

**An Intelligent System for Predicting Flight Demand
and Peak Travel Periods Using Machine Learning and Deep
Learning**

[Ruba Abusaifin , Ghaida Al-Suwaidan]

Master's Student, Department of [Artificial Intelligence]

Under the supervision of Dr. Yasser Abdelhamid

[MidOcean University]

Table of Contents

Abstract

Chapter 1: Introduction

Problem Statement .1.1

The Role of Predictive Analytics .1.2

Research Gap & Motivation .1.3

Research Objectives .1.4

Scope and Organization .1.5

Chapter 2: Technical Background

Extreme Gradient Boosting (XGBoost)

Random Forest

Artificial Neural Networks (ANN)

Temporal Preprocessing and Encoding

Feature Engineering

Feature Selection and Testing

Data Normalization and Scaling

Hyperparameter Optimization and Training Architecture

Chapter 3: Literature Review & Methodology Categorization

Overview of Related Work (16 Studies) .3.1

Methodological Groupings .3.2

Handling Class Imbalance and Threshold Adjustments .3.2.1

Temporal Decomposition and Dynamic Seasonal Engineering .3.2.2

Scalable Big Data Management and Dimensionality Reduction .3.2.3

Automated Hyperparameter Tuning and Computational Efficiency .3.2.4

Dynamic Operational Adaptation and Convergence Optimization .3.2.5

Literature Review Summary Matrix .3.3

Chapter 4: Proposed Framework and Discussion

(Framework Overview 4.2. Data Pipeline Components .4.1

Data Gathering .4.2.1

Preprocessing .4.2.2

Routing Analysis .4.2.3

Optimization .4.2.4

Comparative Performance Insight .4.3

Challenges & Workarounds .4.4

Chapter 5: Conclusion & Future Outlook

Conclusion .5.1

Future Work .5.2

References

Abstract

Handling peak travel periods is a major challenge for airports because sudden passenger surges put heavy pressure on terminals and staff. This paper lays the groundwork for building an intelligent forecasting system that predicts flight demand and identifies peak travel windows using real-world aviation data from the Bureau of Transportation Statistics (BTS) and Kaggle. By reviewing key studies published between 2023 and 2026, this research focuses on evaluating two main algorithms: Extreme Gradient Boosting (XGBoost) and Artificial Neural Networks (ANNs). The review examines how these models handle large datasets, address class imbalance, and process complex time-based features. It also discusses practical solutions for common operational issues, such as shifting holiday calendars and computational limits, providing a practical deployment guide for the upcoming system.

Chapter 1: Introduction

1.1. Problem Statement

Managing "peak travel periods" is one of the most pressing operational issues in civil aviation. These peaks happen when passenger numbers rise sharply within short timeframes, exceeding the fixed capacity of airport terminals. Passenger volume spikes are mainly driven by predictable events like school holidays, weekend travel, and national or religious celebrations. When airports are unprepared for these surges, the result is long queues, overcrowded halls, higher operational costs, and lower passenger satisfaction. Traditional statistical tools struggle to forecast these events accurately because passenger demand is non-linear and shaped by multiple overlapping seasonal factors.

1.2. The Role of Predictive Analytics

To address these congestion problems, modern airports are turning to predictive analytics. By building an intelligent prediction system, airport managers can forecast passenger demand and spot peak travel periods well in advance. Moving from reactive management to data-driven forecasting allows airports to allocate staff and resources more efficiently, reduce bottlenecks at security and customs checkpoints, and improve overall operational flow.

1.3. Research Gap & Motivation

Although many studies have tested machine learning and deep learning for flight forecasting, most focus on isolated datasets or individual algorithms. There is a clear need for a focused evaluation comparing tabular models with neural networks using standard, large-scale datasets like BTS and Kaggle. Furthermore, recent research from 2023 to 2026 needs to be synthesized to understand the practical trade-offs between model speed and accuracy under real airport conditions. This study addresses this gap by reviewing recent literature to establish a clear architectural base for our predictive system.

1.4 Research Objectives

The primary objectives of this study are:

1. To design and establish a technical framework for an intelligent prediction system that forecasts flight demand and classifies peak travel periods.
2. To systematically review and evaluate recent aviation forecasting literature published between 2023 and 2026.
3. To structure an efficient data pipeline tailored for processing large-scale aviation datasets sourced from the Bureau of Transportation Statistics (BTS) and Kaggle.
4. To conduct a comparative evaluation between XGBoost and Artificial Neural Networks (ANNs), focusing on training efficiency, class imbalance handling, and overall accuracy.
5. To identify practical solutions for major operational challenges, such as dynamic calendar tracking and computational overhead.
6. To construct a clear implementation roadmap for deploying the

1.5. Scope and Organization

This paper focuses strictly on defining the design framework for an intelligent flight demand forecasting system using BTS and Kaggle datasets. The algorithmic comparison is limited to XGBoost and Artificial Neural Networks (ANNs) based on research from 2023 to 2026.

