

Flight Arrival Delay Prediction at King Khalid International Airport in Saudi Arabia Using Flight History and Weather Data

Abdullellah Alamer and Tarek Helmy*

Cite <https://doi.org/10.64589/juri/231547>

Submitted: April 06, 2026 Revised: April 22, 2026 Accepted: May 05, 2026

ABSTRACT

Flight arrival delay is a critical operational challenge that affects passenger experience, airline efficiency, and airport resource planning in the aviation industry. This study develops a machine learning-based prediction model for flight arrival delays at King Khalid International Airport in Riyadh, Saudi Arabia, by combining historical flight data with meteorological observations. A dataset of 47,107 flight arrival records collected from Flightera covering the period June 2025–September 2025 was integrated with hourly weather data for the same period obtained from Visual Crossing. The integrated dataset underwent systematic cleaning, preprocessing, and feature engineering, resulting in a final dataset of 42,361 records and 33 features. Three gradient boosting regression models, CatBoost, XGBoost, and LightGBM, were trained and evaluated using an 80:20 train-test split. The model outputs were converted into binary delay classifications using a 15-min threshold, and the performance was assessed using accuracy, precision, recall, and F1-score. LightGBM achieved the best overall performance, recording an accuracy of 94.45%, precision of 77.40%, and F1-score of 60.70%. Feature importance analysis revealed that departure delay, origin airport, and airline identity were the strongest predictors, while weather-related variables such as wind gust and sea-level pressure also contributed meaningfully. These findings indicate that integrating operational flight data with hourly weather information can support more accurate arrival-delay prediction and assist airport operators and airlines in planning resources, managing disruptions, and improving operational decision-making.

Keywords: flight delay prediction, machine learning, LightGBM, King Khalid International Airport, weather data, gradient boosting, feature engineering, arrival delay

1. INTRODUCTION

Air travel is an essential mode of transportation, connecting cities, countries, and continents at an unprecedented scale. As global air traffic continues to grow, flight punctuality has become critical for airlines, airports, and passengers. Flight delays impose high economic and operational costs, including increased fuel consumption, crew scheduling disruptions, missed connections, and reduced passenger satisfaction. Accordingly, flight delays cost the aviation sector billions of dollars annually and affect millions of passengers¹. Understanding and predicting these delays is of considerable importance for aviation stakeholders. The complexity of flight delays is often compounded by delay propagation, where a single localized disruption spreads across the entire network, causing delays for multiple flights. This stochastic nature of air travel makes static scheduling insufficient. The nonlinear interactions between operational infrastructure constraints, air-traffic volume, and environmental conditions further complicate the problem, making the operational environment management highly volatile.

Saudi Arabia has witnessed a rapid expansion of its air transport sector in recent years. King Khalid International Airport (KKIA) in Riyadh serves as the primary gateway to the capital and handles a substantial volume of domestic and international

traffic. Given its strategic importance in the broader ambitions of Saudi Vision 2030 to develop the aviation sector, improving flight operations and punctuality is a matter of national significance².

Machine learning can support flight-delay prediction by identifying complex patterns in large datasets that traditional statistical methods may not capture. A variety of machine-learning techniques ranging from ridge regression to ensemble techniques like random forest and stacking, have been applied to flight-delay prediction^{3,4}. However, the literature is constrained by reliance on static, historical data. Features such as departure delay, origin airport, airline, and time of day are among the strongest predictors of arrival delay. However, most existing research has focused on western airports, with limited attention given to airports in the Middle East or Saudi Arabia in particular². Furthermore, though weather contributes to flight delays, studies explicitly integrating real-time hourly weather observations with flight records at individual-flight level remains limited^{5,6}.

This study addresses these gaps by proposing a machine-learning framework for predicting flight-arrival delays at KKIA using a combination of historical flight data and hourly weather observations collected for the same period. The study makes the following contributions: it constructs a novel integrated dataset combining flight records and weather data matched at the

scheduled arrival hour; it applies systematic feature engineering, including interaction features that capture combined weather effects; and it compares three state-of-the-art gradient boosting models—CatBoost, XGBoost, and LightGBM—to identify the most effective approach for this prediction task.

