Cureus Journal of Computer Science
https://doi.org/10.7759/s44389-026-00043-9

Original Article

A Machine Learning–Based Predictive Framework for Nighttime High-Power Aircraft Engine Run-Up Operations

Somashekar V [1, ✉], Shivakumar PN [1]

1. *Department of Aeronautical Engineering, Nitte Meenakshi Institute of Technology (NMIT) Nitte (Deemed to be University), Bangalore, IND*

Received: February 08, 2026 | Review began: February 16, 2026 | Review ended: March 14, 2026 | Published: March 21, 2026

Abstract

Nighttime high-power aircraft engine run-ups represent a critical operational activity with potential implications for airport management and environmental assessment; yet, their quantitative prediction remains limited in existing literature. The objective of this study is to develop and evaluate a data-driven machine learning framework capable of accurately predicting nighttime high-power engine run-up characteristics using historical operational data. A supervised learning methodology is adopted using an open-access dataset comprising approximately 860 nighttime run-up records. The proposed framework includes data preprocessing, temporal feature engineering (month, season, and quarter), exploratory data analysis, correlation assessment, and comparative model development using Linear Regression, Random Forest, Gradient Boosting, Extreme Gradient Boosting (XGBoost), and an Artificial Neural Network (multilayer perceptron). Model performance is evaluated using coefficient of determination (R^2), root mean square error (RMSE), mean absolute error, and mean absolute percentage error, supported by predicted-versus-actual comparisons and learning-curve analysis. The results demonstrate that ensemble-based models outperform linear and neural network approaches, with XGBoost achieving the highest predictive accuracy on the testing dataset ($R^2 \approx 0.90$, RMSE ≈ 2.15). Learning curves further indicate stable validation performance with increasing training data size. The proposed machine learning framework provides a structured and data-driven approach for predicting nighttime high-power aircraft engine run-up activity within the analyzed operational dataset, offering a valuable analytical tool for data-driven assessment of airport ground operations.

Categories: AI applications, IoT Applications, Machine Learning (ML)

Keywords: nighttime engine run-ups, predictive modeling, machine learning, xgboost, aircraft ground operations

Introduction

Aircraft engine run-up operations constitute an essential component of ground-based aircraft maintenance and operational readiness. These procedures, often involving high engine power settings, are routinely conducted to verify engine performance, troubleshoot anomalies, and ensure compliance with safety requirements prior to flight operations. While run-ups are operationally necessary, their execution - particularly during nighttime hours - has drawn increasing attention due to their variability, operational intensity, and potential implications for airport management and environmental assessment.

Nighttime high-power engine run-ups differ from routine daytime operations in terms of scheduling constraints, operational context, and temporal distribution. Unlike scheduled flight activities, run-ups are irregular and depend on maintenance requirements, aircraft availability, and operational demand. As a result, their occurrence exhibits significant

How to cite this article:
V S, PN S (March 21, 2026) A Machine Learning–Based Predictive Framework for Nighttime High-Power Aircraft Engine Run-Up Operations. Cureus J Comput Sci 3 : es44389-026-00043-9. DOI https://doi.org/10.7759/s44389-026-00043-9

variability across time, making prediction and systematic analysis challenging. Despite the availability of operational data from aviation authorities, quantitative methods for predicting nighttime high-power run-up activity remain limited.

The aviation industry is still an important part of global connectivity, but its rapid growth is bad for the environment and public health [1-6]. During the landing and take-off cycle, aircraft engines release a wide range of pollutants, such as nitrogen oxides, carbon monoxide, and ultrafine particles. These are all known to cause respiratory and cardiovascular problems in communities near airports [7-8]. Historically, urban air quality research has focused on road traffic, but recent studies show that emissions from airports can make up to 24% of the local ambient NO_2 under normal conditions [9-13].

It is hard to measure these emissions because planes move around all the time, and the relationship between engine thrust, fuel composition, and weather is not always straightforward. Conventional dispersion models frequently presume linear additivity and inadequately account for sporadic high-impact events induced by particular runway operations. Machine learning (ML) algorithms have become strong alternatives to deal with these problems. They can learn patterns from high-dimensional datasets without needing detailed physical models [14-20]. Tree-based ensemble methods, such as XGBoost, have shown great accuracy in predicting pollutant levels.

However, the fact that advanced ML models are "black boxes" makes it hard to use them in aviation situations where lives are on the line [21-25]. Model interpretability is urgently required to guarantee that predictions are consistent with recognized atmospheric physics and yield actionable insights for stakeholders [26]. Recent advancements in explainable AI, including the implementation of SHapley Additive Explanations, enable researchers to accurately assign the variance in air quality to particular factors such as wind direction, planetary boundary layer height, and arrival/departure frequencies.

The persistent growth of global air transport has necessitated a focus on operational safety and environmental compatibility. Before take-off, aircraft engines - particularly piston and jet variants - are routinely run up at high power to perform critical checks for potential engine-related malfunctions [27]. While essential for ensuring airworthiness, these operations generate considerable noise pollution, impacting both airport workers and surrounding communities. The challenge is exacerbated during nighttime operations, as noise-induced sleep disturbance is linked to adverse health outcomes, including cardiovascular disease and decreased cognitive performance [28]. Traditional management of such high-power tasks often relies on human-centric paradigms, which are increasingly prone to instability as congestion grows [29].

To address these challenges, there is a burgeoning need for a Computational Air Traffic Management approach that leverages ML to predict and optimize high-power operations [29]. Recent advancements in Cognitive Simulation and closed-loop ML algorithms have demonstrated the ability to optimize high-intensity systems in real-time by processing vast amounts of experimental data to make on-the-fly adjustments [30]. Furthermore, Long Short-Term Memory neural networks have proven highly accurate (up to 99%) in recognizing complex operating cycles in heavy-duty machinery, suggesting their utility in identifying and predicting aircraft engine run-up patterns [31].

ML techniques have demonstrated strong potential in aviation-related applications, including flight delay prediction, air traffic flow management, fuel consumption estimation, and emissions modeling. These data-driven approaches are particularly suited to problems characterized by nonlinear relationships, heterogeneous features, and limited prior assumptions. However, to date, there is a notable gap in the application of ML methods specifically for the prediction of nighttime high-power aircraft engine run-up activity using empirical operational data.

Motivated by this gap, the present study proposes a ML-based predictive framework for analyzing and predicting nighttime high-power engine run-up operations. Using an open-access dataset of historical nighttime run-up records, the study systematically explores data characteristics, temporal variability, and inter-parameter relationships before developing and evaluating multiple supervised learning models. A comparative assessment of linear, ensemble-based, and neural network models is conducted using standard performance metrics, predicted-versus-actual analysis, and learning-curve evaluation.

The primary contribution of this work lies in the development and validation of a structured, data-driven framework that demonstrates the feasibility and effectiveness of ML for predicting nighttime high-power engine run-up activity. By identifying a high-performing predictive model and quantifying its accuracy using multiple evaluation criteria, this study provides a reproducible methodological foundation for future research on aircraft ground operations and data-driven airport analytics.

Materials And Methods

The proposed methodology follows a data-driven, supervised ML framework comprising data preparation, feature engineering, multi-model development, cross-validated performance evaluation, learning-curve-based assessment, and optimal model selection for nighttime high-power aircraft engine run-up prediction (Figure *1*).

The analysis begins with the acquisition of the nighttime high-power engine run-ups dataset obtained from an official government open-access repository [32]. The dataset consists of approximately 860 operational records, representing real-world aircraft engine run-up activities conducted under nighttime conditions. This study employs a quantitative observational research design, where no assumptions are imposed a priori on the underlying physical relationships; instead, patterns are learned directly from measured data.

