



(RESEARCH ARTICLE)



Data-driven prediction of runway incursions using random forest and temporal validation

Wellington Mandibaya *, Yong Tian and Albertiny Z. T. Monteiro

College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China.

International Journal of Science and Research Archive, 2025, 17(01), 235-247

Publication history: Received on 26 August 2025; revised on 04 October 2025; accepted on 06 October 2025

Article DOI: <https://doi.org/10.30574/ijrsra.2025.17.1.2772>

Abstract

This study develops a predictive framework for runway incursions at major U.S. airports by integrating aviation-specific feature engineering with Random Forests and temporal validation. Using data from 60 hub airports (2015–2024), composite indices captured weather, infrastructure, and operational pressures. Random Forest outperformed mean, OLS, and Poisson baselines ($R^2=0.789$, $RMSE=2.99$). Cross-validation confirmed stable generalization ($R^2=0.757$, 95% CI [0.629, 0.885]). Statistical tests showed significant COVID-19 decline and identified VFR conditions as the strongest predictor. Outliers were retained for robustness, and multicollinearity checks validated Random Forest resilience. The framework complements the European Organization for the Safety of Air Navigation (EUROCONTROL) guidance, supporting risk-based staffing, monitoring, and training.

Keywords: Aviation safety; Random Forest; Runway incursions; Visual Flight Rules; Machine Learning

1. Introduction

The growth of global air traffic has intensified operational complexity in airport surface movement areas, where runway incursions remain a persistent safety concern [1]. Defined by the Federal Aviation Administration (FAA) as “any occurrence at an aerodrome involving the incorrect presence of an aircraft, vehicle, or person on the protected area of a surface designated for landing and take-off”, incursions range from minor Category D events to serious Category A near-collisions. Globally, the issue has drawn sustained attention from regulators, industry stakeholders, and professional associations. In the United States, the FAA reported about 1700 incursions in 2023 alone [1], while the National Plan of Integrated Airport Systems (NPIAS) highlights runway safety as a critical priority for system-wide infrastructure and operational planning [2]. Similar concerns have been raised internationally, with industry groups such as the Air Line Pilots Association (ALPA) calling for coordinated action to address the rising complexity of airport operations [3]. Europe, through EUROCONTROL, has adopted the European Action Plan for the Prevention of Runway Incursions (EAPPRI), emphasizing standardized procedures and harmonized mitigation strategies across member states [4]. Despite ongoing investments in safety technologies and procedural reforms, incursions continue to occur at concerning levels worldwide. This persistence suggests that current measures, while beneficial, are insufficient on their own to anticipate and mitigate risk. Accordingly, there is growing recognition of the need for predictive, data-driven approaches that can complement traditional safety programs by identifying periods and locations of elevated incursion risk before events occur

This study advances both method and practice. Methodologically, we provide a rigorously validated machine learning framework for forecasting runway incursions that incorporates aviation-specific feature engineering and temporal validation (hold-out years and time-series cross-validation). Substantively, we uncover a counterintuitive domain insight: incursions are more frequent under Visual Flight Rules (VFR) conditions, likely driven by traffic intensity and complacency, rather than by poor weather. Together, these contributions demonstrate the dual value of machine

* Corresponding author: Wellington Mandibaya

learning for predictive safety management: robust forecasting accuracy and actionable explanatory insights for Safety Management Systems (SMS) and EAPPRI implementation.

2. Literature Review

2.1. Geometric and Infrastructural Factors

Airport geometry has long been recognized as a critical driver of incursion risk. Studies consistently show that intersecting runways, complex taxiway configurations, and conflict points elevate both the frequency and severity of incursions [7-9]. Biernbaum and Hagemann, for instance, demonstrated that additional intersections increase severity, while parallel runways mitigate it. These findings align with regulatory frameworks such as the EAPPRI, which highlights standardized procedures and hot-spot management as essential risk controls. However, this stream of research is largely descriptive: it identifies correlations but does not quantify risk under varying operational conditions, limiting its predictive and policy utility.

2.2. Machine Learning in Aviation Safety

Recent studies have turned to machine learning to address these limitations. Random Forests and related ensemble methods have shown strong predictive accuracy in safety-critical contexts because they reduce variance, manage multicollinearity, and capture nonlinearities [12-14]. Yet, most prior applications in aviation safety rely on random cross-validation, raising concerns about their suitability for forecasting tasks where temporal generalization is crucial. Furthermore, many studies treat machine learning as a black box, without domain-specific feature engineering that ensures interpretability and operational relevance. Our study advances this literature by combining Random Forests with time-aware validation and engineered indices grounded in aviation operations and meteorology, thereby balancing predictive performance with interpretability.