2. METHODOLOGY

This study followed a structured pipeline to develop a machine-learning-based model for predicting flight arrival delays at KKIA in Riyadh, Saudi Arabia. The methodology consisted of five main stages: data collection, data cleaning and preprocessing, data integration, feature engineering, and model development and evaluation. Two primary data sources were used: historical flight arrival records and hourly weather observations corresponding to the same period. The flight and weather datasets were independently preprocessed before being merged into a unified dataset. Following integration, feature engineering was applied to enhance the predictive power of the data. Finally, three gradient-boosting machine-learning models—CatBoost, XGBoost, and LightGBM^{3,4,6}—were trained and evaluated to identify the best-performing model for predicting arrival delay at KKIA. The pipeline is illustrated in Figure 1, which summarizes the sequential stages of the proposed framework, from data collection to model evaluation and best model selection.

2.1. Data Collection.

2.1.1. Flight history data. The flight-history data used in this study were collected from Flightera. The dataset consisted of historical arrival flight records for KKIA, covering the period from June 2025 to September 2025. The dataset contained 47,107 flight records and 11 features, as summarized in Table 1. Among

the variables, `Departure_Delay` and `Arrival_Delay` were defined as the difference between the actual and scheduled departure or arrival times, respectively, measured in minutes. A value greater than zero indicates that the flight was delayed, whereas a value less than zero indicates that the flight arrived or departed earlier than scheduled. Another important variable is `Day_of_Week`, which indicates the day on which the flight operated using abbreviated labels, such as Sat for Saturday, Sun for Sunday, and Mon for Monday.

2.1.2. Weather data. The weather data were collected from Visual Crossing. This dataset provided hourly weather observations recorded at Riyadh and included many variables that may influence flight arrival time. The weather dataset contained 2,928 records and 21 features, and covered the same period as the flight data, from June 2025 to September 2025, to ensure consistency between the two datasets. The main weather-related features used in this study are presented in Table 2.

2.2. Data Cleaning and Preprocessing. Before integrating the datasets, the flight and weather data underwent independent cleaning and preprocessing stages. For the flight dataset, records containing missing values, inconsistencies, or unsuitable entries were removed, reducing the dataset from 47,107 to 42,399 records. The `Flight_Number` column was also removed as it serves only as an identifier with no predictive value. Categorical features, including `Airline_ICAO`, `Origin_IATA`, and `Day_of_Week`, were encoded using label encoding, and the scheduled departure and arrival times were split into separate hour and minute components, yielding four new variables: `Hour_Scheduled_Departure`, `Minute_Scheduled_Departure`, `Hour_Scheduled_Arrival`, and

