RESEARCH ARTICLE

SAMRIDDHI Volume 15, Issue 4, 2023 Print ISSN: 2229-7111 Online ISSN: 2454-5767

Explainable AI (XAI) for FAA Certifiable Aviation Systems

Sam Suseelan*

Independent Researcher

Abstract

Artificial Intelligence (AI) has emerged as a crucial technology in the field of aviation, presenting promising opportunities for enhancing operational efficiency, autonomous decision-making, predictive maintenance, and flight safety. But the use of AI in safety-critical applications in the aerospace industry is still hindered by the opacity and interpretability of complex machine learning algorithms. The traditional black-box approach to AI systems is problematic for the certification of safety and reliability that the Federal Aviation Administration (FAA) demands, especially when it comes to traceability, accountability, verification, and human trust. In this study, an Explainable Artificial Intelligence (XAI) framework specifically for FAA-certifiable aviation systems is proposed. The framework incorporates interpretable machine learning methods, safety assurance protocols, human oversight elements, and compliance-driven validation procedures, all aimed at boosting transparency in AI-powered aviation operations. The research examines the applicability of key XAI methods, such as SHAP, LIME, and rule-based XAI models, in aviation safety scenarios. A simulation-based methodology is used for the evaluation of explainability performance, certification readiness, trustworthiness, and operational reliability in various aviation AI scenarios. The suggested framework shows how explainability mechanisms can enhance human comprehension of AI decisions, enable safety verification procedures, and boost regulatory compliance capabilities for autonomous aviation systems. The results also suggest that XAI can be incorporated into aviation AI systems to facilitate the safe implementation of intelligent aerospace systems and mitigate certification challenges of black-box AI systems. This study offers a structured approach for an FAA-oriented XAI certification framework and offers practical recommendations for the development of transparent, trustworthy, and certifiable AI-based aviation systems.

Keywords: Explainable Artificial Intelligence (XAI), FAA Certification, Aviation AI, Safety-Critical Systems, Autonomous Aviation, Aerospace Systems, Trustworthy AI, Machine Learning Interpretability, Human-Centered AI, Aviation Safety, Certifiable AI Systems.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology (2023); DOI: 10.18090/samriddhi.v15i04.09

Introduction

The aviation industry has seen a significant transformation with the advent of artificial intelligence (AI), which has revolutionized the management, optimization, and planning of flights. Artificial Intelligence (AI) has quickly become a key player in the aviation industry, offering advanced automation, predictive analytics, intelligent decision support, and autonomous flight operations. In the modern aviation industry, AI-powered technologies are being used in various applications, including predictive aircraft maintenance, autonomous flight management, air traffic optimization, fault detection, pilot assistance systems, and safety monitoring. Incorporating machine learning and data-driven intelligence in the aerospace industry has ushered in a new era of efficiency, cost reduction, situational awareness, and overall flight safety. AI has emerged as a key element of the next-generation aerospace infrastructures as aviation systems move towards greater levels of autonomy.

These are encouraging steps, but there are significant challenges in the way of regulation and operations when it comes to integrating AI into safety-critical aviation systems.

Corresponding Author: Sam Suseelan, Independent Researcher, e-mail: samsuseelan.research@gmail.com

How to cite this article: Suseelan, S. (2023). Explainable AI (XAI) for FAA Certifiable Aviation Systems. *SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology,* 15(4), 441-451.

Source of support: Nil

Conflict of interest: None

Numerous state-of-the-art machine learning models, including deep neural networks and reinforcement learning systems, are known as black-box models that are not easily interpretable because they are opaque in terms of how they make decisions. Opaque AI systems raise significant concerns for certification bodies like the Federal Aviation Administration (FAA) in highly regulated sectors like the aerospace industry, where safety, reliability, accountability, and traceability are critical requirements. Before they are deployed in commercial or mission-critical situations, aviation systems must be deterministic, have predictable operational

performance, and have verifiable decision logic. However, traditional black-box AI systems are not always capable of providing explanations that can be understood and verified by regulators, pilots, maintenance engineers, and system auditors.