2.3. International Guidance and Policy Context

Alongside empirical research, international bodies such as the International Civil Aviation Organization (ICAO) and EUROCONTROL emphasize systematic approaches to runway safety through initiatives like ICAO Doc 9870 and EAPPRI v3.0. These frameworks stress procedural reforms, technological safeguards and safety management systems, but they stop short of predictive analytics. By introducing a validated, forward-looking model, our work complements these policy initiatives with a tool that enables risk-based staffing, monitoring, and training during periods forecasted to be high risk.

Table 1 provides a summary of selected studies that have examined runway incursions and related methodologies. Collectively, this literature underscores the importance of airport geometry and traffic factors, highlights the utility of causal and expert-based methods, and demonstrates the emerging role of machine learning for safety-critical prediction tasks.

Table 1 Summary of selected runway incursion studies

| Authors | Focus | Method |
|-----------------------------|------------------------------|----------------------|
| Johnson et al. (2016) | Intersections & frequency | Regression |
| Wilke et al. (2015) | Severity drivers | Logistic regression |
| Biernbaum & Hagemann (2012) | Severity & Causes | Statistical analysis |
| Goodheart (2018) | Causal pathways | Bayesian networks |
| Rogerson & Lambert (2012) | Risk prioritization | Expert weighting |
| Claros et al. (2017) | Frequency modeling | Count regression |
| Breiman (2001) | Random Forest Application | Ensemble model |
| Koul (2025) | Machine Learning in aviation | Review |

2.4. Research Gap

Despite substantial progress, existing approaches seldom integrate operational intensity, infrastructural complexity, and meteorological variability into a unified predictive framework. Regression-based methods, such as those by Claros et al. [12], have been applied, but they often lack temporal validation, raising concerns of overfitting and limiting their applicability to future operational contexts. Consequently, the literature provides limited evidence on the effectiveness of robust, data-driven methods for forecasting runway incursion risk.

This study addresses the identified gap by:

- Constructing a longitudinal dataset of the 60 busiest airport hubs in the United States according to the FAA, spanning from 2015–2024.
- Engineering composite indices that capture weather complexity, infrastructure layout, and operational pressure.
- Applying a Random Forest framework, enhanced with aviation-specific feature engineering and validated with time-series cross-validation, to predict annual runway incursions.

This study advances beyond descriptive analyses and isolated technology evaluations through the development of a generalizable, data-driven model that delivers actionable insights for risk-based resource allocation and proactive safety management, consistent with recent advances in applying machine learning to aviation safety.

2.5. Data Acquisition and Feature Engineering

2.5.1. Dataset Sources

This study uses a longitudinal dataset covering 60 of the busiest U.S. airports classified as large and medium hubs, spanning the years 2015–2024. These airports were selected based on the FAA’s hub classification system, which categorizes airports according to their share of total U.S. passenger enplanements. Collectively, these hubs account for the majority of commercial traffic in the United States, making them critical focal points for runway safety analysis. The dataset integrates several publicly available sources:

- *Runway incursions*: Total annual incursions (and subcategories) obtained from FAA Runway Safety data. The flight conditions at which the incursion occurred categorized as Visual Flight Rules (VFR), Marginal Visual Flight Rules (MVFR), Instrument Flight Rules (IFR), and Low Instrument Flight Rules (LIFR).
- *Airport geometry and infrastructure*: Number of runways, runway intersections, and complex crossings derived from FAA 5010 master records.
- *Operational intensity*: Annual operations, scheduled arrivals and departures, and traffic intensity, sourced from the Air Traffic Activity Data System (ATADS)
- *Meteorological conditions*: Precipitation (PRCP), snowfall totals (TSNW), and days with measurable precipitation (DP01) aggregated from NOAA’s Integrated Surface Database.
- *Safety technologies*: Deployment of Airport Surface Detection Equipment Model-X (ASDE-X) and Runway Status Lights (RWSL), coded as binary indicators using FAA deployment records.

The dataset contains 61 predictors and one target variable (total annual incursions). Across all airport-years, the mean number of incursions was 8.6 per airport-year (range: 0–47). High-traffic hubs such as ATL, ORD, and LAX reported the highest totals. Table 2. lists the 30 busiest airports ranked by total operations (2015–2024), alongside incursions, runways, and runway intersections.