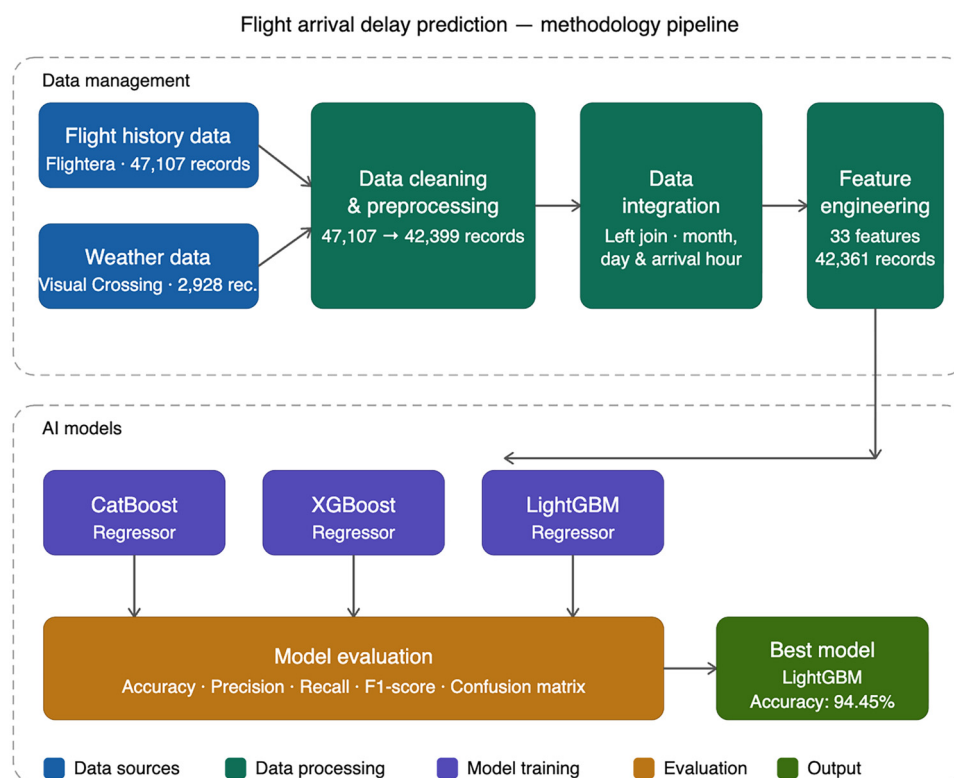


Figure 1. Overall pipeline for flight arrival delay prediction at KKIA

Table 1. *Flight dataset features*

Feature name	Description	Data type
Month	The month in which the flight departed.	float64
Day	Day of the month.	float64
Flight_Number	Unique flight identifier.	object
Airline_ICAO	Airline code that operates the flight.	object
Origin_IATA	Code of the departure airport.	object
Scheduled_Departure_UTC	Scheduled departure time (UTC).	object
Departure_Delay	Difference between actual and scheduled departure time (minutes).	float64
Scheduled_Arrival_UTC	Scheduled arrival time (UTC).	object
Day_of_Week	Day of the week of the flight.	object
Duration	Scheduled flight duration in minutes.	float64
Arrival_Delay	Difference between actual and scheduled arrival time (minutes).	float64

Minute_Scheduled_Arrival. For the weather dataset, no records were removed; however, two variables—precipitation type and severe risk—were excluded owing to the high proportion of null values. Categorical weather features were converted into numerical representations using label and one-hot encoding.

2.3. Data Integration. After dataset preprocessing, the flight and weather data were integrated into a single dataset to

support the prediction analysis. To make the integration possible, the datetime variable in the weather dataset (Table 2) was first converted into a datetime format, from which the month, day, and hour values were extracted and converted to integer type. In the flight dataset, the variables required for matching, including Month, Day, Hour_Scheduled_Departure, and Hour_Scheduled_Arrival, were also converted into integer format to enable merging.

Table 2. *Weather dataset features*

Feature name	Description	Data type
name	Name of the weather station location.	object
datetime	Date and time of the recorded weather observation.	object
temp	Air temperature at the recorded time.	float64
feelslike	Perceived temperature, considering wind and humidity.	float64
dew	Dew-point temperature.	float64
humidity	Relative humidity percentage.	float64
precip	Amount of precipitation recorded.	int64
precipprob	Probability of precipitation (%).	int64
preciptype	Type of precipitation (rain, snow, etc.).	float64
windgust	Maximum wind gust speed recorded.	float64
windspeed	Average wind speed.	float64
winddir	Wind direction in degrees.	int64
sealevelpressure	Atmospheric pressure at sea level.	float64
cloudcover	Percentage of sky covered by clouds.	float64
visibility	Horizontal visibility distance.	float64
uvindex	Ultraviolet radiation-index level.	int64
severerisk	Estimated risk level of severe weather conditions.	float64
conditions	General weather condition description (e.g., clear, cloudy).	object
icon	Weather condition icon label used by the data provider.	object
stations	Weather station identifiers used for the observation.	object
source	Source of the weather data observation.	object

An important step was taken for flights whose scheduled arrival hour was earlier than the scheduled departure hour: the arrival day was moved forward by one day to represent a next-day arrival. Thereafter, the two datasets were merged using a left join based on month, adjusted arrival day, and scheduled arrival hour from the flight dataset, with the corresponding month, day, and hour values from the weather dataset. This approach allowed each flight record to be associated with the weather conditions expected at its scheduled arrival time. It is important to note that the weather data were matched based on the scheduled arrival hour rather than the actual arrival hour. This was done because the actual arrival time was unknown at prediction time, making the scheduled arrival hour the only available reference for weather matching. The merged dataset contained 42,399 records and 31 features. However, it had some null values due to the merging process.

2.4. Data Cleaning After Integration. After the integration step, the merged dataset was further cleaned to identify and resolve issues that appeared after combining the flight and weather records. The process included checking for unmatched entries, missing values created during merging, duplicated observations, and inconsistencies in the final dataset structure. The final integrated dataset consisted of 42,361 records, indicating that only a small number of observations were excluded during this stage.

2.5. Feature Engineering. Feature engineering was performed to enhance the predictive usefulness of the dataset. Four variables were removed: Stations and Source, which contained constant values across all observations; Datetime, whose time components were extracted during integration; and Duration, which was excluded to prevent unrealistic simplification of the prediction task.

A new feature, `num_flight_in_the_hour_arrival`, was created by counting the number of flights scheduled to arrive at KKIA within the same hour, grouped by Month, `Arrival_Day_Adjusted`, and `Hour_Scheduled_Arrival`. This feature was intended to capture airport-traffic concentration and potential congestion effects. Additionally, `Arrival_Delay` values were clipped at zero to ensure the target variable reflected only actual delays rather than early arrivals.

