Title
Go-Around Prediction Modelling: A Comparison of Two Airports

Permalink
https://escholarship.org/uc/item/24h9m71w

Journal
Proceedings of the 2024 International Workshop on ATM/CNS (IWAC2024)

Authors
Hansen, Mark
Liu, Ke
Ding, Kaijing
et al.

Publication Date
2024-11-01

Peer reviewed

Go-Around Prediction Modelling: A Comparison of Two Airports

Dr. Mark Hansen,[2] Ke Liu,[2] Kaijing Ding,[2] Dr. Lu Dai,[2] Abigail Malakun,[2] John Schade,[1)†] Kennis Chan,[1)†]
[1)] *ATAC Corporation, United States*
[2)] *University of California, Berkeley, United States*

†*email: mhansen@ce.berkeley.edu*
†*email: liuke126@berkeley.edu*
†*email: kaijing@berkeley.edu*
†*email: dailu@berkeley.edu*
†*email: abigail.malakun@berkeley.edu*
†*email: jes@atac.com*
†*email: koc@atac.com*

In this paper, we employ the Long Short-Term Memory model (LSTM) to predict the real-time probability of a go-around as an arrival flight approaches an airport within 10 nautical miles of the landing runway threshold. We train the model on two airports–New York JFK and San Francisco SFO. We further develop methods to analyze the causes of go-arounds. Our results indicate that the models for both airports perform quite well and that risk of simultaneous runway occupancy is the primary factor contributing to overall go-around occurrences. The models will be integrated with real-time data streaming and a web-based user interface to provide a real-time alerting capability that can be used by flight crews and other line personnel to identify and manage situations with a high risk of a go-around.

Key Words: Go-around prediction, Real-time prediction, Sequence classification, Long-Short Term Memory Model, Causal analysis

1. Introduction

A go-around is an aborted landing of an aircraft during its final approach or after it has already touched down. This action can either be initiated by the pilot flying or requested by air traffic control due to various reasons, such as an unstable approach or a runway obstruction. While a go-around is implemented as a precautionary measure to enhance flight safety, it poses challenges as it disrupts normal airport operations, increases the workload for air traffic controllers and flight crews,[1)] elevates noise levels and fuel consumption,[2)] and impacts airport throughput[3)] as well as flight on-time performance.[4)]

Despite the high risks and costs associated with go-around occurrence, the decision-making on go-around initiation varies from airline to airline and from person to person. Typically, airline procedures state that if a pilot determines the approach is unstable at specified altitudes according to airline-established policies, they should go around. However, interviews suggest that the decisions on go-arounds are strongly influenced by individual experiences and mental states.[4)] As a result, "the collective industry performance of complying with go-around policies is extremely poor and only about 3% of unstable approaches result in a go-around."[5)] Multiple studies have attempted to examine today's stabilized approach criteria and develop detailed go-around criteria based on interviews and flight simulator experiments.[5)-8)]

Meanwhile, go-arounds have appeared more frequently in recent years. After the pandemic (comparing FY17-19 to FY21-22), the go-around occurrence rate as a percent of arrivals at US Core 30 airport grew from 0.3% to 0.4%, with an increase of 24.9%.[9)] The FAA report attributes this change to "mainly … a large increase in go-arounds relative to a smaller increase in arrival operations" while failing to further elaborate on the reasons behind the "large increase in go-arounds." A New York Times article,[10)] investigated the abnormally high occurrence rate of airline close calls in 2023 and concluded that "mistakes by air traffic controllers — stretched thin by a nationwide staffing shortage — have been one major factor." Since go-arounds and close calls are the two possible results of a loss of separation during landing, we may speculate the same reason to account for the increasing go-around rate.

Considering the high risks and costs, the current highly subjective initiation criteria, and the growing occurrence rate of go-around, this study seeks to enable real-time go-around probability prediction informed by historical data. This study builds upon a series of previous works that develop the go-around detection algorithm and studies the underlying causes of go-around occurrence;[11)] proposes the concept of runway occupancy buffer as a feature to predict go-around occurrence;[12)] and predicts go-around occurrence with Input-Output Hidden Markov Model.[13)] In this study, we will (1) provide a real-time go-around prediction tool based on various spatiotemporal features including operational conditions and environmental measures; (2) predict go-around probabilities using deep sequential neural network models that capture the time-varying dynamic of go-around probability evolution; (3) analyze the features that contribute to high go-around probabilities; and (4) develop a web-based user interface to display the results of the go-around prediction service. The contribution of this work is three-fold. First, we hope that with the help of the prediction service, both pilots and air traffic controllers can be alerted when there is a high risk of a go-around. Second, the feature contribution analysis can inform the different parties about the leading causal and predictive factors for go-arounds and suggest measures to mitigate go-around risk. Finally, by training the model on two different airports we investigate how the predictability and causation of go-arounds differs from one airport to another.