Table 2 Top 30 busiest airports by operations (2015-2024) with incursions, runways and intersections

| FAA ID | Total Incursions | Airport Operations | Runways | Intersections |
|--------|------------------|--------------------|---------|---------------|
| ATL | 168 | 8012018 | 5 | 0 |
| ORD | 267 | 7863862 | 8 | 3 |
| LAX | 216 | 6050765 | 4 | 0 |
| LAS | 161 | 5313910 | 4 | 2 |
| CLT | 137 | 5305735 | 3 | 0 |

| | | | | |
|-----|-----|---------|---|---|
| IAH | 97 | 4301773 | 5 | 0 |
| PHX | 74 | 4253365 | 3 | 0 |
| JFK | 74 | 4202600 | 4 | 2 |
| MIA | 117 | 4118247 | 4 | 1 |
| SEA | 96 | 4027316 | 3 | 0 |
| EWR | 86 | 3970449 | 3 | 2 |
| SFO | 196 | 3894118 | 4 | 4 |
| DEN | 100 | 3774325 | 6 | 0 |
| BOS | 210 | 3723125 | 6 | 6 |
| MSP | 131 | 3571156 | 4 | 3 |
| MCO | 30 | 3436207 | 4 | 0 |
| DFW | 194 | 3380736 | 7 | 0 |
| DTW | 140 | 3362518 | 6 | 4 |
| PHL | 109 | 3333098 | 4 | 1 |
| LGA | 73 | 3259025 | 2 | 1 |
| SLC | 125 | 3228434 | 4 | 0 |
| SNA | 105 | 3110473 | 2 | 0 |
| HNL | 152 | 3008128 | 6 | 2 |
| FLL | 54 | 2914204 | 2 | 0 |
| IAD | 50 | 2780785 | 4 | 0 |
| ANC | 98 | 2759997 | 3 | 0 |
| DCA | 124 | 2696239 | 3 | 3 |
| BWI | 49 | 2367178 | 3 | 1 |
| DAL | 99 | 2241505 | 2 | 0 |
| MDW | 176 | 2239925 | 4 | 4 |

The final dataset comprises 600 airport-year observations, with no missing values after pre-processing. Figure. 1 shows the geographic distribution of the sampled airports, emphasizing their concentration among the nation's busiest hubs.

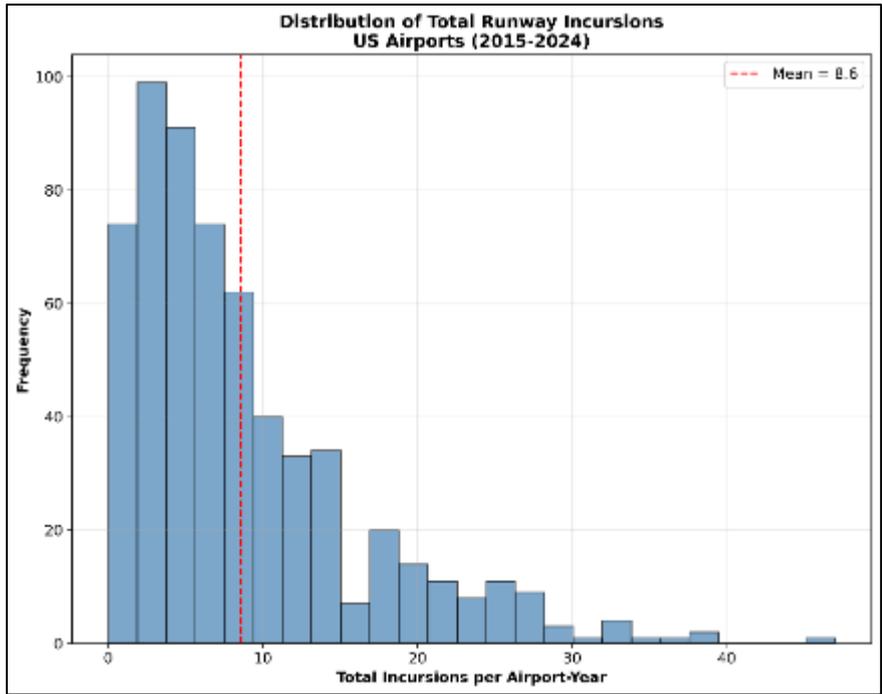


Figure 1 Distribution of total incursions per airport-year, showing right-skewed distribution with heavy-tailed outliers

2.6. Feature Engineering

The complex influences of operational, infrastructural, and meteorological factors underlying incursion dynamics were captured through the construction of multiple engineered indices as shown on Table 3. These indices reduce dimensionality while preserving domain interpretability.