Three interaction features were also generated to capture combined weather effects: `Temp_x_Humidity`, `wind_dir_x_windspeed`, and `icon_fog_x_uvindex`. The final dataset after feature engineering consisted of 42,361 records and 33 features, as summarized in Table 3.

2.6. Development of Models. The selection of gradient boosting models for this study was motivated by empirical and literature-based considerations. Classical machine-learning methods such as random forest and support vector machine were initially evaluated on the dataset; however, they produced inferior performance compared to gradient boosting approaches, particularly in handling the nonlinear relationships and mixed feature types in the dataset. This is consistent with findings of the existing literature, where gradient boosting methods have consistently outperformed classical approaches on structured tabular aviation data. Furthermore, CatBoost, XGBoost, and LightGBM were selected because they have been widely adopted in flight

delay prediction research, offer built-in handling of structured data with heterogeneous features, and provide interpretable feature importance outputs that are valuable for operational decision-making^{3,4,6}.

To predict flight arrival delay, three machine-learning models were developed using the CatBoost⁶, XGBoost⁴, and LightGBM³ Regressors, with `Arrival_Delay` as the target variable; the input features consisted of all remaining variables in the dataset after excluding the target column. The dataset was divided into training and testing sets (80:20) using the `train_test_split` method.

Although the developed models were regression models that predicted arrival delay in minutes, their outputs were further converted into a binary delay classification for evaluation purposes. A flight was considered delayed if its arrival delay was 15 min or more, and not delayed otherwise. This threshold was applied to the actual and predicted delay values so that the regression outputs could be evaluated from an operational classification perspective.

2.7. Experimental Setup. All experiments were implemented in Python 3 using the Google Colab platform. The computational experiments were executed with access to an NVIDIA T4 GPU, which supported the training and evaluation of the machine-learning models. All three models were trained using consistent hyperparameter configurations to ensure a fair comparison. For CatBoost, the number of iterations was set to 1500, with a learning rate of 0.05, a tree depth of 8, and RMSE as the loss function. For XGBoost, the number of estimators was set to 1500, with a learning rate of 0.05, a maximum depth of 8, a subsample ratio of 0.8, and a column sampling ratio of 0.8, using the histogram-based tree method. LightGBM was configured with the same number of estimators, learning rate, maximum depth, subsample ratio, and column sampling ratio as XGBoost. A random seed of 42 was applied to all three models to ensure reproducibility.

2.8. Model Evaluation Metrics. To evaluate model performance, the predicted arrival delay values were converted into two classes using a threshold of 15 min. Flights with arrival delay values greater than or equal to 15 min were classified as delayed, whereas flights with delay values less than 15 min were classified as not delayed. Based on this classification, the developed models were evaluated using accuracy, precision, recall, and F1-score. Additionally, a confusion matrix was used to summarize the classification results.

The evaluation metrics are calculated as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{FP + TP} \quad (2)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$\text{F1 - score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

Table 3. *Features of the merged dataset*

Feature name	Description	Data type
Month	Month of the flight operation.	int64
Airline_ICAO	Encoded airline identifier.	int64
Origin_IATA	Encoded identifier of the departure airport.	int64
Departure_Delay	Difference between actual and scheduled departure time in minutes.	float64
Day_of_Week	Encoded the day of the week of the flight operation.	int64
Arrival_Delay	Arrival delay in minutes after clipping negative values to zero.	float64
Hour_Scheduled_Departure	Scheduled departure hour.	int64
Minute_Scheduled_Departure	Scheduled departure minute.	int64
Hour_Scheduled_Arrival	Scheduled arrival hour.	int64
Minute_Scheduled_Arrival	Scheduled arrival minute.	int64
Arrival_Day_Adjusted	Adjusted arrival day is used for overnight flights arriving on the next day.	int64
temp	Air temperature at the recorded time.	float64
feelslike	Perceived temperature considering surrounding conditions.	float64
dew	Dew-point temperature.	float64
humidity	Relative humidity percentage.	float64
precip	Amount of precipitation recorded.	float64
precipprob	Probability of precipitation.	float64
windgust	Maximum wind gust speed.	float64
windspeed	Average wind speed.	float64
winddir	Wind direction in degrees.	float64
sealevelpressure	Atmospheric pressure at sea level.	float64
cloudcover	Percentage of cloud cover.	float64
visibility	Horizontal visibility distance.	float64
uvindex	Ultraviolet radiation index.	float64
conditions_Partially cloudy	Indicator variable for partially-cloudy weather conditions.	int64
icon_clear-night	Indicator variable for clear-night weather conditions.	int64
icon_fog	Indicator variable for fog weather condition.	int64
icon_partly-cloudy-day	Indicator variable for partly-cloudy day weather condition.	int64
icon_partly-cloudy-night	Indicator variable for partly-cloudy night weather condition.	int64
num_flight_in_the_hour_arrival	Number of flights scheduled to arrive at KKIA during the same hour.	int64
Temp_x_Humidity	Interaction features between temperature and humidity.	float64
winddir_x_windspeed	Interaction features between wind direction and wind speed.	float64
icon_fog_x_uvindex	Interaction features between fog indicator and ultraviolet index.	float64