In the data preparation stage, rigorous preprocessing is carried out to ensure suitability for ML modeling. This includes data cleaning procedures to handle missing values, inconsistencies, and noise inherent in operational datasets. To enhance the explanatory power of the input variables, feature engineering is performed by incorporating temporal indicators such as seasonal and quarterly classifications, enabling the models to capture cyclical and time-dependent operational variations. Following preprocessing, the dataset is partitioned into training and testing subsets using an 80:20 split, which allows independent evaluation of model generalization performance.

The model development phase adopts a comparative multi-model learning strategy, wherein both linear and nonlinear algorithms are implemented to assess their predictive capability. Linear Regression is employed as a baseline model to establish a reference for performance benchmarking. Ensemble-based models such as Random Forest, Gradient Boosting, and Extreme Gradient Boosting (XGBoost) are incorporated to capture nonlinear feature interactions and hierarchical decision structures. In addition, an Artificial Neural Network (Multilayer Perceptron [MLP]) is implemented to evaluate the effectiveness of connectionist learning approaches for this application. To ensure statistical robustness and mitigate overfitting, 10-fold cross-validation is applied uniformly across all models during training.

Model evaluation is conducted using a multi-criteria performance assessment framework, combining both quantitative metrics and graphical diagnostics. Standard regression metrics, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE), are used to quantify prediction accuracy. Learning curves are analyzed to examine model convergence behaviour and relative training-validation performance trends, while predicted-versus-actual plots are employed to visually assess dispersion and systematic deviations. Cross-validation results further provide insight into model stability and consistency across different data partitions.

Based on the evaluation outcomes, the XGBoost model is identified as the most effective predictor, demonstrating superior accuracy and testing-set predictive performance relative to other models. Consequently, a dedicated model validation phase is conducted to verify the statistical soundness of the selected model. This includes repeated 10-fold cross-validation and graphical diagnostic evaluation of prediction errors, as presented in the Results section. These validation procedures confirm the statistical consistency and predictive reliability of the selected model.

The final stage of the methodology focuses on structured research documentation, wherein validated model performance, statistical results, and supporting figures are systematically compiled. The methodology culminates in the identification of an optimal predictive framework, with the XGBoost model achieving a high coefficient of determination ($R^2 \approx 0.90$) and low prediction error ($\text{RMSE} \approx 2.15$). This confirms the suitability of the adopted ML-driven predictive methodology for analyzing nighttime high-power engine run-up operations.

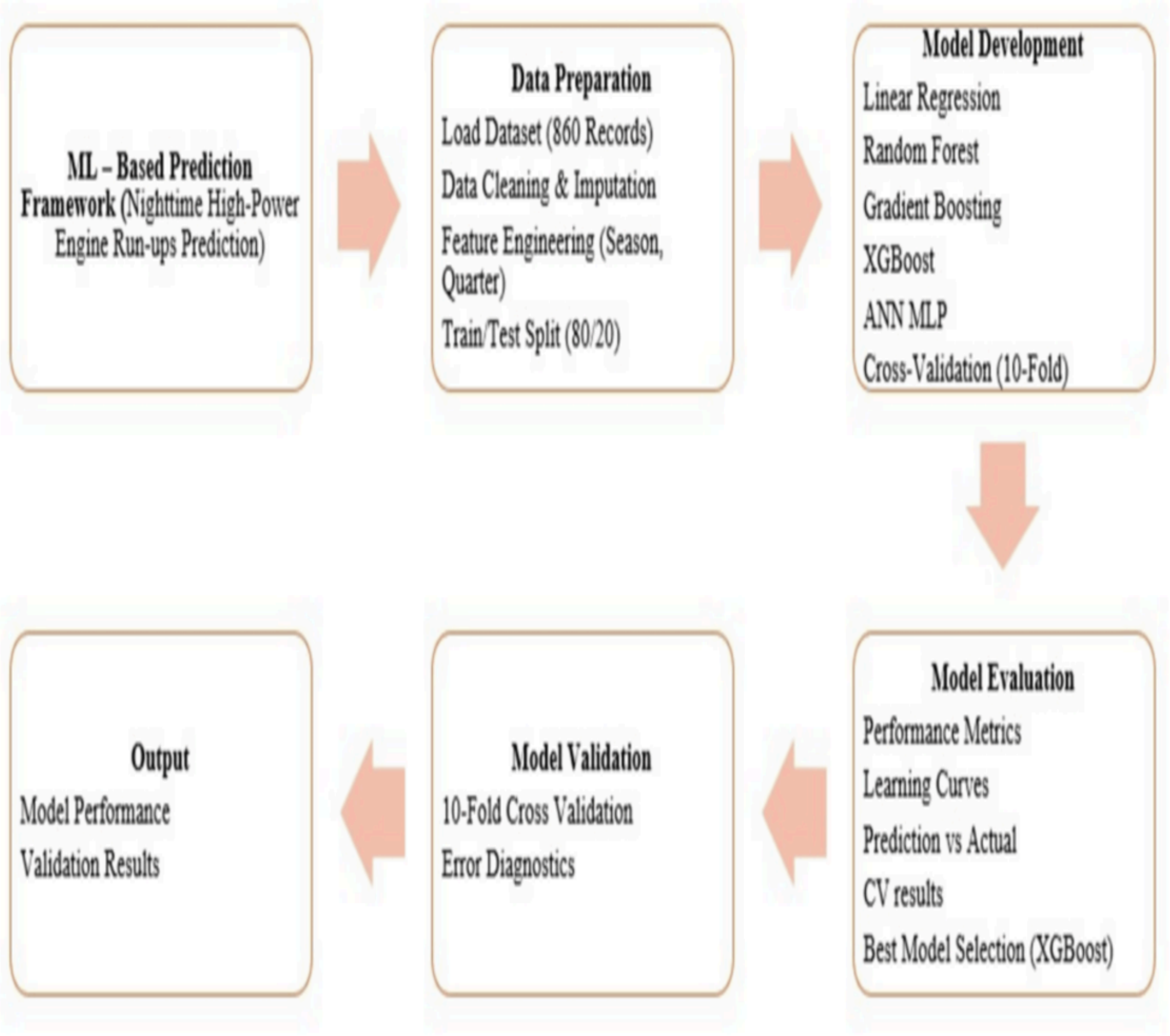

FIGURE 1: Machine Learning–Based Predictive Methodology for Nighttime High-Power Aircraft Engine Run-Up Analysis

ANN, Artificial Neural Network; CV, Cross Validation; ML, Machine Learning; MLP, Multilayer Perceptron; XGBoost, Extreme Gradient Boosting

Results And Discussion

This section presents exploratory data analysis, correlation assessment, comparative model evaluation, and learning-curve-based performance analysis in direct correspondence with the proposed methodology.

Figure *2* presents the statistical distribution and temporal variability of nighttime high-power engine run-up activity using four complementary visualizations, providing an empirical overview of the dataset characteristics prior to ML modeling. These temporal features were retained for subsequent model training.

The distribution of the number of power run-ups shows a strongly right-skewed pattern, with the highest frequency occurring at low run-up counts. A large proportion of observations are concentrated below approximately 5 run-ups, while the frequency decreases progressively as the number of run-ups increases. Observations extending beyond 20 run-ups occur infrequently, with a small number of cases reaching values above 30, indicating the presence of high-intensity operational periods. This wide spread demonstrates substantial variability in nighttime run-up activity across the dataset, emphasizing the non-uniform nature of the operational conditions.

The distribution of run-ups per 1,000 departures further reinforces the presence of skewness and dispersion in the data. Most observations are clustered at relatively low values, typically below 10 run-ups per 1,000 departures, while a long tail extends to values approaching 100. The concentration of data at the lower end, combined with sporadic high values, indicates that while nighttime run-ups are generally infrequent relative to departure volume, there exist specific instances with markedly elevated run-up rates. This variability highlights the need for models capable of handling both typical and extreme operational conditions.