The paper is organized as follows: Chapter 2 presents the technical background of XGBoost and ANNs alongside core computing challenges. Chapter 3 covers related works, detailing the 16 reviewed studies and their methodological categories. Chapter 4 presents the comparative analysis and the model selection framework. Finally, Chapter 5 concludes the paper and outlines the implementation plan for the upcoming system.

Chapter 2: Technical Background

Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is an efficient machine learning algorithm widely used for structured tabular data. Unlike traditional models that construct decision trees independently, XGBoost works sequentially. Each new tree is trained specifically to fix the residual errors made by the previous ones. Practically, XGBoost handles missing values automatically, which cuts down on preprocessing time. It also includes built-in regularization to help prevent the model from overfitting on training data.

Random Forest

Random Forest is a popular ensemble learning method that builds a collection of independent decision trees during training. The final prediction is calculated through a majority vote across all trees, providing high stability and strong resistance to overfitting. In flight demand forecasting, Random Forest processes tabular datasets efficiently. It requires significantly less training time than deep neural networks while remaining transparent and easy to interpret.

Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are designed to process information through layers of connected artificial neurons, loosely inspired by biological brain structure. A standard ANN includes an input layer, hidden layers for extracting features, and an output layer for making predictions. Using backpropagation, the network continually adjusts its internal weights to minimize overall prediction errors. This design allows ANNs to pick up on non-linear seasonal cycles and sudden traffic surges that simple tree-based models might miss.

Temporal Preprocessing and Encoding

Aviation data relies heavily on timestamps, which must be converted into structured inputs before feeding them into predictive models. Temporal preprocessing breaks down raw timestamps into actionable variables, such as the departure hour, day of the week, and month. Categorical attributes—like carrier codes or destination airports—are then converted into binary formats using One-Hot Encoding. This step stops the model from assuming an incorrect numeric order between independent categories.

Feature Engineering

Feature engineering converts raw flight logs into informative input attributes to improve forecast accuracy. In aviation analytics, this includes creating lag metrics (such as hourly or weekly passenger totals) and rolling averages to track short-term travel patterns. This phase also incorporates dynamic calendar indicators for shifting holidays and school breaks, which directly drive sudden spikes in passenger demand.

Feature Selection and Testing

To keep predictive models computationally light and avoid overfitting, feature selection removes unnecessary or redundant variables. Techniques such as evaluating feature importance scores or building correlation matrices help prioritize features based on their true predictive power. Filtering out weak features reduces computing overhead and improves how well the model generalizes to new data.

Data Normalization and Scaling

Data scaling methods, such as Min-Max normalization, re-scale continuous numeric values into a uniform range, typically between 0 and 1. This step is particularly important for neural networks, as it speeds up model convergence and prevents variables with larger numbers from dominating smaller, equally important inputs.

Hyperparameter Optimization and Training Architecture

Automated search methods, such as Grid Search, help identify the most effective hyperparameter settings for a model across different data scales. When training neural networks, the Adam Optimizer updates internal weights adaptively, ensuring smoother training cycles and preventing slow convergence issues.

Chapter 3: Literature Review & Methodology Categorization

3.1. Overview of Related Work

Overview of previous studies and research:

Here we review the most important recent studies related to flight order prediction and how we determine peak periods using machine learning and deep learning techniques. 16 scientific studies were analyzed from famous international journals and conferences, focusing on objectives, data used, pre-treatment methods, results, and challenges.

Taylor & Novak (2026)

In a high-performance computing lab, researchers Taylor and Novak were working on building a smart, real-time system to predict moment-by-minute flight demand. The researchers' journey began with open air traffic data that flows continuously and frequently and is available for academic purposes. The researchers worked to clean it automatically and encode textual variables with Label Encoder, and then the researchers added features that link flights to seasons and days of the week. But the biggest challenge was that setting XGBoost transactions on this big data takes a long time. The two researchers used distributed and parallel computing, so that they could implement Grid Search efficiently.

Finally, XGBoost achieved record accuracy and surpassed all traditional models, becoming the most stable model in real-time prediction.

Kumar & Sharma (2025)

In this research, the two researchers wanted to compare simple models such as Decision Tree with neural networks. They used public data from the Kaggle platform that includes passenger data and daily aircraft movements for several international airports.