The rest of the paper is organized as follows. Section 2 describes our data sources and the study airports. In Section 3, we introduce the detailed methodologies of this study. Section 4 presents the model results for John F. Kennedy International

Airport (JFK) and San Francisco International Airport (SFO). The final section offers the conclusions and discusses future work.

2. Data And Study Area: JFK

This study focuses on arrivals in the years **2019**, **2022**, and **2023** at JFK and SFO.

2.1. Data Sources

- **Integrated Flight Format (IFF), Reduced Data summary (RD), and Events dataset (EV)** from NASA Sherlock Data Warehouse: IFF records provide flight tracks updated every 1-10 seconds depending on the distance between the aircraft and the airport. RD file is a one-line summary for each flight in the facility and EV file records the details of an aircraft event (e.g., landing and takeoff). We use IFF data to obtain the flight trajectories (longitude, latitude, and altitude), ground speed, RD for aircraft type and airline information; and EV data for flight runway threshold crossing time and surface movement.
- **Airport Configuration Data (APTC)** from National Traffic Management Log: it records runway configuration updates of all major airports, including the time of each runway configuration change, the new arrival and departure runway configuration, and aircraft arrival rates and departure rates.
- **METeorological Aerodrome Reports (METAR)** from Aviation Weather Center of NOAA: it provides hourly update on wind, visibility, ceiling, and meteorological conditions in the terminal area.
- **N90 Go-Around Reports** from ATAC corporation: it summarizes the go-arounds that occurred in the New York TRACON (N90). This dataset, although incomplete, served as a supplementary source for cross-validating our go-around detection algorithm.

2.2. JFK airport

JFK International Airport, a major gateway airport in New York City, was the sixth busiest in the U.S. in 2019, with over 460,000 flight operations.[14] The airport has two pairs of parallel runways, all equipped with an Instrument Landing System (ILS), except for runway 13R. In the three study years, runway 22L was the primary arrival pathway, handling 41.1% of traffic, followed by 04R (30.5%) and 31R (10.2%) as shown in Table 1. Notably, runway 13L/31R underwent closure from April to November 2019 for major reconstruction,[15] reducing its usage compared to handling more arrivals in other years.

To identify the go-around events that occurred at JFK from trajectory data, we modified the go-around detection algorithm introduced in 11) by (1) adding another distance check which requires the maximum distance between the go-around flight and runway threshold after the proposed go-around initiation point should be greater than the distance between them at the point; and (2) fine-tuning the parameters to make sure that our algorithm detects each go-around reported in the N90 GoAround Reports that is visually confirmed, using Google Earth, to be real go-arounds. As a result, 2,302 go-arounds are detected in JFK in the three years, accounting for 0.39% of all arrivals.

Table 1. JFK runway usage and go-around occurrence rate

JFK		
Runway	**Usage Percentage in Terms of Arrival Operations**	**Go-around Occurrences per 100 Arrival Operations**
04L	4.7%	0.92
04R	30.5%	0.40
13L	3.1%	0.45
13R	0.1%	1.04
22L	41.1%	0.31
22R	5.2%	0.70
31L	5.0%	0.26
31R	10.2%	0.22

2.3. SFO airport

SFO, a major gateway on the West Coast, was the seventh busiest in the U.S. in 2019, with over 450,000 flight operations.[14] After data cleaning, 516,789 flights that arrived at SFO in 2019, 2022, and 2023, were analyzed and used to train and test our model. Out of the over five hundred thousand flights, 3,611 flights experienced a go-around event, accounting for 0.70% of all arrivals. Figure 7 shows the number of go-arounds initiated at different distances from the targeted landing runway threshold at SFO. In general, more flights initiated a go-around at a distance closer to the runway thresholds. Regarding the three different years, 0.62% of 210 thousand flights that arrived at SFO in 2019, 0.72% of 150 thousand flights in 2022, and 0.91% of 168 thousand flights in 2023, experienced a go-around. While the total number of arrivals decreased after the pandemic, the percentage of go-around flights increased. That is the same trend as the overall go-around occurrence rate at US Core 30 airports.[9]

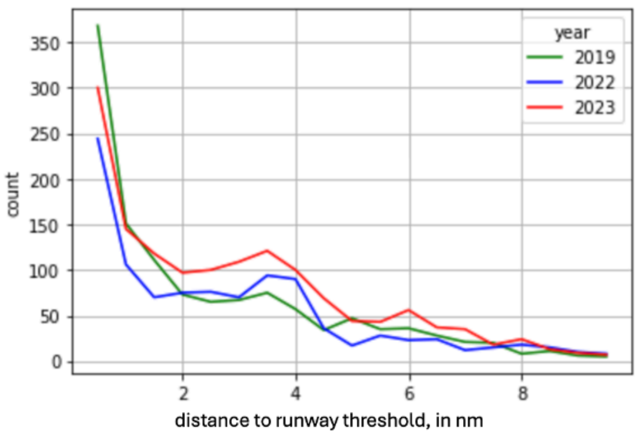

Fig. 1. Number of go-arounds initiated at different distances from runway threshold at SFO.