where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives.

Accuracy measures the overall proportion of correctly classified flights. Precision indicates the proportion of flights predicted as delayed that were delayed. Recall measures the proportion of truly delayed flights that were correctly identified by the model. F1-score is the harmonic mean of precision and recall and provides a balanced measure when both false positives and false negatives are important. The confusion matrix was used to provide a detailed view of the correct and incorrect classifications for delayed and non-delayed flights.

3. RESULTS AND DISCUSSION

3.1. Model Performance. The developed models were evaluated using accuracy, precision, recall, and F1-score after converting the predicted arrival delay values into two classes based on a threshold of 15 min. A flight was classified as delayed if its arrival delay was 15 min or more, and not delayed otherwise. Table 4 summarizes the classification performance of the three developed models.

The results show that all three models achieved high overall accuracy, with LightGBM providing the strongest overall

Table 4. Model Performance

Model	Accuracy	Precision	Recall	F1-score
CatBoost	0.9412	0.7627	0.4748	0.5852
XGBoost	0.9422	0.7413	0.5007	0.5977
LightGBM	0.9445	0.7740	0.4993	0.6070

performance. LightGBM achieved the highest accuracy, precision, and F1-score, while XGBoost produced a slightly higher recall. This indicates that LightGBM offered the best overall balance between identifying delayed flights and minimizing false delay predictions.

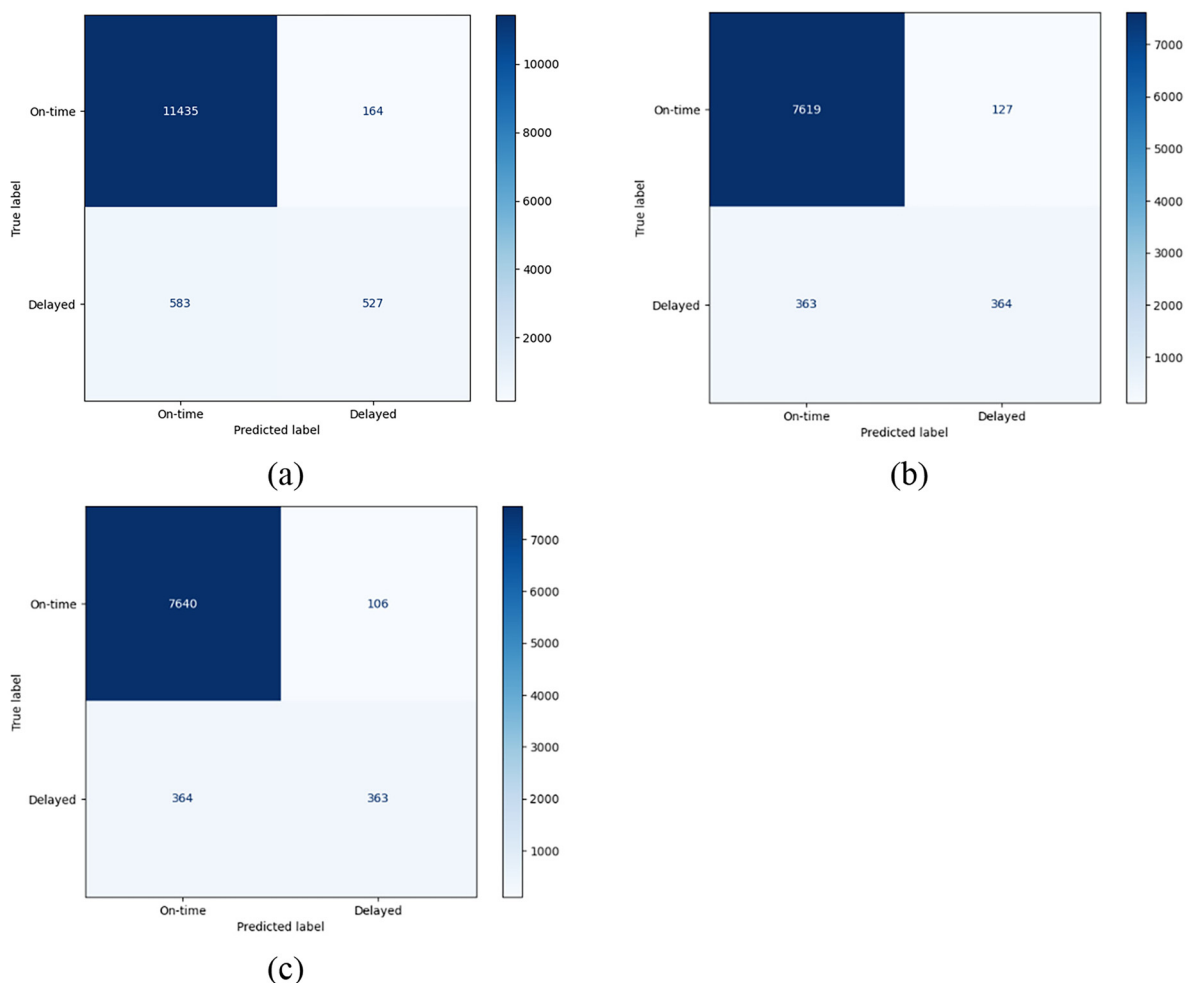
The confusion matrices in Figure 2 show that all three models were more effective at identifying non-delayed flights than delayed flights. This pattern is expected because delayed flights represent a smaller portion of the dataset. The moderate recall values indicate that delayed-flight detection remains challenging, suggesting that the models are useful as decision-support tools but may require further improvement for applications where detecting most delayed flights is the main priority.