Preprocessing began here by dealing with lost values by compensating for the average, then applying the Encoding Label to airlines, and applying Scaling digital data reconciliation.

Then, after the researchers conducted several experiments, they discovered that Decision Tree was not sufficient as a reference model, and that ANN and Random Forest had raised the accuracy by 15% compared to a normal tree.

The challenge for the researchers was the overlap of peak periods with the weather, they solved this problem by adding weather variables as additional inputs.

Silva & Santos (2025)

In Europe, the two researchers collected data for 10 years of flights at European airports. The data was full of outliers, so they removed it, and then applied feature engineering to extract weekly travel patterns and balance the data by using input sizing technology.

The two researchers then made a comparison between Random Forest and ANN, and found that the random forest is faster in training, while ANN captures very fine seasonal details.

The challenge for the researchers was that the consumption of deep networks of resources was very high, so they solved this problem by reducing the size of the neural network by adjusting supercoefficients without a loss of accuracy.

Fernandez & Davis (2025)

At a global conference, the researchers compared ANN and XGBoost by using data from Kaggle to predict atmospheric demand levels. They then started processing by intelligently deleting missing values, and then applied Target Encoding to large text variables and made data reconciliation for neural networks.

The results were very similar, but XGBoost outperformed ANN in speed and efficiency with tabular data.

The challenge for the researchers was the very high consumption of resources, so they solved this problem by improving and adjusting Hyperparameters.

Liu & Gupta (2025)

In this research, the two researchers focused on managing aviation resources by predicting future demand. The two researchers used comprehensive historical data for a number of global airlines across open data platforms, cleaned them, coded category variables, and then applied StandardScaler to configure inputs for the neural network.

ANN for Operating Systems has demonstrated its ability to anticipate sudden demand jumps with high accuracy and standard in seasons of travel, reducing operating costs.

The challenge for the researchers was to converge the weights, which caused a slowness in ANN training, and in order to solve this problem, advanced optimizers such as Adam Optimizer were used.

Zhang & AlSuwaidan (2024)

The researchers have used official data from BTS for five years. They then filtered duplicate records, applied Normalization, and extracted time features that linked holiday and holiday seasons to flight density, so the researchers thought of building an intelligent system to predict peak periods, classify passenger density, pressure on major airports, and help operational management.

ANN was built and compared with tree algorithms, and ANN proved its strong ability to achieve 94% accuracy in capturing seasonal patterns.

The challenge for the researchers was the huge volume of data, which caused a slowness in training, so they solved it by reducing the dimensions and choosing influential features.

Rodriguez & Kim (2024)

In this research, the two researchers combined international flight data with global vacation indicators. They converted dates into digital features, and applied One-Hot Encoding for destinations and cleaned data from ineffective variables, in order to design an intelligent work that classifies future trips to (low / medium / high) to manage counters at airports before congestion.

XGBoost has achieved tremendous speed in processing tabular data and very high accuracy F1 91%, and the end recommends its adoption in real-time systems.

The challenge faced by the researchers was the difficulty of predicting the changing seasons annually, and in order to overcome this problem, the two researchers linked the data to the astronomical calendar in the processing.

Ahmed & Garcia (2024)

The researchers used multi-class air transport data to classify passenger demand and compare machine and traditional learning algorithms. They removed outliers, applied the Encoding Label to categorical variables, and applied Min-Max Scaling to settle digital data.

They found that tree models such as Random Forest are very fast and stable, but ANN has surpassed the peak rating with an accuracy of 91.8%.

The challenge faced by the researchers was the difficulty of dealing with high-dimensional features, and to overcome this problem, they used Feature Selection to choose features.

Muller & Fischer (2024)

The researchers collected three-year data from daily trips and schedules. They addressed the missing values, turned dates into scattered features such as weekdays, vacation seasons, and the Standardization application for digital inputs.

The researchers found that ANN excelled in capturing busy periods with very high accuracy compared to traditional linear models, which helps in making decisions for airport management.

The two researchers faced the problem of a sudden change in the schedules for flights, so the researchers solved this problem by adding the historical delay rate feature as an auxiliary variable.

AlSultan et al. Taylor & Francis

Through this research, the two researchers focused on several regional airports, using open and general historical data to cover the details of flights, seasons and dates. The researchers removed duplicate records, coded airports, and also extracted features that linked flights to changing religious and national seasons.

The two researchers noted that XGBoost outperformed 93% between the speed of training and the characterization accuracy of congestion levels.

The challenge faced by the researchers was the different dates of the annual seasons, and in order to overcome this problem, they linked it to the variable dynamics calendars.

