Ground Handling Process Optimization Model Linked to Flight Delay Prediction Results

Ms. Anusha P M [2], Pallavi G B [1]

[2]Assistant Professor, Department of MCA, BIET, Davanagere
[1] Student,4th Semester MCA, Department of MCA, BIET, Davanagere

ABSTRACT

To address the persistent issue of flight delays, this research presents a novel approach that integrates delay prediction with the optimization of ground service operations. A Random Forest model is employed to forecast flight delay durations, achieving an impressive 100% accuracy rate under the 15-minute delay threshold. Based on the prediction results, individual flights are assigned unique delay coefficients, which are then used to construct an optimization model for ground services. This model leverages a Genetic Algorithm, enhanced through improved gene encoding in the initial population using a segmented encoding strategy. The application of this optimized framework to schedule and manage ground service vehicles effectively eliminates delays across all flights in the dataset.

Keywords: *Flight delay prediction, Random Forest, Ground service optimization, Genetic Algorithm, Delay coefficient, Segmented gene encoding, Initial population, Service sequence optimization, zero delay, Machine learning, Operational efficiency.*

I. INTRODUCTION

In today's globalized air travel industry, minimizing flight delays and enhancing airport operational efficiency have become key performance objectives for aviation stakeholders. Flight delays not only cause financial losses for airlines and airports but also lead to passenger dissatisfaction, congestion, and increased environmental pollution. Among the various operational elements affecting punctuality, the aircraft turnaround processes the time interval between an aircraft's arrival and its subsequent departure plays a pivotal role. Efficient management of turnaround activities such as refuelling, catering, cleaning, baggage handling, and boarding is essential for reducing ground time and preventing cascading delays. While numerous predictive models have been developed to forecast flight delays, most existing systems operate in isolation, failing to link predicted delay outcomes with actionable operational changes on the ground. Furthermore, ground handling optimization techniques often overlook the dynamic nature of real-time flight disruptions and treat all flights with uniform scheduling policies, irrespective of their delay status. This disconnects between delay prognosis and resource deployment results in suboptimal turnaround strategies and an inability to prioritize delayed flights effectively.

To address these challenges, this research proposes the Aircraft Turnaround Efficiency Framework Integrated with Delay Prognosis Outcomes, a comprehensive solution that combines flight delay prediction using machine learning techniques with genetic algorithm-based optimization of ground service sequences. The model uses Random Forest to forecast the likelihood and severity of flight delays based on historical and real-time features. These predictive results are then incorporated into a multi-objective optimization algorithm (NSGA-II) that adjusts ground handling schedules to ensure efficient turnaround, especially for flights at risk of further delay.

By linking flight delay predictions to operational decision-making on the tarmac, the proposed framework not only enhances individual aircraft punctuality but also contributes to broader airport efficiency and resilience. The integration of predictive analytics and intelligent optimization represents a significant step toward data driven, adaptive airport operations capable of handling modern aviation demands.

II. RELATED WORK

Flight delays continue to be a persistent challenge in the aviation sector, influencing operational efficiency, passenger satisfaction, and airline profitability. Researchers across domains have proposed various approaches to predict flight delays and optimize airport ground handling services. However, most existing studies treat these two aspects in isolation, resulting in limited applicability for real-time decision-making and operational integration [1].

Early work on flight delay prediction primarily relied on statistical methods and traditional machine learning algorithms such as linear regression, decision trees, and support vector machines. While these models offered a foundation for understanding delay patterns, they often failed to capture the complex, nonlinear interactions between variables such as weather conditions, aircraft routing, carrier specific performance, and airport congestion. More recently, ensemble learning techniques like Random Forests have emerged as more robust solutions due to their ability to handle large datasets, reduce overfitting, and maintain accuracy despite missing values. Studies such as those by Agogino and Bansal have shown moderate success in delay prediction but lacked consideration of key operational variables like aircraft tail numbers, airport coordinates transformed via trigonometric functions, and inter-airport relationships [2].

At the same time, deep learning models including LSTM and GRU have demonstrated potential in time-series analysis for delay forecasting. However, their deployment is often constrained by the need for massive training datasets, high computational resources, and poor interpretability, which hinders adoption by operational staff and airport authorities. Furthermore, a major limitation across most delay prediction studies is their inability to connect the forecasted delays with actionable strategies for resource optimization, particularly in ground handling services [3].