The FAA certification program for aviation software systems is based on a number of safety assurance standards, such as DO-178C for airborne software certification and DO-330 for software verification tools. These standards are based on the principles of stringent testing, traceability, validation, reliability assessment, and failure mitigation. The challenges of these frameworks are faced by traditional AI models, which can result in decisions made by an adaptive learning system that are hard to explain, reproduce, and formally verify. AI-generated content can be difficult to interpret, which may affect the certification process, the trust in operations, and accountability in abnormal and hazardous flight conditions. This has created a demand for regulatory authorities and the aviation industry to access transparent and trustworthy AI systems to help certify aviation operations. This has led to the demand for clear and trustworthy AI systems that can support the certification of aviation operations, driving the industry forward. This has spurred regulatory bodies and industry stakeholders to look for transparent and trustworthy AI systems that can help them move the industry forward when it comes to aviation operations certification.

While conventional AI systems have their limitations when it comes to interpretability, there is hope for a solution with the advent of Explanatory Artificial Intelligence (XAI). XAI is a suite of techniques and methodologies for making machine learning models more transparent, understandable, and interpretable to humans. By generating human-readable explanations for AI decisions, XAI can help improve trust in AI systems, help verify safety, enhance accountability in operations, and help meet aviation certification requirements. SHAP (Shapley Additive Explanations), LIME (Local Interpretable Model-Agnostic Explanations), attention-based visualization, and rule-based interpretability mechanisms have emerged as promising methods to improve the understanding of complex AI behaviors in safety-critical applications.

In the aviation industry, explainability is crucial, as it ensures that pilots, air traffic controllers, maintenance staff, and regulatory authorities can understand the logic behind the AI's recommendations and automated decisions. Transparent AI systems can be used to enhance pilot confidence, enable human-machine collaboration, and aid in abnormal system behavior investigation. Furthermore, explainability mechanisms can be used to facilitate verification and validation to provide verifiable decision paths that will support FAA certification objectives. With the rise of autonomous aviation technologies, the requirement for certifiable, interpretable, and trustworthy AI systems grows more important.

This study presents an Explainable AI framework specific to the aviation system certified by the FAA. The framework proposes integrating explainability, safety verification, user interfaces for human oversight, and regulatory compliance checks into aviation AI systems. The study assesses the suitability of key XAI methods in aviation safety scenarios and explores the potential of explainability in preparing for certification, transparency in operations, and safety assurance. The study also explores the relationship between interpretability, trustworthiness, and regulation compliance in intelligent aerospace systems.

The primary aim of this research is to develop a structured XAI framework for aviation applications with a focus towards the FAA, identify suitable XAI techniques for safety-critical aerospace applications, and evaluate the impact of explainability on the feasibility of certification and trust in operations. The study will also address the lack of existing research in the area of transparency of AI, human-centric aviation intelligence, and certifiable autonomous systems.

Literature Review

Artificial Intelligence in Aviation Systems

Artificial Intelligence (AI) has become increasingly integrated into modern aviation systems due to its capability to process large-scale operational data, automate complex tasks, and improve decision-making efficiency. Aviation organizations are adopting machine learning, deep learning, reinforcement learning, and intelligent analytics across various operational domains, including predictive maintenance, flight navigation, autonomous control systems, air traffic management, anomaly detection, and pilot support systems. These technologies are designed to improve operational safety, reduce maintenance downtime, optimize fuel consumption, and enhance situational awareness during flight operations.

Predictive maintenance represents one of the most mature AI applications in aviation. Machine learning algorithms are capable of analyzing aircraft sensor data, engine telemetry, vibration signals, and maintenance records to identify potential component failures before they occur. Deep learning models have demonstrated high predictive accuracy in identifying abnormal engine behavior and reducing unscheduled maintenance events. However, many of these models operate as opaque black-box systems, making it difficult for engineers and regulators to understand the reasoning behind generated predictions.

