\]](#)
41. Roh, H.-J. Development and performance assessment of winter climate hazard models on traffic volume with four model structure types. *Nat. Hazards Rev.* **2020**, *21*, 04020023. [\[CrossRef\]](#)
42. Sridhar, B.; Chen, N. Short-term national airspace system delay prediction using weather-impacted traffic index. *J. Guid. Control Dyn.* **2009**, *32*, 657–662. [\[CrossRef\]](#)
43. Adacher, L.; Flamini, M.; Romano, E. Rerouting algorithms solving the air traffic congestion. In Proceedings of the Applied Mathematics and Computer Science: Proceedings of the 1st International Conference on Applied Mathematics and Computer Science, Rome, Italy, 27–29 January 2017; p. 020053. [\[CrossRef\]](#)
44. Rani, S.; Sikka, G. Recent techniques of clustering of time series data: A survey. *Int. J. Comput. Appl.* **2012**, *52*, 1–9. [\[CrossRef\]](#)

45. Dong, Y.; Lu, Z.; Liu, Y.; Zhang, Q.; Wu, D. China's corridors-in-the-sky design and space-time congestion identification and the influence of air routes' traffic flow. *J. Geogr. Sci.* **2019**, *29*, 1999–2014. [\[CrossRef\]](#)
46. Enayatollahi, F.; Atashgah, M.A.A. Wind effect analysis on air traffic congestion in terminal area via cellular automata. *Aviation* **2018**, *22*, 102–114. [\[CrossRef\]](#)
47. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231. [\[CrossRef\]](#)
48. Simaiakis, I. Analysis, Modeling and Control of the Airport Departure Process. Doctor of Philosophy in Aeronautics and Astronautics, Massachusetts Institute of Technology. 2013. Available online: <http://dspace.mit.edu/handle/1721.1/79342> (accessed on 16 June 2016).
49. Zhang, Y. Research on Low Visibility Forecasting and its Correlation with Flight Punctuality. Master's Thesis, Civil Aviation Flight Academy of China, Nanchang, China, 2018.
50. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [\[CrossRef\]](#)
51. Ranzato, M.A.; Huang, F.; Boureau, Y.-L.; LeCun, Y. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In Proceedings of the CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–8. [\[CrossRef\]](#)
52. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
53. Greff, K.; Srivastava, R.K.; Koutnik, J.; Steunebrink, B.R.; Schmidhuber, J. LSTM: A search space odyssey. *IEEE Trans. Neural Networks Learn. Syst.* **2017**, *28*, 2222–2232. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Graves, A. Generating sequences with recurrent neural networks. *arXiv* **2013**, arXiv:1308.0850.
56. Bottou, L. Large-Scale Machine Learning with Stochastic Gradient Descent. *Phys.-Verl. HD* **2010**, *10*, 177–186.
57. Duchi, J.; Hazan, E.; Singer, Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *J. Mach. Learn. Res.* **2011**, *12*, 2121–2159.
58. Tao, H.; Lu, X. On Comparing Six Optimization Algorithms for Network-Based Wind Speed Forecasting. In Proceedings of the 37th Chinese Control Conference (CCC), Wuhan, China, 25–27 July 2018. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.