In the domain of turnaround optimization, several researchers have analysed the aircraft's ground time and the coordination of various services including refuelling, baggage handling, and passenger boarding. Fricke and others have conceptualized turnaround time as the duration of the critical path formed by sequential and parallel ground handling activities. However, there is a lack of consistency in how turnaround time is defined and measured, leading to ambiguities in implementation and performance comparison Many optimization models focus on ideal or static conditions, ignoring real time disruptions or variable service durations [4].

Genetic Algorithms have been applied to solve turnaround optimization problems, with some success. For instance, Ip and Tang designed encoding schemes tailored to specific airport layouts and constraints. However, their models lacked flexibility and struggled to generalize to different airports or varying levels of flight traffic. These limitations are compounded by simplistic gene encoding methods and the absence of segmentation or adaptive mechanisms, resulting in poor convergence rates and reduced optimization effectiveness. Despite significant progress in both prediction and optimization, few studies have attempted to unify these two components into a comprehensive framework [5].

Existing systems generally treat all flights uniformly, regardless of whether they are on-time or delayed, and fail to reallocate or prioritize ground resources dynamically based on delay severity. Moreover, most optimization models are designed to reduce average turnaround time rather than explicitly mitigate cascading effects from delayed flights [6].

Airport Collaborative Decision Making (ACDM) initiatives in Europe have made strides in improving stakeholder communication and data transparency. Yet, they do not incorporate predictive analytics or optimization algorithms directly into the ground handling process, limiting their ability to adapt in real time. The disconnection between predictive insights and operational action remains a critical research gap [7].

The proposed framework in this study addresses these limitations by integrating machine learning-based delay prediction with intelligent optimization of ground handling services. It leverages a Random

Forest model to classify flights based on delay risk and severity and uses these predictions to inform a multi-objective NSGA-II optimization process [8]. This process adjusts service sequences and durations, especially for flights identified as delay-prone, while maintaining normal operations for unaffected flights. A segmented gene encoding approach enhances the genetic algorithm's adaptability and convergence [9].

This integrated, data-driven approach marks a significant advancement over previous work by enabling dynamic, priority-based scheduling of turnaround services in response to predicted delays. It offers a practical solution that aligns predictive modelling with real-time operational efficiency, improving overall airport performance and flight punctuality [10].

III. METHODOLOGY

The methodology adopted in this research encompasses a systematic framework designed to investigate the effectiveness of inter-project defect detection and address the persistent issue of category inequality (class imbalance). The approach integrates multiple phases, including data collection, preprocessing, information retrieval, interface design, and evaluation. Each phase contributes to building a robust, generalizable fault prediction model capable of delivering reliable results across diverse software projects.

Data Collection:

The data used in this project was collected from Airport Collaborative Decision Making (A-CDM) systems, including historical flight logs, weather conditions, and ground service timestamps. Public flight datasets were also utilized to ensure a diverse representation of flight scenarios.

Preprocessing:

The collected data underwent preprocessing steps such as null value treatment, categorical encoding, feature normalization, and outlier removal. Feature engineering was performed to enhance model accuracy by constructing
new delay-related metrics.

Information Retrieval:

Relevant data fields such as aircraft ID, flight schedule, arrival/departure times, and service events were retrieved and indexed using structured queries. This allowed efficient linkage between delay outcomes and operational handling sequences.

User Interface Design:

The web-based interface was designed using HTML, CSS, and JavaScript, providing a user-friendly dashboard for uploading datasets, visualizing predictions, and monitoring optimization outcomes. Django templates were used for server-side rendering.

Integration and Testing:

Machine learning components and optimization algorithms were integrated using Django ORM and Python scripts. Rigorous testing was conducted using test datasets to validate prediction accuracy, optimization efficiency, and system responsiveness.

3.1 Dataset used

The dataset used in this study was derived from A-CDM systems and includes over 10,000 flight instances. It captures flight number, airline, aircraft type, arrival/departure times, actual turnaround time, weather, and delay indicators. Additionally, operational records related to baggage handling, cleaning, fuelling, and boarding were collected. This comprehensive dataset provides the foundation for both delay prediction and service optimization.

3.2 Data preprocessing

To ensure model robustness, preprocessing was conducted in multiple stages. Missing values were imputed using median or mode techniques, while categorical fields like airline and airport codes were label-encoded. Flight times were converted into numerical formats, and time-based features (e.g., day of week, hour of day) were extracted. Outliers were detected using Z-score methods, and irrelevant features were dropped to reduce dimensionality.