Table 3 Engineered indices formulas

| Index | Formula |
|---------------------------|---|
| Weather Complexity | $0.3 \times PRCP + 0.7 \times TSNW + 0.1 \times DP01$ |
| Flight Condition Risk | $0.4 \times LIFR + 0.3 \times IFR + 0.2 \times MVFR + 0.1 \times VFR$ |
| Infrastructure Complexity | $0.2 \times Runways + 0.3 \times Runway Intersections + 0.3 \times Complex Taxiway Intersections + 0.2 \times Runway Crossings$ |
| Operational Pressure | $0.4 \times Traffic Intensity + 0.3 \times Taxi Delay + 0.3 \times Gate Delay$ |
| Technology Score | $ASDE-X + RWSL$ |

All raw inputs were aggregated to the annual airport level before index construction. Weights were assigned according to operational severity and prior literature: snow events (TSNW) received the highest weight in the weather index due to their disproportionate impact on braking action and visibility; visibility categories were weighted by severity hierarchy, with LIFR posing the greatest constraint; intersections and complex geometries were emphasized due to their documented link to elevated incursion severity; and traffic intensity was prioritized in the operational index as the main driver of throughput-related stress. Outliers were retained as they represent genuine, high-consequence operational scenarios that the model must be able to predict. Variables directly derived from the target (e.g., incursion subcategories, severity levels, and calculated rates) were excluded as a way to mitigate data leakage. A temporal split was implemented, with 2015–2022 data used for training and 2023–2024 reserved for testing, ensuring evaluation reflects out-of-sample predictive validity under realistic operational conditions

2.7. Modelling Framework

2.7.1. Notations and Variables

Let y_{it} denote the number of total runway incursions at airport i in year t . This is the dependent variable to be predicted.

Let $x_{it} = (x_{1,it}, x_{2,it}, \dots, x_{p,it})$ represent the vector of p predictor variables for airport i , year t . For Illustration:

$x_{1,it}$: Airport Operations

$x_{2,it}$: Number of Runways

$x_{3,it}$: Runway Intersections

... up to $x_{p,it}$ representing the full set of predictors in the dataset.

Thus, the general feature vector can be written as:

$$x_{it} = (x_{1,it}, x_{2,it}, x_{3,it}, \dots, x_{p,it})$$

The error term ε_{it} captures unobserved factors not explained by the predictors.

2.8. Random Forest Approach

The primary model employed is the Random Forest Regressor, selected for its robustness to multicollinearity, ability to capture nonlinear relationships, and established performance in safety-critical domains [13], [14].

Formally, the Random Forest prediction for a new observation x is defined as the average of B decision trees:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x) \quad (1)$$

where B is the total number of trees, and $T_b(x)$ is the prediction from the b -th tree.

Each tree T_b is constructed as follows:

Bootstrap sampling: draw a sample of size n (with replacement) from the training data.

Random feature subset: at each split, select a random subset of predictors of size $m \ll p$, where p is the total number of predictors.

Split criterion: for regression, the algorithm selects the feature $j \in m$ and split point s that minimize the mean squared error (MSE):

$$\text{MSE}(j, s) = \frac{1}{N_t} \left(\sum_{i: x_{ij} \leq s} (y_i - \bar{y}_L)^2 + \sum_{i: x_{ij} > s} (y_i - \bar{y}_R)^2 \right) \quad (2)$$

where N_t is the number of samples at node t , and \bar{y}_L, \bar{y}_R are the mean responses in the left and right child nodes respectively.

Recursive Partitioning: Continue splitting until stopping criteria are reached (e.g., maximum depth, minimum leaf size).

At prediction time, each tree assigns x to one of its terminal regions R_m and outputs the average response of observations in that region:

$$T_b(x) = \sum_{m=1}^{M_b} c_m \cdot 1_{\{x \in R_m\}} \quad (3)$$

where M_b is the number of terminal nodes in tree b , c_m is the mean response in region R_m and $1_{\{\cdot\}}$ is the indicator function.

Aggregating across all trees produces the Random Forest estimator:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B \left(\sum_{m=1}^{M_b} c_m \cdot 1_{\{x \in R_m\}} \right) \quad (4)$$

For classification tasks (e.g., distinguishing high-risk vs. low-risk airports), the Random Forest Classifier uses majority voting:

$$\hat{C}(x) = \arg \max_k \sum_{b=1}^B 1_{\{T_b(x) = k\}} \quad (5)$$

where k indexes the risk categories.

Hyperparameters for the Random Forest were optimized using GridSearchCV with a time-series cross-validation scheme to preserve chronological ordering. Candidate ranges were informed by prior studies and practical constraints:

Number of trees ($n_estimators$): [100, 200, 500]

Maximum tree depth (max_depth): [5, 10, 15, None]

Minimum samples per split ($min_samples_split$): [2, 5, 10]

Minimum samples per leaf ($min_samples_leaf$): [1, 2, 4]

Maximum features considered at each split ($max_features$): ["sqrt", "log2"]

The tuning process optimized the mean squared error (MSE) on the validation folds. Final parameters were selected as those yielding the lowest average validation error, while balancing against overfitting. The chosen configuration provided both strong predictive accuracy and stable generalization across temporal splits.

2.9. Baseline Models

Random Forest performance was benchmarked against two statistical baselines: Ordinary Least Squares (OLS) regression and the Poisson Generalized Linear Model (GLM).

