aerospace

MDPI

Article

DDA-SIM-ATT: A Synergistic Multi-Module Fusion Model for High-Precision Prediction of Departure Flight Taxi-Out Time

Yue Lu [1], Yanzhi Li [2,*], Qingwei Zhong [2], Jifei Zhong [2], Yingxue Yu [3] and Yuxin Zhang [4]

1 Ordos Low-Altitude Economy Development Co., Ltd., Ordos 017200, China; dkgs0701@163.com
2 College of Air Traffic Management, Civil Aviation Flight University of China, Chengdu 641400, China; qingweizhong@cafuc.edu.cn (Q.Z.); zhongjifei@cafuc.edu.cn (J.Z.)
3 College of Electromechanical Engineering, Guangzhou City Construction College, Guangzhou 510925, China; yyx_cafuc@163.com
4 China Academy of Civil Aviation Science and Technology, Beijing 100083, China; zhangyuxin@mail.castc.org.cn
* Correspondence: lyzlyl@cafuc.edu.cn

Abstract

Accurate prediction of departure flight taxi-out time is critical for enhancing airport surface efficiency and reducing flight delays. However, existing methods often struggle with data sparsity, inadequate representation of complex spatio-temporal interactions among aircraft, and imbalanced sample distributions. To address these challenges, this paper proposes a synergistic multi-module fusion model named DDA-SIM-ATT-CatBoost. The model integrates three core modules: a Dynamic Data Augmentation (DDA) module that expands the training distribution through operationally consistent perturbations to mitigate data imbalance; a Similarity Theory (SIM) module employing K-Prototypes clustering and Mahalanobis distance to achieve precise matching of historical operational patterns; and an Attention Mechanism (ATT) module that dynamically recalibrates feature weights to emphasize critical influencing factors. These modules work synergistically to provide a robust and discriminative input representation for the CatBoost regressor, which excels at handling categorical features and complex nonlinearities. Using real-world departure data from a major hub airport, the proposed model achieves prediction accuracies of 74.57%, 89.12%, and 97.76% within error margins of ±120 s, ±180 s, and ±300 s, respectively, with a Mean Absolute Percentage Error (MAPE) of 10.34%, Mean Absolute Error (MAE) of 87.55 s, and Root Mean Square Error (RMSE) of 125.61 s. Ablation studies validate the positive contribution and synergistic effect of each module, while comparative experiments demonstrate that our model significantly outperforms baseline models such as XGBoost and Random Forest. The DDA-SIM-ATT framework provides a generalizable and high-precision solution for taxi-out time prediction, offering reliable decision support for airport surface operations.

Keywords: air transportation; taxi-out time prediction; similarity theory; attention mechanism; dynamic data augmentation

Academic Editor: Miroslav Kelemen

Received: 14 February 2026
Revised: 16 March 2026
Accepted: 23 March 2026
Published: 27 March 2026

1. Introduction

The taxi-out time of a departing flight is defined as the duration from the aircraft's off-block time at the parking stand to its arrival at the runway holding point. As a core segment of ground operations, taxi-out time directly influences flight punctuality and is closely related to airport surface operational efficiency, scheduling coordination, and passenger experience. With continuous growth in global air traffic, major hub airports face increasing surface operation pressures, making the taxiing phase a significant contributor to flight delays. Statistics indicate that ground taxiing accounts for approximately 20–30% of total flight operation time, and its volatility significantly affects optimal airport resource allocation and overall network stability. Therefore, accurate prediction of departure flight taxi-out time is crucial for providing essential data support for gate assignment, push-back sequence optimization, and runway scheduling, while also reducing excessive fuel consumption and carbon emissions.

However, taxi-out time is subject to complex coupling effects from multiple dynamic factors, including aircraft type, traffic flow, taxiway route, weather conditions, air traffic control regulations, and unexpected disruptive events. Traditional methods relying on fixed empirical values or simple statistical models struggle to adapt to highly dynamic and stochastic real-world environments, often resulting in significant prediction bias. Consequently, in-depth analysis of influencing mechanisms and the development of intelligent prediction models capable of integrating multi-source heterogeneous data are imperative for advancing "Smart Airport" initiatives. This study aims to explore key influencing factors and evolutionary patterns of taxi-out time and to establish a high-precision prediction method, providing a theoretical basis and technical reference for enhancing the predictability and controllability of flight ground operations.

Research on taxi-out time prediction for departing flights, as a key technology for improving airport surface operational efficiency, optimizing ground resource scheduling, and reducing flight delays, has evolved alongside traffic growth and advancements in data-driven technologies. Current research primarily falls into two categories: investigations into the influencing factors of taxi-out time and data-driven predictive modeling based on machine learning.

In the study of influencing factors, Gulding [1] analyzed Eurocontrol and FAA data, highlighting the significant impact of simultaneous departure and arrival counts. Idris [2] developed a multiple linear regression model using departure queue length as a key variable, pioneering data-driven prediction. Simaiakis [3] decomposed taxi-out time into unimpeded taxi-out, departure queuing, and congestion delay, accurately reflecting surface congestion status. Clewlow [4] analyzed factors affecting departure taxi time and fuel consumption, noting a strong correlation with arriving flights, particularly under interactive runway configurations. Ravizza [5] incorporated aircraft characteristics, weather, and operational timing into a unified modeling framework. Yin [6] proposed a taxiing situation awareness framework using a multi-dimensional indicator system. Wang [7] systematically compared models and factors, noting significant effects from taxi route, aircraft category, and weather. Xia [8] developed a weather-integrated prediction model using METAR data. Park [9] comprehensively considered weather, flight volume, and runway occupancy time. Tang [10] proposed a micro-feature analysis method based on refined taxiway networks.

In a data-driven predictive modeling study based on machine learning, Balakrishna [11] used non-parametric reinforcement learning, achieving 60% accuracy within ±5 min. George [12] employed linear regression, neuro-fuzzy systems, and Q-learning, with reinforcement learning showing superior accuracy in complex environments. Lee [13] applied various machine learning methods, enhancing accuracy by incorporating terminal concourses and gate information. Lian [14] proposed a Support Vector Regression (SVR) model optimized by swarm intelligence for high-congestion conditions. Herrema [15] applied neural networks, regression trees, and Multilayer Perceptrons (MLP), with regression trees achieving an average error of 1.6 min. Diana [16] compared ensemble methods, Ordinary Least Squares (OLS), and regularized regression, noting the Gradient Boosting model performed best in later NextGen phases. Li [17] proposed a Wide and Deep Neural

https://doi.org/10.3390/aerospace13040314

Network, categorizing features into sparse and dense groups, significantly outperforming traditional methods. Gu [18] used Gradient Boosted Trees (GBT) for feature selection and Gradient Boosted Regression Trees (GBRT) for modeling, achieving 94.31% accuracy within ±5 min. Yuan [19] constructed a multiple linear regression model and used the northern goshawk optimization algorithm to fine-tune an XGBoost model.

Beyond traditional research that focuses on influential factors and data-driven modeling for taxi-out time prediction, the intrinsic characteristics of civil aviation operations—such as data scarcity and class imbalance caused by unevenly distributed operational conditions (e.g., peak hours, specific aircraft types, or adverse weather)—pose significant challenges to model generalization. In response, advanced techniques in data augmentation, similarity modeling, and attention mechanisms have gained increasing attention.

In the area of data augmentation, Zhang [20] proposes a framework to address data scarcity and class imbalance in aircraft trajectory prediction under severe weather conditions. By incorporating an improved CTGAN, the method enhances prediction accuracy. Yoon and Lee [21] introduce the ATRADA framework, which employs a Transformer encoder to learn latent patterns in trajectory data, providing an effective example of data augmentation applications in the aviation domain.

Regarding similarity-based approaches, AZ Naser [22] develops the SPINEX algorithm, which leverages similarity and multi-scale temporal interactions to achieve superior forecasting accuracy and interpretability, consistently ranking among top performers across diverse datasets. P. Yang [23] presents a comprehensive review of learning-based approaches for similarity computation across sets, sequences, and graphs, proposing a novel taxonomy based on learning targets—similarity learning, cost matrix learning, and search heuristic learning—to address efficiency and quality challenges in similarity metrics.

In terms of attention mechanisms, Viresh Pati [24] proposes CAPS, a structured attention mechanism for time series forecasting that effectively disentangles global trends, local shocks, and seasonal patterns within a single layer. Yang [25] introduces the MT-DSTGAN model, which captures dynamic spatio-temporal correlations in airport traffic data through multi-task learning of airport-wide and segment flows, significantly improving taxi time prediction accuracy.

These studies collectively underscore the growing importance of data augmentation, similarity modeling, and attention mechanisms in addressing the unique challenges of aviation operations and provide valuable references for the methodological framework adopted in this work.

However, a systematic review of the literature reveals the following core research gaps. First, existing studies lack a deep integration and efficient representation mechanism for multi-source heterogeneous data—including highly dynamic surface traffic, flight attributes, and operational environment—which limits the model's comprehensive perception of complex operational situations. Second, the capability to model spatio-temporal dependencies and dynamic interactions among flights during surface operations remains limited; current models struggle to accurately capture conflicts, following behaviors, and queuing dynamics during taxiing. Third, in the presence of imbalanced sample distributions in historical data (e.g., scarce samples under extreme weather or peak congestion scenarios), there is a lack of effective data augmentation or adaptive learning mechanisms. This leads to unstable prediction performance in key operational scenarios, particularly under high-accuracy requirements such as the ±2 min interval. These limitations collectively undermine the reliability, generalization capability, and practical engineering value of existing prediction models in real-world operational environments.

Therefore, to systematically address these challenges, this paper proposes an integrated prediction model named DDA-SIM-ATT-CatBoost, which fuses Dynamic Data

Augmentation (DDA), Similarity Theory (SIM), and an Attention Mechanism (ATT). The model aims to expand the distribution space of training samples through dynamic data augmentation, alleviating data sparsity and imbalance issues; utilize Similarity Theory to achieve high-accuracy matching of historical operational patterns, mining inherent data patterns; and employ the attention mechanism to adaptively calibrate the weights of multi-source features, strengthening the representation of key influencing factors. Finally, precise regression prediction is achieved using the CatBoost algorithm, with the expectation of comprehensively improving the prediction accuracy, stability, and generalization ability of the model in complex real operational environments, thereby providing more reliable decision support for refined scheduling of airport ground operations.