3.2. Discussion of Results. The results show that LightGBM outperformed CatBoost and XGBoost in the flight-arrival delay prediction task. This suggests that LightGBM was better able to capture the relationships between the flight, scheduling,

and weather-related variables included in the dataset. Since the data used in this study are structured tabular data with both original and engineered features, LightGBM appears to have benefited from its ability to model nonlinear interactions and feature contributions efficiently.

The feature importance results for LightGBM are presented in Figure 3. As shown in Figure 3, *Departure_Delay* was the dominant feature in the LightGBM model, followed by *Origin_IATA* and *Airline_ICAO*, confirming that operational flight variables were the strongest predictors of arrival delay^{5,7}. This is operationally logical, as a flight that departs late is highly likely to arrive late regardless of weather or routing conditions.

Of the weather-related variables, wind gust ranked fifth with an importance score of 2395, followed closely by the engineered interaction feature *winddir_x_windspeed* (2035) and sea level pressure (2030), while humidity (1809), *Temp_x_Humidity* (1706), wind direction (1382), and temperature (1342) also appeared among the top 15 features. The strong presence of

**Figure 2.** Confusion matrices of the three developed models: (a) CatBoost, (b) XGBoost, and (c) LightGBM

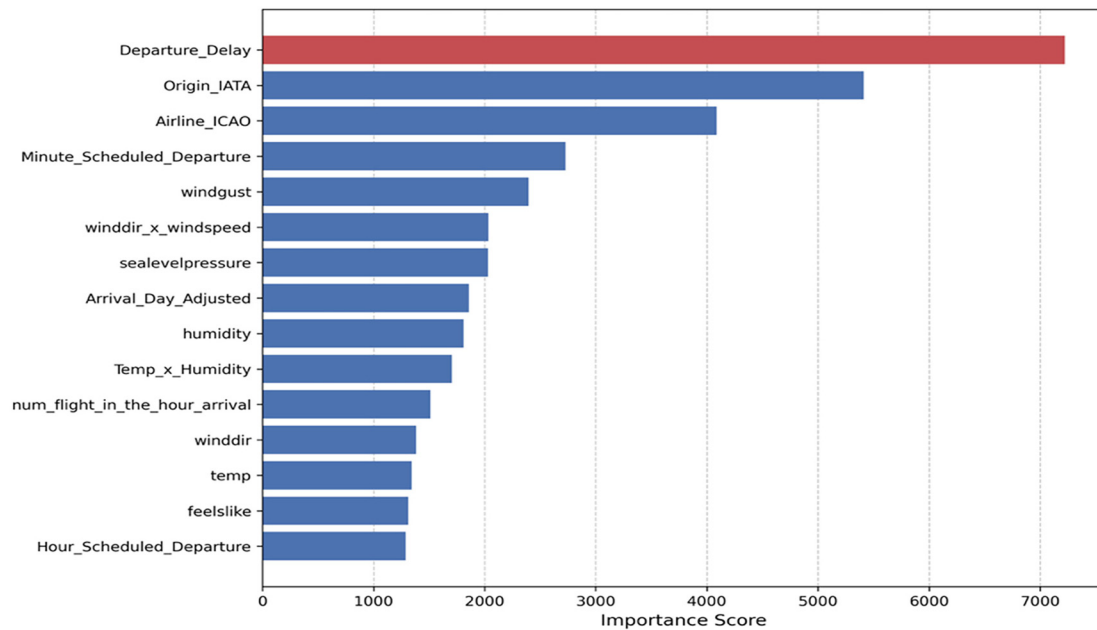


Figure 3. Feature importance of the LightGBM model (top 15 features); Departure_Delay is highlighted in red as the most influential feature