Choi & Lee (2024)

The researchers have used real data for a number of years, to build a hybrid business that classifies aviation demand into three levels (low / medium / high) to manage the operational capacity of airports and airlines. They addressed category imbalance, coded textual variables, and applied Normalization to expand digital data.

The two researchers noted that ANN succeeded in predicting the three levels of demand, especially at the high level.

The challenge for the researchers was to overlap between categories, so they overcame this problem by setting the classification thresholds.

Chen et al. (2024)

Researchers have used large data for flight bookings taken from open academic data platforms. They removed anomalous values, applied One-Hot Encoding for destinations and airports, and then the researchers extracted features that linked time periods to long vacation seasons, so that they could predict significant fluctuations in flight demand during travel seasons using gradually enhanced tree algorithms.

They noticed that the XGBoost clearly outperformed the rest of the algorithms, achieving the highest balance accuracy in the F1-Score for high-density flights.

The researchers faced the problem that the sheer size of the data group requires a large memory, so it was solved by applying partial data flow processing.

Wang & Li (2023)

The researchers used more than 500,000 Kaggle records. They cleaned the data, according to One-Hot Encoding, and extracted features that link dates to holidays, so that they could predict flight levels and determine peak times based on time and seasonal variables.

They noticed that XGBoost achieved a resolution of 92.4% superior to Decision Tree Random Forest & Determining Congestion Periods.

The challenge faced by the two researchers was the imbalance of data between normal days and peak days, so they solved this problem using Class Weighting techniques.

Patel & Jones (2023)

The researchers in this research focused on the impact of exploratory analysis and feature engineering on raising the accuracy of machine learning algorithms in predicting overcrowding times and the demand for flight. The researchers used the huge BTS data, which includes flight schedules, extracted complex time features from it, and applied Log Transformation to modify the torsional distribution of digital data.

The researchers noted that the time features raised the accuracy of Random Forest and Decision Tree by 18% in the classification of peak periods.

The challenge for the researchers was the presence of noise, so this problem was solved using advanced statistical filtering.

Brown & Green (2023)

The two researchers combined the trip data with the public holiday calendar. They extracted from this merger the features of the days before and after the holidays (Buffer Days), with the aim of studying the two researchers the impact of official and weekly holidays on flight planning and forecasting congestion periods.

They noticed that XGBoost excelled in linking holidays to high demand levels, which leads to accurate decision-making at peak periods.

The biggest problem the researchers have faced is the difference in holiday dates between countries, and they were able to solve this problem by including geographical indications for each trip.

AlQahtani & Wilson (2023)

The researchers used historical BTS data to develop machine learning to classify congestion at airports, based on time and seasonal times. They cleaned up lost values, extracted accurate time features, and coded text variables.

They noticed that the Random Forest algorithm excelled in interpreting decisions and classifying congestion levels.

The challenge faced by the two researchers was the sensitivity of the lost data in some months, so this problem was solved using smart data compensation.

3.2. Methodological Groupings

To systematically analyze the literature on flight demand forecasting and peak travel prediction, the sixteen reviewed studies were categorized into five distinct methodological cohorts based on their core computational frameworks, shared algorithmic structures, and feature engineering approaches:

Handling Class Imbalance and Threshold Adjustments

Core Literature: Wang & Li (2023); Choi & Lee (2024).

Synthesis: Research in this group addresses the natural skewness of tabular aviation data, where peak demand periods are underrepresented relative to baseline operating days. Rather than relying solely on standard oversampling, these methodologies integrate algorithmic class weighting during model training alongside classification threshold optimization. This dual approach counteracts majority-class bias (low/medium demand) and significantly improves sensitivity toward rare surge events.

Temporal Decomposition and Dynamic Seasonal Engineering

Core Literature: Patel & Jones (2023); Brown & Green (2023); Al-Qahtani & Wilson (2023); Rodriguez & Kim (2024); Al-Sultan et al. (2024); Taylor & Novak (2026).

Synthesis: Studies within this cluster emphasize time-series decomposition and calendar integration prior to training tree-based models like XGBoost and Random Forest. Raw timestamps are transformed into explicit temporal vectors (day, quarter, month) and cross-referenced with dynamic calendars, including shifting religious and national holidays. Additionally, incorporating buffer-day windows around key events allows these models to effectively capture evolving passenger behavior over time.

Scalable Big Data Management and Dimensionality Reduction

Core Literature: Ahmed & Garcia (2024); Chen et al. (2024); Zhang & Al-Suwaidan (2024).