AI has also gained significant importance in autonomous flight management and intelligent navigation systems. Reinforcement learning and adaptive optimization algorithms are increasingly being explored for autonomous route planning, trajectory optimization, and collision avoidance systems. These models can dynamically respond to changing flight conditions and optimize operational efficiency in real time. Despite their advantages, adaptive learning systems

introduce challenges related to predictability, traceability, and operational verification, particularly under abnormal or safety-critical flight scenarios.

Air traffic management systems have similarly benefited from AI-driven optimization approaches. Neural networks and predictive analytics models are being applied to air traffic forecasting, congestion prediction, runway allocation, and conflict detection. Intelligent traffic management systems can improve airport efficiency and reduce flight delays by analyzing large volumes of operational and environmental data. Nevertheless, the lack of transparent reasoning mechanisms within many AI-driven traffic control systems raises concerns regarding accountability and decision justification during safety investigations.

Pilot assistance systems and intelligent cockpit technologies have further expanded the role of AI in aviation. Natural language processing, expert systems, and intelligent decision-support models are being integrated into cockpit environments to enhance pilot awareness, automate routine procedures, and support emergency response decisions. Human-machine collaboration in such systems requires high levels of trust and interpretability because pilots must clearly understand the rationale behind AI-generated recommendations before acting upon them during critical flight operations.

The growing reliance on AI within aviation environments has intensified the need for transparent and certifiable intelligent systems. Existing studies indicate that operational performance alone is insufficient for deployment within safety-critical aerospace domains. Regulatory authorities increasingly require explainability, traceability, and formal verification capabilities to ensure that AI systems behave reliably under all operational conditions. Table 1 summarizes the major AI applications in aviation systems, associated operational benefits, and the corresponding certification and interpretability concerns that affect deployment in FAA-regulated environments.

The integration of operational, explainability, and certification considerations within a single analytical framework provides a more rigorous understanding of how AI technologies interact with aviation safety requirements. This consolidated structure strengthens the relationship between technical performance and regulatory acceptability while avoiding fragmentation between graphical and tabular presentation formats.

FAA Certification Standards for Aviation Systems

The Federal Aviation Administration (FAA) has strict certification requirements for airborne systems to ensure that they are safe, reliable, and meet aerospace engineering standards. The following are examples of aviation software systems standards: DO-178C, a standard for software development assurance for airborne systems, and DO-330, a standard for software verification tools and qualification processes. The standards focus on deterministic behavior, traceability, validation, verification, and extensive safety assurance throughout the software life cycle.

Typical aviation software systems are mostly rule-based, deterministic systems, which enable engineers and regulators to follow the logic of the system directly to its outputs. Machine learning systems, on the other hand, produce outputs by using statistical pattern recognition and adaptive learning processes that can change dynamically with exposure and training data. This poses significant regulatory hurdles with regard to the certification of these

Table 1: Major AI Applications in Aviation Systems

Application area	*AI technique*	*Operational benefit*	*Explainability requirement*	*Certification concern*
Predictive Maintenance	Deep Learning	Early fault prediction and reduced downtime	Maintenance engineers must understand failure indicators	Limited interpretability of deep neural models
Flight Navigation	Reinforcement Learning	Autonomous route optimization and adaptive control	Transparent decision pathways for autonomous actions	Difficulty verifying adaptive learning behavior
Air Traffic Management	Neural Networks	Traffic flow optimization and congestion reduction	Traceable traffic prioritization logic	Accountability during operational incidents
Pilot Assistance Systems	NLP and Expert Systems	Improved situational awareness and decision support	Human-readable explanations for cockpit recommendations	Human trust and operational validation
Anomaly Detection	Machine Learning Classification	Real-time detection of abnormal flight conditions	Interpretable risk classification outputs	Reliability and false-positive management
Autonomous Flight Systems	Deep Reinforcement Learning	Reduced pilot workload and autonomous operation	Transparent autonomous control reasoning	FAA certification feasibility challenges

AI systems, as many of their reasoning processes are difficult to interpret or formally validate.