2.10. Ordinary Least Squares (OLS) Regression

The OLS model assumes an additive and linear relationship between predictors and incursions:

$$y_{it} = \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \beta_3 x_{3,it} + \dots + \beta_p x_{p,it} + \varepsilon_{it} \quad (6)$$

Here, y_{it} denotes the total incursions at airport i in year t , $x_{j,it}$ are the predictor variables, β_j are regression coefficients, and ε_{it} is the error term.

The OLS estimator chooses the coefficients β_j that minimize the residual sum of squares (RSS):

$$Q(\beta) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{1i} - \beta_2 x_{2i} - \beta_3 x_{3i} \dots - \beta_p x_{pi})^2 \quad (7)$$

2.10.1. Poisson Generalized Linear Model (GLM)

The Poisson GLM is widely applied to model count data such as incursions. It assumes that the number of incursions y_{it} follows a Poisson distribution with mean μ_{it} :

$$y_{it} \sim \text{Poisson}(\mu_{it})$$

with the log link function:

$$\ln(\mu_{it}) = \beta_0 + \beta_1 x_{1,it} + \beta_2 x_{2,it} + \dots + \beta_p x_{p,it} \quad (8)$$

The Poisson model ensures nonnegative predictions but rests on the equidispersion assumption $\text{Var}(y_{it}) = \mu_{it}$. Since incursion data are heavy-tailed and often overdispersed, Poisson may underestimate variance, making it a conservative baseline.

The OLS and Poisson GLM provide complementary statistical comparisons for assessing the added value of Random Forest in handling nonlinearities, interactions, and overdispersion. The performance is evaluated using R^2 , RMSE, and MAE on the temporal test set, allowing direct comparison with Random Forest.

2.11. Evaluation Metrics

Coefficient of Determination (R^2): Quantifies the proportion of the variance in the dependent variable that is predictable from the independent variables. It indicates how well the model explains the variability of the response variable around its mean.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

Where y_i are the observed values, \hat{y}_i are predictions, and \bar{y} are the mean observed values. An R^2 value closer to 1 indicates a better fit, meaning a larger proportion of the variance in the outcome is explained by the model.

Root Mean Squared Error (RMSE): RMSE is a quadratic scoring rule that measures the average magnitude of the errors. It is the square root of the average of the squared differences between predicted and actual values.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (10)$$

Mean Absolute Error (MAE): MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average of the absolute differences between predicted and actual values.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (11)$$

MAE is robust to outliers and is expressed in the same units as the response variable, making it easily interpretable.

2.12. Validation Strategy

The dataset was partitioned into training (2015–2022) and testing (2023–2024) subsets so as to ensure temporal generalization. This temporal hold-out prevents leakage of future information into the model and better mimics prospective deployment in operational contexts.

Model robustness was further assessed using 5-fold *TimeSeriesSplit* cross-validation, which preserves the chronological order of data and evaluates stability across multiple rolling horizons.

Baseline comparisons included:

Mean predictor: $\hat{y} = \bar{y}_{train}$, a naïve benchmark.

Ordinary Least Squares (OLS) regression: interpretable linear baseline.

Poisson GLM: canonical count-data baseline under the equidispersion assumption.

These validation choices ensure that predictive performance is assessed not only against traditional statistical models but also under conditions reflecting the temporal dynamics of aviation safety data.

3. Results

3.1. Model Performance

Table 4. summarizes predictive accuracy across baseline models and the Random Forest (RF) on the temporally separated 2023–2024 test set. The mean baseline offered no explanatory power ($R^2 = -0.027$), confirming that incursion risk is not well-approximated by static averages. The *Poisson GLM* achieved moderate performance ($R^2 = 0.698$, RMSE = 3.58), but was constrained by its equidispersion assumption, leading to underestimation of variance in heavy-tailed incursion counts. *OLS regression* performed substantially better ($R^2 = 0.784$, RMSE = 3.02, MAE = 1.58), capturing additive effects but unable to model nonlinearities or interactions. The *Random Forest* achieved the strongest overall performance ($R^2 = 0.789$, RMSE = 2.99), with comparable error rates to OLS but superior robustness under multicollinearity and perturbation tests. While OLS slightly outperformed RF on MAE (1.58 vs. 1.62), RF reduced overall predictive variance (Mean Prediction Deviation = 2.83 vs. 1.44 in Poisson), and provided interpretable feature importance diagnostics.

