Team 25: Flight Dashboard Final Report

Navya Bingi, Andrew Guerrero, Agha Yusuf Khan, Zone Li, Ariana Richardson, Tyra Silaphet

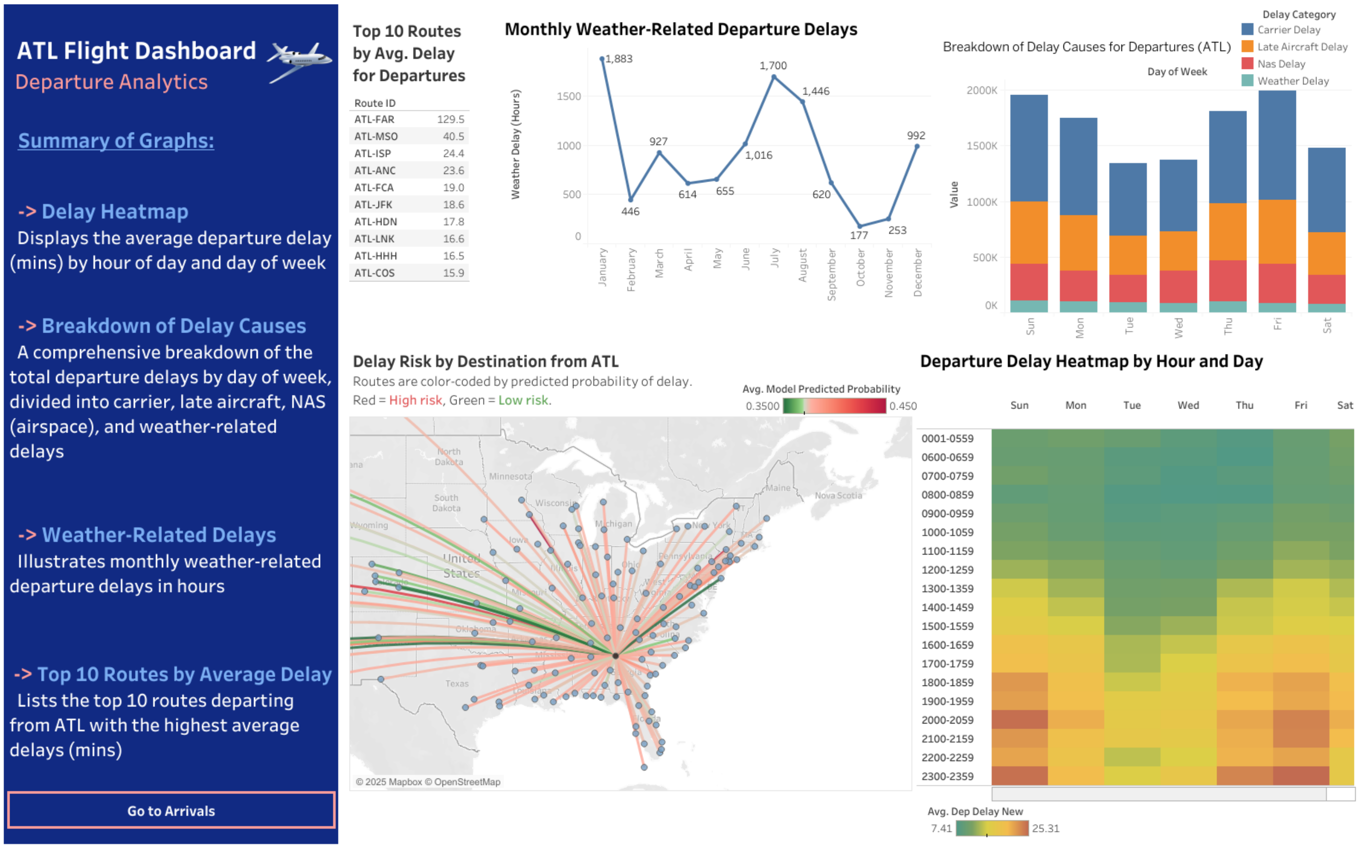

Figure 1: Flight Dashboard

1 Introduction

Currently, practices within the domestic air travel industry are under immense public scrutiny. For an industry that has seen technological innovations alongside business transformations, air travel has never been more safe and efficient. However, increased public awareness of airline incidents has led to greater dissatisfaction among consumers. Due to the increasing demand for air travel after COVID-19, airlines need to streamline their operations and system robustness. To do so, airlines must emphasize the use of artificial intelligence tools. Our aim is to operationalize machine learning models and subsequent visualizations to provide greater insights in such operations, specifically flight delays. To support this initiative, we successfully developed a machine learning-powered flight delay prediction dashboard, visualized in Figure 1, to enhance operational insights and decision-making. The project was developed over a 1.5-month timeline with no external or cloud-related costs incurred.