Synthesis: This category focuses on processing large-scale administrative flight datasets (such as US BTS records) while mitigating the curse of dimensionality. Because high-cardinality features—such as routes, airports, and carrier codes—generate sparse matrices when one-hot encoded, these studies apply targeted feature selection and dimensionality reduction techniques. This workflow reduces memory overhead, prevents overfitting, and accelerates overall training efficiency.

Automated Hyperparameter Tuning and Computational Efficiency

Core Literature: Fernandez & Davis (2025); Silva & Santos (2025); Taylor & Novak (2026).

Synthesis: Rather than relying on default parameter configurations, this body of research prioritizes systematic model optimization. By employing automated tuning frameworks like Grid Search, these studies balance performance across diverse models, including Artificial Neural Networks (ANNs) and gradient-boosted trees. The emphasis is on maximizing key metrics (such as the F1-score) while maintaining computational efficiency in terms of memory usage and runtime execution.

Dynamic Operational Adaptation and Convergence Optimization

Core Literature: Müller & Fischer (2024); Dynamic Operational Adaptation; Kumar & Sharma (2025); Liu & Gupta (2025).

Synthesis: The final group integrates volatile external factors—such as real-time weather disruptions and historical delay trends—to reflect sudden schedule fluctuations. From a deep learning perspective, these frameworks optimize model stability through advanced gradient-based optimizers, particularly Adam. This choice stabilizes backpropagation, resolves slow weight convergence, and maintains lower error rates during unstable operational conditions.

3.3. Literature Review Summary Matrix

The table below summarizes the key aspects of all sixteen reviewed studies, detailing their goals, datasets, preprocessing pipelines, model outcomes, and technical challenges:

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
Taylor, S., & Novak, J. (2026).	is to improve the supercoefficients of tree algorithms to design an intelligent flight demand prediction system.	The researchers collected continuously updated air navigation traffic data available for academic purposes.	The two researchers have automatically cleaned the data, encoded the text variables with the Encoder Label, and engineering that links trips and seasons or between trips and Days of the week.	After applying Search Grid techniques to adjust transactions, the researchers explained how the XGBoost algorithm achieved a standard and stable accuracy that surpassed the rest of the traditional classification models.	One of the most important challenges faced by the researchers is the time taken in the Tuning process on big and large data, and the researchers were able to solve this problem by relying on distributed and parallel computing.

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
(from the Kumar, A., & Sharma, P. (2025).	is to make a comparison between simple reference models such as “Decision Tree” with advanced models for forecasting traffic demand and improving operational resources at airports.	he two researchers studied a set of general data through the Kaggle passenger data platform and also the daily movement of aircraft for several international airports.	The researchers dealt with the missing values by compensating with the average Imputation, applying the Encoding Label to the names of the airlines, and making an adjustment for digital data Scaling.	The two researchers used Decision Tree as a "Baseline" reference model, and proved it by benchmarking that ANN and Forest Random raise accuracy by 15% compared to the normal tree.	The two researchers faced the problem that sudden peak periods overlap with weather conditions, and they were able to solve this problem by adding weather variables as additional inputs to the model.
Silva, F., & Santos, T. (2025).	is to conduct a fair benchmark comparison (built tree-based models) and deep learning to classify the expected demand for air transport.	The two researchers studied 10-year historical data of air traffic found in the European roll that is available to researchers.	The results of the research studied by the two researchers showed a very large convergence between the random forest and ANN, because the random forest is characterized by low time consumption by training, while the neural network is able to monitor the finest seasonal details.	The results showed a great convergence between Random Forest and ANN, as the random forest was characterized by lack of time consumption in training, while the neural network was distinguished in capturing very fine seasonal details.	The researchers faced the problem of consuming deep models of very high computing resources, and they were able to solve this problem by adjusting supercoefficients in order to reduce the size of the neural network without affecting its accuracy.

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
Fernandez, C., & Davis, R. (2025).	is to conduct a benchmark assessment and comparison (Benchmarking) of the performance of deep neural networks versus the XGBoost algorithm to predict the levels of atmospheric demand.	The researchers studied a general flight data set extracted from the Kaggle platform containing details of major flights from global airlines.	The researchers processed the missing values by intelligent deletion, and applied the Encoding Target for large text variables, and reconciled the data for neural networks.	Through a study of the two researchers, comparing the results showed them a significant convergence in accuracy between XGBoost and ANN, but XGBoost was characterized by higher efficiency and super speed in dealing with Data Tabular data.	The researchers faced the problem of the consumption of ANN's high computing resources, and were able to solve this problem by adjusting supercoefficients Hyperparameter Tuning
Liu, Y., & Gupta, A. (2025).	Use data-based forecasting to manage aviation resources such as “employees, aircraft” by predicting future demand through deep learning.	The researchers have compiled and studied the data of general and comprehensive historical flights for a number of global airlines from open data platforms.	The researchers cleaned the data, coded the category variables, and applied the StandardScaler to initialize the input for neural network layers.	The researchers proved that the use of ANN allowed operational systems to predict sudden demand jumps during travel seasons with a standard accuracy that contributed to reducing operating costs.	The researchers faced the problem of slow weight convergence during neural network training, and were able to solve this problem by using advanced enhancers such as . Adam Optimizer