2. Influencing Factors of Departure Flight Taxi-Out Time and Their Correlation Analysis

2.1. Definition of Taxi-Out Time

Taking the current departure flight i as the research object, the surface taxiing interval between its Actual Off-Block Time (AOBT) and its Actual Take-Off Time (ATOT) is [AOBT(i), ATOT(i)]. According to the definition of departure flight taxi-out time, its calculation formula is as follows:

$$T = T_{\text{ATOT}} - T_{\text{AOBT}} \tag{1}$$

where T represents taxi-out time, T_{ATOT} represents the flight's actual take-off time, and T_{AOBT} represents the flight's actual off-block time.

Within the studied time interval, departing and arriving aircraft at the airport are categorized into 8 types: departing flights d_1, d_2, d_3, d_4, and arriving flights a_1, a_2, a_3, a_4. Any type of aircraft may influence the departure taxi-out time of the current aircraft i, with specific definitions provided in Table 1.

Table 1. Airport surface operation situation analysis.

Flight Category	Operational Constraints
d_1	{AOBT (j) < AOBT (i), AOBT (i) < ATOT (j) < ATOT(i)}
d_2	{AOBT (j) < AOBT (i), ATOT(j) > ATOT(i)}
d_3	{AOBT (i) < AOBT (j) < ATOT (i), AOBT (j) < ATOT (j) < ATOT (i)}
d_4	{AOBT (i) < AOBT (j) < ATOT (i), ATOT (j) > ATOT (i)}
a_1	{ALDT(j) < AOBT (i), AOBT (i) < AIBT(j) < ATOT(i)}
a_2	{ALDT(j) < AOBT (i), AIBT(j) > ATOT(i)}
a_3	{AOBT (i) < ALDT(j) < ATOT(i), AOBT (i) < AIBT(j) < ATOT(i)}
a_4	{AOBT (i) < ALDT(j) < ATOT(i), AIBT(j) > ATOT(i)}

For any departure flight j, its off-block time and take-off time are denoted as AOBT(j) and ATOT(j), respectively. A flight j is defined as type d_1 if its off-block time is earlier than that of flight i and its take-off time is also earlier than that of flight i. The classification rules for the remaining types d_2, d_3, d_4 and a_1, a_2, a_3, a_4 follow a similar logic.

2.2. Influencing Factors of Taxi-Out Time

Through literature analysis, it is known that departure flight taxi-out time is significantly influenced by factors such as surface traffic flow, 30 min average taxi-out time, taxiing distance, and weather, which have been fully demonstrated in previous studies. Comprehensively considering existing research findings and data availability, this paper selects 13 influencing factors: number of aircraft pushing back in the same period (x_1), number of aircraft taking off in the same period (x_2), departure queue (x_3), arrival queue (x_4), 30 min average taxi-out time (x_5), average taxi-out time 30 min before off-block (x_6), average taxi-out time 60 min before off-block (x_7), airline code (x_8), taxiing distance (x_9), aircraft type (x_{10}), apron (x_{11}), weather (x_{12}), and number of turns (x_{13}). The definition of each influencing factor is as follows:

(1) Number of aircraft pushing back in the same period x_1 (flights): Its expression is shown in Equation (2):

$$x_1 = d_2 + d_3 \tag{2}$$

In the equation, x_1 represents the number of all departure flights j that push back within a certain range before and after the current departure flight i pushes back.

(2) Number of aircraft taking off in the same period x_2 (flights): Its expression is shown in Equation (3):

$$x_2 = d_1 + d_2 + d_3 + d_4 \tag{3}$$

In the equation, x_2 represents the number of all flights whose surface operations influence the current departure flight i within the defined period.

(3) Departure queue x_3: Its expression is shown in Equation (4):

$$x_3 = d_1 + d_3 \tag{4}$$

In the equation, x_3 represents the number of departure flights j that have already taken off during the taxiing process of the current departure flight i. A longer departure queue leads to increased waiting time for aircraft before the runway.

(4) Arrival queue x_4: Its expression is shown in Equation (5):

$$x_4 = a_3 + a_4 \tag{5}$$

In the equation, x_4 represents the number of arriving flights j that have already landed during the taxiing process of the current departure aircraft i. Since aircraft must comply with landing priority regulations during taxiing, the arrival queue also increases the taxi-out time of departure aircraft.

(5) Thirty min average taxi-out time x_5: Its expression is shown in Equation (6):

$$x_5 = \frac{1}{n}\sum_{i=1}^{n} t_i \tag{6}$$

In the equation, x_5 represents the average taxi-out time over a 30 min window, measured in seconds (s); t_i is the average taxi-out time of the i-th departure flight; and n is the number of departure flights that have taxied out within the 30 min.

(6) Average taxi-out time 30/60 min before off-block x_6/x_7: Refers to the average taxi-out time of all departure flights at the airport within 30/60 min before the target flight's off-block time. This metric reflects the congestion level and overall operational efficiency of the surface just before the target flight's departure, serving as a characteristic variable describing real-time surface traffic conditions.

$$x_6 = \frac{1}{N}\sum_{i=1}^{N} t_i \tag{7}$$

$$x_7 = \frac{1}{M}\sum_{i=1}^{M} t_i \tag{8}$$

In Equations (7) and (8): N, M are the numbers of departure flights that take off within 30 and 60 min before the target flight's off-block time, respectively, and t_i is the taxi-out time of the i-th departure flight; the logic for Equation (8) is similar.

(7) Airline x_8: Differences in apron distribution, operational characteristics, and flight organization methods among airlines may lead to systematic differences in taxi-out time. First, the average taxi-out time for each airline is calculated. Then, the K-means clustering method is used to group the airlines into four categories: A, B, C, and D. The classification results are shown in Figure 1. Category A $\leq$ 875 s, cluster center 638 s; 875 s < Category B $\leq$ 1062 s, cluster center 987 s; 1062 s < Category C $\leq$ 1225 s, cluster center 1107 s; Category D > 1225 s, cluster center 1288 s.

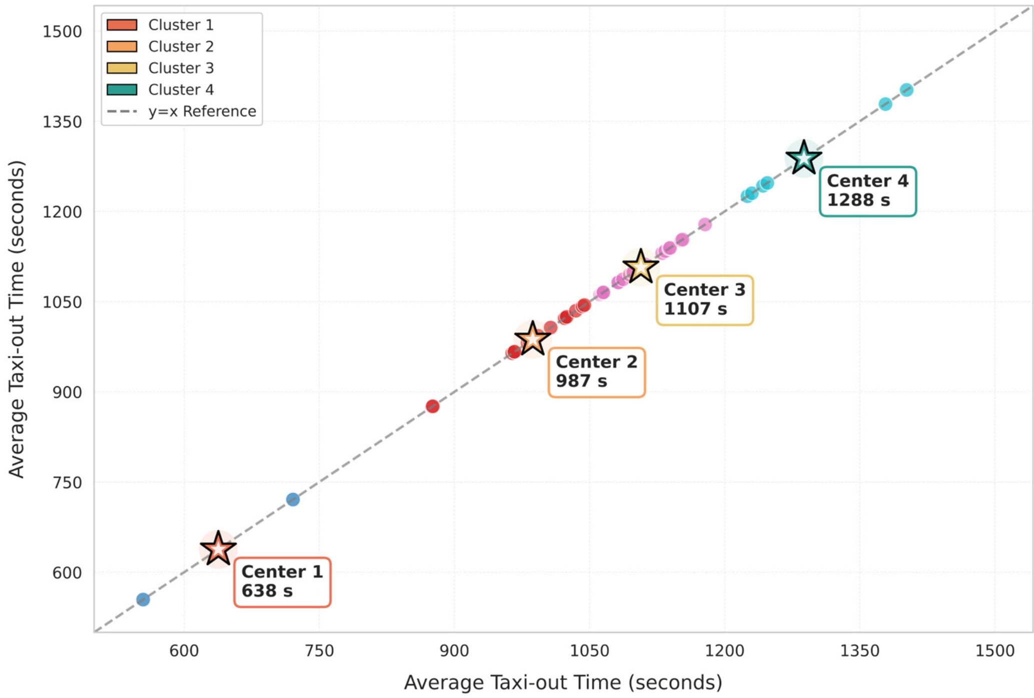

Figure 1. Airline classification clustering results.

(8) Taxiing distance x_9: Based on published stand locations, taxiway layouts, and standard taxi routes, the taxiing distance x_9 is measured using a proportional scaling method, unit: meters (m). Its expression is shown in Equation (9):

$$x_9 = 3600\frac{d_i}{s} \tag{9}$$

In the equation, d_i is the measured length of the taxiing distance for flight i; s is the measured length of the runway; 3600 is the runway length in meters.

(9) Aircraft type x_{10}: Different aircraft types have different taxiing performances and wake turbulence separation requirements, leading to differences in taxi-out time. According to ICAO Doc 8643, aircraft are classified based on Maximum Take-Off Weight (MTOW) and Wake Turbulence Category (WTC): medium and below aircraft are coded as 0, heavy aircraft are coded as 1.

(10) Apron x_{11}: The location of different aprons affects the taxiing distance and taxi-out time of flights. Generally, the farther the apron location and the more complex the taxi path, the longer the taxi-out time may be. First, the average taxi-out time for each apron is calculated. Then, the K-means clustering method is used to group the aprons into three categories: A, B, C. Category A $\leq$ 919 s, cluster center (814, 814);

919 s < Category B $\leq$ 1140 s, cluster center (967, 967); Category C > 1140 s, cluster center (1256, 1256). As shown in Figure 2.