2 Problem Definition

Leveraging US Bureau of Transportation (BTS) On-Time Performance data, our aim is to develop a tool that models the behaviors of US carriers. This tool will use predictive modeling to anticipate impediments to on-time air travel such as flight delays and cancellations due to internal and external factors. These modeling tasks will be the basis for an initial visualization of air activity. A qualitative review has explored analytical approaches to flight delay data, emphasizing the prevalence of predictive analytics in this domain [1]. This paper highlights the prevalence of predictive analytics towards this endeavor; while that remains a focus of our approach, we extend this notion by building a complete visualization tool for stakeholders.

3 Literature Survey

Examining research within this domain and adjacent ones, our solution has some foundational merit based on prior approaches. Existing resources for flight delay, though informational, pose certain limitations. Currently, people rely on various sources to view flight schedules, real-time air traffic, and flight delay information, such as airline apps, booking websites, and other travel platforms. For example, users must visit multiple sources to get a comprehensive picture of routes and airport demand/delays; This information is therefore scattered, making the user experience more disjoint. A combined AI and visual tool will resolve this challenges leading to an enhancement of such experiences for both industry representatives and passengers alike. Present tools used by airlines involve live flight tracking rather than historical data, focusing primarily on real-time information derived from the Automatic Dependent Surveillance Broadcast (ADS-B). Research shows that the accuracy of demand forecasting and delay predictions can be improved with machine learning models [1, 2]. In addition, these tools rarely provide a visualization of the potential flight path, focusing primarily on real-time plane position rather than true flight tracking. Adopting time-based methods can help make such tools more dynamic [3].

3.1 Traditional Flight Delay Studies

Our inspirations extend from an existing machine learning (ML) pipeline which predicted flight delays through scheduling and historical flight data [2]. This feature engineering framework, entitled FDPP-ML, would be significant for our project, provided we account for this approach's omission of weather parameters. The inclusion of weather patterns would be available through FAA and BTS data. Several studies have showcased flight delays based upon weather conditions, flight characteristics, and traffic congestion, offering a variety of classification approaches to address extreme class imbalances [3, 4]. Research also suggests ground support service plays a key role in departure delays, emphasizing the importance of turnaround time [5]. As our data extends to performance data, this research would be out of scope for our analysis due to its focus on ground support nodes such as door opening, closing, and boarding times, however it shows context for potential results analysis and conclusions. All three research papers fail to provide accessible visualizations; a problem that is the foundation of our approach.

3.2 Related Research Areas

Rendering flight patterns using Markov Chains could allow for unique analyses in understanding flight delay data. Previously, an improved Time Limited Dispatch (TLD) model based on Markov processes was utilized to identify dispatch strategies, average safety level, and dispatch reliability [6]. Similarly, air travel demand forecasting has been explored through network analysis and deep learning [7, 8]. Our project will expand on these ideas by creating more robust predictive models and incorporating a digestible dashboard for monitoring delays and cancellations.

Another research area related to flight delays is aircraft maintenance. While the BTS On-Time Performance data does not hold this information directly, additional data sources within BTS and FAA can be appended to include the maintenance component. Our project intends to build on an existing multivariate time-series dataset for aviation maintenance, and reviewed predictive maintenance aviation techniques [9, 10]. This area of maintenance is intriguing due to its relationship with flight delay behavior. Assuming that such data can be appended to the On-Time Performance dataset, we will integrate this data into our machine learning models. However, in interest of model simplification, the maintenance data lens will be primarily experimental rather than a core feature of our analysis. Another experimental possibility is the inclusion of passenger demand data into flight delay models. As a standalone analysis, passenger demand forecasting based on economic indicators has been explored by aviation researchers [11]. Leveraging real-time data is critical for such passenger demand approaches; one means is to model passenger flow through airline or airport systems [12]. Existing methods to do so have been reviewed by the team in the hopes of applicability to our flight delay problem. These approaches can potentially inform our understanding of flight delay as instead of tracking passenger flow, we can analyze the counts of scheduled flights per day. Lastly, while not core to our approach, limited research has been conducted in the area of cost analyses [13]. While our efforts extend primarily to the development of the model and visualization, if able to extend this approach to establish costs, then our methodology would have other perceived benefits.