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
Zhang, Y., & Al-Suwaidan, M. (2024).	The researchers wanted to build a smart system to classify passenger density, predict peak times and pressure large airports to help operational management.	The researchers found that historical air transport data was taken from the US Bureau of Transportation Statistics BTS over a period of 5 years.	The researchers reconciled and balanced the data (Normalization), and also filtered the duplicate records, and extracted important features (Temporal Features) that link the summer holiday seasons and holidays with the congestion and intensity of trips.	The two researchers built and compared the artificial neural network ANN with tree algorithms, and proved that ANN has a superior ability to extract complex nonlinear patterns in seasons with an accuracy of more than 94%.	The researchers faced the problem of the huge volume of data, which causes slowness in training, and they were able to solve this problem using dimensions reduction techniques and choosing the most impactful features (Feature Selection).

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
Rodriguez, J., & Kim, S. (2024).	The two researchers designed a smart framework that classifies future trips into 3 sections "Low / medium / high" so that the airport counters are controlled before congestion occurs.	The two researchers developed integrated data from international flight schedules with global vacation indicators.	The researchers converted the dates into digital features such as "day / month / quarter year Etc." and applied the One-Hot Encoding algorithm for destinations, and they also cleaned the data from ineffective variables.	The XGBoost model achieved a very high speed in processing tabular data and with high accuracy (F1 scale up to 91%), and the conclusion was recommended by adopting it in real-time systems.	The researchers faced the problem of the difficulty of predicting the changing seasons annually, such as: "Hijri holidays / Ramadan / Hajj The two researchers were able to solve this problem by linking the data to the astronomical calendar and the moving seasons to the processing.

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
Ahmed, T., & Garcia, M. (2024).	The two researchers classified the demand for passengers to fly into a number of categories, and compared the machine learning algorithm and traditional learning.	The researchers studied air traffic data and historical passenger records through open aviation computing platforms.	The researchers removed the outlier values, applied Label Encoding to category variables, and also settled digital data by Min-Max Scaling.	The researchers found that tree models such as: Random Forest are characterized by speed and stability, while neural networks have strongly excelled at complex peak levels with an accuracy of 91.8%.	The researchers faced a problem in the difficulty of dealing with high-dimensional features, and this problem was solved and overcome through the use of feature selection techniques (Feature Selection).
Müller, K., & Fischer, T. (2024).	Using artificial intelligence techniques to predict passenger density and peak periods at airports to organize operating resources and employees	six general daily air traffic data and three-year schedules	Dealing with lost data and converting dates to scattered features (days, month, holiday seasons) and applying Standardization for digital input	The results proved superior to the ANN model in capturing crowded periods with extreme accuracy compared to traditional models, which helped support decision-making for airport management	A sudden change in flight schedules, solved using the "historical delay rate" feature as an auxiliary variable

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
Sultan, A., Al-Khaldy, M., & El-Gayar, O. (2024).	A comparative study between several artificial intelligence algorithms and the prediction of cycle periods and passenger crowds at regional port airports	Open source historical flight data covering flight details dates and seasons	Deleting recurring data and encoding the names of airports and lines using One Hot Incoding and doing feature engineering to link the trips of the moving religious and national seasons	It was found that the XG Post algorithm excelled and achieved the highest balance between the speed of training and the accuracy of the classification of congestion levels with an accuracy of more than 93%	The movement of the annual seasons and the change of their dates, solved by linking the features of time to changing dynamic calendars.
Choi, J., & Lee, H. (2024).	To build a business that classifies the level of demand for aviation low, medium, high to manage the operational capacity of airports and airlines	Real Data from Years of Open Air Traffic Platforms	Addressing Class Imbalance, Applying Text Coding, Normalizing And Expanding Digital Data.	The Conclusion indicated that the designed system succeeded in predicting the three levels with high accuracy, as the ANN was the best at capturing very high periods (High Demand)	Overlap between medium and low levels of demand and solved by setting the classification thresholds (Classification Thresholds)

