Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 12, No. 2, 2025

Research on Airport Traffic Situation Based on Multi-Source Data

Jie Ren [1], Weijia Hao [2]

[1] Guangzhou Civil Aviation College, Guangzhou Guangdong, 510403, China
[2] Guangdong Nanhua Vocational College OF Industry And commerce, Guangzhou Guangdong, 510000, China

Abstract: With the rapid development of the aviation industry, the complexity and uncertainty of airport traffic situations have increased significantly. Existing prediction methods for airport operations often rely on single data sources, which fail to comprehensively and accurately reflect dynamic changes in airport operations. This study proposes a multi-source data fusion framework for airport traffic situations and establishes an intelligent prediction model based on the fused data to achieve precise forecasting. Furthermore, optimization strategies for airport operations, including resource allocation and flight scheduling, are proposed based on prediction results. The effectiveness of these strategies is validated through simulation experiments. The results demonstrate that the intelligent prediction method based on multi-source data fusion significantly improves the accuracy and reliability of airport operation forecasting, providing a scientific basis for airport management and enhancing operational efficiency and service quality. The innovation of this research lies in proposing a hierarchical fusion architecture that combines deep learning with attention mechanisms to address spatiotemporal alignment challenges of heterogeneous data, as well as employing multi-objective optimization algorithms to balance resource utilization and passenger satisfaction metrics.

Keywords: Multi-source Data Fusion; Airport Traffic Situation; Intelligent Prediction; Optimization Strategies; Deep Learning.

1. Introduction

1.1. Research Background

Airports, as critical infrastructure in aviation networks, face escalating operational complexities due to rapid growth in air traffic [1][2]. Airport operations are influenced by multiple factors, including flight data (e.g., flight numbers, schedules, delays), meteorological data (e.g., weather conditions, wind speed, visibility), ground traffic data (e.g., vehicle flow, road congestion), and other related factors (e.g., facility status, staffing) [3]. Existing studies predominantly focus on single data sources, leading to incomplete or inaccurate predictions of dynamic operational changes [4]. Therefore, integrating multi-dimensional information through multi-source data fusion is essential to construct an intelligent prediction and optimization framework for addressing real-time operational challenges.

1.2. Research Significance

Accurate prediction and optimization of airport traffic situations hold significant value:Operational Efficiency: Optimized resource allocation reduces delays and improves runway utilization;Passenger Satisfaction: Dynamic scheduling minimizes waiting times and enhances service quality; Safety Assurance: Early warnings for adverse weather conditions mitigate operational risks.

For instance, Singapore Changi Airport improved gate turnover rates by 18% through data-driven scheduling [5]. This study provides both theoretical and practical insights for advancing smart airport development.

1.3. Research Content

This study encompasses four key components:Factor Correlation Modeling: Quantifying the contributions of flight data, meteorological data, and ground traffic data using Principal Component Analysis (PCA) and Granger causality tests [6];Multi-Source Data Fusion Framework: Designing a feature-level fusion method based on deep learning to resolve spatiotemporal alignment challenges;

Dynamic Prediction Model: Developing a multi-timescale prediction model (short-term 15 minutes, medium-term 1 hour, long-term 24 hours) integrating Long Short-Term Memory (LSTM) and attention mechanisms;Multi-Objective Optimization: Balancing resource utilization and passenger satisfaction via the NSGA-II algorithm.

2. Analysis of Factors Influencing Airport Traffic Situations

2.1. Flight Data

Flight data, including flight numbers, schedules, aircraft types, and delay causes, are critical drivers of airport operations. For example, mechanical failures cause abrupt delays, while air traffic control-induced delays exhibit temporal clustering. Analysis of historical data from a major hub airport revealed a 15% higher delay rate during morning peaks (6:00–9:00) due to runway saturation and insufficient ground support [7]. Aircraft type differences (e.g., A380 vs. B737 turnaround times) also impact gate allocation efficiency. Singapore Changi Airport improved gate turnover rates by 18% through aircraft-type-specific scheduling strategies [5].

2.2. Meteorological Data

Meteorological factors nonlinearly affect operations. For instance, runway closure probability rises significantly when visibility drops below 800 meters, and crosswinds exceeding 15 m/s necessitate directional adjustments for certain aircraft. Analysis of Guangzhou Baiyun International Airport’s 2021 data shows typhoon season (July–September) flight cancellation rates 3.2 times higher than non-typhoon periods. Localized convective weather (e.g., microbursts) requires integration of radar data and numerical forecasting models.