Table 2. SFO runway usage and go-around occurrence rate.

SFO		
Runway	**Usage Percentage in Terms of Arrival Operations**	**Go-around Occurrences per 100 Arrival Operations**
10L	0.1%	3.79
10R	0.1%	4.02
19L	3.6%	0.64
19R	0.3%	0.74
28L	38.5%	0.80
28R	57.4%	0.65

Like the case for JFK runways 13L and 04R, despite being the least utilized for arrivals, SFO runways 10L and 10R exhibit the highest go-around rate at over 3 per 100 operations. In contrast, the most used runways for arrivals, such as 04R, 22L, and 31R for JFK, or 28L and 28R for SFO, report go-around rates below 0.40% and 0.80% respectively. Several factors might explain these variations. Firstly, runway selection often hinges on weather conditions; less-used runways might be selected under adverse weather, which may impact approach stability and cause go-arounds. Second, the less-used arrival runways, when used as arrival runways, are usually used as mix-use runways that accommodate both arrival and departure flights at the same time, which may generate additional risk (e.g., simultaneous runway occupancy and failed separation) for safe landings and lead to go-arounds. Third, pilots may be less experienced to land on the less-used runways, which leads to more frequent unsuccessful landing attempts. We integrate these considerations in feature engineering section to assess their influence on go-around probabilities more comprehensively.

3. Methodology

3.1. Model specification

This section details our modeling method for predicting sequential go-around probabilities during landing approaches. It involves data processing, feature engineering, model training with hyperparameter tuning, and performance evaluation.

Table 3. Model variable description.

Group	Variable	Description
(i) Flight specific information	Type_[x] +	1 if flight is operated by an international airline, 0 otherwise
	WC_[x] +	Dummy variable for aircraft weight class
	Body_[x] +	1 if flight is wide-body aircraft, 0 otherwise
	month_[x] +	Dummy variable for flight month
	dow_[x] +	Dummy variable for flight day of the week
	tod_[x] +	Dummy variable for flight time of the day
(ii) Final Approach Stability	groundspeed gs_neg/pos	Flight ground speed, in knots Negative/ Positive components of groundspeed after standardization
	energy eg_neg/pos	Kinetic energy height, in feet Negative/ Positive components of energy after standardization
	horiz	Horizontal deviation from extended runway centerline, in meters
	alt_dev_amp alt_dev_neg/pos	Altitude deviation from 3-degree glideslope starting from revised landing threshold, in 100 feet Negative/ Positive components of altitude deviation after standardization
(iii) In-trail Relation	lead_ent_D	1 if leading aircraft is a departure, 0 otherwise
	lead_off_rwy	1 if leading aircraft has got off runway, 0 otherwise
	separation	Separation from leading aircraft, in nautical miles
	hat_ROB	Predicted runway occupancy buffer, in seconds
	speed_diff	Ground speed difference from leading aircraft, in knots
	alt_diff	Altitude difference from leading aircraft, in 100 feet
	closing	1 if separation from leading aircraft is decreasing, 0 otherwise
	lead_rwy_time	0 if leading aircraft has not yet landed, otherwise runway occupancy time of leading aircraft, in seconds
	WC_trail	Weight class of trailing aircraft
	WC_lead_arr	Weight class of leading arrival aircraft
(iv) Airport/ Runway Conditions	rwy_[x] +	Dummy variable for landing runway
	mix_rwy	1 if landing runway is also used for departure when the aircraft is landing, 0 otherwise
	crs_rwy	1 if the landing runway intersects with other runways in use, otherwise
	dep_rwy	1 if the arrival runway configuration is 22L/22R, 4R/4L or staggered 31L/31R when the aircraft is landing, 0 otherwise
	ADR	Airport departure rate
	AAR	Airport arrival rate
	arr_ratio	Ratio between number of flights whose altitude is below 200 FL in TRACON and airport arrival rate
	#Obj_rwy	Number of objects on landing runway
(v) Weather	weather_[x] +	Dummy variable for meteorological condition
	visibility	Airport visibility, in miles
	ceiling	Airport ceiling condition, in 100 feet
	head	Head wind, in knots
	tail	Tail wind, in knots
	cross	Cross wind, in knots
	gust	Gust wind, in knots

+ Variables are one-hot encoded.