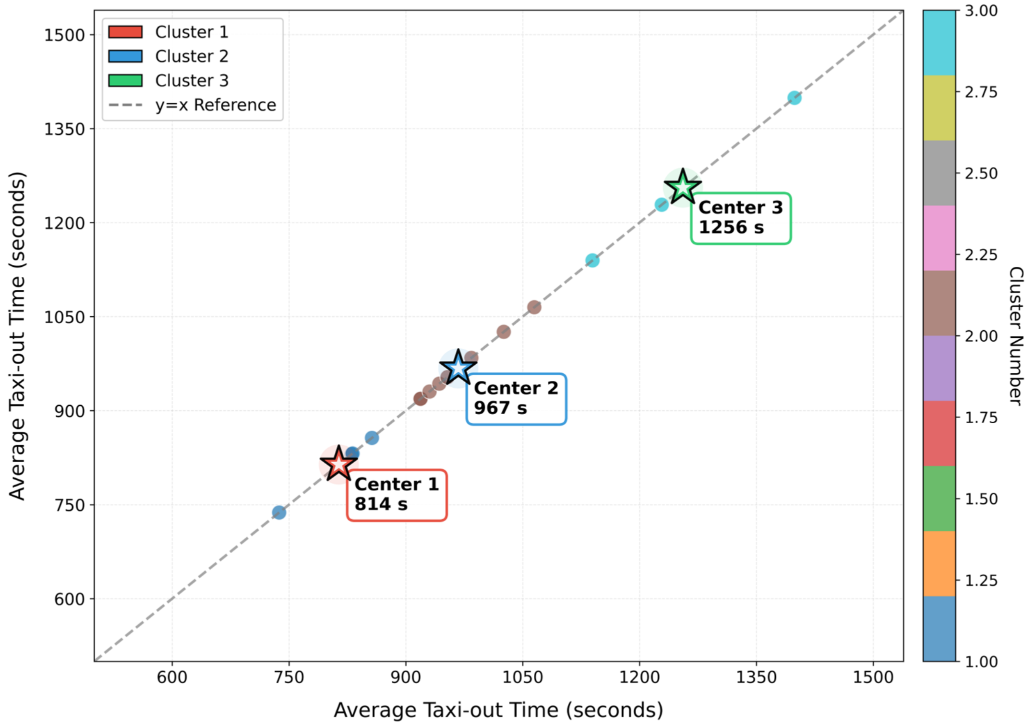

Figure 2. Apron classification clustering results.

(11) Weather x_{12}: From METAR reports, eight weather features are screened and retained: light rain (y_1), rain (y_2), light rain and light fog (y_3), rain and light fog (y_4), weak thunderstorm and rain (y_5), thunderstorm and rain (y_6), severe thunderstorm and rain (y_7), and no impact (y_8). These weather features are mapped to the 30 min average taxi-out time to construct a weather feature classification scheme. The 30 min average taxi-out time is clustered using the K-means method into three categories: A, B, C. Category A $\leq$ 979 s, cluster center 832 s; Category B is 979–1416 s, cluster center 1126 s; Category C > 1416 s, cluster center 1716 s. Flights with average taxi-out time $\leq$ 979 s and unaffected by weather are assigned a value of 1; flights with average taxi-out time between 979 and 1416 s and affected by y_1–y_6 are assigned a value of 2; flights with average taxi-out time > 1416 s and affected by y_7 are assigned a value of 3. Screening revealed no cases of significant changes in taxi-out time due to low visibility takeoffs or temporary runway changes. Therefore, a weather factor quantification scheme is constructed based on METAR reports: Category A (no impact y_8) assigned value 1, Category B (y_1–y_6) assigned value 2, Category C (severe thunderstorm and rain y_7) assigned value 3. Finally, the quantified weather factor indicator x_{12} corresponding to 5442 flights is obtained, as shown in Figure 3.

(12) Number of turns x_{13}: The number of turn angles is determined by the aircraft's taxi path. Based on the standard taxi route map of Shenzhen Airport, the number of turn angles is counted. The number of turn angles for each taxi path is counted separately and matched with flights one by one, ultimately obtaining the number of turns variable x_{13}.

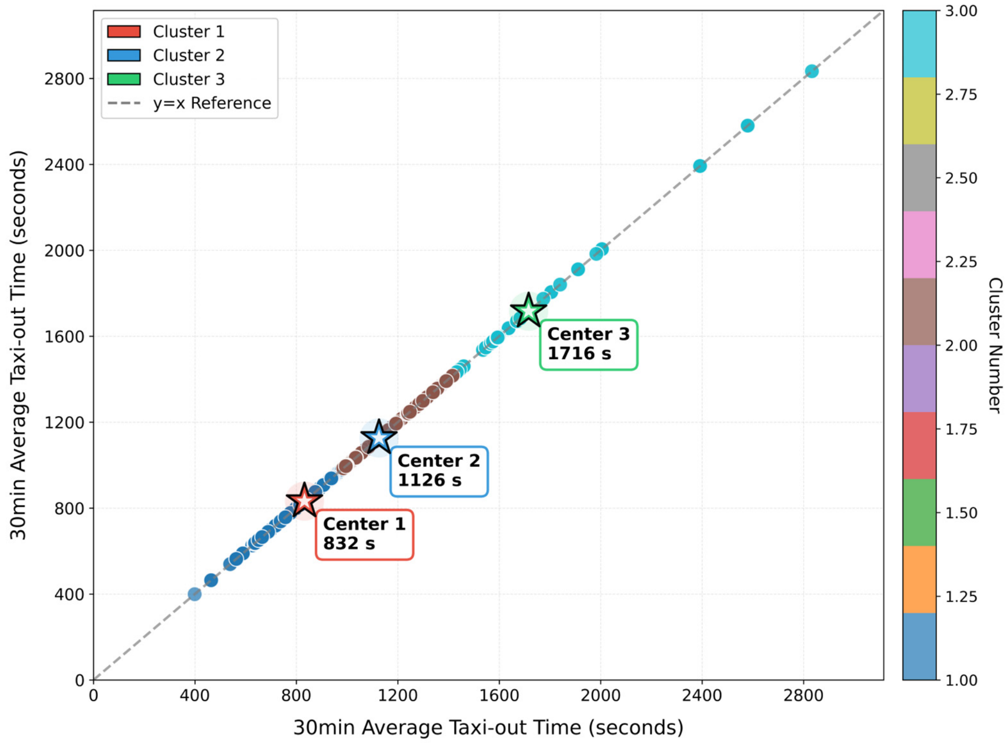

Figure 3. Weather classification clustering results.

Considering that taxi-out time is jointly influenced by multiple sources of factors, they are categorized into three groups:

(1) Airport Surface Traffic Flow: Reflects the operational load and flow status of the airport operational area, embodying the impact of surface congestion on taxi-out time. Mainly includes: number of aircraft taking off in the same period, departure queue, arrival queue, 30 min average taxi-out time, average taxi-out time 30 min and 60 min before off-block, number of aircraft pushing back in the same period. When surface traffic flow is relatively sparse, aircraft can usually complete taxiing, queuing, and take-off clearance smoothly after push-back. Conversely, it leads to prolonged taxi-out time.

(2) Flight Intrinsic Attributes: Characterize the operational performance and basic features of individual flights, mainly including airline code and aircraft type. Different airlines have varying scheduling processes, corporate cultures, and taxiing strategies, which may lead to systematic differences in taxi-out time; aircraft type determines performance parameters such as thrust level, taxiing speed, and turning radius.

(3) Flight Operational Environment: Describes the indirect impact of external meteorological and operational conditions on the taxiing process, mainly including taxiing distance, apron, weather, and number of turns. Among them, taxiing distance determines the physical path length from the gate to the runway; apron location reflects the complexity of the flight's area and taxiing path; weather conditions reduce taxiing speed and increase safety intervals; the number of turns reflects the complexity of the taxiing path and the number of potential conflict points. Overall, the operational environment also influences taxi-out time through meteorological constraint effects and path complexity effects.

2.3. Correlation Analysis of Influencing Factors on Taxi-Out Time

This paper conducts a systematic statistical analysis based on flight operation data and METAR reports from a major hub airport in South Central China over 14 days from

26 May to 8 June 2019. A total of 12,640 flight operation records and 672 weather reports were collected. Among these, there were 5986 departure flight records, with original fields covering key operational parameters such as flight number, aircraft type, runway used, parking stand, actual off-block time, actual on-block time, actual take-off time, and actual landing time. To ensure data quality, the 3σ principle was applied to clean the raw data, eliminating duplicate records and abnormal values, resulting in 5442 valid departing flight records. Based on this, 13 potential influencing factors were selected, and Pearson correlation coefficient analysis was employed to explore their relationships with flight taxi-out time. The output is presented as an information tree diagram in Figure 4.

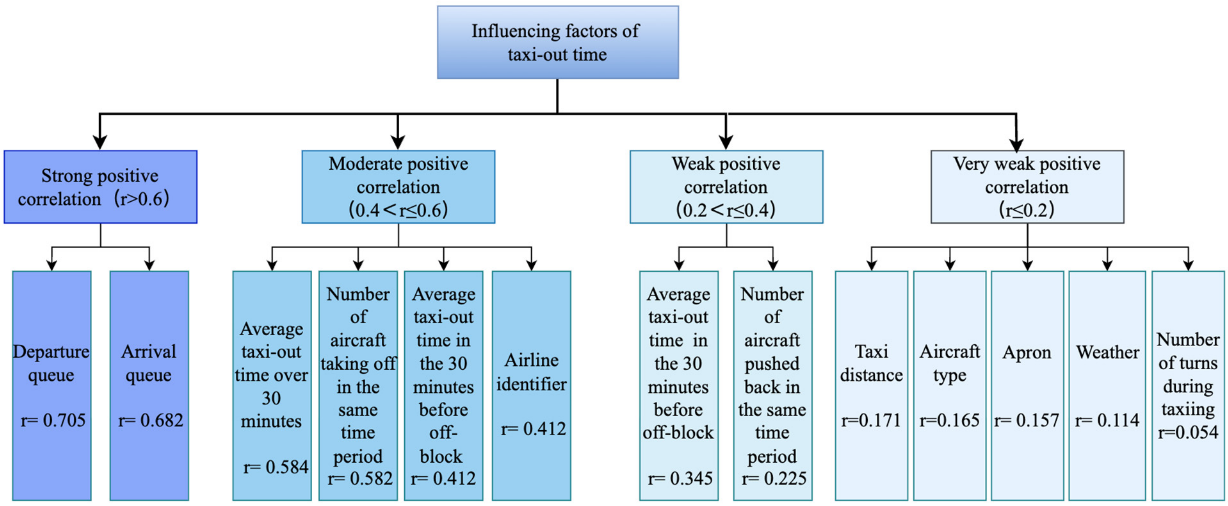

Figure 4. Information tree diagram of factors affecting taxi-out time.

As shown in Figure 4, the departure queue and arrival queue exhibit a strong positive correlation with taxi-out time ($r > 0.6$). The 30 min average taxi-out time, number of simultaneous departing flights, average taxi-out time 30 min before off-block, and airline code show a moderate positive correlation ($0.4 < r \leq 0.6$). The average taxi-out time 60 min before off-block and the number of simultaneous push-back aircraft show a weak positive correlation ($0.2 < r \leq 0.4$). Taxi distance, aircraft type, apron, weather, and number of turns show very weak correlations ($r \leq 0.2$).

