International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

Intelligent Flight Delay Prediction Using Machine Learning

P. Anusha[1], Syed Mannan Uddin[2], T.Sree Chandana[3], V.Vaishnavi[4]
[1]Assistant Professor, Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Hyderabad, Telangana, India
[2,3,4]Student, Department of Computer Science and Engineering, Guru Nanak Institute of Technology, Hyderabad, Telangana, India

Abstract- Accurate flight delay prediction is fundamental to establish the more efficient airline business. Recent studies have been focused on applying machine learning methods to predict the flight delay. Most of the previous prediction methods are conducted in a single route or airport. This paper explores a broader scope of factors which may potentially influence the flight delay, and compares several machine learning-based models in designed generalized flight delay prediction tasks. To build a dataset for the proposed scheme, automatic dependent surveillance broadcast (ADS-B) messages are received, pre-processed, and integrated with other information such as weather condition, flight schedule, and airport information. The designed prediction tasks contain different classification tasks and a regression task. Experimental results show that long short-term memory (LSTM) is capable of handling the obtained aviation sequence data, but overfitting problem occurs in our limited dataset. Compared with the previous schemes, the proposed random forest-based model can obtain higher prediction accuracy (90.2% for the binary classification) and can overcome the overfitting problem.

Keywords- Long short-term Memory, Deep Learning, Machine Learning.

I. INTRODUCTION

Air transportation plays an important role in connecting people and goods around the world. With the rapid increase in air traffic, flight delays have become a common problem that affects airline operations, increases costs, and causes inconvenience to passengers. Flight delays can happen due to several reasons such as weather conditions, air traffic congestion, technical issues, and airport operational problems . Traditional methods of predicting flight delays mainly rely on past records and manual analysis, which may not always provide accurate results because delay factors are complex and continuously changing. Today, large amounts of aviation data are available through technologies like Automatic Dependent Surveillance–Broadcast (ADS-B), weather systems, and airline databases, which provide useful information for analysis . Machine Learning, a part of Artificial Intelligence, helps in analyzing large datasets and finding patterns that influence flight delays.

By using historical flight data, machine learning models can predict delays more accurately and help airlines make better operational decisions . Therefore, Intelligent Flight Delay Prediction Using Machine Learning is an effective approach to improve prediction accuracy, reduce delays, and enhance the overall efficiency of air transportation systems.

II. LITERATURE SUVEY

Sarah Ahmed A et. al.,: This recent work evaluates different machine learning algorithms for enhanced flight delay prediction. The study compares models such as Random Forest, Decision Tree, and other classification techniques to identify the most effective approach. Results show improved prediction accuracy, supporting better airline scheduling, passenger management, and operational efficiency.

G.W.I. Dilshani, Muditha Tissera:This study compares machine learning and deep learning models for airline delay prediction. Various predictive models were evaluated based on their accuracy and performance. The findings show that advanced ML and DL models significantly enhance prediction capability, helping airlines reduce operational uncertainty.

Min Dai:This paper proposes a hybrid machine learning-based model for predicting flight delays through aviation big data. The method combines feature selection techniques with hybrid

ML algorithms to improve prediction performance. Results demonstrate higher accuracy and better handling of large-scale aviation datasets compared to traditional approaches.

Swati Dhadake et al.: This study presents a machine learning approach for predicting flight delays using weather conditions and flight schedule data. Different ML algorithms were analyzed to improve prediction accuracy. The results show that considering environmental and operational factors helps airlines predict delays more effectively and improve scheduling decisions.

III. EXISTING SYSTEM

Existing flight delay prediction systems mainly use historical flight data, weather information, and airport traffic details to estimate possible delays. Traditional models use machine learning techniques such as Linear Regression, Decision Trees, and Support Vector Machines for prediction. Recent systems also use deep learning models like RNN to improve accuracy by analyzing large amounts of flight data. Some advanced systems use real-time aviation data from Automatic Dependent Surveillance–Broadcast (ADS-B) to track aircraft movement and improve delay prediction. However, flight delays are influenced by many factors such as weather changes, route congestion, technical issues, and airport operations, making accurate prediction difficult.

Limitations of Existing System

- Depends mostly on historical data and may fail in unexpected situations.
- Does not fully consider real-time flight conditions.
- Important factors like air route traffic and operational changes are often ignored.
- Deep learning models require large amounts of data for better accuracy.
- Prediction accuracy may decrease due to complex delay causes and limited data availability.