3.3 Current Limitations

Throughout our literature study we found several limitations in the current body of research. For example, some research studies on flight delay and related areas focused on a given airline rather than extending analysis across the whole industry [14]. In addition, a research study on the application of AI within the aviation industry found a primary challenge of such tools is operationalizing them for broader use in industry [15]. While development has gone into such tools, they are underutilized in day-to-day operations contributing to persistent inconsistencies. Thus, any solution needs to be agnostic to company infrastructure.

3.4 Methodology

Several resources that we reviewed provided ample justification for our intended approach. Regarding flight delay research, binary logistic regression has been used in the past encompassing complex passenger sentiments and for evaluating contradiction against perceived optimism [16]. Additionally, our intentions extend beyond current delay prediction efforts in that we seek to deploy information to the benefit of passenger end users [17]. Lastly, we found in current literature that similar AI applications have been developed for highlighting airline safety [18]. Our methodology builds off of those same approaches in an extension to flight delay.

4 Proposed Methods

Our approach includes several novel elements: (1) combining machine learning prediction models with a user-interactive dashboard for visualizing delay forecasts; (2) incorporating both historical and real-time data (planned) into a unified platform; (3) visualizing predicted delays and risks in a route-specific and airline-specific way for user accessibility; and (4) experimenting with feature engineering that accounts for contextual variables such as weather, time of day, and airport congestion.

4.1 Innovations

Our project introduces a user-focused, data-driven solution to a persistent challenge in air travel: understanding and anticipating flight delays. While many existing tools provide real-time tracking or raw delay data, few offer predictive insights in an accessible, visual format. Our approach fills that gap by combining machine learning models with a highly interactive dashboard, making delay prediction both practical and personalized. This system is designed for both airline stakeholders and everyday travelers. Unlike prior research that focuses solely on algorithm performance or industry use, we emphasize interpretability, usability, and integration of contextual variables. We are also innovating in the data pipeline itself. By enhancing flight-level data with contextual information—such as weather, airport congestion, and temporal patterns—we aim to improve the accuracy and relevance of our models. The dashboard's design prioritizes clarity and flexibility, allowing users to explore delay risk by airline, airport, route, or time of day. Furthermore, on the model-development front, utilizing techniques such as SMOTE-Tomek links to deal with class imbalance will lead to a more balanced dataset and improved model generalization. Collectively, these innovations move beyond academic modeling and toward a deployable, user-centered tool that supports informed travel decisions and operational transparency.

4.2 Data Collection and Processing

We began processing with loading and cleaning two raw data sources: METAR weather observations and on-time performance (OTP) flight records, both data sources being collected in Hartsfield-Jackson airport. For the METAR dataset, we filtered to only include METAR reports, converted data types, handled missing values using various imputation methods, created binary indicator variables, and removed unnecessary columns. For the OTP dataset, we removed columns collected post-takeoff (and thus unusable in our model), filtered out cancelled and diverted flights containing missing response variable values, converted data types, and created a datetime column for merging with the METAR dataset.

Then, we merged the two datasets together based on timestamps to align weather conditions with flights, as closely as possible. We analyzed which routes had the most cancelled or diverted flights, displaying the most and least problematic routes. Finally, we performed some final processing, including extracting and cyclically encoding circular time-related features, removed highly correlated columns to prevent multicollinearity issues, manually selected the columns to be used for modeling flight delays, then split the data into Hartsfield-Jackson departures and arrivals to enable specialized models to predict each type of delay.