3.1.1. Flight sequences and features

We apply linear interpolation to the raw flight tracks, which update every 1-10 seconds, to create subsampled points every 0.5 nautical mile (nm) within a 10 nm approach to the landing runway threshold. This generates a maximum of 20 data points per flight, ending at the start of a go-around initiation or at the runway threshold for normal landings. Each point, termed a cutoff gate, ranges from gate 0 (at 10 nm out) to gate 19 (at runway threshold). Go-around flights are assigned with a consistent binary label of 1 up to the go-around initiation gate; subsequent points are excluded. For instance, a flight initiating a go-around at 4.4 nm will have all points from 10 to 4.5 nm labeled as 1. Points beyond this, from 4 to 0.5 nm, are omitted. The result is a dataset with up to 20 labeled gates per flight, each gate featuring a compiled vector of attributes and a go-around indicator. Based on prior research and data availability in Section 2, we derive five groups of features at each cutoff gate, as summarized in Table 3; and they can be further categorized into either static or dynamic features. Static features, such as flight-specific information and airport configurations, remain invariant during the final approach. In contrast, dynamic features such as altitude and speed will change through the approach phase. The entire dataset undergoes a Train-Validation-Test split (70%, 20%, 10%) for model training, early-stopping validation, and testing, respectively, with a pre-split random shuffle to ensure representativeness. Numerical features are standardized by gate based on training set statistics, the same statistics are then applied to the validation and test sets to maintain consistency. More details on the feature engineering process are available in 11) and 12).

3.1.2. Model architecture and training

Utilizing trajectory data with environmental and airport operational conditions, our goal is to convert the multivariate time series of the approach phase into sequential go-around probability predictions. Hence, we employ a Long Short-Term Memory (LSTM) network, a type of recurrent neural network designed to capture temporal dependencies in sequential data.

To accommodate the variable lengths of flight sequences of go-around flights and normal flights, we implement padding to streamline the input data, facilitating the encoding of sequences into contiguous batches for the LSTM network. The network features an input layer shaped to process the temporal sequence data, followed by a masking layer designed to manage the instances of early go-arounds. For flights undergoing a go-around prior to the terminal gate, all ensuing data points are masked, ensuring that model predictions rely solely on pertinent pre-go-around data, thus preserving the temporal sequence integrity and reflecting the constraints of operational practice.

At the core of the model, LSTM units are deployed to navigate the temporal dynamics of the aircraft's approach and landing phase. A dropout layer is added to prevent overfitting by randomly deactivating neurons during training iterations. Subsequent layers include a time-distributed dense layer, which fosters a robust connection at each timestep; and finally, an output layer with a sigmoid activation function that computes the likelihood of a go-around at each gate.

In the training phase, we choose the Adam optimizer for its adaptive learning rate capabilities, and an early stopping protocol to optimize convergence and prevent overfitting. Given the extreme class imbalance in our dataset, with a minimal go-arounds fraction (0.39%), we adopted the binary focal cross-entropy loss (BFL) function, which is designed to enhance model sensitivity to the underrepresented class.[16] The BFL function is particularly advantageous for our problem as it integrates a focusing parameter(γ) and a balancing term(α), effectively accentuating the correct classification of the infrequent yet critical go-around events. It is computed as:

$$L\left(y, \hat{y}\right) = -\alpha y\left(1-\hat{y}\right)^{\gamma} log\left(\hat{y}\right) - (1-y)\hat{y}^{\gamma} log\left(1-\hat{y}\right) \quad (1)$$

where $y \in \{0, 1\}$ is a true go-around label; $\hat{y} \in [0, 1]$ is an estimate of go-around probability; α is set at 0.95 to underscore the importance of the minority class; γ is set at 2, echoing the recommendations from the focal loss literature.[16)]

To obtain the optimal model configuration, we employ a grid search to tune hyperparameters. This process iterates through an expansive pool of 1,296 candidate permutations, derived from an array of LSTM units (128, 256), dense units (64, 128, 256), batch sizes (64, 128, 256, 512), activation functions (ReLU, tanh, swish), learning rates (0.01, 0.001), and dropout rates (0.0, 0.1, 0.2) to regularize the network. In addition, the loss functions within this tuning matrix encompass the standard binary cross-entropy and the binary focal cross-entropy, with an alpha value of 0.95 to account for class imbalance. Training incorporated early stopping callbacks to curb overfitting, while precision, recall, and AUC metrics gauged model performance. Each model iteration, along with its history, was preserved for analysis.

3.1.3. Model evaluation

To evaluate the LSTM models for go-around prediction, we first use the test set to identify the optimal probability threshold for each model candidate. Each candidate model generates a probability $\hat{P}_{ij}$ indicating the go-around probability from gate j to gate 20 at the runway threshold for each test flight. A flight with $\hat{P}_{ij}$ higher than the picked threshold will be viewed as a predicted go-around, and lower than that as a predicted non-go-around. We employ a grid search with a range of 0.1 to 1.0 to select the threshold that maximizes the F2 score (2) at each gate. This fine-tuning ensures the model sensitivity is finely calibrated in detecting go-arounds at each gate along the approach path.