3. Construction of the Departure Flight Taxi-Out Time Prediction Model

The construction process of the DDA-SIM-ATT-CatBoost based departing flight taxi-out time prediction model is shown in Figure 5.

Step 1: Data preprocessing and correlation analysis. Clean the raw departing flight data by removing missing, abnormal, and duplicate samples. Subsequently, the Pearson correlation coefficient is used for feature correlation analysis to screen feature variables significantly correlated with taxi-out time, providing a high-quality data foundation for model construction.

Step 2: Dynamic data augmentation. To improve model generalization ability and mitigate overfitting, a dynamic data augmentation module is introduced. The original data is expanded in a controlled manner through feature perturbation strategies aligned with operational logic, doubling the dataset size and constructing an augmented training sample set.

Step 3: Tuple similarity clustering analysis. The K-Prototypes clustering algorithm is applied to the augmented mixed-type data for clustering, effectively integrating numerical and categorical features. The similarity coefficient between a sample and cluster centers is calculated based on Mahalanobis distance, and cluster assignment is determined based on similarity, providing a structural basis for subsequent feature weighting.

Step 4: Attention-based feature weighting. A single-head attention mechanism is introduced to dynamically calibrate the weights of input features. Learnable parameters are used to enhance the representation of key features and suppress noise interference, generating a more discriminative weighted feature representation.

Step 5: Model construction and validation. The aforementioned modules are integrated to construct the DDA-SIM-ATT-CatBoost prediction model. The data is split into training and testing sets in a 7:3 ratio. The augmented training data is used to train the CatBoost model with attention weighting. Finally, model performance is validated on the test set using metrics such as Mean Squared Error and Mean Absolute Error.

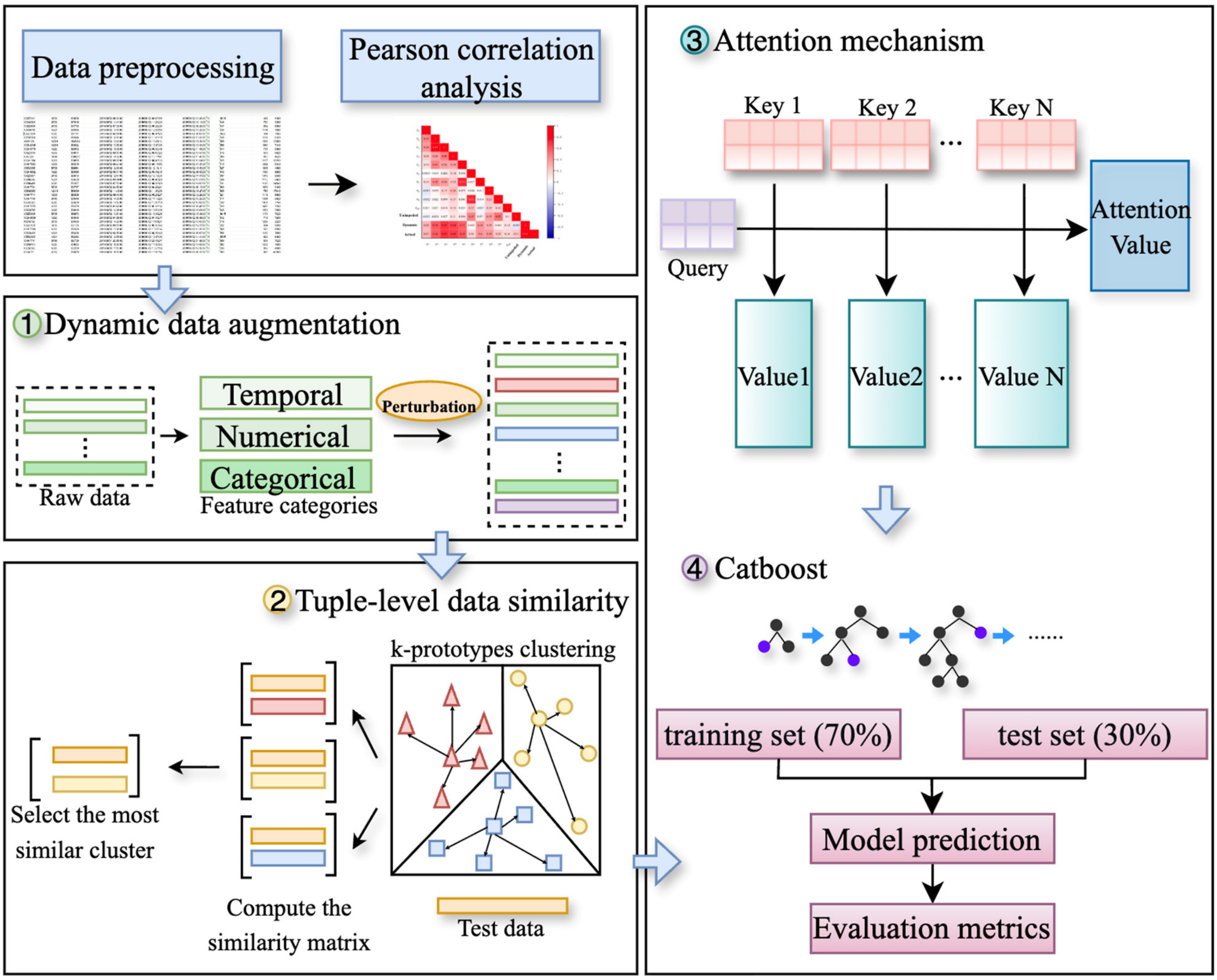

Figure 5. Flowchart for the prediction model of departing flight taxi-out time.

3.1. Dynamic Data Augmentation Module

Given that civil aviation operation data involves flight safety and operational privacy, its acquisition is subject to strict data governance policies and information security regulations, leading to generally limited availability of publicly usable datasets. This constraint can easily cause model training to struggle with convergence due to insufficient samples, thereby affecting the model's generalization performance. To address this problem, we introduce data augmentation techniques, which are applied after data preprocessing and before the training/testing split. Specifically, we generate high-quality synthetic samples by introducing reasonable perturbations to the original data based on various influencing factors, ensuring no cross-set mixing between augmented and original samples, and generating high-quality synthetic samples to expand the training set size, thereby improving model stability and prediction accuracy. This strategy helps enhance the modeling capability for the dynamic evolution patterns of key operational indicators like taxi-out time and constructs a flexible and efficient data augmentation mechanism for complex time series prediction tasks.

Dynamic data augmentation, as an adaptive data augmentation method, can dynamically adjust augmentation strategies and intensity based on the model's learning state, loss changes, and sample distribution characteristics during training. By analyzing historical time series data, it predicts future data trends and generates synthetic data with similar temporal dependencies and distribution characteristics. Its formula is expressed as follows:

$$X' = f(X, \theta t) \tag{10}$$

where X' is the original input sample, θ_t is the augmentation parameter (e.g., noise intensity, scaling ratio) that changes dynamically with the training phase t, and $f(\,)$ is the augmentation transformation function.

As shown in Figure 6, the original data is perturbed separately according to time features, categorical features, and numerical features, generating new data with a certain degree of randomness or minor variations. Subsequently, the perturbed data is merged with the original data to construct an augmented dataset, providing richer sample information for model training. This process also reduces the model's dependency on specific data distributions, improving the stability and generalization ability of taxi-out time prediction, as shown in Figure 6.

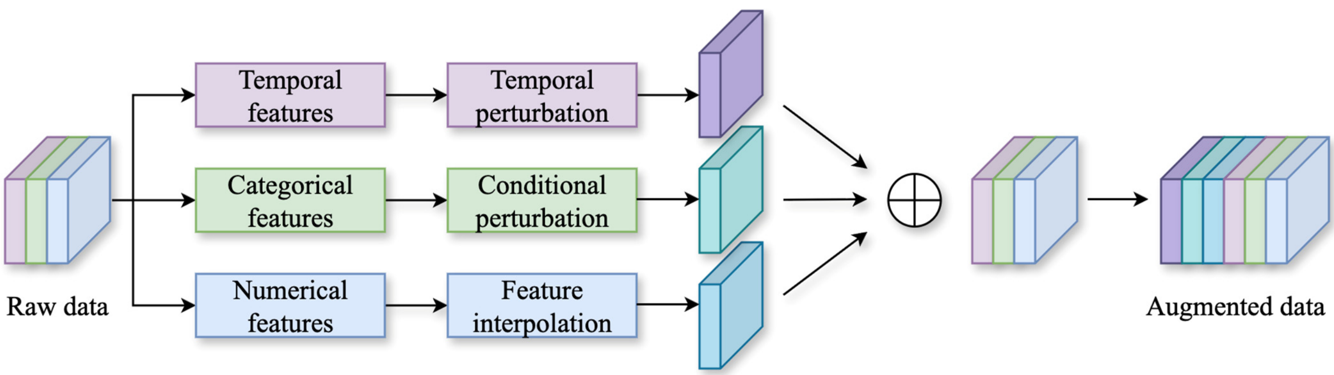

Figure 6. Flowchart for the dynamic data augmentation module.

3.2. *Tuple Data Similarity Module*

Similarity is a metric used to evaluate the closeness between two samples across multiple dimensions, including temporal correlation, statistical distribution consistency, and feature structure matching. In departing flight data, numerical proximity across multi-dimensional features often leads to similar regularities in taxi-out time. For example, flights with similar surface traffic flow, parking stands, and taxi routes typically operate in similar ground environments, experiencing comparable surface traffic conditions and clearance rhythms. Under the combined influence of these factors, taxi-out time exhibits strong similarity patterns. In this paper, each complete record of departing flight information in the dataset is defined as a tuple, containing 13 feature variables and their corresponding taxi-out time, serving as the basic unit for similarity calculation. To effectively quantify the similarity between tuples, Mahalanobis distance is introduced as the measurement basis. It describes the difference between an individual observation point and a certain

https://doi.org/10.3390/aerospace13040314

data distribution, while considering the covariance structure and scale differences across dimensions, thus possessing excellent properties such as scale invariance and dimension independence. Through normalization of the Mahalanobis distance, it can be transformed into a similarity coefficient, more intuitively reflecting the similarity relationship between tuple data (Figure 6).

$$DM = \sqrt{(X-\mu)^T \sum^{-1} (X-\mu)} \tag{11}$$

$$S = \frac{1}{(1 + D_M)} \tag{12}$$

where μ represents the mean vector and Σ represents the covariance matrix; the range of the similarity coefficient S is (0, 1].