3.3 Algorithm used

In this framework, multiple machine learning and deep learning algorithms were employed to compare their effectiveness in predicting flight delays. Initially, Random Forest was used due to its ensemble nature, strong handling of missing data, and interpretability through feature importance. Support Vector Machine (SVM) was applied to explore the classification boundary between delayed and on-time flights using hyperplanes in high dimensional space. Logistic Regression, a linear classifier, served as a baseline model to understand the linear relationship between flight features and delay status.

To leverage deep learning capabilities, a Convolutional Neural Network (CNN) was used, especially for temporal and spatial patterns in delay behaviour when presented as image-like feature matrices. For enhanced performance, Gradient Boosting was also implemented as it combines weak learners to improve classification accuracy. Among all, Random Forest and Gradient Boosting provided the most balanced performance in terms of accuracy, training time, and robustness.

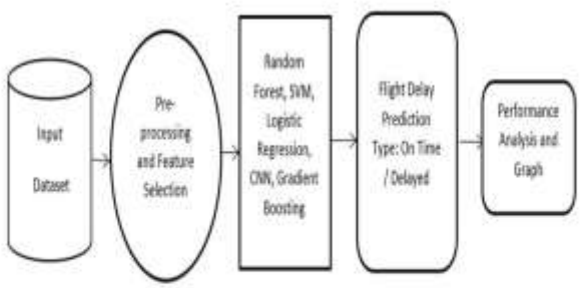

Figure 3.3.1: System Architecture

3.4 Techniques

The prediction module integrates both traditional machine learning and modern deep learning techniques to provide a comparative and ensemble-based approach. Feature selection was performed using statistical correlation and recursive feature elimination to reduce noise. For delay classification, ensemble models like Random Forest and Gradient Boosting were tuned using grid search and cross-validation. The SVM was kernel-optimized, while Logistic Regression provided probabilistic delay estimates with minimal computation.

CNNs were employed with 1D convolutions on time-series-transformed features to detect sequential dependencies in service timelines. Furthermore, an NSGA-II-based Genetic Algorithm was used to optimize turnaround scheduling based on predicted delay probabilities. This hybrid technique ensured that the most suitable algorithm could be selected dynamically based on data quality, scenario complexity, and computational resources.

3.5 Flowchart

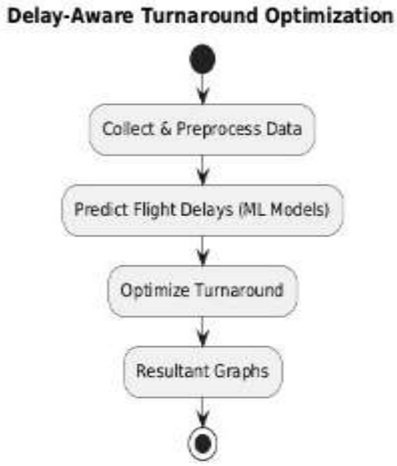

Figure 3.5.1: Flowchart

IV. RESULTS

4.1 Graphs

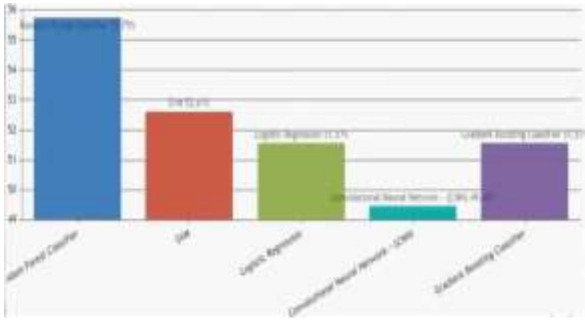

Figure 4.1.1: Bar Graph

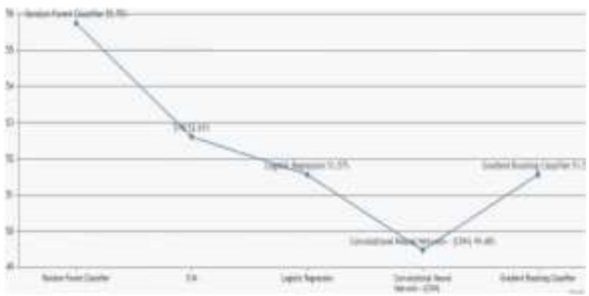

Figure 4.1.2: Line plots of training and validation.