these meteorological variables confirms that weather conditions during the scheduled arrival hour contributed meaningfully to the predictions, despite operational variables dominating the top three positions. These findings highlight the role of individual meteorological parameters in flight delay. Wind speed and wind gusts have been specifically identified as disruptive factors during approach and landing phases, as strong crosswinds reduce runway capacity and force aircraft into holding patterns or missed approaches. Sea level pressure is closely related to air density and affects aircraft lift and engine performance, with low-pressure conditions associated with increased taxi and turnaround times.

High humidity levels have been linked to reduced visibility and increased likelihood of convective weather activity, both of which contribute to ground and airborne delays. Therefore, the presence of these variables in the LightGBM model reflects genuine physical relationships between atmospheric conditions and arrival performance at KKIA.

Figure 4 presents a comparison of all evaluation metrics across the three models, further highlighting LightGBM as the best-performing model.

The results also suggest that the engineered features contributed positively to model performance. Particularly,

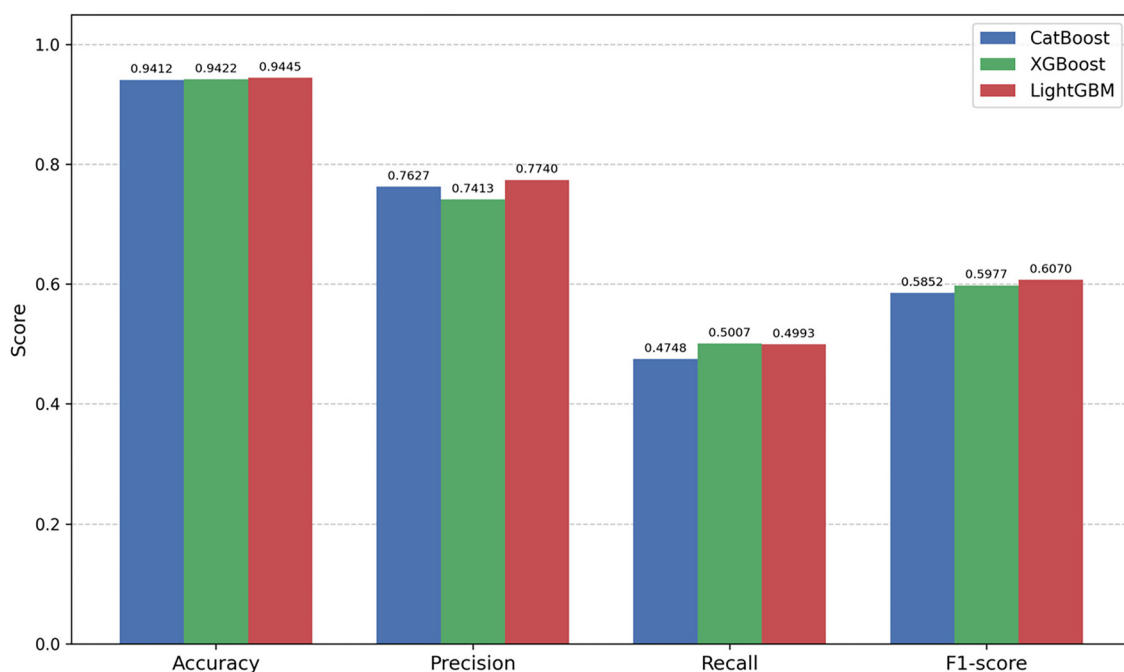


Figure 4. Comparison of model performance across evaluation metrics (Accuracy, Precision, Recall, and F1-score) for CatBoost, XGBoost, and LightGBM

num_flight_in_the_hour_arrival may have helped represent airport-traffic concentration, and interaction variables such as Temp_x_Humidity and winddir_x_windspeed may have helped the models capture more complex weather effects. These variables support the utility of combining domain-based feature engineering with machine-learning techniques.