The distribution of the percentage of power run-ups exhibits a broad range spanning from values close to 0% up to nearly 100%. The frequency is highest at low percentages, with a gradual decline as the percentage increases. A substantial number of observations fall within the 20-60% range, indicating that in many records, power run-ups constitute a significant fraction of total operations. The continuous spread across the full percentage range reflects the heterogeneity of nighttime operational practices captured in the dataset.

The temporal analysis of average power run-ups by month reveals moderate seasonal variation across the calendar year. Monthly average values range approximately between 6.5 and 7.8 run-ups, with the highest average observed in February and another elevated period around July. Lower average values are observed during April, September, and December. Although the overall variation remains within a relatively narrow numerical band, the presence of systematic month-to-month differences indicates that temporal factors contribute measurably to run-up activity levels.

Figure *3* presents kernel density estimates of the number of nighttime power run-ups, stratified by season (left panel) and calendar quarter (right panel). These distributions provide a non-parametric representation of how run-up frequencies are distributed across different temporal groupings within the dataset.

In the seasonal analysis, all four distributions - Winter, Spring, Summer, and Fall - exhibit a pronounced right-skewed profile. The highest density for each season occurs at low run-up counts, with peak density values concentrated approximately in the range of 0-3 power run-ups. Across seasons, the majority of probability mass is located below 10 run-ups, while a long tail extends toward higher values, reaching beyond 30 run-ups in isolated instances. The substantial overlap among the seasonal density curves indicates that the overall distributional shape remains broadly similar across seasons. Minor variations are observed in the mid-range (approximately 8-15 run-ups), where Summer and Spring show slightly higher density compared to Winter and Fall; however, these differences are modest relative to the overall spread.

The quarterly distributions show a closely analogous pattern. All quarters display peak density at low run-up counts, again centered roughly between 0 and 3 run-ups, followed by a gradual decline as the number of run-ups increases. The distributions for Q1 through Q4 overlap extensively across the entire range, particularly below 10 run-ups, where the majority of observations are concentrated. At higher run-up values (above approximately 15), the density decreases rapidly for all quarters, with low but non-zero density extending to values near 35-40 run-ups. The similarity in tail behaviour across quarters suggests that high run-up events, while infrequent, occur in all quarters represented in the dataset.

From a quantitative distributional perspective, the absence of sharp shifts in peak location or spread between seasons and quarters indicates that temporal stratification does not introduce distinct multimodal behaviour in the number of nighttime power run-ups. Instead, the distributions are characterized by consistent right-skewness, comparable dispersion, and substantial overlap across all temporal categories. These observed characteristics highlight that temporal grouping modifies the density magnitude locally but does not fundamentally alter the underlying distributional structure of the run-up data.

Figure *4* presents the Pearson correlation matrix among key operational and temporal variables associated with nighttime power run-up activity, including year, month, number of power run-ups, number of power run-ups per 1,000 departures, and percentage of power run-ups for the period. The heatmap provides a quantitative assessment of linear associations between pairs of variables, with correlation coefficients ranging from −1 to +1.

The strongest linear association observed in the matrix is between the number of power run-ups and the percentage of power run-ups for the period, with a correlation coefficient of $r = 0.781$. This value indicates a high positive linear relationship, reflecting that these two variables vary consistently in the same direction across the dataset. All other correlations involving the number of power run-ups are comparatively weaker, including a low positive correlation with the number of power run-ups per 1,000 departures ($r = 0.120$) and a moderate negative correlation with year ($r = -0.214$).

Temporal variables show generally weak linear relationships with operational metrics. The correlation between month and number of power run-ups is near zero ($r = -0.019$), while the correlation between month and percentage of power run-ups is effectively negligible ($r = 0.001$). Similarly, the year variable exhibits weak correlations with most parameters, including a small positive correlation with the percentage of power run-ups ($r = 0.151$) and weak negative correlations with the number of power run-ups per 1,000 departures ($r = -0.095$).

The number of power run-ups per 1,000 departures shows minimal linear association with other variables in the matrix. Its correlation with the percentage of power run-ups is nearly zero ($r = 0.003$), and its correlation with month ($r = -0.077$) and year ($r = -0.095$) remains weak. These values indicate limited linear dependence between normalized run-up frequency and the other variables considered.

Overall, the correlation matrix demonstrates that, with the exception of the strong association between absolute run-up counts and their corresponding percentage representation, most pairwise linear correlations are weak, with absolute correlation coefficients generally below 0.25. This indicates limited linear interdependence among the majority of variables in the dataset. The observed correlation structure highlights the multivariate nature of nighttime power run-up activity and underscores the importance of modeling approaches capable of capturing relationships beyond simple linear dependence, without imposing restrictive assumptions on variable interactions.

Figure *5* presents a comparative evaluation of five regression models - Linear Regression, Random Forest, Gradient Boosting, XGBoost, and Artificial Neural Network (MLP) - using four performance metrics: R^2, RMSE, MAE, and mean absolute percentage error (MAPE). For each metric, results are shown separately for training and testing datasets, enabling direct assessment of predictive accuracy and generalization.

In terms of R^2, the XGBoost model achieves the highest performance, with a training R^2 value approaching 1.00 and a testing R^2 of approximately 0.90. Although the XGBoost model achieves a near-perfect training R^2, the testing performance remains consistently high ($R^2 \approx 0.90$), and cross-validation results together with the learning-curve analysis indicate stable predictive behaviour across different data partitions, suggesting that the model maintains strong generalization performance rather than severe overfitting. Random Forest and Gradient Boosting also demonstrate strong explanatory power, each attaining testing R^2 values in the range of 0.87-0.88. In contrast, Linear Regression exhibits substantially lower testing performance, with a testing R^2 of approximately 0.39, while the ANN (MLP) achieves a moderate testing R^2 of about 0.49. These results indicate clear quantitative differences in variance explained across models.

The RMSE comparison further differentiates model performance. XGBoost yields the lowest RMSE values among all models, with a training RMSE of approximately 0.3 and a testing RMSE near 2.2. Random Forest and Gradient Boosting follow, with testing RMSE values of approximately 2.4 and 2.5, respectively. Linear Regression and ANN (MLP) display higher error magnitudes, with testing RMSE values exceeding 4.8, indicating larger deviations between predicted and observed values for these models.

A similar ranking is observed in the MAE results. XGBoost produces the lowest mean absolute error, with testing MAE close to 1.0, followed by Random Forest at approximately 1.2 and Gradient Boosting at around 1.4. Linear Regression and ANN (MLP) record higher testing MAE values of approximately 3.2 and 2.9, respectively. These values quantitatively confirm that XGBoost minimizes average absolute prediction error more effectively than the alternative models.

The MAPE metric further highlights differences in relative prediction accuracy. XGBoost again exhibits the lowest testing MAPE, approximately 35%, whereas Random Forest and Gradient Boosting show higher testing MAPE values of roughly 40% and 50%, respectively. Linear Regression and ANN (MLP) display substantially larger testing MAPE values, exceeding 95%, reflecting higher relative errors when predictions are expressed as percentages.

The Figure *5* demonstrates consistent quantitative agreement across all four evaluation metrics. XGBoost outperforms all other models on the testing dataset, achieving the highest R^2 and the lowest RMSE, MAE, and MAPE. Random Forest and Gradient Boosting show competitive but comparatively lower performance, while Linear Regression and ANN (MLP) exhibit notably higher prediction errors. The convergence of results across multiple independent metrics provides strong numerical evidence supporting the selection of XGBoost as the most accurate predictive model within the evaluated framework.

Figure *6* presents scatter plots of predicted values versus actual values for five regression models - Linear Regression, Random Forest, Gradient Boosting, XGBoost, and Artificial Neural Network (MLP) - evaluated on the testing dataset. In each subplot, the dashed diagonal line represents perfect agreement between predicted and observed values, allowing direct visual and quantitative assessment of model accuracy and dispersion.