4.3 Modeling Approach

To predict flight delays more effectively, we trained two separate binary classification models using **Random Forest Classifiers**: one for Atlanta **departures** and another for **arrivals**. Our objective was to predict whether a flight would be delayed by more than 15 minutes, treating it as a binary classification problem. Given a feature vector $X \in R^n$ representing flight, scheduling, and weather-related attributes, we aim to learn a function:

$$f(X) = P(y = 1 \mid X)$$

where $y = 1$ indicates a delay greater than 15 minutes.

Each model outputs both a binary label and a predicted probability, which are used in Tableau to color-code risk levels on the flight maps. To ensure fair representation of the minority class (delayed flights), we applied **class weighting** to handle the approximate 4:1 imbalance between on-time and delayed flights. The dataset was split using an **85/15 stratified train-test split**, maintaining the original class proportions. We performed a randomized hyperparameter search with 5-fold cross-validation to identify optimal model parameters. The final model was trained on the entire training data and evaluated using the test set.

Feature Engineering

To capture non-linear patterns in time-based behavior and seasonality, we used sine and cosine transformations for the scheduled hour and month:

$$\text{Hour}_{\text{sin}} = \sin\left(\frac{2\pi \cdot \text{hour}}{24}\right),$$
$$\text{Hour}_{\text{cos}} = \cos\left(\frac{2\pi \cdot \text{hour}}{24}\right)$$

In addition, we joined hourly METAR weather observations with flight records based on timestamp alignment. This included variables such as wind speed, visibility, humidity, sea-level pressure, and precipitation. We ensured alignment by rounding timestamps and selecting the weather data closest to the scheduled departure or arrival time..

4.4 Visualization

To translate our model insights into an intuitive and actionable format, we developed an interactive Tableau dashboard focused on Atlanta's Hartsfield-Jackson (ATL) airport. The dashboard is split across two linked views—Departures and Arrivals— which users can toggle between using a navigation button. Each view presents several visualizations that highlight temporal, spatial, and causal trends in delays. One core feature is a route-level risk map that displays predicted delay risk for each destination served from ATL.

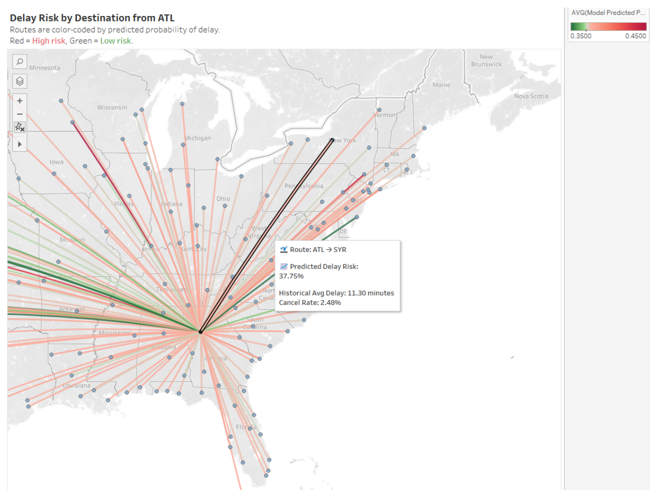

Figure 2: Departure Delay Risk Flight Path Map

Flight paths are color-coded from green to red based on the model's predicted probability of delay. As seen in Figure 2, which portrays the Departure modeling delay specifically, hovering over a route reveals detailed model output and historical statistics, including average delay duration and cancellation rates. This geospatial visualization helps users intuitively compare risk across destinations. To showcase the arrival model implementation as well, the following visuals, discussed will be arrival-specific, however, modeling was implemented for both departures and arrivals. We also visualize temporal trends using a heatmap that displays average delays by hour of day and day of week, as seen in Figure 3 3.