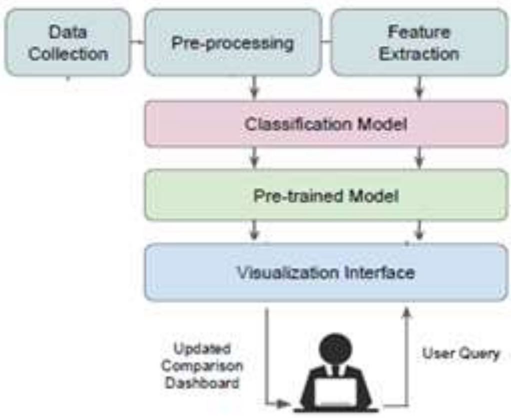
IV. PROPOSED SYSTEM

We explore a broader scope of factors which may potentially influence the flight delay and quantize those selected factors. Thus, we obtain an integrated aviation dataset. Our experimental results indicate that the multiple factors can be effectively used to predict whether a flight will delay . Several machine learning based-network architectures are proposed and are matched with the established aviation dataset. Traditional flight prediction problem is a binary classification task. To comprehensively evaluate the performance of the architectures, several prediction tasks covering classification and regression are designed .Conventional schemes mostly focused on a single route or a single airport. However, our work covers all routes and airports which are within our ADSB platform.

Advantages Oof Proposed System

- Considers multiple delay factors.
- Uses airport information.
- Includes weather data.
- Considers airport traffic flow.
- Includes route traffic flow.
- Uses an integrated aviation dataset.
- Random Forest accuracy: 90.2%.
- Demonstrates strong ensemble learning performance.
- Improves flight delay prediction accuracy

V. SYSTEM ARCHITECTURE



The architecture of Intelligent Flight Delay Prediction Using Machine Learning begins with data collection, where historical flight information such as departure and arrival times, weather conditions, airline details, airport traffic, and previous delay records are gathered from different sources. This raw data is then sent for pre-processing, where missing values, duplicate records, and irrelevant information are removed, and the data is cleaned and converted into a suitable format for analysis. After preprocessing, feature extraction is performed to identify important factors that influence flight delays, such as weather, departure delay, route distance, airport congestion, and airline performance.

These extracted features are given as input to the classification model, where machine learning algorithms like Decision Tree, Random Forest, or Logistic Regression are used to analyze patterns and classify whether a flight is likely to be delayed or on time. The model is then trained as a pre-trained model using historical datasets, enabling it to make accurate predictions for new flight information. Finally, the results are displayed through a visualization interface, where users can enter flight details as queries and view updated prediction results, delay comparisons, and dashboard insights, helping passengers and airline authorities make better operational decisions.

VI. METHODOLOGY

Modules

- **Data Collection:** In this module, flight-related data is collected from available sources. It includes details such as flight number, departure time, arrival time, weather conditions, airline information, and delay records.
- **Dataset:** The collected data is stored in a structured dataset. This dataset contains all the necessary input features required for training the machine learning model.
- **Dataset Cleaning:** In this step, unwanted or incorrect data is removed. Missing values, duplicate records, and inconsistent entries are handled to improve the quality of the dataset.
- **Data Preparation:** The cleaned data is converted into a suitable format for machine learning. This includes feature selection, data transformation, and splitting the dataset into training and testing data.
- **Model Selection:** A suitable machine learning algorithm is selected for prediction, such as Random Forest, Decision Tree, or Logistic Regression, based on performance and accuracy.
- **Analysis & Prediction:** The selected model analyzes the input data patterns and predicts whether a flight will be delayed or arrive on time.
- **Accuracy Evaluation:** The performance of the model is tested using evaluation metrics such as accuracy, precision, recall, and confusion matrix to check how well the model predicts delays.
- **Saving Model:** After achieving good accuracy, the trained model is saved so it can be reused later for predicting flight delays with new input data.

VI. IMPLEMENTATION

The flight delay prediction system starts by collecting historical flight data, including flight number, departure time, arrival time, weather conditions, airline name, source, destination, and delay status. This collected data is then preprocessed by removing missing values, eliminating duplicate records, and converting text-based information into numerical format so that it can be used by machine learning models. After preprocessing, important features such as weather conditions, departure time, traffic congestion, airline performance, and travel distance are selected because they directly influence flight delays.

Next, the dataset is divided into two parts: 80% training data and 20% testing data. The training data is used to train machine learning algorithms such as Decision Tree, Random Forest, and Logistic Regression to learn patterns from past flight records. Once the model is trained, the testing data is given to the model to predict whether a flight will be on time or delayed. The predicted results are then compared with actual flight data to evaluate the model's accuracy. Finally, the system displays the prediction result as "On Time" or "Delayed", and if delayed, it can also estimate the possible delay time. This process helps airlines and passengers make better travel decisions by predicting delays in advance.