To reduce the computational cost of calculating similarity coefficients one by one and to improve measurement accuracy, a clustering analysis method based on multi-dimensional features (K-Prototypes) is adopted, with its calculation formula as follows:

$$d(x_i, c_j) = \sum_{r=1}^{p} (x_{ir} - c_{jr})^2 + \gamma \sum_{r=1}^{q} \delta(x_{ir}, c_{jr}) \tag{13}$$

where the first part calculates Euclidean distance (numerical features) and the second part calculates Hamming distance (categorical features); γ represents the weight of categorical features; δ represents the indicator function; p represents the number of numerical features; q represents the number of categorical features; i corresponds to the i-th sample x_i in the sample set $x = \{x_1, x_2, \ldots, x_n\}$; j represents the number of clusters; and r indicates that each sample x_i contains p attributes $\{x_{i1}, x_{i2}, \ldots, x_{ip}\}$.

It can be seen from Figure 7 that the augmented data is used to perform K-Prototypes clustering on all samples, so that data within a cluster exhibits high similarity in the multi-dimensional feature space. For a sample to be predicted, its similarity coefficient with each data point within a cluster is calculated, and the average similarity coefficient within the cluster is taken. Subsequently, the cluster with the highest average similarity coefficient is selected as the cluster most similar to that sample, and taxi-out time prediction is based on the data within that cluster. This method can fully utilize the similarity of multi-dimensional features within clusters to improve prediction accuracy and robustness.

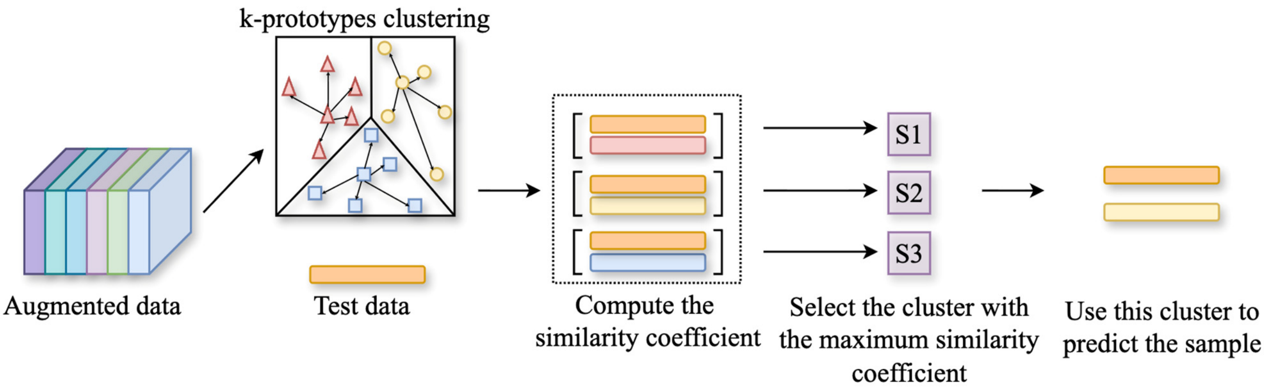

Figure 7. Flowchart for the tuple data Similarity module.

3.3. Attention Mechanism Module

In departing flight taxi-out time prediction, raw data contains multi-dimensional features such as surface traffic flow, aircraft type, apron, and weather. Complex dependencies and nonlinear relationships exist among these features. Moreover, the degree of influence of each feature on taxi-out time varies, making accurate prediction difficult through direct modeling. Therefore, the attention mechanism, which mimics the human cognitive process of selective focus, is chosen. It can adaptively and accurately capture the features most critical for taxi-out time prediction, effectively suppressing redundant information and noise interference, thereby enhancing the accuracy and robustness of the prediction model.

The attention mechanism implements its function based on an interaction framework constructed from Query (Q), Key (K), and Value (V) vectors. Its calculation formula is as follows:

$$Attention(Q, K, V) = \text{SoftMax}(\frac{QK^T}{\sqrt{d_k}})V \tag{14}$$

where $\sqrt{d_k}$ is a scaling factor used to control the magnitude of the dot product.

The single-head attention mechanism adopted in this paper is a basic and intuitive form of attention. Its core is to learn the importance weights of input features through a single attention head, thereby highlighting key features and suppressing secondary ones. The single-head attention mechanism has fewer parameters, a simpler model, and stronger generalization ability. The attention mechanism process is shown in Figure 8. First, the model calculates the similarity between the taxi-out time prediction demand (Query) and each input feature. The SoftMax function normalizes these similarities, yielding probabilistic attention weights that highlight features most relevant to taxi-out time and suppress redundant or irrelevant information. Subsequently, the value vector (Value) corresponding to each feature is multiplied by its attention weight and summed, producing a final representation that comprehensively considers feature importance, i.e., the attention-weighted representation. Through this mechanism, a more precise input expression is provided for taxi-out time prediction. The model can effectively identify structural dependencies and important patterns in the input data, filtering out noise interference while improving the accuracy and focus of prediction or classification tasks, as shown in Figure 8.

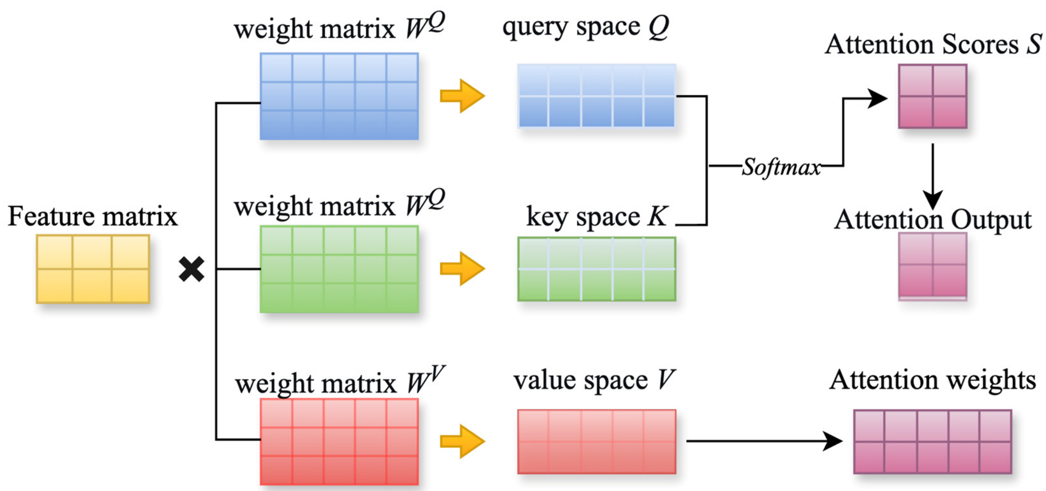

Figure 8. Flowchart for the Attention Mechanism module.

3.4. CatBoost Algorithm

CatBoost, as an optimized gradient boosting decision tree algorithm, exhibits significant advantages in handling categorical features. The algorithm uses the Ordered Target Statistics method to encode categorical variables, effectively mitigating overfitting risk; meanwhile, by introducing the Ordered Boosting strategy and feature crossing mechanism, it further reduces prediction bias. Its structure is based on symmetric binary trees, improving training and prediction efficiency while enhancing model stability. Additionally, CatBoost has built-in efficient categorical feature processing modules and combines greedy feature crossing methods, effectively alleviating the dimensionality explosion problem that may arise from high-dimensional feature combinations.

In the taxi-out time prediction task, a large number of categorical features (e.g., airline, parking stand, meteorological conditions) are typically involved, along with complex inter-feature interaction effects. Leveraging its excellent categorical feature encoding capabilities and automatic feature combination mechanisms, CatBoost can fully capture nonlinear relationships and high-order interactions among features. Therefore, this algorithm holds significant advantages in constructing taxi-out time prediction models, enabling high prediction accuracy while enhancing the model's generalization ability, making it suitable for high-dimensional data analysis needs in complex real-world scenarios.

4. Experimental Results and Analysis

4.1. Experimental Procedure

(1) Research object and data preprocessing: The taxi-out time of departing flights from Runway 15 at Shenzhen Bao'an International Airport is taken as the research object. There are initially 5986 departing flight records. After removing incomplete, duplicate, and abnormal data, 5442 valid departing flight records are obtained, as shown in Table 2.

Table 2. Preprocessing results for factors affecting taxi-out time.

Taxi-Out Time	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}
1482	0	4	0	0	1432.8	0	0	0	4137	2	2	1	2
1230	0	4	1	0	1432.8	1482	1482	0	2615	1	2	1	3
1320	1	5	3	1	1432.8	1356	1356	1	2843	1	2	1	5
2077	1	8	4	0	1432.8	1344	1344	1	3728	2	2	1	4
1055	2	4	2	1	1432.8	1527.3	1527.3	2	2279	1	2	1	3
1068	0	6	2	0	1133.1	1432.8	1432.8	0	2160	1	2	1	4
...	...	...	...	...	...	...	...	...	...	...	...	...	...

(2) Dynamic data augmentation: First, time windows are divided at 30 min intervals, using the taxi-out time and the average taxi-out time within 30 and 60 min before off-block as constraints to ensure temporal consistency of the augmented data. Based on this, Gaussian noise with a standard deviation of 5% is introduced to numerical features to enhance model robustness. Simultaneously, for categorical features (including airline, aircraft type, and apron), they are randomly replaced with other values from the same time window with probabilities of 20%, 15%, and 10%, respectively, to increase feature distribution diversity. Finally, a random perturbation of ± 5 min is applied to the target variable (taxi-out time), and the time window and its average taxi-out time are recalculated accordingly to ensure that deviations between augmented samples and the overall statistical characteristics of the window remain within a reasonable range.

(3) Tuple similarity data determination: This paper performs K-Prototypes clustering on the 10,884 departing flight records obtained after data augmentation. Based on the evaluation results of the silhouette coefficient and Davies-Bouldin Index (DBI), the optimal number of clusters is determined to be 4 (Figure 9). A silhouette coefficient closer to 1 indicates better clustering; a smaller DBI indicates tighter clusters and better separation between clusters. At k = 4, the two metrics achieve the best balance. Subsequently, within each cluster, the K-Nearest Neighbors (KNN) algorithm combined with Euclidean distance is used to select the 5 historical data records most similar to the sample to be predicted. Further, the covariance matrix is calculated using Mahalanobis distance to derive the average similarity coefficient. Finally, the cluster with the highest average similarity coefficient is selected as the basis for prediction.