We then assess model performance using five key metrics: AUC-ROC, precision, recall, accuracy, and F2 score. Finally, the optimal model is selected based on the highest mean F2 scores across all gates. As the cost of missing a go-around detection highly outweighs the cost of getting a false alarm, we select the F2 score for its weighted emphasis on recall, prioritizing the detection of actual go-around instances over avoiding false alarms, computed as (2).

$$F2 = \frac{\left(1+2^2\right)\times precision+recall}{2^2\times precision+recall} \quad (2)$$

where $precision = \frac{TP}{TP+FP}$ and $recall = \frac{TP}{TP+FN}$ with TP: true positive, FP: false positive, FN: false negative.

3.2. Model explainability

This section presents methods for identifying key factors influencing go-around decisions through a global analysis to determine the main predictors.

We conducted a permutation-based importance analysis on the test dataset to pinpoint dynamic features—those varying during flight arrival as detailed in Section 3.1.1.—that significantly affects go-around probability, indicating strong predictive power within the model. Our methodology, outlined in Table 4, emphasizes conducting N (e.g., 20) iterations for each feature at each gate to counteract random data fluctuations and solidify confidence in the identified feature weights.[17,18] The BFL metric, chosen for its sharp sensitivity to class imbalances, serves as a criterion for assessing feature importance, focusing specifically on dynamic features crucial for predicting go-arounds.

Table 4. Algorithm on feature importance using permutation.

INPUT: Trained LSTM model (m), test set features (X), go-around labels (y)
Compute original model performance using F2 score: Ori = BFL(y, m(X))
for each cutoff gate j in {0, ..., 19} do
for each dynamic feature i in X do
    Initialize an empty list to store importance scores: IptScores = []
        for iteration in {1, ..., 20} do
            Randomly shuffle the values of feature i at gate j to create a perturbed feature matrix X_perm
            Compute the F2 score on the perturbed dataset: Perm = F2(y, m(X_perm))
            Calculate the importance score for the current iteration: PI_{ij} = Ori - Perm
            Append PI_{ij} to IptScores
            end for loop
        Compute the average importance score for feature i at gate j: β_{ij} = Average(IptScores)
end for loop
end for loop
OUTPUT: A matrix of averaged importance scores β, where each entry β_{ij} represents the importance of feature i at gate j.

4. Model Results

This section discusses the results of training and validating the LSTM model for go-around detection for JFK and SFO, respectively.

4.1. Model training and performance

After data cleaning, the final dataset includes **585,680 flights with 2,302 go-arounds (0.393%) for JFK**, and **516,789 flights with 3,611 go-arounds (0.699%) for SFO**. For JFK, the data split results in a training set with 1,611 go-arounds, a validation set with 460 go-arounds, and a test set with 231 go-arounds. For SFO, the data split results in a training set with 2,527 go-arounds, a validation set with 722 go-arounds, and a test set with 362 go-arounds.

The model, selected from 1,296 candidates via grid search for its best average F2 score of 0.7922, features 128 LSTM units and a dense layer with 256 units. The 'ReLU' activation function, with a Binary Focal Loss (BFL) function weighted with an alpha of 0.95, improves the model's ability to predict sparse go-around events.

The initial pilot model, selected from 1,296 candidates in the GAP 2.0 model, was based on the 2019 data from JFK Airport. Using a grid search method, the model was chosen for its optimal average F2 score of 0.2910. The architecture includes 256 LSTM units and a dense layer with 128 units. The 'tanh' activation function, with a Binary Focal Loss (BFL) function weighted with an alpha of 0.95, improves the model's ability to predict sparse go-around events. Currently, the optimal model candidate for the combined 3-year datasets is undergoing hyperparameter tuning on the Berkeley High Performance Computing (HPC) service. The extensive dataset and large candidate pool necessitate the use of HPC resources to efficiently explore and refine the model parameters.

Table 5 and Table 6 report the testing set performance of the pilot models, including the F2 score, recall, and precision. Emphasizing recall is crucial because the primary goal is to ensure that go-around predictions are not missed. The "omitted go-around" column indicates the number of go-around flights omitted after they have been initiated, which ensures the performance metrics, specifically recall and precision, not to be inflated by subsequent timestamps after a go-around event has already been detected.