For the Linear Regression model, the scatter shows substantial deviation from the perfect prediction line, particularly at higher actual values. The quantitative performance metrics reported in the plot indicate an R^2 of 0.3914 and an RMSE of 5.2900, reflecting limited alignment between predicted and observed values. The dispersion of points around the diagonal increases as actual values increase, indicating larger absolute errors across a broad range of observations.

The Random Forest model demonstrates a markedly improved alignment with the perfect prediction line. Data points cluster more closely around the diagonal across the full range of actual values. This improvement is supported by an R^2 of 0.8720 and an RMSE of 2.4257, indicating substantially higher variance explanation and lower prediction error compared to Linear Regression.

Similarly, the Gradient Boosting model exhibits strong agreement between predicted and actual values, with an R^2 of 0.8737 and an RMSE of 2.4100. The distribution of points closely follows the diagonal trend, with moderate scatter at higher actual values. The quantitative metrics indicate performance comparable to that of the Random Forest model.

The XGBoost model shows the highest level of agreement among all evaluated models. The majority of data points lie very close to the perfect prediction line over the entire range of actual values. This is quantitatively reflected in the highest R^2 value of 0.8994 and the lowest RMSE of 2.1512 among all models. The reduced scatter around the diagonal indicates consistently lower prediction error across both low and high actual values.

In contrast, the ANN (MLP) model displays wider dispersion relative to the diagonal line, particularly for higher actual values. The reported metrics, $R^2 = 0.4886$ and RMSE = 4.8492, indicate moderate explanatory power and higher prediction error compared to ensemble-based models. The scatter pattern reflects greater variability in predictions across the dataset.

Overall, Figure *6* provides a direct visual and quantitative comparison of predictive accuracy across models. The progression from Linear Regression to ensemble-based models is accompanied by increasing R^2 values and decreasing RMSE values, with XGBoost achieving the highest predictive accuracy and lowest error. The consistency between visual alignment and numerical performance metrics across all subplots reinforces the reliability of the comparative evaluation presented in this study.

Figure *7* presents learning curves illustrating the variation of training and validation R^2 scores as a function of training set size. The observed stabilization of validation performance with increasing training data suggests that the models achieve consistent predictive behaviour despite the moderate dataset size used in this study for four models: XGBoost, Gradient Boosting, Random Forest, and Artificial Neural Network (MLP). The curves provide a quantitative visualization of how predictive performance evolves with increasing data availability, based solely on observed model scores.

For XGBoost, the training R^2 remains consistently close to 1.00 across all training set sizes, while the validation R^2 increases from approximately 0.28 at the smallest training size to about 0.84 at the largest size. The validation curve shows a monotonic increase with diminishing increments beyond roughly 200-250 samples, indicating a stabilization of validation performance. The shaded region around the validation curve narrows with increasing sample size, reflecting reduced variability in validation R^2 estimates.

The Gradient Boosting model exhibits training R^2 values decreasing slightly from approximately 0.99 to 0.93 as the training size increases. In parallel, the validation R^2 rises from about 0.58 to approximately 0.82. The gap between training and validation curves remains relatively stable across the training range, with validation performance converging toward a plateau beyond roughly 300 samples. The variability band also contracts gradually as the training size increases.

For Random Forest, the training R^2 increases modestly from approximately 0.94-0.97, while the validation R^2 improves from about 0.57 to approximately 0.81. The validation curve shows gradual convergence and reduced variability with larger training sizes, particularly beyond 200 samples. The relative ordering of training and validation scores remains consistent throughout the training range.

The ANN (MLP) learning curves display greater variability compared to the ensemble-based models. The training R^2 fluctuates between approximately 0.69 and 0.83, while the validation R^2 increases from around 0.15 to approximately 0.70 as the training size increases. The shaded uncertainty region for the validation curve is wider at smaller sample sizes and narrows progressively with additional data, indicating higher dispersion at lower training sizes and increased stability at larger sizes.

Across all four models, the learning curves show that validation R^2 consistently increases with training set size, while the rate of improvement diminishes at larger sample sizes. The ensemble-based models (XGBoost, Gradient Boosting, Random Forest) achieve higher validation R^2 values than the ANN (MLP) throughout the training range. The convergence behaviour and relative separation between training and validation curves are quantitatively distinct for each model, providing a data-supported basis for comparing learning behaviour without invoking distributional or structural assumptions.