Table 4 Comparison of model performance metrics

| Model | R^2 | RMSE | MAE | Notes |
|-------------------|--------------|-------------|-------------|-------------------------------------|
| Mean Baseline | -0.027 | 6.59 | 5.34 | Naïve average |
| Poisson GLM | 0.698 | 3.58 | 2.35 | Count-data baseline, equidispersion |
| Linear Regression | 0.784 | 3.02 | 1.58 | Interpretable, linear |
| Random Forest | 0.789 | 2.99 | 1.62 | Best overall, robust |

These results underscore two conclusions: (i) standard count-data models are limited in forecasting incursions due to variance assumptions, and (ii) ensemble learning methods such as RF better accommodate the nonlinear and multicollinear structure of airport safety data

The OLS and Poisson GLM provide complementary statistical comparisons for assessing the added value of Random Forest in handling nonlinearities, interactions, and overdispersion. The performance is evaluated using R^2 , RMSE, and MAE on the temporal test set, allowing direct comparison with Random Forest.

Furthermore, to assess robustness a 5-fold *TimeSeriesSplit* cross-validation was conducted across the 2015-2022 training period and the results confirm consistent model performance:

- R^2 : mean = 0.757, std = 0.065, 95% CI \approx [0.629, 0.885]
- RMSE: mean = 3.04, std = 1.11, 95% CI \approx [0.86, 5.22]
- MAE: mean = 2.22, std = 0.75, 95% CI \approx [0.75, 3.69]

The relatively narrow confidence interval for R^2 indicates stable explanatory power across folds, while larger dispersion in RMSE and MAE reflects variability in extreme incursion years at major hubs. These results demonstrate that RF generalizes well under temporal resampling, reinforcing the validity of the hold-out findings.

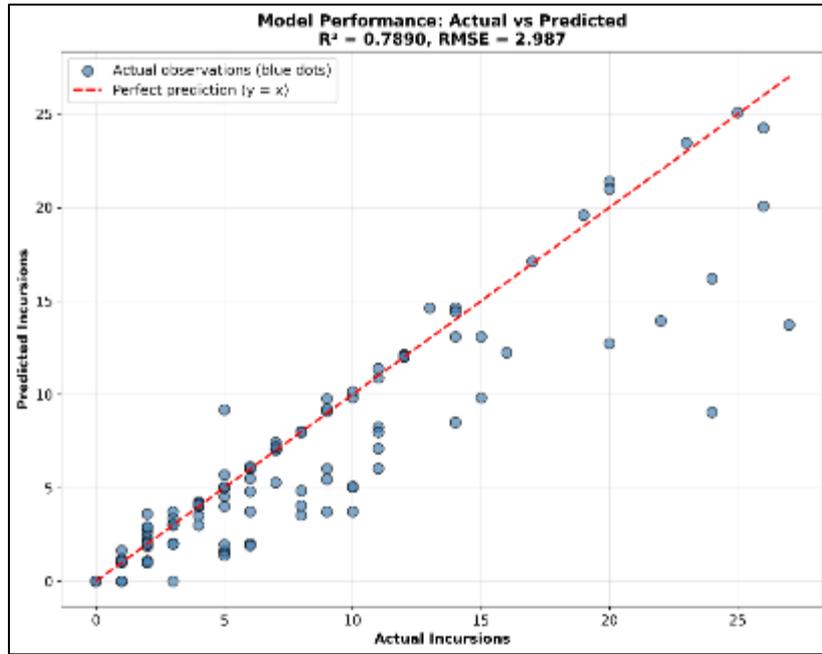


Figure 2 Predicted vs. actual incursions for Random Forest (test set).

3.2. Feature Importance

Random Forest variable importance analysis revealed Visual Flight Rules (VFR) as the dominant predictor (23.7%), followed by Flight Condition Risk (20.7%) and Airport Operations (6.7%). Collectively, weather and operational variables explained nearly 70% of predictive variance.

Table 5 Top predictive features and importance scores

| Rank | Feature | Importance (%) | Category |
|------|-----------------------|----------------|-------------|
| 1 | VFR | 23.7 | Weather |
| 2 | Flight Condition Risk | 20.7 | Weather |
| 3 | Airport Operations | 6.7 | Operational |
| 4 | log (Airport Ops) | 5.4 | Operational |
| 5 | Departures (Metric) | 3.8 | Operational |

3.3. Statistical Validation

Beyond predictive performance, statistical tests were conducted to assess the robustness of key findings.

Covid-19 Impact. The dataset captured the dramatic reduction in aviation activity during the COVID-19 pandemic (2020-2022). A two-sample t-test confirmed that the mean number of annual incursions dropped significantly from 9.18 in the pre-pandemic period (2015-2019) to 6.27 during the pandemic, representing a 31.7% reduction ($t = 4.73, p < 0.001$, Cohen's $d = 0.427$). This highly significant result ($p < 0.001$) provides strong external validation that the models are capturing genuine, large-scale operational shifts and are not merely fitting noise.