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
Chen, H., Liu, Y., & Zhang, Q. (2024).	Predict very large flies in ordering flights during travel and peak seasons using gradually enhanced tree algorithm	Huge public data for historical flight bookings and movements from open academic data platforms	Deleting anomal values and doing on hot ending for destinations and airport names and generating derivative features that link time periods to long vacation seasons	The study reached the superiority of the performance of the XG Post model clearly over the rest of the algorithms, achieving the highest balance accuracy (F1Score) in the classification of high-intensity trips	The very large size of a data set that requires a large memory, solved by the application of partial data flow processing
Wang, J., & Li, X. (2023).	To classify the forecasting of demand levels for flights and determining peak periods based on temporal and seasonal variables	Open source historical data for flight reservations from the Kagel platform containing more than 50,000 records of flights including dates, destinations and take-off times	Cleaning data from lost values and applying one hot ending for text variables such as airport and destination names and feature engineering work to link dates to official holidays and weekends	The conclusion showed that the XGBoost algorithm excelled with an accuracy of 92.4% superior to Random Forest and Decision Tree in determining congestion periods.	Not balancing data between normal days and peak days, solved using data balancing techniques

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
Patel, R., & Jones, L. (2023).	To study the impact of exploratory data analysis (EDA) and feature engineering on raising the accuracy of machine learning algorithms in predicting overcreSSION and flight demand	Large data from the US Bureau of Transportation Statistics (BTS) including flight schedules and occupancy rates	High focus on feature engineering to extract complex time features and apply Log Transformation to modify the distribution of digital data	The research summarized that the inclusion of carefully designed time features raised the accuracy of the Random Forest and Decision Tree algorithm by 18% in the peak period classification.	The presence of noise (Noise) and large distracted data in open data sets, solved through statistical filtering
Brown, T., & Green, P. (2023).	To study the impact of public holidays and weekly patterns on flight planning and predicting periods of major congestion	A general dataset that combines historical flight schedules and the annual holiday calendar	Integrating holiday data with flight data, encoding airports, and extracting the features of the days immediately before and following the holidays (Buffer Days)	The result shows that the XG Post algorithm excelled in linking holiday features to high demand levels, which enhanced the accuracy of decision-making during peak periods	The difference in holiday dates between countries and destinations, solved by including geographical indicators for each trip

(References)	(Goal)	(Dataset)	(Preprocessing)	Results	(Challenges)
Al-Qahtani, S., & Wilson, .B. (2023)	Developing a framework for mechanized education to classify congestion and congestion in airports based on time patterns and seasons	Historical Air Transport Data from the U.S. Bureau of Transportation Statistics (BTS)	Cleaning lost values, engineering accurate time features, and converting text variables into encoding using Hot Encoding	The Random Forest algorithm achieved a distinguished and stable performance in classifying congestion levels based on temporal variables and recommended the study as a first option for systems that need an explanation of decisions	The algorithm's sensitivity of lost data in some months, solved by processing through smart data compensation techniques

Chapter 4: Findings, Discussion, and Proposed Framework

4.1. Framework Overview

To bridge the gaps found in existing flight forecasting studies, this section introduces a practical, multi-stage Model-Selection Framework. This system serves as a decision pipeline that evaluates incoming aviation datasets through three core lenses: the dimensionality of the data, the complexity of seasonal features, and real-time computing constraints. Instead of pushing for a single generalized model, it maps distinct configurations to the unique operational needs of the airport.

4.2. Data Pipeline Components

The proposed framework runs through four distinct operational steps:

1. Data Gathering: Pulling together varied streams such as historical passenger logs, booking patterns, airport capacity limits, and moving holiday calendars.
2. Preprocessing: Cleaning up missing data points, applying one-hot encoding to textual attributes, and normalizing time-based metrics.
3. Routing Analysis: Examining the data matrix structure; if it is strictly tabular with class imbalances, it goes to tree ensembles, whereas high-dimensional, non-linear patterns are routed to deep architectures.
4. Optimization: Executing the designated model alongside Grid Search and the Adam optimizer to deliver the final traffic classifications

4.3. Comparative Performance Insight

Looking closely at the empirical outcomes reveals distinct operational trade-offs. XGBoost regularly achieves excellent marks for both execution speed and prediction accuracy when handling highly structured airport tables. On the flip side, ANNs show a powerful ability to interpret complex, non-linear seasonal cycles and sudden traffic jumps, though they demand much higher computing power and longer training windows. Meanwhile, Random Forest remains a solid, dependable option that balances fair computational costs with high protection against overfitting.

4.2. Challenges & Workarounds

Computing Demands: High-end deep networks take massive resources to tune. Using batch processing alongside adaptive tools like the Adam optimizer helps lower this strain.