For the JFK model in Table 5, the probability thresholds, varying from 0.25 to 0.40 across different gates, reflect the adaptive strategy to capture the varying likelihoods of go-around events. The model performance shows F2 scores ranging from approximately 0.63 to 0.81 across different gates, with a gate-average value of 0.765. Recall (i.e., $\frac{TP}{TP+FN}$) values vary between 0.79 and 0.93, demonstrating the model's strong capability to detect go-arounds, albeit with some variability. Precision (i.e., $\frac{TP}{TP+FP}$) values at JFK range from 0.34 to 0.61, indicating a higher rate of false positives compared to SFO. Accuracy values consistently hover around 0.99, as is expected for a highly imbalanced data set.

For the SFO model in Table 6, the probability thresholds, varying from 0.32 to 0.42 across different gates, indicate a slightly narrower range for go-around detection. The model performance shows F2 scores ranging from approximately 0.53 to 0.70 across different gates, with a gate-average value of 0.652. Recall values are consistently high, ranging from 0.69 to 0.82, indicating the model's robust ability to detect go-around events. Precision values for SFO range from 0.28 to 0.50, reflecting a better balance between capturing as many go-arounds as possible and the occurrence of false positives. Accuracy values for SFO are also around 0.99, similar to JFK, but with a generally higher recall.

The comparative analysis highlights the strengths and weaknesses of the model at both airports. At JFK, the model achieves higher F2 scores and recall values, indicating its robustness in detecting go-arounds while maintaining reasonable precision. In contrast, the model at SFO shows variability in performance, with lower F2 scores and precision, suggesting the need for further refinement to reduce false positives and improve the balance between recall and precision. These insights are crucial for understanding the operational effectiveness of the go-around detection system and guiding future enhancements. By focusing on improving the model's performance at SFO, we can achieve more consistent and reliable go-around detection across different airport environments. The ongoing hyperparameter tuning and data analysis will be key to refining the model and enhancing its predictive capabilities.

Table 5. Pilot LSTM model performance for JFK (2019, 2022, 2023 data).

gate	omitted go-around	thd	f2_score	precision	recall	accuracy
0	0	0.315	0.7175	0.4638	0.8312	0.9955
1	0	0.365	0.7803	0.5904	0.8485	0.9971
2	0	0.380	0.7909	0.5958	0.8615	0.9971
3	2	0.380	0.7876	0.5806	0.8646	0.9970
4	4	0.385	0.7859	0.5782	0.8634	0.9970
5	5	0.385	0.7903	0.5833	0.8673	0.9971
6	9	0.315	0.7885	0.4976	0.9234	0.9962
7	11	0.315	0.7899	0.5012	0.9227	0.9963
8	15	0.395	0.7924	0.5918	0.8657	0.9973
9	19	0.410	0.7902	0.6186	0.8491	0.9976
10	24	0.305	0.7954	0.5066	0.9275	0.9965
11	35	0.390	0.8092	0.6070	0.8827	0.9977
12	48	0.380	0.8093	0.5927	0.8907	0.9977
13	58	0.365	0.7884	0.5588	0.8786	0.9976
14	74	0.345	0.7748	0.5167	0.8854	0.9975
15	93	0.310	0.7565	0.4713	0.8913	0.9974
16	113	0.315	0.7410	0.4619	0.8729	0.9977
17	130	0.330	0.7095	0.4468	0.8317	0.9979
18	145	0.325	0.6849	0.4192	0.8140	0.9981
19	164	0.305	0.6250	0.3397	0.7910	0.9980

Table 6. Pilot LSTM model performance for SFO (2019, 2022, 2023 data).

gate	omitted go-around	thd	f2_score	precision	recall	accuracy
0	0	0.320	0.5346	0.2826	0.6878	0.9855
1	0	0.375	0.6508	0.3791	0.7928	0.9894
2	1	0.370	0.6580	0.3768	0.8089	0.9893
3	3	0.375	0.6643	0.3784	0.8189	0.9894
4	7	0.410	0.6749	0.4492	0.7718	0.9919
5	11	0.405	0.6870	0.4526	0.7892	0.9921
6	16	0.410	0.6809	0.4589	0.7746	0.9923
7	29	0.405	0.6765	0.4487	0.7748	0.9924
8	38	0.400	0.6724	0.4360	0.7778	0.9923
9	53	0.390	0.6616	0.4081	0.7832	0.9919
10	60	0.395	0.6527	0.4152	0.7616	0.9923
11	76	0.390	0.6538	0.4048	0.7727	0.9924
12	96	0.375	0.6323	0.3680	0.7707	0.9920
13	123	0.365	0.6238	0.3412	0.7866	0.9919
14	150	0.370	0.6109	0.3354	0.7689	0.9928
15	163	0.390	0.6304	0.4033	0.7337	0.9948
16	185	0.405	0.6673	0.4698	0.7458	0.9962
17	213	0.405	0.6953	0.4978	0.7718	0.9971
18	235	0.370	0.6809	0.4232	0.8032	0.9968
19	273	0.415	0.6323	0.4114	0.7303	0.9977

4.2. Model explainability

This section presents the interpretability of our LSTM network, focusing on both global and local factor contributions. The goal is to identify key predictors of go-arounds and understand how specific flight conditions affect predictions in real-time.