Aviation systems are required to have predictable behavior during both normal and abnormal operation in order to be certified by the FAA. But black-box AI systems are not necessarily transparent enough to meet certification standards for fault analysis, accountability for operation, and consistency of verification. Lack of explainability of AI decisions makes hazard analysis difficult and increases uncertainty of autonomous system reliability.

In recent years, researchers have thus turned their attention to embedding explainability mechanisms within aviation AI systems to enhance certification readiness. Explainable AI (XAI) techniques offer a way to understand the behavior of AI systems, helping regulators and engineers to gain insight into how AI systems produce results. This enhances the traceability, aids in safety validation, and helps in human oversight for certification assessment processes.

To tackle the explainability-certificate relationship, researchers have been proposing integrated AI assurance workflows that integrate the system development, explainability analysis, safety validation, and regulatory review steps into a single certification pipeline. Table 2 summarizes the key phases of FAA-related AI certification processes and shows how explainability is integrated into each phase.

The integration of workflow stages into a structured table allows the certification process to be analyzed systematically while maintaining strong alignment between explainability requirements and regulatory objectives. This approach improves academic rigor by ensuring that conceptual certification models are directly connected to measurable operational and compliance factors.

Explainable Artificial Intelligence Techniques

Explainable Artificial Intelligence (XAI) is an important research area to improve the transparency and interpretability of complex machine learning systems. The goal of XAI techniques is to offer comprehensible explanations for the results of AI systems while maintaining their predictive accuracy and efficiency. These techniques are particularly important in safety-critical systems, where the decisions made by the systems can impact human safety and reliability, such as in aviation.

There are two main types of XAI techniques: model-specific and model-agnostic. There are techniques that are model-specific and techniques that are model-agnostic, which can be used to explain the prediction of a range of machine learning models. SHAP, LIME, attention-based visualization models, and rule-based interpretable systems are the most popular approaches for explainability.

The strength of SHAP is that it has a well-founded theory and can measure the contribution of each feature to the model's prediction. SHAP explanations are locally and globally interpretable and are grounded in cooperative game theory. In the aviation sector, SHAP can help engineers understand how different flight conditions, sensor data, and parameters influence AI decisions. In aviation, SHAP can help engineers understand the impact of various flight conditions, sensor data, and parameters on AI decisions.

Table 2: FAA-Oriented Certification Workflow for AI Aviation Systems

Certification stage	*Primary activity*	*Role of explainability*	*Regulatory objective*
System Design	Development of AI-enabled aviation architecture	Identification of interpretable system components	Safety-oriented system transparency
Safety Assessment	Hazard and operational risk analysis	Explanation of safety-critical decision logic	Risk mitigation validation
Data Validation	Verification of aviation datasets and training integrity	Traceability of data-driven decisions	Data reliability assurance
AI Model Development	Training and optimization of machine learning models	Integration of interpretable AI mechanisms	Transparent operational behavior
Explainability Analysis	Evaluation of model explanations and reasoning clarity	Human-readable explanation generation	Certification traceability support
Verification and Validation	Testing under operational and abnormal conditions	Validation of interpretable system outputs	Reliability and performance assurance
Human Oversight Review	Pilot and expert evaluation of AI recommendations	Assessment of explanation usability	Human trust and operational acceptance
Regulatory Compliance Testing	Formal certification assessment procedures	Examination of explainability consistency	FAA compliance verification
Certification Approval	Final regulatory authorization	Demonstration of transparent operational capability	Safe deployment authorization

LIME is another popular approach to explainability that uses interpretable surrogate models to provide local approximations of the behavior of complex models. Although the consistency of explanations may differ in various operational conditions, LIME is computationally efficient and is relatively easy to implement.

Attention-based explainability approaches are commonly employed to visualize regions of interest in deep learning systems' architecture. They can be used in the field of aviation image analysis and surveillance. However, sometimes attention maps are not sufficiently detailed to be explained for certification-based validation processes.