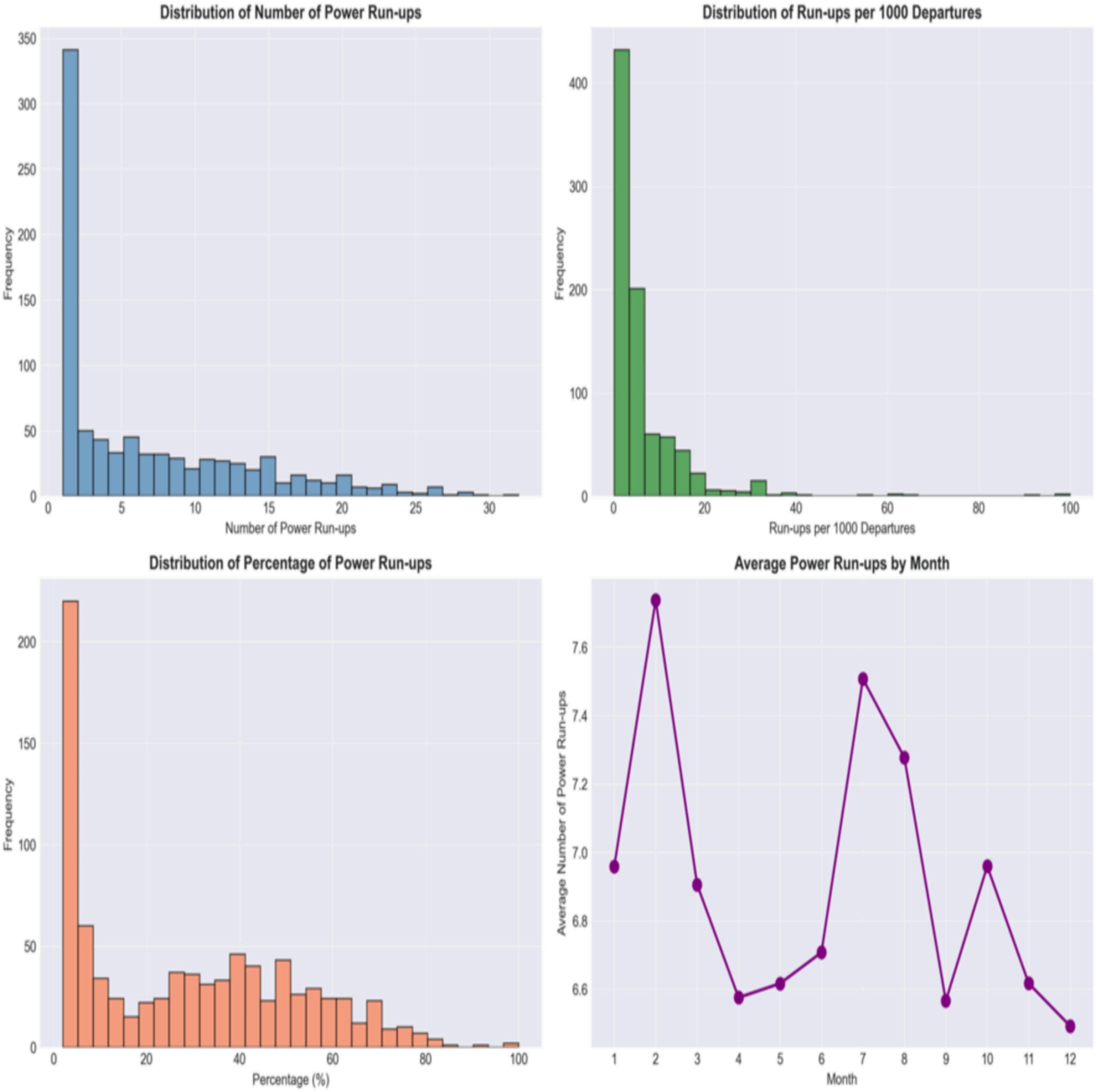

FIGURE 2: Descriptive Statistics of Nighttime High-Power Engine Run-Up Activity

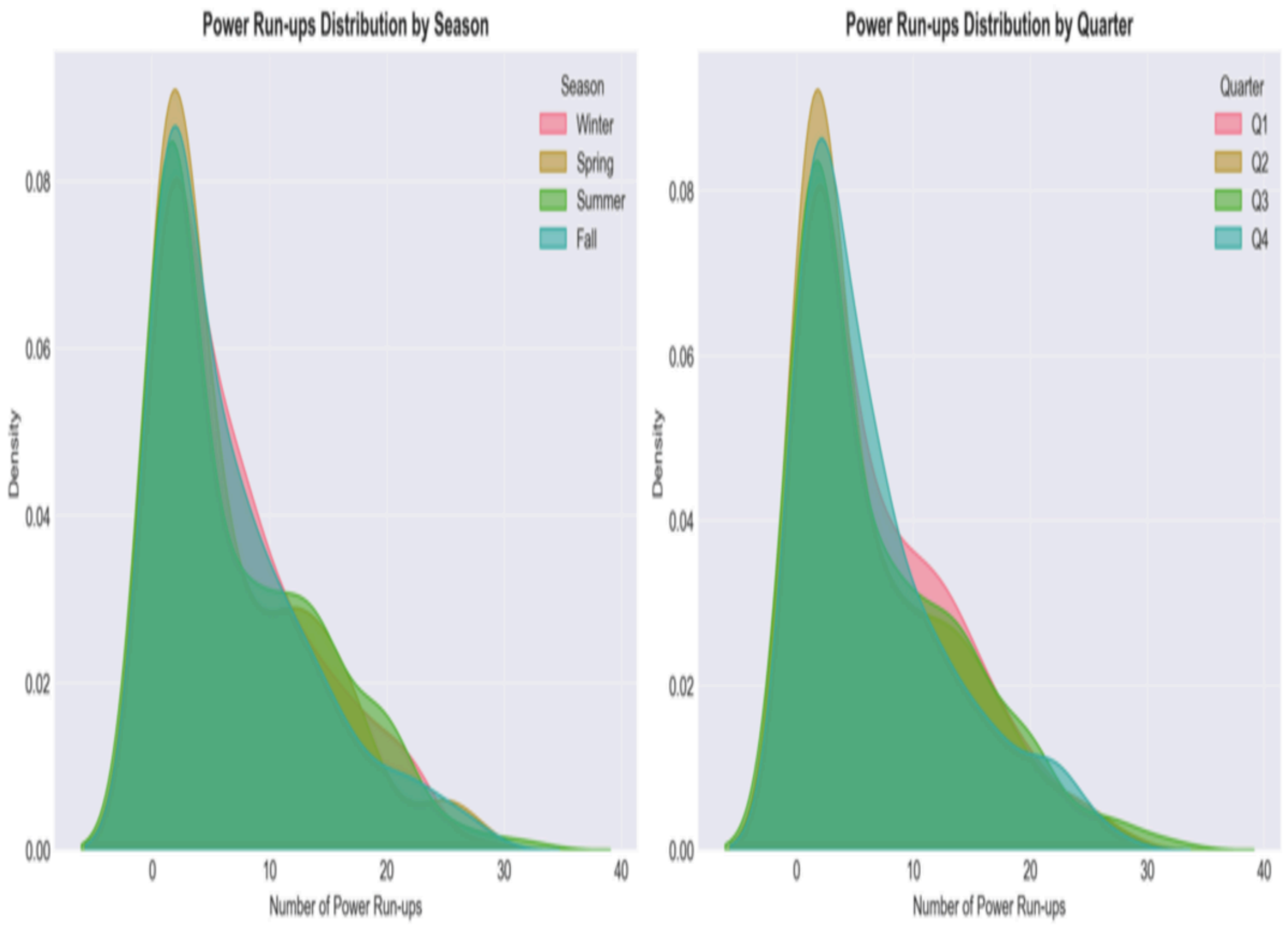

FIGURE 3: Seasonal and Quarterly Distributions of Nighttime Power Run-Ups

How to cite this article:
V S, PN S (March 21, 2026) A Machine Learning–Based Predictive Framework for Nighttime High-Power Aircraft Engine Run-Up Operations. Cureus J Comput Sci 3 : es44389-026-00043-9. DOI https://doi.org/10.7759/s44389-026-00043-9