Denver International Airport reduced delays by 12% using the Weather Research and Forecasting (WRF) model for thunderstorm path prediction [8].

2.3. Ground Traffic Data

Ground traffic data encompass passenger vehicles, cargo trucks, and service vehicles. At Shanghai Pudong Airport, peak-hour vehicle inflows reach 5,000 vehicles/hour, with parking saturation extending passenger walking distances. GPS trajectory analysis reveals that each 100-meter increase in taxi pick-up distance prolongs boarding time by 4 minutes. London Heathrow Airport reduced passenger misconnections by 10% through real-time traffic monitoring and shuttle route optimization [9].

2.4. Other Factors

Infrastructure status (e.g., runway wear, jet bridge malfunctions) and staffing (e.g., security personnel shifts) critically impact efficiency. Real-time maintenance data can preempt risks, as demonstrated by a case where hydraulic system failure caused delays. Policy changes, such as pandemic-related terminal closures at Shanghai Hongqiao Airport in 2022, reduced passenger diversion efficiency by 25%, highlighting the need for dynamic adaptability in models [10].

2.5. Factor Correlation Analysis

Pearson correlation and grey relational analysis [11] reveal:Flight delay rates and visibility exhibit a correlation coefficient of -0.73 ($P < 0.01$);Ground traffic congestion and passenger misconnection rates show a grey relational degree of 0.68;Facility failures impact on-time performance with a 3-hour lag.Structural Equation Modeling (SEM) confirms meteorological factors account for 66% of delay effects (direct: 0.45; indirect via ground traffic: 0.21).

3. Multi-Source Data Fusion Methodology Operations

3.1. Data Preprocessing

3.1.1. Data Cleaning

Outliers in flight data (e.g., -30-minute delays) are identified via boxplots and 3σ criteria, then corrected using neighboring flight interpolation. Missing meteorological data (e.g., wind speed gaps) are filled via ARIMA time-series forecasting. Ground traffic outliers (e.g., 500% traffic spikes) are detected via Isolation Forest and smoothed using moving averages.

3.1.2. Data Standardization

Heterogeneous data are standardized hierarchically: Flight delays normalized to [0,1] via Min-Max scaling; Wind speeds standardized via Z-score; Categorical variables (e.g., weather types) encoded via One-Hot Encoding; Radar images enhanced via histogram equalization.

3.2. Feature Extraction

Random Forest-based feature importance analysis identifies key features: Flight Dimension: Rolling delay rate (6-hour window), aircraft turnaround time; Meteorological Dimension: 1-hour precipitation probability, vertical wind shear; Ground Traffic Dimension: Parking occupancy rate, taxi arrival peaks. Principal Component Analysis (PCA) reduces 47 features to 15 dimensions, retaining 95% variance. t-SNE visualizations validate feature separability.

3.3. Data Fusion Framework

3.3.1. Fusion Technique Selection

Deep Neural Networks (DNNs) outperform Kalman filters in nonlinear fusion. A hierarchical architecture is adopted: Bottom Layer: CNN extracts spatial features (e.g., regional traffic distribution); Middle Layer: LSTM captures temporal dependencies; Top Layer: Attention mechanisms dynamically weight multi-source features.

3.3.2. Framework Design

The four-layer framework includes: Data Ingestion Layer: Real-time APIs for ADS-B signals, weather radar, and magnetic vehicle flow data; Feature Extraction Layer: Parallel processing via CNN (images), LSTM (time-series), and Embedding (categorical data);Fusion Layer: Multi-head attention aligns multimodal features into 128-dimensional vectors; Output Layer: Airport Traffic Index (0–100; lower values indicate higher operational stress).Distributed computing ensures sub-200ms inference latency for real-time applications.

4. Intelligent Prediction Model Development and Validation

4.1. Model Selection

A hybrid Temporal Convolutional Network (TCN)-LSTM model addresses long-sequence gradient vanishing:Input Layer: 15 fused features;TCN Layer: Dilated convolutions (dilation=1,2,4) extract multi-scale temporal patterns;LSTM Layer: 64 neurons with tanh activation;Output Layer: Linear regression for 1-hour delay prediction.Bayesian optimization determines hyperparameters: learning rate=0.001, batch size=64, dropout=0.3.

4.2. Model Training and Validation

4.2.1. Dataset Partitioning

A rolling window approach (7-day window, 1-day step) generates 52 training-validation-test subsets to capture seasonal variations. Data from Beijing, Shanghai, and Guangzhou airports (2019–2022) comprise 1.2 million records.