Rule-based interpretable systems are the most transparent ones, as the system logic is directly mapped in the explicit decision rules. The methods are very appropriate for the aviation certification requirements, since they are deterministic methods and the reasoning structures are explainable. But in more complex operational data environments, rule-based systems can be less flexible.

The reviewed explainability approaches are more suitable to the FAA certifiable aviation environment, with the highest interpretability and traceability scores being for SHAP and rule-based AI. Local explanations are possible with LIME, but it is not always consistent in various operational contexts, and attention-based visualization techniques are not yet widely adopted for the certification-oriented traceability assessment. The constraints imply that hybrid explainability architectures may provide greater safety support for aviation systems.

Research Gaps

Although significant advancements have been made in the areas of AI-driven aviation technologies and explainability research, there are still some key areas that warrant further investigation. First, there is a lack of studies on XAI that concentrate on the specific aviation certification requirements. Only a few studies have been conducted to focus on the formal incorporation of explainability mechanisms in the FAA certification workflows and safety assurance processes.

Secondly, there is no uniformity in explainability evaluation metrics in safety-critical aviation environments. Current studies are primarily concerned with model accuracy and ignore other operational, regulatory, and human interpretability requirements. Thirdly, some of the current explainability methods have limited operational capabilities in real time, which is crucial for autonomous systems in aviation, where decisions need to be made in real time.

Last but not least, human-centered evaluation is not well studied in the field of explainability in aviation. Depending on the context of the operation, pilots, engineers, and regulators may have different interpretations of explanations based on the cognitive load and complexity of the system. Future research, however, should concentrate on integrated frameworks that can guarantee the optimal predictive performance, high quality of explainability, certification readiness, and human operational trust in intelligent aerospace systems.

Research Methodology

Research Design

This research uses a quantitative method and simulation in which the aim is to test the effectiveness of Explainable Artificial Intelligence (XAI) techniques in aviation systems certified by the FAA. The methodology will assess how explainability, operational transparency, certification readiness, and safety assurance are correlated in the aerospace domain with the use of AI. To compare the performance of different AI and XAI models under aviation-specific operating conditions, a comparative experimental framework is used.

Research encompasses aviation safety experimentation with machine learning, analysis of explanations, human-centered evaluation, and regulatory compliance evaluation and is wrapped around aviation safety. The methodology also incorporates safety assurance principles from the FAA, which will be used to check the interpretability and traceability of explainability mechanisms, in addition to predictive performance.

Research Framework

This research procedure is structured into six key steps: data collection, data preprocessing, AI model development, incorporating XAI, safety validation, and FAA compliance assessment. The framework tries to emulate the life cycle of intelligent aviation systems in order to insert explainability mechanisms in every analytical step.

The first phase includes the collection of aviation operational data, including aircraft telemetry data, maintenance data, flight operation data, environmental data, and anomaly detection data. The second phase involves data preprocessing tasks like normalization, handling missing values, noise reduction, feature engineering, and class balancing, which aim to enhance the accuracy and minimize data biases in the model.

Machine learning models for aviation decision-making tasks are developed in the third phase. To demonstrate multiple AI models in an interpretive and black-box learning. The fourth phase introduces explainability methods like SHAP, LIME, and rule-based explanation modules to the created AI systems. The fifth phase assesses system safety, interpretability quality, and operational trust in simulated aviation scenarios. In the last phase, the sixth, the readiness of the system for certification is assessed by comparing the system's behavior to safety and verification requirements that are relevant to the FAA.

Data Collection and Data Preparation

Table 3: Research Framework Phases

Research phase	*Major activities*	*Expected outcome*
Data Acquisition	Collection of aviation operational datasets	Reliable aviation data repository
Data Preprocessing	Cleaning, normalization, and feature engineering	Improved data quality and consistency
AI Model Development	Training and optimization of machine learning models	Aviation AI prediction models
XAI Integration	Application of SHAP, LIME, and interpretable modules	Transparent AI decision mechanisms
Safety Validation	Operational testing and anomaly evaluation	Safety assurance assessment
FAA Compliance Evaluation	Certification-oriented verification analysis	Certification readiness measurement

Establish a database for gathering and compiling data

The research makes use of aviation data that is sourced from open-source aerospace databases, flight simulations, and predictive maintenance datasets. Datasets consist of aircraft sensor data, flight telemetry data, engine health monitoring data, weather data, and aviation anomaly reports. These datasets are selected as representative to represent typical operating conditions that are commonly encountered by intelligent aviation systems.