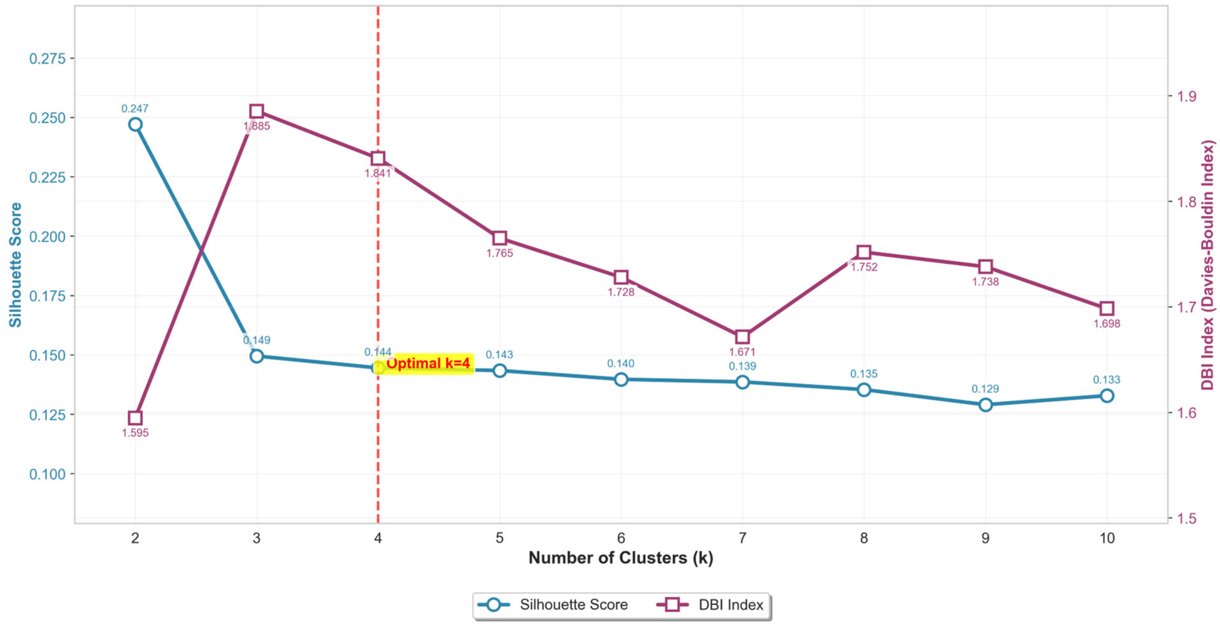

Figure 9. Combined plot of silhouette coefficient and Davies-Bouldin Index (DBI).

(4) The key hyperparameters of the CatBoost regressor are presented in Table 3. These parameters are adapted according to the sample size of each cluster and include early stopping to prevent overfitting.

Table 3. Model training parameters.

Parameter	Value
Iterations (based on sample size)	<100:500; 100–500:800; ≥500:1000
Tree depth (based on sample size)	<100:4; 100–500:6; ≥500:8
Early stopping rounds	50
Learning rate	0.05
Gradient clipping	max_norm = 1.0

4.2. Feature Selection

Based on constructing the DDA-SIM-ATT-CatBoost prediction model, this paper proposes a strategy to adaptively adjust model parameters according to the training sample size: when samples are limited, the depth and iteration count of the tree model are appropriately restricted to mitigate overfitting risk; when samples are sufficient, model complexity is gradually increased to enhance its fitting capability. Furthermore, based on the pre-trained attention model, input features are importance-weighted to construct a weighted feature matrix, which is then used as input for training the CatBoost model. The final prediction results for departing flight taxi-out time are shown in Table 4.

This paper selects Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and accuracy within specified error margins as evaluation metrics for model performance. In the task of predicting surface taxi-out time for departing flights at Bao'an International Airport, the model achieves accuracies of 74.57%, 89.12%, and 97.76% within error tolerances of ±120 s, ±180 s, and ±300 s, respectively. Simultaneously, MAPE is 10.34%, MAE is 87.55 s, and RMSE is 125.61 s. Compared with previous research results from the same research group [8], all evaluation metrics show significant improvement.

Table 4. Taxi-out time prediction results.

Feature Selection	±120 s/(%)	±180 s/(%)	±300 s/(%)	MAPE/(%)	MAE/(s)	RMSE/(s)
$x_1 \sim x_{13}$	72.58	86.27	96.78	10.94	93.38	132.74
$x_1 \sim x_{12}$	73.36	88.1	97.49	10.65	89.99	129.19
$x_1 \sim x_{11}$	74.57	89.12	97.76	10.34	87.55	125.61
$x_1 \sim x_{10}$	73.76	88.08	96.13	10.39	89.94	127.51
$x_1 \sim x_9$	73.5	87.12	97.26	10.65	90.77	131.55
$x_1 \sim x_8$	72.77	87.3	96.99	10.75	91.95	130.35

As shown in Table 4, after removing the weakly correlated features (number of turns and weather), the model achieves the highest accuracy across all time tolerances, with the best MAPE, MAE, and RMSE. This indicates that this feature combination helps improve the model's generalization ability and prediction accuracy. Further reduction in features leads to performance degradation, suggesting that the remaining features still hold important predictive value. Although weather factors accounted for the highest proportion (58.31%) in national civil aviation flight irregularity situations in 2024, due to the limited sample size of special weather events and the uncertainty of its impact, this feature did not stably contribute to the model. Additionally, aircraft taxiing is primarily straight-line, and the number of turns has limited impact on the total taxi time, with routes typically optimized by the ATC system. In summary, using the first 11 features with higher correlation achieves optimal results in predicting departing flight taxi-out time.

4.3. Ablation Experiment

To systematically evaluate the contribution of each module to the overall model performance, this paper designed and conducted multiple ablation experiments. A total of seven combination models were constructed with the following specific settings:

(1) CatBoost model: Serves as the baseline model for performance comparison with subsequent improved models incorporating different modules.
(2) CatBoost + SIM model: Introduces the Similarity module (SIM) on top of the baseline model to assess the impact of similarity information on taxi-out time prediction.
(3) CatBoost + ATT model: Integrates the Attention Mechanism module (ATT) into CatBoost to examine the role of the attention mechanism in improving prediction accuracy.
(4) CatBoost + DDA model: Adds the Data Augmentation module (DDA) to explore the improvement effect of data augmentation strategy on model prediction performance.
(5) CatBoost + DDA + SIM model: Further introduces the Similarity module based on CatBoost + DDA to evaluate the additional contribution of the Similarity module under data augmentation conditions.
(6) CatBoost + DDA + ATT model: Integrates the Attention Mechanism module into the CatBoost + DDA model to analyze the performance gain when data augmentation and attention mechanism work together.
(7) CatBoost + DDA + SIM + ATT model: Integrates all three components—data augmentation, the Similarity module, and the attention mechanism—to comprehensively evaluate the improvement in taxi-out time prediction accuracy through multi-module combination.

All models are trained and tested under the same dataset and parameter settings. The experimental results are shown in Table 5 and Figure 10.

Table 5. Ablation study results.

Module	±120 s/(%)	±180 (s)/(%)	±300 (s)/(%)	MAPE/(%)	MAE/(s)	RMSE/(s)
CatBoost (Baseline)	66.42	84.44	95.51	11.91	106.02	144.82
CatBoost + SIM	69.63	86.72	97.29	11.64	99.4	134.1
CatBoost + ATT	73.32	88.66	97.59	10.83	91.41	126.71
CatBoost + DDA	70.73	85.7	96.62	10.75	100.12	146.11
CatBoost + DDA + ATT	73.05	87.96	97.13	10.74	93.26	133.15
CatBoost + DDA + SIM	73.24	89.77	97.67	10.6	92.2	127.5
CatBoost + DDA + SIM + ATT	74.57	89.12	97.76	10.34	87.55	125.61

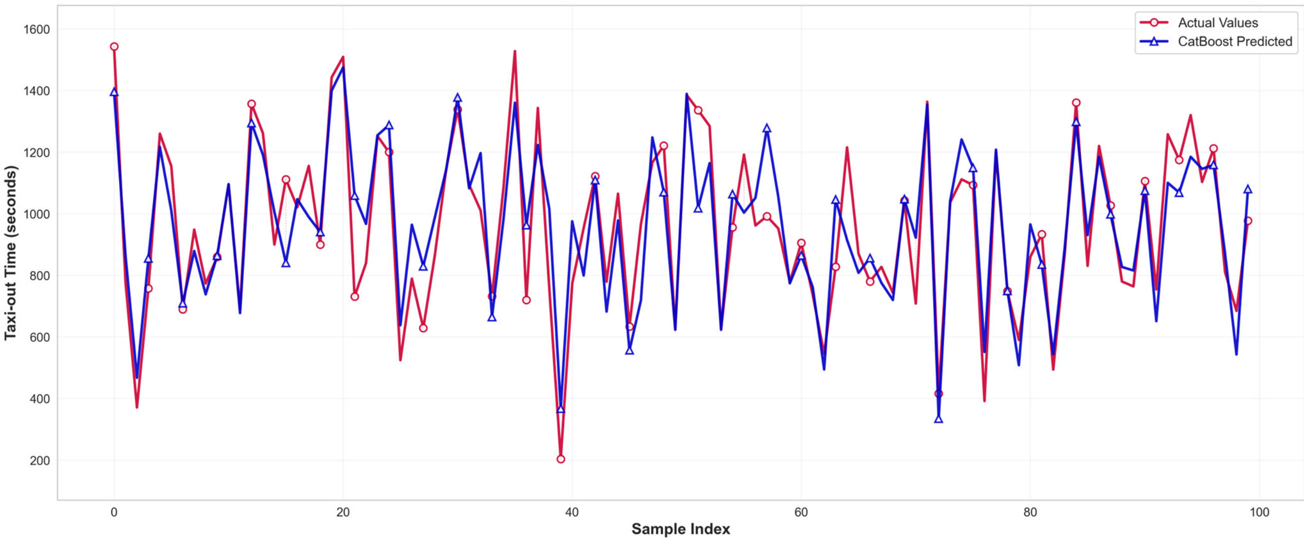

Figure 10. Taxi-out time prediction results based on CatBoost.