Although XGBoost achieved a slightly higher recall than LightGBM, its lower precision and F1-score indicate that it produced a weaker overall balance in classification performance. CatBoost also performed well, but its lower recall and F1-score made it less effective at identifying delayed flights than the other two models. Therefore, LightGBM was considered the most suitable model for this study because it achieved the best overall classification performance and demonstrated a strong ability to utilize both operational and weather-related variables.

4. CONCLUSIONS

This study investigated flight arrival delay prediction at KKIA in Riyadh, Saudi Arabia, using a machine-learning framework that integrates historical flight records with hourly weather observations. Three gradient-boosting models, CatBoost, XGBoost, and LightGBM, were developed and evaluated. Based on the experimental results, the following key conclusions were drawn:

- LightGBM achieved the best overall performance, recording an accuracy of 94.45%, precision of 77.40%, recall of 49.93%, and F1-score of 60.70%.
- Departure delay was the most dominant predictor, confirming that operational disruptions at departure propagate directly to arrival.
- Origin airport and airline identity were the second- and third-most influential features, highlighting the role of route- and carrier-specific factors.
- Weather variables, including wind gust, sea level pressure, and humidity, contributed meaningfully, confirming that meteorological conditions should not be excluded from delay prediction models.
- Engineered features, particularly winddir_x_windspeed, Temp_x_Humidity, and num_flight_in_the_hour_arrival, supported model performance, validating the value of domain-informed feature engineering.
- Integrating flight and weather data via a left join on the scheduled arrival hour proved effective in enriching the prediction environment.

The findings contribute to data-driven aviation delay management research by providing evidence from a Middle Eastern airport, which remains underrepresented in existing flight-delay prediction studies. From a practical perspective, the proposed framework can support airport operators and airlines in resource planning, disruption monitoring, and operational decision-making. Nevertheless, this study has a notable limitation: for flights experiencing significant delays of several hours, the matched weather record may not accurately reflect the

atmospheric conditions at the time the flight landed. Future work may address this limitation by incorporating real-time weather forecasts, extending the dataset across longer periods, applying time-based validation, and testing the framework on other major Saudi airports.

AFFILIATIONS AND AUTHOR DETAILS

Undergraduate Author

Abdullellah Alamer – Department of Information and Computer Science, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0009-0004-7889-8575
Email: s202265520@kfupm.edu.sa

Corresponding Author

Tarek Helmy – Research Mentor, Professor, Department of Information and Computer Science, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia;  0000-0001-9683-682X
Email: helmy@kfupm.edu.sa

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support provided by the Information and Computer Science Department and IRC-ISS at King Fahd University of Petroleum and Minerals for providing the computing resources.

REFERENCES

- (1) Dai, M. A hybrid machine learning-based model for predicting flight delay through aviation big data. *Sci. Rep.* **14**, 4603 (2024).
- (2) Alfarhood, M. *et al.* Predicting Flight Delays with Machine Learning: A Case Study from Saudi Arabian Airlines. *Int. J. Aerosp. Eng.* **2024**, 3385463 (2024).
- (3) Ke, G. *et al.* LightGBM: A Highly Efficient Gradient Boosting Decision Tree. in *Advances in Neural Information Processing Systems* vol. 30 (Curran Associates, Inc., 2017).
- (4) Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794 (Association for Computing Machinery, New York, NY, USA, 2016). doi:10.1145/2939672.2939785.
- (5) Yuan, Y., Wang, Y. & Lai, C. S. Multi-Attribute Data-Driven Flight Departure Delay Prediction for Airport System Using Deep Learning Method. *Aerospace* **12**, 246 (2025).
- (6) Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. & Gulin, A. CatBoost: unbiased boosting with categorical features. in *Advances in Neural Information Processing Systems* vol. 31 (Curran Associates, Inc., 2018).
- (7) Chowdhury, F. T. *et al.* Flight delay prediction using machine learning and explainable AI: a case study on Hazrat Shahjalal International Airport, Dhaka. *Case Stud. Transp. Policy* **23**, 101663 (2026).