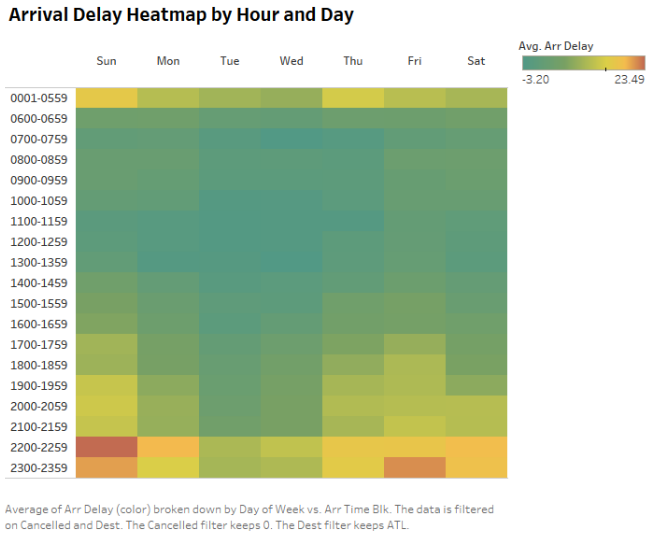

Figure 3: Arrival Delay Heatmap by Hour and Day

This enables users to identify when delays are most likely, such as late evenings and weekends. A separate bar chart breaks down delays by day of week and causes: carrier-related issues, late-arriving aircraft, NAS (National Airspace System), and weather, as seen in Figure 4.

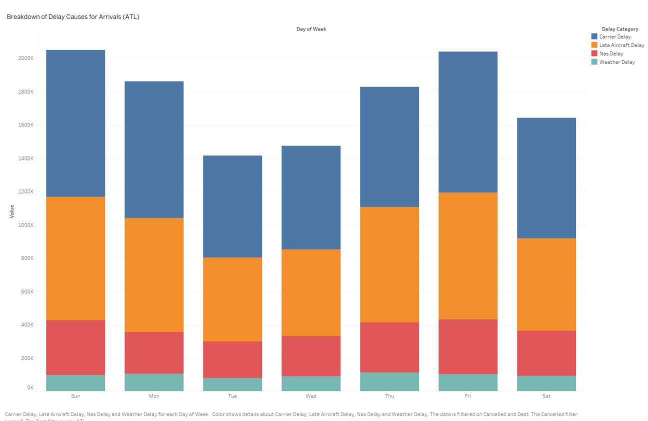

Figure 4: Breakdown of Delay Causes for Arrivals

To capture seasonal trends, we added a line chart, seen below in Figure 5 showing total monthly weather-related delays. This view highlights spikes in delays during certain months, such as winter storms or summer thunderstorms. In addition, a simple ranked list shown in Figure 6, visualizes the top 10 routes from ATL with the highest average delay for arrivals and departures, providing a quick reference for the most problematic routes.

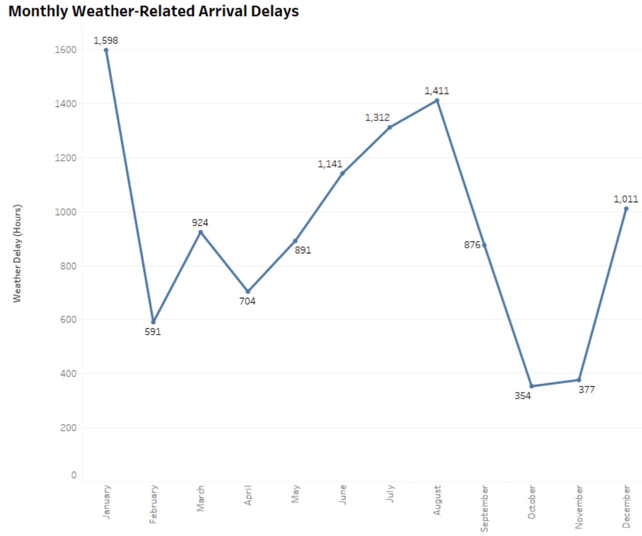

Figure 5: Monthly Weather-Related Arrival Delays ATL
Note: Data is filtered to include only non-cancelled flights (`Cancelled = 0`) with origin at ATL.

Top 10 Routes by Avg. Delay for Arrivals

Route ID	
MSO-ATL	76.80
RNO-ATL	43.87
ANC-ATL	31.71
PIA-ATL	27.74
OGG-ATL	26.44
SNA-ATL	21.33
PWM-ATL	20.66
MTJ-ATL	18.48
JFK-ATL	16.27
ASE-ATL	15.79

Sum of Avg Delay broken down by Route ID. The view is filtered on Route ID, which has multiple members selected.