Algorithm used for existing system:

Step 1: Start
Step 2: Collect Historical Flight Data – Depends upon limited factors such as arrival time , departure time and previous delay time.
Step 3: Clean the Dataset - Remove missing values, duplicate records, and incorrect data entries to improve dataset quality.
Step 4: Prepare Data - Convert categorical data (airline name, weather type, airport name) into numerical format and normalize the dataset for model training.
Step 5: Select Important Features - Choose factors that influence delay prediction, such as departure time, weather conditions, airline performance, traffic congestion, and travel distance.
Step 6: Split the Dataset - Divide the dataset into training data and testing data

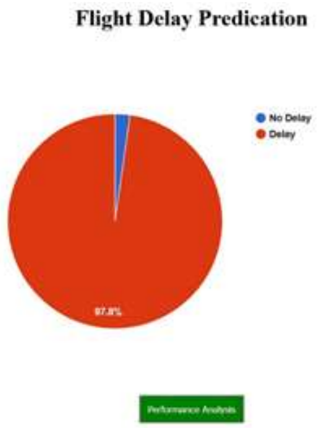
International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 3, May-Jun-2026, ISSN (Online): 2395-566X

Step7: Apply Machine Learning Model - Train the system using algorithms like LSTM &RNN on historical flight data.
Step 8: Predict Flight Delay - Use the trained model to predict whether the flight status is On Time or Delayed.
Step 9: Evaluate Performance - Compare predicted results with actual results and calculate model accuracy.
Step 10: Display Output - Show the prediction result to the user as On Time or Delayed which is having lower predicting accuracy.
Step 11: Stop

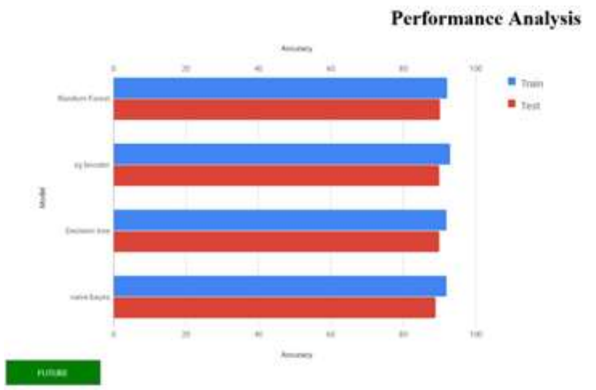
Algorithm used for proposed system:
Step 1: Start
Step 2: Collect Historical Flight Data – Gather flight information including departure time, arrival time, airline details, weather conditions, airport traffic, travel distance, and previous delay records.
Step 3: Data Preprocessing – Remove missing values, duplicate records, and inconsistent data entries to improve data quality.
Step 4: Feature Engineering – Extract and create useful features such as delay trends, peak-hour traffic indicators, weather impact factors, and airline performance metrics.
Step 5: Data Transformation – Convert categorical attributes into numerical values and normalize the dataset for efficient model training.
Step 6: Feature Selection – Identify the most influential factors affecting flight delays using statistical and machine learning techniques.
Step 7: Split Dataset – Divide the dataset into training, validation, and testing datasets.
Step 8: Train Advanced Machine Learning Model – Apply optimized machine learning algorithms such as Random Forest, and LSTM to learn flight delay patterns.
Step 9: Hyperparameter Optimization – Fine-tune model parameters to improve prediction accuracy and reduce errors.
Step 10: Predict Flight Delay – Use the trained model to classify flights as On Time or Delayed.
Step 11: Evaluate Model Performance – Measure performance using accuracy, precision, recall, F1-score, and confusion matrix.
Step 12: Display Prediction Result – Provide users with accurate flight delay predictions along with the probability of delay.
Step 13 : Stop

VII. EXPERIMENTAL RESULTS

Delay Analysis
The pie chart visualizes the distribution of delayed and non-delayed flights. The analysis indicates that delayed flights constitute the majority of the dataset.



Performance Analysis
The performance comparison graph shows the accuracy of different machine learning algorithms. Among the evaluated models, Random Forest and XGBoost achieved higher prediction accuracy, demonstrating their effectiveness for flight delay prediction.



The system successfully predicts flight delays using machine learning techniques with high accuracy. Future enhancements include integrating real-time weather, air traffic information, and advanced deep learning models to further improve prediction performance.

VIII. CONCLUSION

In this paper, random forest-based and LSTM-based architectures have been implemented to predict individual flight delay. The experimental results show that the random forest based method can obtain good performance for the binary classification task and there are still room for improving the multi-categories classification tasks. The LSTM-based architecture can obtain relatively higher training accuracy, which suggests that the LSTM cell is an effective structure to handle time sequences. However, the overfitting problem occurred in the LSTM-based architecture still needs to be solved. In summary, the random forest-based architecture presented better adaptation at a cost of the training accuracy when handling the limited dataset.