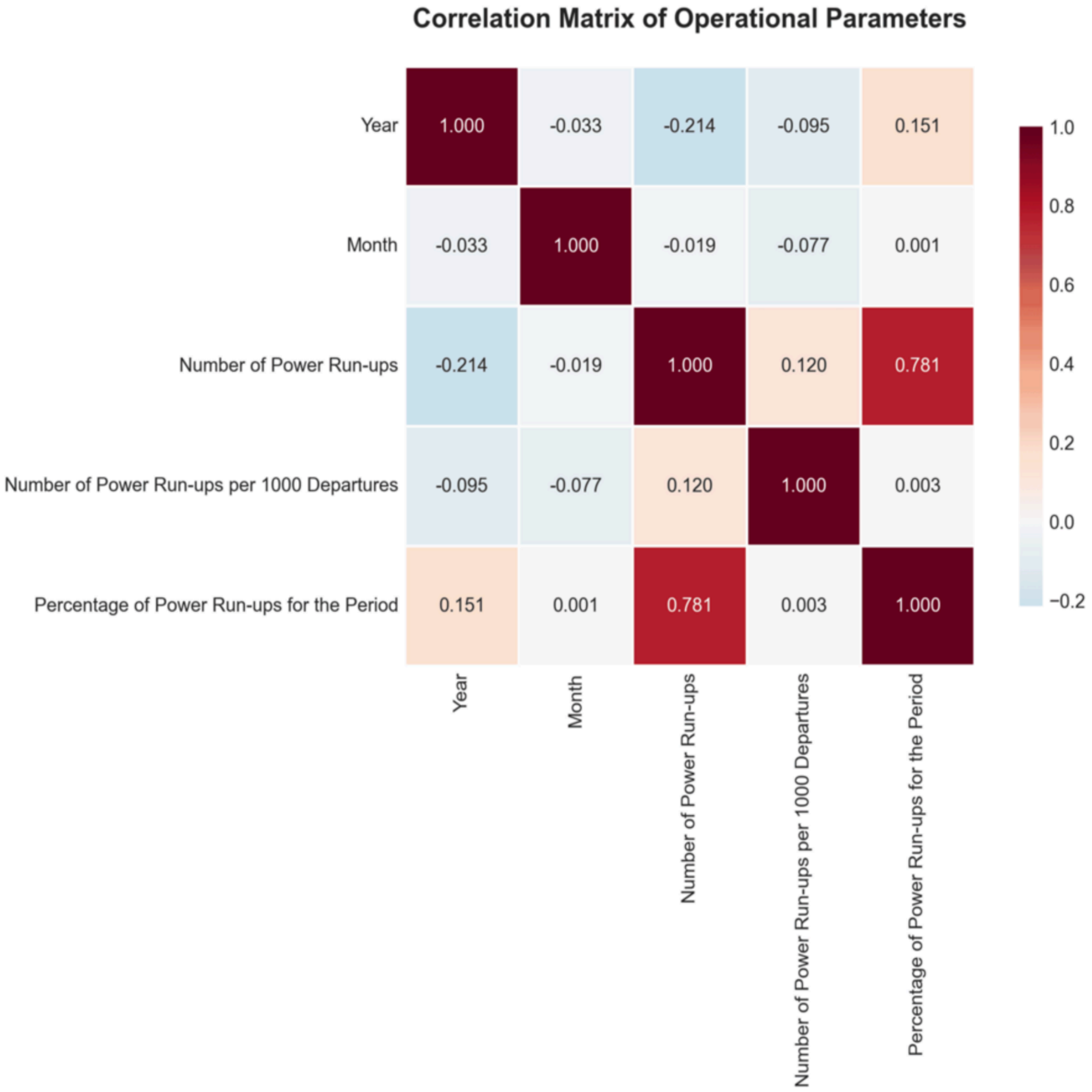

FIGURE 4: Correlation Structure of Operational Parameters

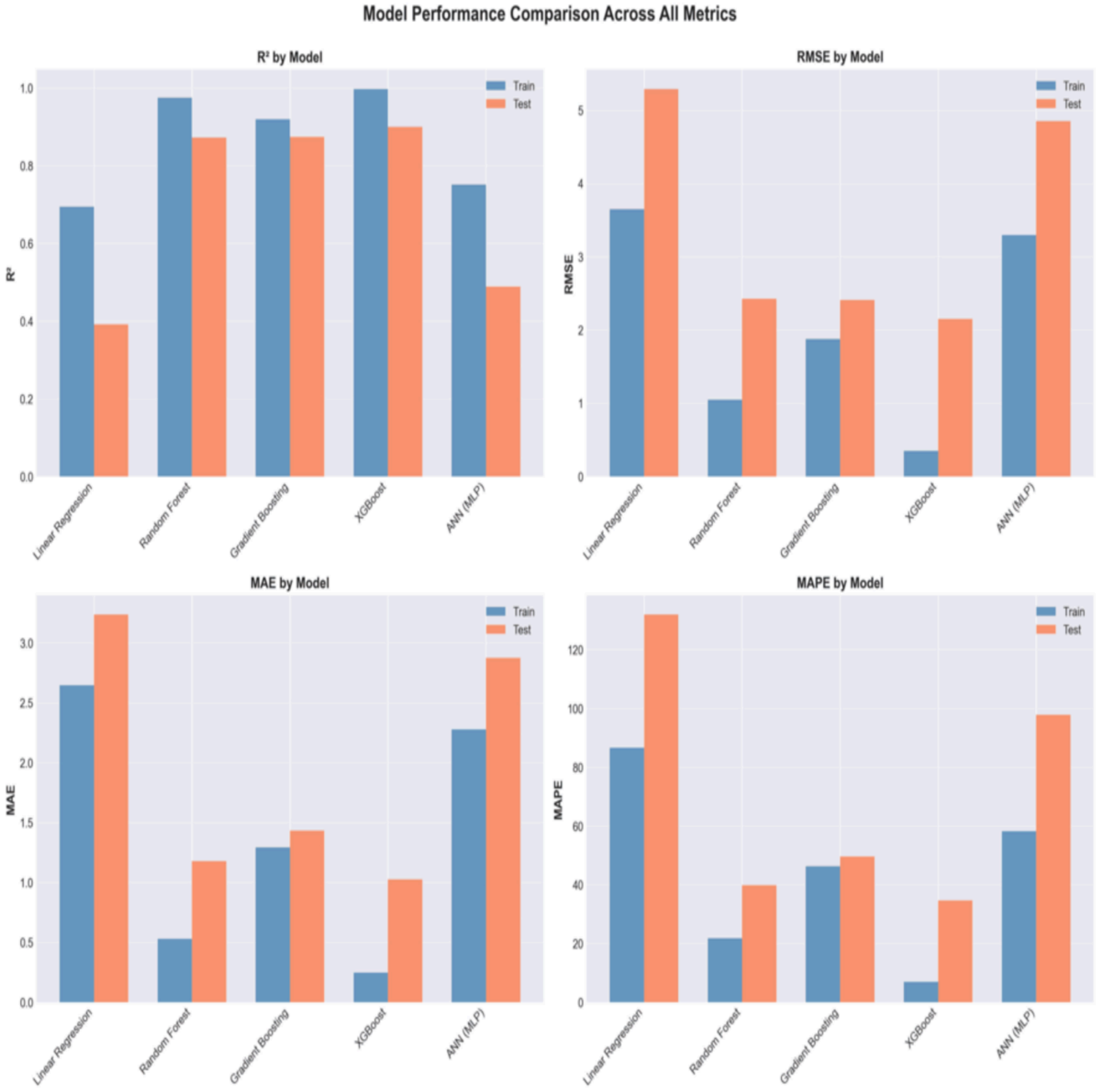

FIGURE 5: Comparative Performance Evaluation of Machine Learning Models

MAE, Mean Absolute Error; MAPE, Mean Absolute Percentage Error; RMSE, Root Mean Square Error; ANN, Artificial Neural Network; MLP, Multilayer Perceptron; XGBoost, Extreme Gradient Boosting

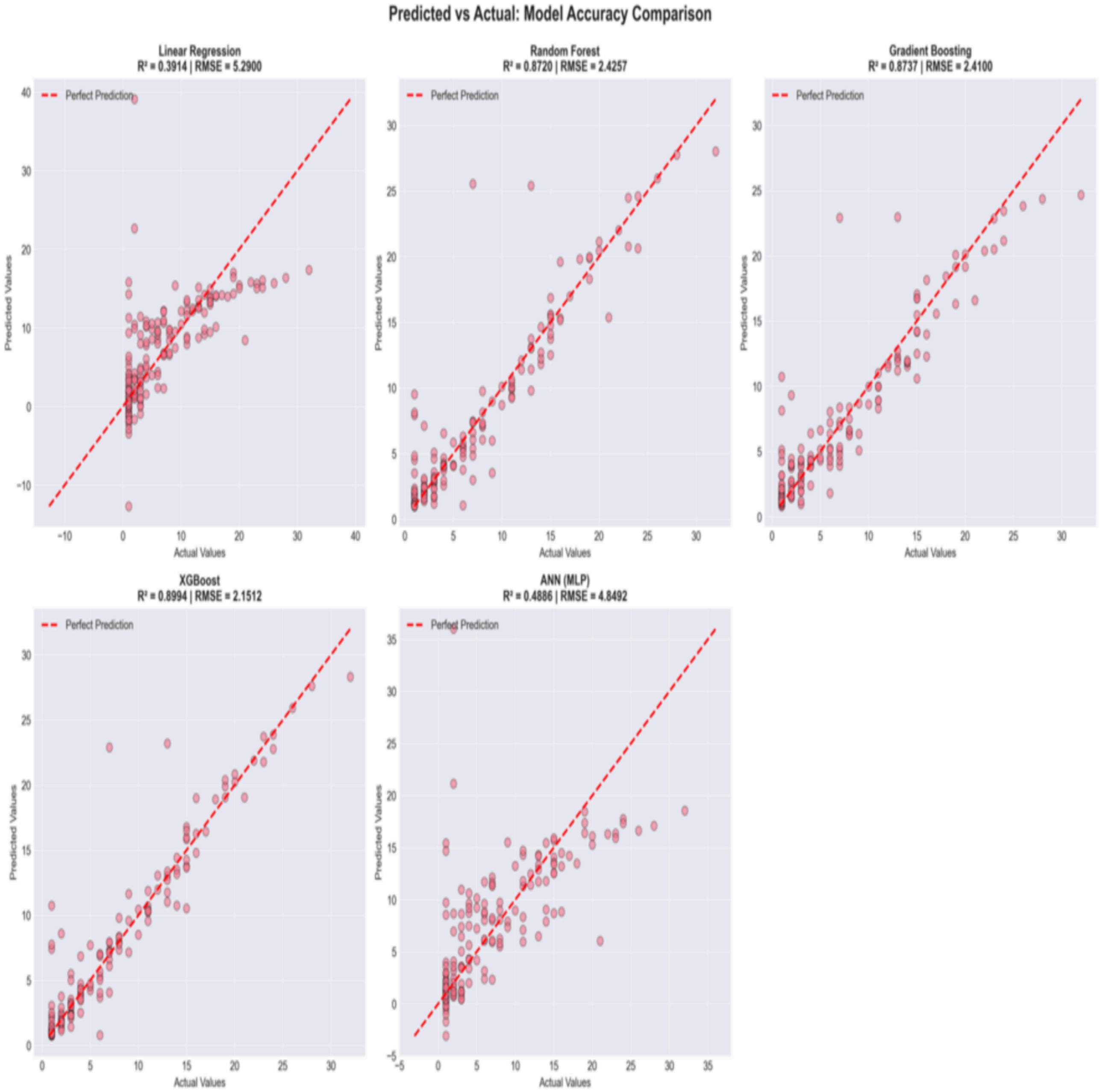

FIGURE 6: Predicted versus Actual Comparison of Model Accuracy

RMSE, Root Mean Square Error; ANN, Artificial Neural Network; MLP, Multilayer Perceptron; XGBoost, Extreme Gradient Boosting