Compared to the CatBoost baseline, each module introduced individually brings performance gains, with the Attention Mechanism (ATT) module showing the most significant improvement. The CatBoost + ATT model achieves prediction accuracies of 73.32%, 88.66%, and 97.59% within the ±120 s, ±180 s, and ±300 s intervals, respectively. The main reason is that the ATT strengthens the role of key operational features through feature weighting, enhancing the model's ability to discern taxi-out time dynamics in complex operational scenarios. It is worth noting that Table 5 shows a slight decrease in the accuracy of CatBoost + DDA + SIM + ATT compared to the version without the attention mechanism in the ±180 s interval. This phenomenon may be attributed to the following reasons. First, the introduction of the attention mechanism enhances the model's ability to capture critical time steps but also increases model complexity. Under limited sample conditions, this may introduce a slight risk of overfitting, resulting in minor fluctuations in generalization performance on specific intervals such as ±180 s. Second, the attention mechanism tends to focus more on critical or extreme patterns, potentially allocating less attention to intermediate intervals. In contrast, the model without attention maintains a more balanced representation of global features, which may explain its slightly better performance on intermediate intervals. In addition, the baseline model already achieves 89.77% accuracy on the ±180 s interval, which is close to the upper bound of performance for this threshold, leaving limited room for further improvement. Therefore, although the full model shows a slight drop on the ±180 s interval, it achieves significant improvements on other evaluation metrics, indicating that the attention mechanism still contributes positively to the overall model performance. This also suggests that in complex time series prediction tasks, model improvements may lead to non-monotonic changes in specific metrics, highlighting the need for multi-dimensional evaluation.

Module combinations can further optimize model performance, with the combination of Data Augmentation (DDA) and the Similarity module (SIM) being particularly effective. CatBoost + DDA + SIM improves accuracy to 73.24%, 89.77%, and 97.67% within the ±120 s, ±180 s, and ±300 s intervals, respectively, representing an improvement of over 7 percentage points compared to the baseline, and performs best in the ±180 s interval. The DDA expands sample distribution through perturbation and resampling, improving model generalization; the SIM enhances prediction stability by measuring feature similarity between samples using Mahalanobis distance. The two exhibit good synergistic effects.

Finally, the model integrating all modules achieves optimal performance across all intervals, with accuracies of 74.57%, 89.12%, and 97.76% for ±120 s, ±180 s, and ±300 s, respectively. MAPE and RMSE are reduced to 10.34% and 125.61 s. Compared with previously published literature from the research team, the model proposed in this paper shows significant improvement in prediction accuracy within the ±120 s and ±180 s intervals, demonstrating notable engineering application value.

4.4. Comparative Experiment

To systematically evaluate the performance of different base models in the flight taxi-out time prediction task, this paper conducts a horizontal comparative analysis under a controlled experimental framework, selecting various representative models covering classical machine learning methods and cutting-edge deep learning architectures. The comparison is based on the same dataset and standardized evaluation system, aiming to systematically assess the adaptability and performance differences in each model in the taxi-out time prediction task. Specific results are shown in Table 6.

Table 6. Comparative experiment results.

Base Model	±120 s/(%)	±180 (s)/(%)	±300 (s)/(%)	MAPE/(%)	MAE/(s)	RMSE/(s)	*p*-Value
BP	65.22	82.41	94.59	12.59	113.16	159.91	$p < 0.001$
LSTM	63.93	81.39	94.73	12.84	114.47	158.27	$p < 0.001$
RF	65.99	83.58	95.63	12.22	108.86	149.63	$p < 0.001$
GRU	64.8	82.13	94.56	12.55	113.32	157.69	$p < 0.001$
XGBoost	65.62	84.07	95.45	12.33	109.12	150.4	$p < 0.001$
CatBoost	66.42	84.44	95.51	11.91	106.02	144.82	$p < 0.001$
OURS	74.57	89.12	97.76	10.34	87.546	125.61	$p < 0.001$

As shown in Table 6, there are significant performance differences among different models in departing flight taxi-out time prediction. Overall, ensemble learning models such as CatBoost, XGBoost, and RF outperform deep learning models like BP, LSTM, and GRU across all metrics. This indicates that under conditions of multi-source heterogeneous, noisy operational data, ensemble learning methods exhibit stronger robustness in nonlinear feature representation and noise resistance.

Taking the RMSE metric as an example, the model performance ranking is: OURS < CatBoost < RF < XGBoost < GRU < LSTM < BP. The RMSE of the OURS model is 125.61 s, approximately 13.2% lower than that of the best base model, CatBoost (144.82 s), indicating optimal error convergence and demonstrating clear advantages in overall prediction accuracy and stability. From the prediction accuracy metric (±180 s), the OURS model achieves an accuracy of 89.12%, about 4.68 percentage points higher than CatBoost. The accuracy ranking of the models is: OURS > CatBoost > XGBoost > RF > BP > GRU > LSTM.

https://doi.org/10.3390/aerospace13040314

This result further validates the effectiveness of the proposed method in improving prediction accuracy.

To statistically validate the performance improvements achieved by the proposed DDA-SIM-ATT-CatBoost model, we conducted independent-sample Mann–Whitney U tests comparing it against each baseline model. As presented in Table 6, all comparisons yield p-values below 0.001, indicating that the observed improvements are statistically highly significant.

Prediction error distribution plots are drawn for six base learner models (as shown in Figure 11). The results show that the prediction errors of all models are mainly concentrated within the -2 min to 2 min interval, indicating that all six base models can achieve relatively accurate predictions for most samples. However, compared to other models, the error distribution curve of the CatBoost model is more concentrated with less deviation, indicating higher prediction accuracy and stability in the flight operation time prediction task.

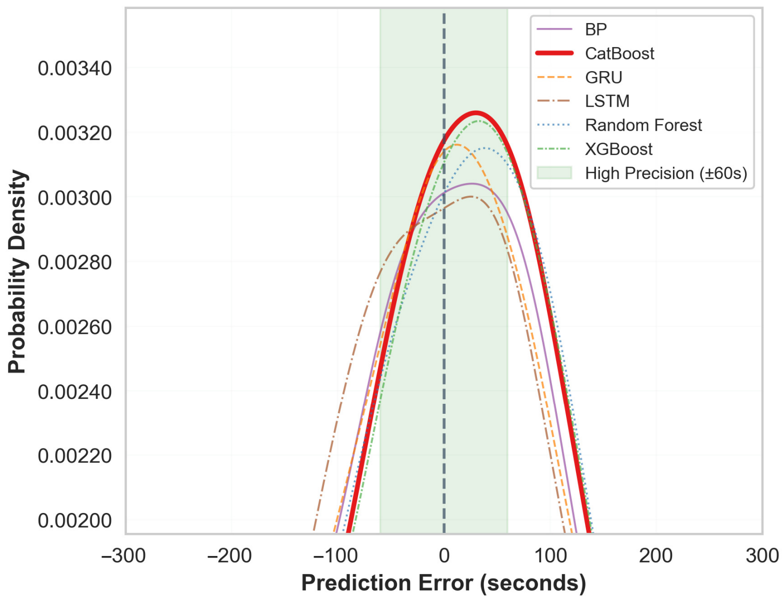

Figure 11. Prediction error distribution of baseline models for taxi-out time.

In summary, the CatBoost model performs best across all metrics, demonstrating strong nonlinear fitting capability and stability. The proposed OURS model, by integrating tuple similarity calculation and attention mechanism strategies, significantly outperforms traditional machine learning and deep learning single models in both error control and prediction accuracy, exhibiting strong generalization ability and engineering applicability.

5. Discussions

5.1. Interpretation of Major Findings and Underlying Mechanisms

The principal finding of this study is that the integrated DDA-SIM-ATT-CatBoost model achieves significant and robust improvements in taxi-out time prediction accuracy. Specifically, the model attained accuracies of 74.57%, 89.12%, and 97.76% within error margins of $\pm$120 s, $\pm$180 s, and $\pm$300 s, respectively, with substantial reductions in MAPE, MAE, and RMSE compared to baseline models. The underlying mechanisms for this enhanced performance can be attributed to the synergistic operation of its constituent modules.

First, the Dynamic Data Augmentation (DDA) module mitigates the data sparsity and imbalance inherent in operational aviation datasets, which often lack sufficient samples for rare but impactful scenarios (e.g., extreme weather, peak congestion). By applying controlled, operationally consistent perturbations to both numerical and categorical features, DDA effectively expands the training distribution. This enhances the model's ability to generalize to unseen conditions, thereby improving stability and reducing overfitting—a common limitation of models trained on limited, skewed historical data.

Second, the Similarity Theory (SIM) module, utilizing K-Prototypes clustering and Mahalanobis distance, enables precise matching of a target flight to historically similar operational contexts. This mechanism allows the model to leverage patterns from analogous situations (e.g., similar traffic flow, apron location, and taxi route complexity), effectively capturing the complex spatio-temporal dependencies and interaction effects between aircraft that are often oversimplified in traditional models. This approach moves beyond treating flights in isolation and better represents the queuing and conflict dynamics of the airport surface.

Finally, the Attention Mechanism (ATT) module dynamically recalibrates the importance weights of the 13 multi-source input features. Our ablation study confirmed that ATT yielded the most significant individual performance gain. This is because the module intrinsically learns to amplify the influence of highly predictive, real-time dynamic factors (e.g., departure/arrival queue length, 30 min average taxi-out time) while suppressing noise from less correlated or static features (e.g., number of turns, weather in this dataset). This adaptive feature weighting provides a more discriminative input representation for the subsequent CatBoost regressor.

5.2. Comparison with Existing Studies and Contribution of This Work

Our findings must be contextualized within the broader landscape of taxi-out time prediction research. Previous studies have made substantial progress by identifying key influencing factors [1–10] and applying various machine learning and deep learning techniques [11–20]. For instance, ensemble methods like Gradient Boosting and XGBoost have shown strong performance [16,19] and deep learning approaches have been explored for feature representation [18]. However, as outlined in the Introduction, persistent challenges remain regarding handling data imbalance, modeling complex inter-flight interactions, and achieving high accuracy in tight tolerance intervals (e.g., ±2 min).

Compared to these studies, our work makes several distinct contributions that directly address the limitations of existing approaches and enhance practical applicability:

(1) Holistic Integration for Data Challenges: While data augmentation is widely used in fields such as computer vision, its application in aviation time series prediction—particularly in a dynamic and operationally aware manner (DDA)—remains novel. Likewise, the explicit use of Similarity Theory (SIM) for pattern matching enables more direct modeling of spatio-temporal dependencies than conventional sequence models like LSTM. This integrated framework systematically tackles key data challenges including sparsity, imbalance, and missing values, thereby improving model robustness in real-world operational settings.