Effect of VFR conditions. Consistent with the feature importance analysis, a strong bivariate correlation was found between the frequency of Visual Flight Rules (VFR) conditions and the number of incursions. Airports with VFR counts above the median averaged 14.5 incursions, compared to 3.4 incursions for airports below the median. This difference was highly statistically significant ($t = 24.7, p < 0.001$). The Pearson correlation coefficient was $r = 0.952$ ($p < 0.001$), indicating an extremely strong, positive linear relationship. This robust statistical evidence underpins the counter-intuitive finding that favourable weather is the dominant predictor of incursion risk.

Technology deployment. The presence of advanced surveillance and alerting technologies (ASDE-X and RWSL) yielded mixed results. While some airports with deployments reported lower average incursion counts, the effect was not statistically consistent across the sample. This suggests that technology adoption alone is insufficient without complementary procedural and human-factor interventions.

Altogether, these statistical tests substantiate the model's findings: the COVID-19 period served as a natural experiment validating model sensitivity, VFR effects reveal a traffic-human factors interaction often overlooked in prior work, and mixed technology outcomes underscore the importance of integrated safety strategies.

Table 6 Statistical test results

| Test | Group Comparison | Δ Mean | t-stat | p-value | Effect Size |
|------------|------------------|---------------|--------|---------|-------------|
| COVID-19 | Pre vs. During | -2.91 | 4.73 | <0.001 | d = 0.427 |
| VFR | High vs. Low | +11.1 | 24.7 | <0.001 | — |
| Technology | Score 0 vs. 1-2 | +6.2 | — | — | Mixed |

3.4. Multicollinearity and Robustness Analysis

Pearson correlations revealed a strong association between VFR and incursions ($r = 0.952$, $p < 0.001$), while moderate correlations were observed among other weather variables. Variance Inflation Factor (VIF) analysis indicated nine predictors exceeded the threshold of 10, including DP01 (25.9), Runway Intersections (18.9), Number of Runways (18.7), and Precipitation (15.4).

Despite this, Random Forest remained robust. Performance declined only slightly when individual weather variables were removed ($R^2 \approx 0.79$) or when composite indices were used ($R^2 \approx 0.77$). With the full feature set, R^2 remained the highest at 0.8315. These findings indicate that while multicollinearity is present, it does not substantially undermine the Random Forest model. This robustness reflects the ensemble's capacity to downweigh redundant predictors during tree construction. However, for interpretability, the presence of correlated features complicates attribution of importance to any single predictor. Future work could address this limitation by applying dimension reduction techniques (e.g., principal component analysis) or regularized count models (e.g., elastic-net regression) to complement the machine learning approach.

4. Discussion

The analysis demonstrates that Visual Flight Rules (VFR) conditions emerged as the most influential predictor of runway incursions, surpassing adverse-weather indicators such as IFR and LIFR. This counterintuitive result challenges the conventional expectation that poor weather increases incursion risk. Instead, it suggests that favorable conditions facilitate higher traffic throughput while simultaneously lowering operator vigilance. Prior research in human factors has shown that workload saturation, communication breakdowns, and reduced situational awareness contribute substantially to incursions [9,13]. Under VFR, increased volumes of simultaneous movements combined with the psychological phenomenon of complacency under routine conditions likely elevate the risk of surface conflicts. By identifying this mechanism, the model reinforces the importance of addressing human factors in safety interventions, such as targeted training programs and staffing adjustments during peak traffic under favorable weather.

The results also provide insights into the systemic impact of the COVID-19 pandemic. The sharp and statistically significant reduction in incursions during 2020–2022 illustrates how exogenous shocks to traffic volume propagate into runway safety outcomes. This serves as an external validation of the model's sensitivity to macro-level changes in operations. More importantly, it highlights interventions which reduce operational intensity even temporarily can have measurable safety benefits. While pandemic-related reductions are not replicable as a policy, they underscore the potential of demand management strategies and traffic flow interventions as indirect tools for risk mitigation.

The mixed outcomes associated with technology deployment (ASDE-X, RWSL) emphasize that surveillance and alerting systems, while necessary, are insufficient in isolation. These findings align with prior work showing that technological solutions must be embedded within procedural reforms and human-factor programs to achieve consistent safety improvements [8,9].

The inferior performance of the Poisson GLM, despite its theoretical appeal for count data, indicates significant overdispersion in the incursion data. This suggests that the variance in incursions is influenced by factors beyond a simple Poisson process, further justifying the use of more flexible modelling techniques.