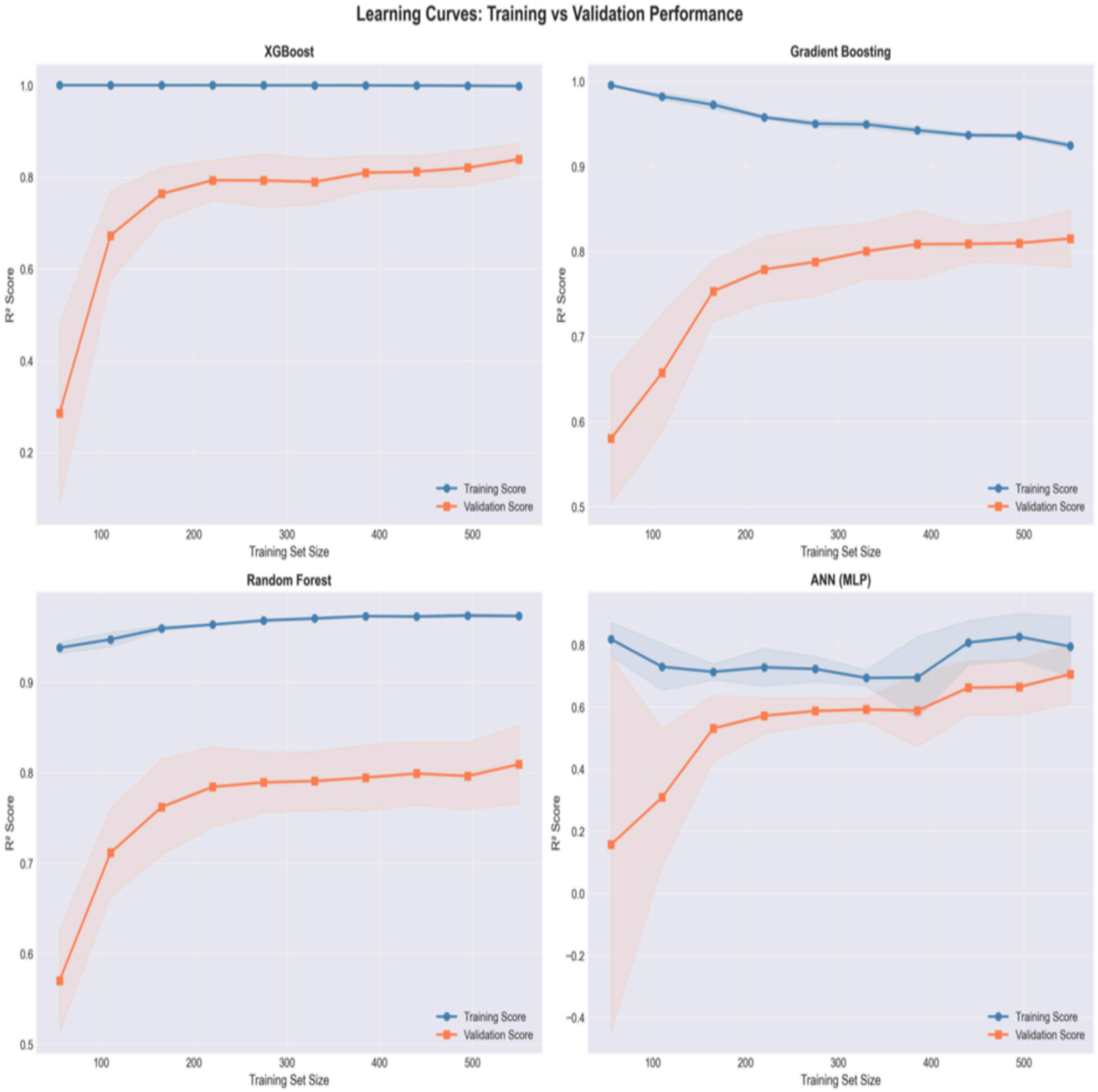

FIGURE 7: Learning Curves Showing Training and Validation Performance

ANN, Artificial Neural Network; MLP, Multilayer Perceptron; XGBoost, Extreme Gradient Boosting

Conclusions

This study developed and evaluated a supervised machine learning framework for predicting nighttime high-power aircraft engine run-up activity using historical operational data. A comparative analysis of five regression models - Linear Regression, Random Forest, Gradient Boosting, XGBoost, and Artificial Neural Network (MLP) - demonstrated clear quantitative differences in predictive performance. Among the evaluated models, XGBoost achieved the highest accuracy on the testing dataset, with a coefficient of determination of approximately $R^2 = 0.90$, an RMSE of about 2.15, a MAE close to 1.0, and a MAPE of approximately 35%. Random Forest and Gradient Boosting exhibited comparable but slightly lower performance, with testing R^2 values in the range of 0.87-0.88 and RMSE values between 2.4 and 2.5. In contrast, Linear Regression and ANN (MLP) showed significantly higher prediction errors, with testing R^2 values below 0.50 and RMSE values exceeding 4.8. Learning curve analysis further indicated that validation performance for ensemble-based models stabilized as the training set size increased, confirming consistent predictive behaviour across varying data volumes. The results indicate that nighttime high-power engine run-up activity is characterized by substantial variability, right-skewed distributions, and weak linear correlations among key operational and temporal variables. Exploratory data analysis and correlation assessment demonstrated that simple linear relationships are insufficient to represent the observed data structure. The superior performance of ensemble-based machine learning models highlights the effectiveness of data-driven approaches in capturing complex, multivariate patterns present in real-world aviation operational datasets. This study presented a supervised machine learning framework for predicting nighttime high-power aircraft engine run-up activity using historical operational data. Comparative evaluation of five regression models demonstrated that ensemble-based algorithms provide substantially higher predictive accuracy than linear and neural network approaches. Among the evaluated models, XGBoost achieved the best performance on the testing dataset ($R^2 \approx 0.90$, RMSE ≈ 2.15), indicating strong predictive capability within the analyzed dataset. The results demonstrate that machine learning techniques can effectively model variability in nighttime run-up activity based on available operational features.