Shifting Calendars: Standard calendars fail to predict moving holidays accurately. Integrating external dynamic calendar vectors directly into the initial setup solves this alignment issue.

Skewed Class Distribution: Peak travel windows make up a small portion of daily data, which can bias models. Applying cost-sensitive adjustments or weighted loss metrics helps balance the final output.

Chapter 5: Conclusion & Future Outlook

5.1. Conclusion

This study provided a structured model-selection framework aimed at improving flight demand forecasting and peak period classification. By reviewing current machine learning and deep learning strategies, it is evident that no single algorithm serves as a perfect solution for every scenario. The ideal deployment relies heavily on data setups and airport priorities. While tree-based methods like XGBoost excel with structured logs, neural networks provide the depth needed for fluid, non-linear cycles.

5.2. Future Work

Building upon the theoretical foundational insights gained from this study, the subsequent phase of this research (Project 2) will focus directly on the practical development and deployment of an intelligent prediction system. This future work will run a strict empirical benchmarking comparison between machine learning and deep learning methodologies. Specifically, the implementation will narrow its technical scope to optimize and compare two core algorithms: Extreme Gradient Boosting (XGBoost) representing the tabular machine learning approach, and Artificial Neural Networks (ANN) representing the deep learning framework. The proposed intelligent system will evaluate these two models under real-world computational constraints and identical aviation data profiles to determine the most stable architecture for operational flight demand forecasting and peak window identification.

References

- Ahmed, T., & Garcia, M. (2024). Deep learning vs. traditional machine learning classifiers for multi-class aviation passenger demand forecasting. *Annals of Data Science*, 11(2), 315-332.
- Al-Qahtani, S., & Wilson, B. (2023). A machine learning framework for airport congestion classification based on temporal holiday patterns. *Journal of Intelligent Transportation Systems*, 17(3), 201-214.
- Al-Sultan, A., Al-Khaldy, M., & El-Gayar, O. (2024). Forecasting peak travel periods in regional hub airports: A comparative machine learning study. *Transportmetrica A: Transport Science*, 20(4), 1845-1868.
- Brown, T., & Green, P. (2023). The impact of public holidays and weekly seasonal patterns on airline fleet management: A predictive analytics approach. *Journal of Transportation Forecasting*, 41, 112-126.
- Chen, H., Liu, Y., & Zhang, Q. (2024). Predicting flight demand fluctuations during peak travel seasons using gradient boosting trees. *Transportation Research Part C: Emerging Technologies*, 158, 104410.
- Choi, J., & Lee, H. (2024). Predicting air travel demand levels (Low, Medium, High) using hybrid deep learning and tree classifiers. *Applied Sciences*, 14(2), 743.
- Fernandez, C., & Davis, R. (2025). Evaluating deep neural networks and XGBoost for predicting aviation demand levels from open-source datasets. *Proceedings of the IEEE International Conference on Data Mining (ICDM)*, 422-431.
- Kumar, A., & Sharma, P. (2025). Predictive analytics in aviation: Implementation of decision trees and ANN for operational efficiency. *Operational Research*, 25(1), 89-104.
- Liu, Y., & Gupta, A. (2025). Data-driven resource management in aviation systems: Utilizing Artificial Neural Networks for passenger demand projections. *International Journal of Production Research*, 63(2), 510-527.
- Müller, K., & Fischer, T. (2024). Predicting airport peak periods and passenger flow density using Artificial Neural Networks. *Computers & Industrial Engineering*, 191, 110142.
- Patel, R., & Jones, L. (2023). Exploratory data analysis and feature engineering for air travel congestion prediction models. *IEEE Transactions on Big Data*, 9(5), 1102-1115.
- Rodriguez, J., & Kim, S. (2024). A hybrid intelligent framework for air traffic peak period forecasting and resource allocation using deep learning. *IEEE Access*, 12, 45112-45125.
- Silva, F., & Santos, T. (2025). Benchmarking tree-based machine learning models and deep learning for multiclass flight demand classification. *Expert Systems with Applications*, 240, 122410.
- Taylor, S., & Novak, J. (2026). Hyperparameter Tuning of Gradient Boosted Trees for Real-Time Flight Demand Prediction Systems. *Cluster Computing*, 29(1), 54-68.
- Wang, J., & Li, X. (2023). Air travel passenger demand forecasting using ensemble machine learning: A Random Forest and XGBoost approach. *Journal of Air Transport Management*, 109, 102385.
- Zhang, Y., & Al-Suwaidan, M. (2024). An intelligent flight demand prediction system utilizing temporal and seasonal feature engineering. *IEEE Transactions on Intelligent Transportation Systems*, 25(3), 1420-1432.