4.2.2. Model Training

Adam optimizer with Huber Loss ($\delta=1.0$) balances MSE and MAE. Training converges at 50 epochs (validation loss=0.12). Learning rate decay (0.1 per 10 epochs) avoids local optima.

4.2.3. Model Validation

Five-fold cross-validation yields an average MAE of 8.2 minutes, 14% lower than standalone LSTM. The model achieves 87% recall for severe delays (>60 minutes) and outperforms ARIMA in extreme weather (22% lower MAE).

4.3. Prediction Result Analysis

4.3.1. Accuracy Evaluation

Test set performance:Clear Weather: MAE=6.3 minutes, $R^2=0.91$;Rainy Weather: MAE=9.8 minutes, $R^2=0.85$.Long-term (24-hour) predictions show 18% lower MAE than Prophet models.

4.3.2. Visualization

An interactive Dash platform visualizes regional heatmaps and delay curves. Case studies show 92% accuracy in predicting gate adjustments during thunderstorms. Multi-dimensional data drilling (e.g., by airline or aircraft type) aids decision-making.

5. Airport Operation Optimization Strategies

5.1. Resource Allocation Optimization

5.1.1. Gate Assignment Optimization

A multi-objective model minimizes total delays while maximizing gate utilization, subject to aircraft-gate compatibility and minimum turnaround constraints. Improved NSGA-II with dynamic crossover probabilities (Pc=0.7–0.9) generates Pareto-optimal solutions. Simulations show 22% higher gate turnover and 18% shorter delays. At Shenzhen Bao'an Airport, peak-hour gate conflicts decreased from 15% to 5%.

5.1.2. Boarding Gate Optimization

Hungarian algorithm-based matching reduces average walking distance by 28%. Amsterdam Schiphol Airport reported 12% higher passenger satisfaction using similar strategies.

5.2. Flight Scheduling Optimization

Dynamic departure sequencing prioritizes high-risk delays and weather-affected flights based on alternate airport capacities. Simio simulations show 15% shorter departure waits and 12% higher runway throughput. Chicago O'Hare Airport reduced evening peak departure delays from 45 to 32 minutes.

6. Conclusion

This study establishes an intelligent prediction and optimization framework for airport traffic situations using multi-source data fusion and deep learning. Experimental results demonstrate a 23% improvement in prediction accuracy over traditional methods, with optimization strategies effectively mitigating resource conflicts. Future work will focus on cross-airport collaboration via federated learning, high-fidelity digital twin environments, and reinforcement learning for real-time dynamic scheduling.

References

[1] Shi, D. (2022). Construction and Implementation of a Comprehensive Evaluation Index System for Airport Operations. Journal of Civil Aviation, 6(02), 26–29.

[2] Guo, R., Li, L., & Lei, Z. (2024). Evaluation of Airport Operations. Journal of Transportation Science and Management, 5(21), 9–12.

[3] Du, J. (2023). Dynamic Assessment and Optimization Control Methods for Airport Surface Traffic Situations (Doctoral dissertation, Nanjing University of Aeronautics and Astronautics). DOI:10.27239/d.cnki.gnhhu.2023.000099.

[4] Zhang, Z., & Cha, Z. (2024). Flight Delay Prediction Based on Multi-Airport Terminal Area Traffic Situations. Science, Technology and Engineering, 24(12), 5220–5226.

[5] Xu, Y., Wang, H., & Xiong, M. (2021). Civil Aviation Risk Assessment and Prediction Model Based on PCA-GA-BP. Ship Electronic Engineering, 41(02), 77–81.

[6] Dong, D., Liao, M., & Li, L. (2023). Research on Aviation Security Technical Inspection in ATSQ Construction Based on Grey Relational Analysis. China Science and Technology Information, (24), 117–123.

[7] Ma, X., & Zhao, D. (2023). Entity Extraction Method in Civil Aviation Based on Multi-Model Fusion. Computer Engineering and Design, 44(08), 2516–2522. DOI:10.16208/j.issn 1000-7024.2023.08.036.

[8] Denver International Airport Case Study. (2022). Aviation Weather Integration Report.

[9] Heathrow Airport Traffic Management. (2023). Annual Operational Review.

[10] Shanghai Hongqiao Airport Pandemic Response Report. (2022).

[11] Grey Relational Analysis in Transportation Systems. (2021). Journal of Advanced Transportation, 45(3), 112–125.