In the global analysis of our LSTM model, we employed Permutation Importance to assess the impact of dynamic features across different cutoff gates. The F2 score was used as the evaluative metric due to its robustness against the class imbalance prevalent in our dataset. To ensure stability in our findings, the permutation process for each feature at each gate was iterated 20 times. The resulting measures of feature importance are visualized as boxplots in Fig. 1 and 2, which show the top 10 features for each specific gate. The x-axis of these plots lists the feature names, while the y-axis quantifies their importance based on the observed variation in the F2 score from the permutations.

In Fig. 1 of JFK, the plots indicate that *lead_rwy_time* and separation consistently emerge as the most significant predictors across multiple gates. *lead_rwy_time* captures the duration the leading aircraft occupies the runway, which is critical for understanding potential go-around scenarios due to runway occupancy. separation measures the distance between aircraft, a crucial factor in determining whether a go-around might be necessary for safety. Some features, such as *#Obj_rwy*, *altitude/horizontal deviation*, and *energy level*, show relatively modest importance at specific gates, suggesting the potential contribution of these features.

In Fig. 2 of SFO, Similar to JFK, *lead_rwy_time* and *separation* are highlighted as significant predictors, indicating their robust predictive power across different airports. The consistency of these features' importance at SFO underscores their universal relevance in go-around detection models. Features like *altitude deviation* and *energy* also show significant importance at various gates, reflecting their roles in the stability and performance of aircraft during landing approaches. As with JFK, certain features such as *groundspeed* exhibit minimal importance, reaffirming the potential for noise or less predictive value in these variables.

The permutation importance analysis provides valuable insights into which features most significantly influence the LSTM model's predictions of go-arounds. The prominence of *lead_rwy_time* and *separation* across both airports suggests that these factors are universally critical in predicting go-arounds. These features likely represent the immediate operational environment's constraints and demands, directly affecting the decision-making process for pilots and air traffic controllers.

The result displays the ability of our prediction model in real-time risk assessment and its usefulness as a decision-support tool in the landing phase. The feature contribution analysis offers explanations for the predicted go-around probabilities, which can potentially help the pilots and air traffic controllers develop measures to mitigate the go-around risks.

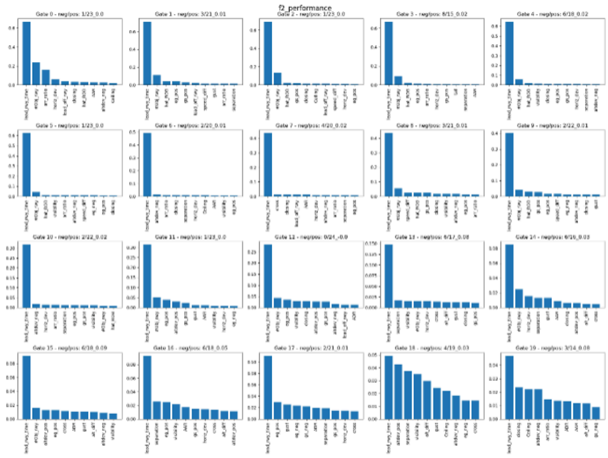

Fig. 1. JFK: Global interpretation by 20 gates.

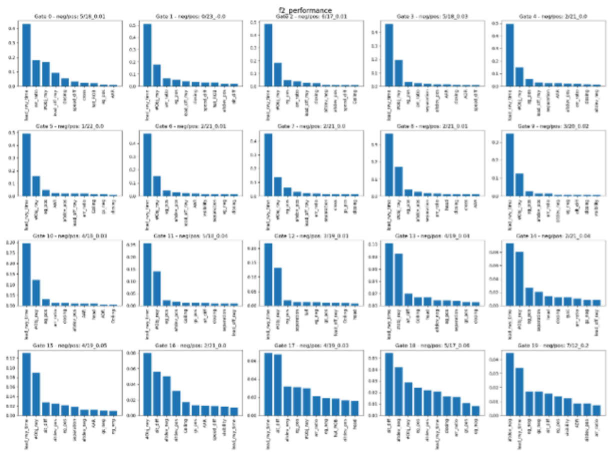

Fig. 2. SFO: Global interpretation by 20 gates.

5. Conclusion and Discussion

In this paper, we employ the LSTM network to predict the real-time go-around probability as a flight is approaching the airport within 10 nm of the landing runway threshold. LSTM model is chosen for its ability to capture the time dependency in the flight sequence and provide an evolutionary and consistent go-around probability. We further develop methods to analyze the feature contribution to go-around occurrence. From our results, *runway occupancy time* of the leading aircraft, as well *#obj_rwy* appear the most important predictive features, although other features become moderately important for certain gates and airports. We have integrated these pre-trained models and analyses with real-time data streaming, and finally develop a demo web-based user interface that integrates the different components designed previously into a real-time tool that can eventually be used by flight crews and other line personnel to identify situations in which there is a high risk of a go-around.

