Title: Securing aviation networks: A generalized and robust machine learning solution for monitoring abnormalities in commercial aircraft

Authur names: Md. Atiqur Rahman , Md. Nawab Yousuf Ali , Fatema Tuj Johora , Aliza Ahmed Khan , Mohammad Nurul Huda

Abastact: Accurate abnormality prediction in commercial aircraft is critical for aviation safety, yet existing approaches often compromise reliability due to limited datasets or suboptimal feature selection. While several machine learning models have been proposed, most suffer from small sample sizes or inadequate handling of imbalanced data, limiting real-world applicability. This study addresses these gaps by developing an XGBoost-based predictive model trained on 9,949,927 flight samples, significantly larger than prior research. Our methodology incorporates SMOTEENN for data balancing and employs rigorous feature selection using Gradient Boosting Regressor with Bonferroni correction, identifying five key predictors: timestep, baroaltitude, velocity, longitude, and latitude. The model achieved robust performance with an accuracy of 0.9363 (95% CI: 0.9360-0.9366), precision of 0.9222, recall of 0.9719, PR-AUC of 0.9755, Balanced Accuracy of 0.9295, Log Loss of 0.1902, and Brier Score of 0.0516, indicating scalability, generalizability, and robustness, thereby outperforming existing approaches. Nested cross-validation confirmed robustness with accuracies of 0.927 ± 0.010 and 0.934 ± 0.008 for 5-fold and 10-fold schemes. SHAP and LIME analyses validated the decision logic, showing predictions driven by meaningful and logically consistent feature patterns. These results demonstrate that comprehensive data utilization, proper feature selection, and interpretability yield more reliable abnormality detection compared to models trained on filtered datasets. The study provides aviation safety systems with a scalable, transparent framework for large-scale anomaly monitoring and establishes a practical benchmark for future research and deployment of data-driven aircraft monitoring systems.

Keywords Automatic dependent surveillance-broadcast (ADS-B) Real-time flight anomaly detection XGBoost Gradient boosting regressor-based on bonferroni correction Large-scale flight data analysis

10.1016/j.rineng.2025.108884