The scope of this work can be extended in several directions. Future studies may incorporate additional explanatory features such as aircraft type, engine rating, maintenance category, or airport-specific operational constraints to enhance predictive capability. The integration of meteorological variables and operational context indicators may further improve model performance. Extending the framework to multi-airport datasets would allow assessment of spatial variability and model transferability. In addition, future work could explore hybrid modeling approaches that combine machine learning with physics-informed or rule-based constraints, as well as probabilistic modeling techniques to quantify predictive uncertainty. The proposed framework may also be adapted for real-time or near-real-time decision-support applications in airport ground operations management. Future studies should evaluate the proposed framework using multi-airport datasets and extended temporal validation to further assess model transferability across different operational environments.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Somashekar V, Shivakumar PN

Acquisition, analysis, or interpretation of data: Somashekar V, Shivakumar PN

Drafting of the manuscript: Somashekar V, Shivakumar PN

Critical review of the manuscript for important intellectual content: Somashekar V, Shivakumar PN

Supervision: Somashekar V

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

https://catalog.data.gov/dataset/nighttime-power-runups

References

1. Han B, Yao T, Li G, Song Y, Zhang Y, Dai Q, Yu J: Marginal reduction in surface NO2 attributable to airport shutdown: A machine learning regression-based approach. Environmental Research. 2022, 214:114117. 10.1016/j.envres.2022.114117
2. Bo X, Xue X, Xu J, Du X, Zhou B, Tang L: Aviation's emissions and contribution to the air quality in China. Atmospheric Environment. 2019, 201:121-131. 10.1016/j.atmosenv.2019.01.005
3. Wan J, Zhang H, Zhang Q, Li MZ, Xu Y: Deep learning framework for forecasting en route airspace emissions considering temporal-spatial correlation. Science of The Total Environment. 2023, 905:166986. 10.1016/j.scitotenv.2023.166986
4. Zhu C, Wu J, Liu M, Wang L, Li D, Kouvelas A: Recovery preparedness of global air transport influenced by COVID-19 pandemic: Policy intervention analysis. Transport Policy. 2021, 106:54-63. 10.1016/j.tranpol.2021.03.009
5. Rezo Z, Steiner S, Škurla Babić R: Comparative review of ICAO and EUROCONTROL flight carbon emission approximators. Sustainability. 2025, 17:6329. 10.3390/su17146329
6. Fichert F, Forsyth P, Niemeier HM: Aviation and Climate Change: Economic Perspectives on Greenhouse Gas Reduction Policies. Routledge, London; 2020. 10.4324/9781315572406
7. Mueller SC: Measuring and Modeling Aviation-Related Air Pollution in Near-Airport Communities [Doctoral dissertation]. Boston University, 2025.
8. Wen C, Lang J, Fu Y, et al.: Refined aircraft landing-takeoff activity modeling to improve the estimation of aviation CO2 and pollutants emissions. npj Climate and Atmospheric Science. 2025, 8:316. 10.1038/s41612-025-01195-6
9. Prashanth P, Eastham SD, Speth RL, Barrett SR: Aerosol formation pathways from aviation emissions. Environmental Research Communications. 2022, 4:021002. 10.1088/2515-7620/ac5229
10. Grobler C, Wolfe PJ, Dasadhikari K, et al.: Marginal climate and air quality costs of aviation emissions. Environmental Research Letters. 2019, 14:114031. 10.1088/1748-9326/ab4942
11. Eastham SD, Barrett SR: Aviation-attributable ozone as a driver for changes in mortality related to air quality and skin cancer. Atmospheric Environment. 2016, 144:17-23. 10.1016/j.atmosenv.2016.08.040
12. Cameron MA, Jacobson MZ, Barrett SRH, et al.: An intercomparative study of the effects of aircraft emissions on surface air quality. Journal of Geophysical Research: Atmospheres. 2017, 122:8325-8344. 10.1002/2016jd025594
13. Yim SHL, Lee GL, Lee IH, et al.: Global, regional and local health impacts of civil aviation emissions. Environmental Research Letters. 2015, 10:034001. 10.1088/1748-9326/10/3/034001
14. Kalafatelis AS, Nomikos N, Giannopoulos A, Alexandridis G, Karditsa A, Trakadas P: Towards predictive maintenance in the maritime industry: A component-based overview. Journal of Marine Science and Engineering. 2025, 13:425. 10.3390/jmse13030425
15. Chati YS, Balakrishnan H: Statistical Modeling of Aircraft Engine Fuel Flow Rate. 30th Congress of the International Council of the Aeronautical Sciences, Daejeon, Korea; 2016.
16. Kavallieratos G, Diamantopoulou V, Katsikas SK: Shipping 4.0: Security requirements for the cyber-enabled ship. IEEE Transactions on Industrial Informatics. 2020, 16:6617-6625. 10.1109/tii.2020.2976840
17. Gómez Ruiz MÁ, de Almeida IM, Pérez Fernández R: Application of machine learning techniques to the maritime industry. Journal of Marine Science and Engineering. 2023, 11:1820. 10.3390/jmse11091820
18. Angelopoulos A, Giannopoulos A, Nomikos N, Kalafatelis A, Hatziefremidis A, Trakadas P: Federated learning-aided prognostics in the Shipping 4.0: Principles, workflow, and use cases. IEEE Access. 2024, 12:6437-6454. 10.1109/access.2024.3350777
19. Achouch M, Dimitrova M, Ziane K, Sattarpanah Karganroudi S, Dhouib R, Ibrahim H, Adda M: On predictive maintenance in industry 4.0: Overview, models, and challenges. Applied Sciences. 2022, 12:8081. 10.3390/app12168081
20. Kalafatelis AS, Trochoutsos C, Giannopoulos AE, Angelopoulos A, Trakadas P: A stacking ensemble learning model for waste prediction in offset printing. ICIEAEU '23: Proceedings of the 2023 10th International Conference on Industrial Engineering and

Applications. 2023, 267-272. 10.1145/3587889.3588210

21. Zheng H, Paiva AR, Gurciullo CS: Advancing from predictive maintenance to intelligent maintenance with AI and IIoT. arXiv. 2020, 10.48550/arXiv.2009.00351
22. Dong Y, Tao J, Zhang Y, Lin W, Ai J: Deep learning in aircraft design, dynamics, and control: Review and prospects. IEEE Transactions on Aerospace and Electronic Systems. 2021, 57:2346-2368. 10.1109/taes.2021.3056086
23. Dong Y: Implementing deep learning for comprehensive aircraft icing and actuator/sensor fault detection/identification. Engineering Applications of Artificial Intelligence. 2019, 83:28-44. 10.1016/j.engappai.2019.04.010
24. Dong Y: An application of Deep Neural Networks to the in-flight parameter identification for detection and characterization of aircraft icing. Aerospace Science and Technology. 2018, 77:34-49. 10.1016/j.ast.2018.02.026
25. Deepika J, Reddy PM, Murari K, Rahul B: Predictive maintenance of aircraft engines: machine learning approaches for remaining useful life estimation. 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Goathgaun, Nepal. 2025, 1679-1686. 10.1109/ICMCSI64620.2025.10883526
26. Gajghate SS, Noor MM, Kumar S, et al.: A transformer guided multi modal learning framework for predictive and causal assessment of thermal runaway in high energy batteries. Scientific Reports. 2025, 15:37054. 10.1038/s41598-025-20886-x
27. Serafino G: Aircraft Trajectory Optimization for Weather Avoidance and Emission Reduction Applications [Doctoral dissertation]. Università degli Studi di Trieste, 2017.
28. Elechi CE, Chika C: Implication of aircraft noise on workers in Port Harcourt International Airport Omuagwa Rivers State. World Journal of Innovation and Modern Technology. 2021, 4:46-69.
29. Azzopardi MA, Whidborne JF: Computational air traffic management. 2011 IEEE/AIAA 30th Digital Avionics Systems Conference, Seattle, WA, USA. 2011, 1B6-1-1B6-22. 10.1109/DASC.2011.6095967
30. Mosley A: FY24 LDRD Annual Report PDF (No. LLNL--TR-2006342). Lawrence Livermore National Laboratory, Livermore, CA; 2025. 10.2172/2584734
31. Goodenough B: Electrification of a Heavy-Duty Off-Road Material Handler: Energy Savings and Emission Reductions [Doctoral dissertation]. Michigan Technological University, 2024.
32. Nighttime High Power Run-ups. (2026). https://catalog.data.gov/dataset/nighttime-power-runups.

How to cite this article:
V S, PN S (March 21, 2026) A Machine Learning–Based Predictive Framework for Nighttime High-Power Aircraft Engine Run-Up Operations. Cureus J Comput Sci 3 : es44389-026-00043-9. DOI https://doi.org/10.7759/s44389-026-00043-9