Future work can enhance our predictive model by expanding the feature space with more detailed datasets, incorporating granular aircraft dynamics (e.g., deceleration rates, descent rates, heading degrees) and weather conditions (e.g., wind shear). Improving data quality and aligning feature engineering with aviation standards can also increase model accuracy. Structurally, introducing gate-specific weighted dense layers could refine the temporal analysis of flight approaches, and adding more dense layers may deepen insights into feature interactions. The go-around detection algorithm and prediction model have been adapted for other airports, specifically SFO in our case, with minimal modification and similar accuracy. Further avenues for development include expanding the dataset with additional years of data and adapting the models for more airports.

References

1) R. C. Jou, C. W. Kuo, and M. L. Tang, “A study of job stress and turnover tendency among air traffic controllers: The mediating effects of job satisfaction,” Transportation research part E: logistics and transportation review, vol. 57, pp. 95-104, 2013.
2) X. Prats, V. Puig, J. Quevedo, and F. Nejjari, “Multi-objective optimisation for aircraft departure trajectories minimising noise annoyance,” Transportation Research Part C: Emerging Technologies, vol. 18(6), pp. 975-989, 2010.
3) J. Shortle, and L. Sherry, “A Model for Investigating the Interaction between Go-Arounds and Runway Throughput,” in 2013 Aviation Technology, Integration, and Operations Conference, pp. 4235, 2013.
4) T. Blajev, and W. Curtis, “Go-around decision-making and execution project,” Final Report to Flight Safety Foundation, 2017.
5) A. Campbell, P. Zaal, J. A. Schroeder, and S. Shah, “Development of possible go-around criteria for transport aircraft,” in 2018 Aviation Technology, Integration, and Operations Conference, pp. 3198, 2018.
6) P. Zaal, A. Campbell, J. A. Schroeder, and S. Shah, “Validation of Proposed Go-Around Criteria Under Various Environmental Conditions”, In AIAA Aviation 2019 Forum, pp. 2993, 2019.
7) A. Campbell, P. Zaal, S. Shah, and J. A. Schroeder, “Pilot Evaluation of Proposed Go-Around Criteria for Transport Aircraft”, in AIAA Aviation 2019 Forum, pp. 3610, 2019.
8) A. M. Campbell, P. M. Zaal, S. R. Shah, and J. A. Schroeder, “Go-around criteria refinement for transport category aircraft,” Journal of Air Transportation, vol. 30(1), pp. 3-22, 2022.
9) FAA Air Traffic Organization, “Air Traffic By the Numbers,” pp. 19, 2023.https://www.faa.gov/air_traffic/by_the_numbers/media/Air_Traffic_by_the_Numbers_2023.pdf.
10) S. Ember, and E. Steel, “Airline Close Calls Happen Far More Often Than Previously Known,” The New York Times, August 2023. Link: https://www.nytimes.com/interactive/2023/08/21/business/airline-safety-close-calls.html.
11) L. Dai, Y. Liu, and M. Hansen, “Modeling go-around occurrence using principal component logistic regression,” Transportation Research Part C: Emerging Technologies, vol. 129, pp. 103262, 2021
12) L. Dai, and M. Hansen, “Real-Time Prediction of Runway Occupancy Buffers,” in 2020 International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT), IEEE, pp. 1-11, 2020.
13) L. Dai, Y. Liu, and M. Hansen, “Predicting go-around occurrence with input-output hidden Markov model,” in International Conference on Research in Air Transportation, September 2020.
14) Bureau of Transpiration Statistics, “Airport Rankings 2019”, 2020. https://www.bts.dot.gov/airport-rankings-2019.
15) International Airport Review, “Port Authority board approves $355 million runway revamp at JFK Airport”, 2018, https://www.internationalairportreview.com/news/76160/runway-revamp-jfk-airport/.
16) T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in Proceedings of the IEEE international conference on computer vision, pp. 2980-2988, 2017.
17) X. Cheng, H. Shen, Y. Huang, Y. -L. Cheng and J. Lin, "Using Mobile Charging Drones to Mitigate Battery Disruptions of Electric Vehicles on Highways," 2024 Forum for Innovative Sustainable Transportation Systems (FISTS), Riverside, CA, USA, 2024, pp. 1-6, doi: 10.1109/FISTS60717.2024.10485588.
18) K. Liu, M. Hansen, “Excess Delay from GDP: Measurement and Causal Analysis,” in International Conference on Research in Air Transportation, 2022.