Figure 6: Top 10 Routes for Arrivals by Average Delay

5 Evaluation

Model performance was assessed using standard evaluation techniques for classification models. As seen in Table 1, model performance was strong across both departure and arrival predictions, with ROC-AUC scores above 0.73 and overall accuracy nearing 80%, as seen in. Despite the 4:1 class imbalance in the data, our model is able to provide meaningful flight delay predictions, correctly flagging nearly half of all delays, while still maintaining a strong overall accuracy score. These results indicate the model effectively balances precision and

Model	Accuracy	Precision	Recall	ROC AUC	PR AUC	KS	Macro F1	Weighted F1
Arrival Delay	0.7995	0.4123	0.4628	0.7643	0.4385	0.3670	0.66	0.80
Departure Delay	0.7712	0.4339	0.4792	0.7316	0.4677	0.3449	0.66	0.788

Table 1: Delay Prediction Metrics for Modeling

recall, offering practical value in flagging high-risk flights without overwhelming users with false positives.

We also explored operationalizing our pipeline through a “snapshot” experiment, simulating real-time data updates by testing the model-dashboard flow on time-based subsets (e.g., Q1 2022). Results showed the pipeline and visualizations remained stable as data grew. However, these tests were done locally by manually overwriting data sources. In practice, industry deployment would require integration with cloud or SQL-based systems for seamless updates. While our approach shows promise, full automation would depend on access to such infrastructure.

To investigate the effectiveness of the visualizations, a product survey was sent out to lay-users and airline industry professionals, resulting in 28 responses. Questions followed the Likert Scale and assessed four key areas: dashboard accessibility, aesthetics, interpretability, and usefulness in a travel context. From an aesthetics standpoint, our product performed satisfactorily. Users cited difficulty interpreting the visuals and had varying degrees of difficulty navigating the interactive elements including filters and drop-down lists. Overall, there was consensus on the potential usefulness of the product as there is demand for information being presented in this manner, but this perceived value is diminished due to a lack of easily digestible visualizations. Based on the feedback, we prioritized improving visual interpretability and enhancing functional accessibility, while attempting to maintain the current design, culminating to the final version being presented.

6 Conclusions and Discussion

In this project, we developed a data-driven, user-centric tool for understanding and predicting flight delays at Hartsfield-Jackson Atlanta International Airport. Our contributions span the full data science pipeline—curating and merging weather and flight datasets, applying predictive modeling with a Random Forest classifier, and deploying results through an interactive Tableau dashboard. We built separate models for arrivals and departures, each demonstrating strong accuracy (0.80 for arrivals and 0.77 for departures) while maintaining reasonable recall for identifying delayed flights, despite the underlying class imbalance.

Our dashboard enables users to explore temporal trends (e.g., time-of-day and day-of-week delays), route-level risk visualizations, delay causes, and real-time flight delay predictions. These features enhance interpretability and real-world usefulness, assisting both travelers and airline stakeholders informed decisions. User feedback from our survey suggested high perceived usefulness, but highlighted opportunities to improve visual clarity and interactive accessibility, which we addressed in the final version.

Limitations. While we have expanded on previous work by integrating predictive modeling with an interactive visual tool, there remains room for improvement. Notably, our weather data was limited to conditions at Hartsfield-Jackson airport, assuming that delays were solely influenced by weather at the origin (for departures) or destination (for arrivals). Additionally, during model optimization, we balanced performance between predicting delays and non-delays rather than maximizing accuracy for one outcome. Depending on the intended use of the tool, this trade-off could shift—for instance, a user focused solely on identifying as many delays as possible would require a different optimization approach.

Future Work. Future improvements may include incorporating additional airports into our modeling, integrating true real-time data pipelines, and refining the dashboard interface for broader deployment. With additional data, such as aircraft maintenance logs or airport congestion metrics, we could improve both model accuracy and explanatory power. All team members contributed equally to the success of the project. Members divided into modeling and visualization sub-teams based on interest and skill set, with ongoing collaboration during all stages of development.

References

[1] Leonardo Carvalho, Alice Sternberg, Leandro Maia Goncalves, Ana Beatriz Cruz, Jorge A Soares, Diego Brandão, Diego Carvalho, and Eduardo Ogasawara. On the relevance of data science for flight delay research: a systematic review. *Transport Reviews*, 41(4):499–528, 2021.