(2) Superior Performance in Operationally Critical Intervals: Many existing models report high accuracy within broader tolerances (e.g., ±5 min) [17]. In contrast, our model achieves a marked improvement in the more operationally stringent ±2 min (±120 s) interval, with an accuracy of 74.57%. This level of precision is essential for tactical decision-making tasks such as push-back scheduling, runway sequencing, and conflict prediction, directly supporting airport surface management under dynamic conditions.

(3) Synergistic Module Validation: Through systematic ablation experiments, we quantitatively demonstrate not only the individual contributions of DDA, SIM, and ATT but also their synergistic effect. The CatBoost + DDA + SIM + ATT configuration

consistently delivers optimal performance, confirming that the modules complement one another: DDA enriches the training data with plausible variability, SIM ensures contextual relevance through pattern matching, and ATT focuses the model on the most salient temporal signals. This synergy results in a more reliable and generalizable prediction framework.

Together, these contributions advance the field by proposing and validating a multi-module fusion framework that systematically addresses the core limitations identified in prior research. The model demonstrates strong potential for deployment in real-world airport traffic management systems, offering both improved accuracy and operational interpretability.

5.3. Limitations and Future Research Directions

This study acknowledges several limitations inherent to its scope and methodology. Firstly, the model's development and validation were based on data from a single major hub airport over a limited 14-day period. While the proposed multi-module framework is general in design, its direct performance at airports with significantly different layouts, traffic patterns, or air traffic control procedures requires further verification to assess true generalizability. Secondly, the representation and impact of complex, dynamic factors such as severe weather and explicit aircraft-to-aircraft interactions were simplified in the current modeling approach, potentially leaving nuanced operational effects unaccounted for.

Future research should prioritize multi-airport validation to establish the framework's transferability, potentially leveraging transfer learning techniques. Furthermore, enhancing the model's capability to incorporate more granular, real-time data streams—such as specific air traffic control instructions, refined meteorological impacts, and real-time surface surveillance—could significantly improve its predictive accuracy and operational relevance. Advancing the interpretability of the model through explainable AI techniques is also crucial to foster trust and facilitate its integration into practical airport decision-support systems.

6. Conclusions

(1) The proposed DDA-SIM-ATT-CatBoost model, which synergistically integrates dynamic data augmentation, Similarity Theory, and an attention mechanism, achieves prediction accuracies of 74.57%, 89.12%, and 97.76% within error margins of ±120 s, ±180 s, and ±300 s, respectively. Most notably, the substantial improvement within the critical 2–3 min window—a range directly relevant to tactical decision-making in airport surface operations—significantly enhances the model's engineering practicality and operational applicability.

(2) Ablation studies confirm that each constituent module contributes positively to overall performance. The attention mechanism exerts the most significant influence through adaptive feature weighting; dynamic data augmentation effectively alleviates sample imbalance; and Similarity Theory improves pattern matching accuracy. The synergistic integration of these three modules reduces MAPE, MAE, and RMSE to 10.34%, 87.55 s, and 125.61 s, respectively, surpassing those of all baseline configurations.

(3) Compared with conventional models such as XGBoost and Random Forest, the proposed model demonstrates superior performance in both prediction accuracy and stability. Relative to the second-best CatBoost model, it achieves reductions in MAE and RMSE of 17.4% and 13.2%, respectively, highlighting its enhanced nonlinear fitting capability and robustness. These results underscore its effectiveness in capturing complex operational patterns and support its potential for deployment in real-world traffic management systems.

Author Contributions: Conceptualization, Y.L. (Yue Lu); Methodology, Y.L. (Yue Lu); Software, Q.Z.; Validation, Q.Z.; Investigation, Y.Y.; Resources, J.Z.; Writing—original draft, Y.L. (Yanzhi Li); Writing—review and editing, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the National Key R&D Program of China (No. 2024YFC3014400) and the Sichuan Provincial Science and Technology Program Project (2026YFHZ0286).

Data Availability Statement: All data included in this study are available upon request by contacting the corresponding author.

Conflicts of Interest: Author Yue Lu was employed by the Ordos Low-Altitude Economy Development Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Gulding, J.; Knorr, D.; Rose, M.; Chen, X.; Enaud, P.; Hegendoerfer, H. US/Europe comparison of ATM-related operational performance. *Air Traffic Control Q.* **2010**, *18*, 5–27. [CrossRef]
2. Idris, H.; Clarke, J.P.; Bhuva, R.; Kang, L. Queuing model for taxi-out time estimation. *Air Traffic Control Q.* **2001**, *10*, 1–22. [CrossRef]
3. Simaiakis, I.; Pyrgiotis, N. An analytical queuing model of airport departure processes for taxi out time prediction. In Proceedings of the 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Fort Worth, TX, USA, 13–15 September 2010; p. 9148. [CrossRef]
4. Clewlow, R.; Simaiakis, I.; Balakrishnan, H. Impact of arrivals on departure taxi operations at airports. In *AIAA Guidance, Navigation, and Control Conference*; AIAA: Reston, VA, USA, 2010; p. 7698. [CrossRef]
5. Ravizza, S.; Chen, J.; Atkin, J.A.; Stewart, P.; Burke, E.K. Aircraft taxi time prediction: Comparisons and insights. *Appl. Soft Comput.* **2014**, *14*, 397–406. [CrossRef]
6. Yin, J.; Hu, M.; Ma, Y.; Han, K.; Chen, D. Airport taxi situation awareness with a macroscopic distribution network analysis. *Netw. Spat. Econ.* **2019**, *19*, 669–695. [CrossRef]
7. Wang, X.; Brownlee, A.E.; Woodward, J.R.; Weiszer, M.; Mahfouf, M.; Chen, J. Aircraft taxi time prediction: Feature importance and their implications. *Transp. Res. Part C Emerg. Technol.* **2021**, *124*, 102892. [CrossRef]
8. Xia, Z.; Wang, C.; Fang, P. Prediction of departure flight taxi-out time considering the influence of weather factors. *Sci. Technol. Eng.* **2023**, *23*, 11892–11899. [CrossRef]
9. Park, D.K.; Kim, J.K. Influential factors to aircraft taxi time in airport. *J. Air Transp. Manag.* **2023**, *106*, 102321. [CrossRef]
10. Tang, X.; Zhang, W.; Zhang, S.; Wu, C.L. A Road-Level Transport Network Model with Microscopic Operational Features for Aircraft Taxi-Out Time Prediction. *Aerospace* **2025**, *12*, 721. [CrossRef]
11. 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 Emerg. Technol.* **2010**, *18*, 950–962. [CrossRef]
12. George, E.; Khan, S.S. Reinforcement learning for taxi-out time prediction: An improved Q-learning approach. In *2015 International Conference on Computing and Network Communications (CoCoNet)*; IEEE: New York, NY, USA, 2015; pp. 757–764. [CrossRef]
13. Lee, H.; Malik, W.; Jung, Y.C. Taxi-out time prediction for departures at Charlotte airport using machine learning techniques. In *16th AIAA Aviation Technology, Integration, and Operations Conference*; AIAA: Reston, VA, USA, 2016; p. 3910. [CrossRef]
14. Lian, G.; Zhang, Y.; Desai, J.; Xing, Z.; Luo, X. Predicting Taxi-Out Time at Congested Airports with Optimization-Based Support Vector Regression Methods. *Math. Probl. Eng.* **2018**, *2018*, 7509508. [CrossRef]
15. Herrema, F.; Curran, R.; Visser, H.; Huet, D.; Lacote, R. Taxi-out time prediction model at Charles de Gaulle Airport. *J. Aerosp. Inf. Syst.* **2018**, *15*, 120–130. [CrossRef]
16. Diana, T. Can machines learn how to forecast taxi-out time? A comparison of predictive models applied to the case of Seattle/Tacoma International Airport. *Transp. Res. Part E Logist. Transp. Rev.* **2018**, *119*, 149–164. [CrossRef]
17. Li, N.; Jiao, Q.Y.; Zhang, L.; Wang, S.C. Using deep learning method to predict taxi time of aircraft: A case of Hong Kong airport. *J. Aeronaut. Astronaut. Aviat.* **2020**, *52*, 371–386. [CrossRef]
18. Gu, M.; Lu, C.; Dai, M.; Qian, J.J. Departure aircraft taxi time prediction based on gradient boosting regression tree. In Proceedings of the 2020 the 10th International Workshop on Computer Science and Engineering (WCSE 2020), Shanghai, China, 19–21 June 2020; pp. 76–80. [CrossRef]
19. Yuan, L.; Liu, J.; Chen, H. Additional Taxi-Out Time Prediction for Flights at Busy Airports by Fusing Flow Control Information. *Appl. Sci.* **2024**, *14*, 9968. [CrossRef]

20. Zhang, H.; Liu, Z. 4D aircraft trajectory prediction considering severe weather effects. *Aeronaut. J.* **2025**, *129*, 3619–3643. [CrossRef]
21. Yoon, S.; Lee, K. Aircraft Trajectory Dataset Augmentation in Latent Space. *Int. J. Aeronaut. Space Sci.* **2025**, *27*, 1960–1971. [CrossRef]
22. Naser, A.Z.; Naser, M.Z. SPINEX-TimeSeries: Similarity-based predictions with explainable neighbors exploration for time series and forecasting problems. *Comput. Ind. Eng.* **2025**, *200*, 110812. [CrossRef]
23. Yang, P.; Wang, H.; Yang, J.; Qian, Z.; Zhang, Y.; Lin, X. Deep learning approaches for similarity computation: A survey. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 7893–7912. [CrossRef]
24. Pati, V.; Kim, Y.; Pham, V.; Twitty, J.; Yang, S.; Lu, J. CAPS: Unifying Attention, Recurrence, and Alignment in Transformer-based Time Series Forecasting. *arXiv* **2026**, arXiv:2602.02729. [CrossRef]
25. Yang, X.; Yang, H.; Mao, Y.; Wang, Q.; Yin, S. Multi-Task Dynamic Spatio-Temporal Graph Attention Network: A Variable Taxi Time Prediction Model for Airport Surface Operation. *Aerospace* **2024**, *11*, 371. [CrossRef]