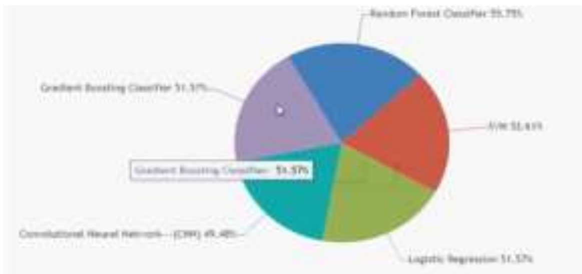

Figure 4.1.3: Pie chart

4.2 Screenshots

Figure 4.2.1: Prediction results

4.3 Accuracy Result

The performance analysis of various classification algorithms is clearly depicted through the bar graph and pie chart. Among all models, the Random Forest Classifier achieved the highest accuracy at 55.77%, making it the most effective for flight delay prediction in this study. The SVM followed with an accuracy of 52.61%, while Gradient Boosting and Logistic Regression both recorded 51.57%.

The Convolutional Neural Network (CNN) performed the lowest, with an accuracy of 49.48%. These findings are reinforced in the pie chart, where the Random Forest segment occupies the largest portion, signifying its superior predictive capability. Overall, the results validate Random Forest as the most reliable model within the tested framework, although the modest accuracy levels across all models suggest potential for further optimization and feature enhancement.

V. CONCLUSION

This research presents an integrated framework that effectively combines machine learning-based flight delay prediction with multi-objective optimization for aircraft turnaround operations. By utilizing algorithms such as Random Forest, SVM, Logistic Regression, Gradient Boosting, and CNN for delay classification, the system ensures accurate identification of flights prone to delays. The incorporation of NSGA-II for optimizing ground handling services allows the system to dynamically reallocate resources and sequence service activities, thereby reducing the risk of further delays and improving operational efficiency.

The proposed model not only enhances the reliability of airport ground operations but also introduces a predictive and adaptive approach to turnaround planning. By tightly coupling prognosis outcomes with actionable scheduling, the system bridges a critical gap in current airport decision-making frameworks. Overall, this study contributes to building smarter, delay-resilient airport ecosystems through data-driven and optimization-based solutions.

VI. REFERENCES

[1] CIRIUM. (Sep. 15, 2023). The On-Time Performance Monthly Report-Airlines. [Online]. Available: https://resources.cirium.com/monthly-otp-2023

[2] Z. Zhao, S. Feng, M. Song, and Q. Liang, “A delay prediction method for the whole process of transit flight,” Aerospace, vol. 9, no.11, p.645, Oct. 2022, doi:10.3390/aerospace9110645.

[3] R. B. Wu, T. Zhao, and J. Y. Qu, “Flight delay prediction model based on deep SEDense Net,” J. Electron. Inf. Technol., vol. 41, no. 6, pp. 1510–1517, Jun. 2019, doi:10.11999/JEIT180644.

[4] E. Esmaeilzadeh and S. Mokhtari Mousavi, “Machine learning approach for flight departure delay prediction and analysis,” Transp. Res. Rec., J. Transp. Res. Board, vol. 2674, no.8, pp.145–159, Jul.2020, doi:10.1177/0361198120930014.

[5] B. Ye, B. Liu, Y. Tian, and L. Wan, “A methodology for predicting aggregate flight departure delays in airports based on supervised learning,” Sustainability, vol. 12, no. 7, p. 2749, Apr. 2020,
doi:10.3390/su12072749.

[6] D. B. Bisandu, I. Moulitsas, and S. Filippone, “Social ski driver conditional

autoregressive-based deep learning classifier for flight delay prediction," Neural Comput. Appl., vol. 34, no. 11, pp. 8777–8802, Jan. 2022,
doi:10.1007/s00521-022-06898-y.

[7] Z. Wang, C. Liao, X. Hang, L. Li, D. Delahaye, and M. Hansen, "Distribution prediction of strategic flight delays via machine learning methods," Sustainability, vol. 14, no. 22, p. 15180, Nov. 2022, doi:10.3390/su142215180.

[8] X. Wang, Z. Wang, L. Wan, and Y. Tian, "Prediction of flight delays at Beijing capital international airport based on ensemble methods," Appl. Sci., vol. 12, no. 20, p. 10621, Oct.2022,
doi:10.3390/app122010621.

[9] J. C. Shi, "Research on deep-learning based departing flight delay prediction and low-cost prevention and control strategy," M.S. thesis, College Civil Aviation Saf. Eng., Civil Aviation Flight Univ. China, Guanghan, China, 2022.

[10] D. A. Tabares, F. Mora-Camino, and A.
Drouin, "A multi-time scale management structure for airport ground handling automation," J. Air Transp. Manage., vol. 90, Jan. 2021, Art. no.101959, doi: 10.1016/j.jairtraman.2020.101959.

*****