[2] Maged Mamdouh, Mostafa Ezzat, and Hesham A. Hefny. A novel intelligent approach for flight delay prediction. *Journal of Big Data*, 10(1):179, 2023.

[3] Meshal Alfarhood, Rakan Alotaibi, Bassam Abdulrahim, Ahmad Einieh, Mohammed Almousa, and Abdulrhman Alkhanifer. Predicting flight delays with machine learning: A case study from saudi arabian airlines. *International Journal of Aerospace Engineering*, 2024(1):3385463, 2024.

[4] Seongeun Kim and Eunil Park. Prediction of flight departure delays caused by weather conditions adopting data-driven approaches. *Journal of Big Data*, 11(1):11, 2024.

[5] Zhao Yang, Yixin Chen, Jinbiao Hu, Yilu Song, and Yi Mao. Departure delay prediction and analysis based on node sequence data of ground support services for transit flights. *Transportation Research Part C: Emerging Technologies*, 153:104217, 2023.

[6] Zhong Lu, Xihui Liang, Ming J Zuo, and Jia Zhou. Markov process based time limited dispatch analysis with constraints of both dispatch reliability and average safety levels. *Reliability Engineering & System Safety*, 167:84–94, 2017.

[7] Hoai Nguyen Huynh, Kuan Luen Ng, Roy Toh, and Ling Feng. Understanding the impact of network structure on air travel pattern at different scales. *Plos one*, 19(3):e0299897, 2024.

[8] Lijuan Liu and Rung-Ching Chen. A novel passenger flow prediction model using deep learning methods. *Transportation Research Part C: Emerging Technologies*, 84:74–91, 2017.

[9] Hong Yang, Aidan LaBella, and Travis Desell. Predictive maintenance for general aviation using convolutional transformers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 12636–12642, 2022.

[10] Izaak Stanton, Kamran Munir, Ahsan Ikram, and Murad El-Bakry. Predictive maintenance analytics and implementation for aircraft: Challenges and opportunities. *Systems Engineering*, 26(2):216–237, 2023.

[11] Karina A Lundaeva, Zakhar A Saranin, Kapiton N Pospelov, and Aleksei M Gintciak. Demand forecasting model for airline flights based on historical passenger flow data. *Applied Sciences*, 14(23):11413, 2024.

[12] Xiaojia Guo, Yael Grushka-Cockayne, and Bert De Reyck. Forecasting airport transfer passenger flow using real-time data and machine learning. *Manufacturing & Service Operations Management*, 24(6):3193–3214, 2022.

[13] Judith Rosenow, Philipp Michling, Michael Schultz, and Jörn Schönberger. Evaluation of strategies to reduce the cost impacts of flight delays on total network costs. *Aerospace*, 7(11):165, 2020.

[14] Cen Song, Xiaoqian Ma, Catherine Ardizzone, and Jun Zhuang. The adverse impact of flight delays on passenger satisfaction: An innovative prediction model utilizing wide & deep learning. *Journal of Air Transport Management*, 114:102511, 2024.

[15] Michael J Pritchard Ph D, Austin T Walden Ph D, and Paul J Thomas Ph D. Integrated organiza-

tional machine learning for aviation flight data. *Journal of Aviation/Aerospace Education & Research*, 33(2):10, 2024.

[16] Gülay Demir, Sarbast Moslem, and Szabolcs Duleba. Artificial intelligence in aviation safety: Systematic review and biometric analysis. *International Journal of Computational Intelligence Systems*, 17(1):1–30, 2024.

[17] Daniel Y Suh and Megan S Ryerson. Forecast to grow: aviation demand forecasting in an era of demand uncertainty and optimism bias. *Transportation Research Part E: Logistics and Transportation Review*, 128:400–416, 2019.

[18] Feng Cao, Tieqiao Tang, Yunqi Gao, Oliver Michler, and Michael Schultz. Predicting flight arrival times with deep learning: A strategy for minimizing potential conflicts in gate assignment. *Transportation Research Part C: Emerging Technologies*, 169:104866, 2024.