REFERENCES

1. M. Leonardi, “Ads-b anomalies and intrusions detection by sensor clocks tracking,” IEEE Trans. Aerosp. Electron. Syst., to be published, doi: 10.1109/TAES.2018.2886616.
2. Y. A. Nijsure, G. Kaddoum, G. Gagnon, F. Gagnon, C. Yuen, and R. Mahapatra, “Adaptive air-to-ground secure communication system based on ads-b and wide-area multilateration,” IEEE Trans. Veh. Technol., vol. 65, no. 5, pp. 3150–3165, 2015.
3. J. A. F. Zuluaga, J. F. V. Bonilla, J. D. O. Pabon, and C. M. S. Rios, “Radar error calculation and correction system based on ads-b and business intelligent tools,” in Proc. Int. Carnahan Conf. Secur. Technol., pp. 1–5, IEEE, 2018.
4. D. A. Pamplona, L. Weigang, A. G. de Barros, E. H. Shiguemori, and C. J. P. Alves, “Supervised neural network with multilevel input layers for predicting of air traffic delays,” in Proc. Int. Jt. Conf. Neural Networks, pp. 1–6, IEEE, 2018.
5. S. Manna, S. Biswas, R. Kundu, S. Rakshit, P. Gupta, and S. Barman, “A statistical approach to predict flight delay using gradient boosted decision tree,” in Proc. Int. Conf. Comput. Intell. Data Sci., pp. 1–5, IEEE, 2017.
6. L. Moreira, C. Dantas, L. Oliveira, J. Soares, and E. Ogasawara, “On evaluating data preprocessing methods for machine learning models for flight delays,” in Proc. Int. Jt. Conf. Neural Networks, pp. 1–8, IEEE, 2018.
7. J. J. Rebollo and H. Balakrishnan, “Characterization and prediction of air traffic delays,” Transp. Res. Part C Emerg. Technol., vol. 44, pp. 231–241, 2014.
8. L. Hao, M. Hansen, Y. Zhang, and J. Post, “New york, new york: Two ways of estimating the delay impact of new york airports,” Transp. Res. Part ELogist. Transp. Rev., vol. 70, pp. 245–260, 2014.
9. ANAC, “The Brazilian National Civil Aviation Agency.” anac.gov, 2017. [online] Available:http://www.anac.gov.br/.
10. S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, “Efficient knn classification with different numbers of nearest neighbors,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 5, pp. 1774–1785, 2017.
11. J. Sun, Z. Wu, Z. Yin, and Z. Yang, “Svm-cnn-based fusion algorithm for vehicle navigation considering atypical observations,” IEEE Signal Process. Lett., vol. 26, no. 2, pp. 212–216, 2018.
12. Y. J. Kim, S. Choi, S. Briceno, and D. Mavris, “A deep learning approach to flight delay prediction,” in Proc. Digit. Avion. Syst. Conf., pp. 1–6, IEEE, 2016.
13. Y. Cong, J. Liu, B. Fan, P. Zeng, H. Yu, and J. Luo, “Online similarity learning for big data with overfitting,” IEEE Trans. Big Data, vol. 4, no. 1, pp. 78–89, 2017.
14. F. Tang, Z. M. Fadlullah, B. Mao, and N. Kato, “An intelligent traffic load prediction-based adaptive channel assignment algorithm in sdn-iot: A deep learning approach,” IEEE Internet Things J., vol. 5, pp. 5141–5154, Dec 2018.
15. N. Kato, Z. M. Fadlullah, B. Mao, F. Tang, O. Akashi, T. Inoue, and K. Mizutani, “The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective,” IEEE Wireless Commun., vol. 24, pp. 146–153, June 2017.
16. J. Wang, J. Liu, and N. Kato, “Networking and communications in autonomous driving: A survey,” IEEE Commun. Surveys Tuts., vol. 21, pp. 1243–1274, April 2019.
17. Y. Kawamoto, H. Nishiyama, N. Kato, F. Ono, and R. Miura, “Toward future unmanned aerial vehicle networks: Architecture, resource allocation and field experiments,” IEEE Wireless Commun., vol. 26, pp. 94–99, February 2019.
18. D. Takaishi, Y. Kawamoto, H. Nishiyama, N. Kato, F. Ono, and R. Miura, “Virtual cell-based resource allocation for efficient frequency utilization in unmanned aircraft systems,” IEEE Trans. Veh. Technol., vol. 67, no. 4, pp. 3495–3504, 2018.