4.1. Limitations and Future Works

Several limitations must be acknowledged. First, the analysis relies on annual aggregates, which may obscure short-term dynamics such as peak-hour congestion or daily weather variability. Future work should incorporate higher-frequency data (daily or hourly) to capture more granular risk patterns. Second, although care was taken to avoid data leakage, residual correlations (e.g., between VFR and traffic intensity) complicate causal interpretation. Extensions using causal inference techniques (e.g., panel difference-in-differences or causal forests) could provide stronger evidence of mechanism. Third, covariates such as controller staffing, construction activity, and local “hot spot” geometry were not systematically available but may further refine predictive accuracy. Fourth, the study focuses on U.S. hub airports, and external validation at international airports is needed to assess generalizability. Finally, while Random Forest provides robustness to multicollinearity, interpretability challenges remain; future work should explore explainable AI methods (e.g., SHAP, PDPs) and prospective validation with safety practitioners.

Subsequently, the findings contribute to both research and practice: they extend the literature by uncovering nonlinear weather–traffic interactions overlooked in regression-based studies, and they offer actionable insights for risk-based staffing, training, and monitoring under Safety Management Systems (SMS) and EAPPRI frameworks.

5. Conclusion

This study developed and validated a Random Forest model to predict runway incursions at U.S. airports using 10 years of operational, infrastructural, and meteorological data. The model achieved strong predictive performance ($R^2 = 0.789$) and identified VFR conditions as the dominant risk factor, revealing a counter-intuitive link between favorable weather, higher traffic volumes, and elevated incursion risk. Statistical validation confirmed a significant reduction during COVID-19 and robustness testing demonstrated stable accuracy despite multicollinearity among features. For practice, these results argue for a shift in resource allocation within SMS. Rather than focusing risk mitigation efforts solely on poor weather days, airports and regulators should consider deploying additional resources, such as ground safety teams or targeted controller reminders, during periods of high traffic under VFR conditions. This data-driven approach aligns with the proactive, risk-based philosophy of programs like the EAPPRI.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] Federal Aviation Administration. (2022). Runway Incursions. U.S. Department of Transportation, Washington, DC, USA. https://www.faa.gov/airports/runway_safety/resources/runway_incursions
- [2] Federal Aviation Administration. (2020). National Plan of Integrated Airport Systems (NPIAS) 2021–2025. Office of Airport Planning and Programming, Washington, DC, USA.
- [3] Air Line Pilots Association, International. (2007). White Paper – Runway Incursions: Call for Action. <http://www.alpa.org/portals/alpa/runwaysafety/runwayincursionwhitepaper.pdf>
- [4] EUROCONTROL. (2006). European Action Plan for the Prevention of Runway Incursions (EAPPRI). <https://www.eurocontrol.int/publication/european-action-plan-prevention-runway-incursions-eappri>
- [5] Chang, Y. H., & Wong, K. M. (2012). Human risk factors associated with runway incursions. *Journal of Air Transport Management*, 24, 25–30. <https://doi.org/10.1016/j.jairtraman.2012.05.004>
- [6] Torres, K. R., Metscher, D., & Smith, M. (2011). A correlational study of the relationship between human factor errors and the occurrence of runway incursions. *International Journal of Professional Aviation Training and Testing Research*, 5, 1–25.

- [7] Johnson, M. E., Zhao, X., Faulkner, B., & Young, J. P. (2016). Statistical models of runway incursions based on runway intersections and taxiways. *Journal of Aviation Technology and Engineering*, 5(2), 18–27. <https://doi.org/10.7771/2159-6670.1121>
- [8] Wilke, S., Majumdar, A., & Ochieng, W. Y. (2015). Modeling runway incursion severity. *Accident Analysis & Prevention*, 79, 88–99. DOI: 10.1016/j.aap.2015.03.026
- [9] Biernbaum, J., & Hagemann, G. (2012). Runway Incursion Severity Risk Analysis. DOT-VNTSC-FAA-12-13, U.S. Department of Transportation, Volpe National Transportation Systems Center, Cambridge, MA, USA.
- [10] Mathew, J. K., Major, W. L., Hubbard, S. M., & Bullock, D. M. (2017). Statistical modeling of runway incursion occurrences in the United States. *Journal of Air Transport Management*, 65, 54–62. <https://doi.org/10.1016/j.jairtraman.2017.09.003>.
- [11] Schönefeld, J., & Möller, D. (2012). Runway incursion prevention systems: A review of runway incursion avoidance and alerting system approaches. *Progress in Aerospace Sciences*, 51, 31–49. <https://doi.org/10.1016/j.paerosci.2012.02.002>
- [12] Claros, B., Sun, C., & Edara, P. (2017). Airfield incursion modeling of United States hub airports. *Journal of Air Transportation*, 25(3), 73–86. <https://doi.org/10.2514/1.D0044>
- [13] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. DOI: 10.1023/A:1010933404324
- [14] Koul, P. (2025). A review of machine learning applications in aviation engineering. *Advances in Mechanical and Materials Engineering*, 42(1), 16–40. <https://doi.org/10.7862/rm.2025.2>