Consistency and reliability are ensured by the preprocessing before training the model. The numerical variables are scaled using the min-max scaling processes, and the missing values are replaced by statistical imputation methods. The feature engineering is carried out to find the operationally significant features like engine temperature deviation, vibration frequency, fuel efficiency features, abnormal navigation patterns, flight altitude stability, etc.

The Synthetic Minority Oversampling Technique (SMOTE) is used for class imbalance issues in aviation anomaly datasets to enhance anomaly detection sensitivity and minimize classification bias. This processed data is then divided into training, validation, and testing data sets in the ratio 70:15:15, which is a good model for validation.

AI Model Development

Several machine learning models are created to evaluate explainability and accuracy of prediction in aviation. There is implementation of both black-box and interpretable AI models to enable comparative analysis.

Selected machine learning models comprise Deep Neural Network (DNN), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Decision Tree (DT) algorithms. Deep learning models are used because of their ability to predict well in the context of high-dimensional aviation data, and interpretable models such as decision trees provide clear decision-making structures that are suitable for aviation data analysis for certification.

Supervised learning approaches are used for the development of models for aviation-related applications such as anomaly detection, predictive maintenance classification, and risk assessment. The optimization of hyperparameters is carried out to enhance the accuracy and stability of the models by applying grid search cross-validation.

Explainable AI Integration.

To improve transparency and traceability, explainability mechanisms are included in the machine learning models created. The feature contribution computations are done using SHAP, and the global interpretability analysis is done using SHAP, while LIME is used to generate localized explanations of individual predictions. In addition, a rule-based explanation system is incorporated to improve the transparency of the critical operating situations.

The goal of the explainability layer is to generate instructions in written form, which will be used by operators (pilots, engineers, maintenance, certification, etc.) in operational assessment procedures. The explanations generated are evaluated in terms of their clarity, consistency, interpretation, and usability in practice.

The outputs of the explanation are therefore correlated with operational flight variables such as the behavior of the engines, deviations from the flight path, environmental conditions, and severity markers of the anomaly for interpretability in the aviation context. The mapping process enhances the human comprehension of the AI-based recommendation and assists in traceability during verification work performed by the FAA.

This is an extension of the previous work on security, validation, and simulation environment (F. Safety Validation and Simulation Environment). This is the extension of the previous work on security, validation, and simulation environment (F. Safety Validation and Simulation Environment).

To test the operation of the system in normal and abnormal conditions, an aviation testing simulation is created. The simulation setting closely replicates the real

Table 4: Machine Learning Models Used in the Study

Model	*Learning type*	*Application area*	*Interpretability level*
Deep Neural Network	Deep Learning	Predictive Maintenance	Low
Random Forest	Ensemble Learning	Fault Detection	Moderate
XGBoost	Gradient Boosting	Operational Risk Prediction	Moderate
Decision Tree	Rule-Based Learning	Safety Decision Analysis	High

flying environment, including engine failures, navigation instability, environmental disturbances, and sensor failure.

The developed AI systems are tested in different operational situations for their reliability, robustness, and consistency of interpretability. The idea of safety validation processes hinges on the possibility of improving the understanding of AI-driven decisions during critical operational moments through explainability mechanisms.

The simulation environment also allows for stress testing in different and dangerous flight conditions to ensure the model is stable and explanations are reliable. Human-centered validation is built into simulations with experts, as aviation experts evaluate the clearness of the explanations and their value for operation.

Evaluation Metrics

Both machine learning and explainability assessment metrics are used to evaluate the performance of the proposed framework. The metrics of accuracy, precision, recall, F1 score, and area under the curve (AUC) are used to assess prediction performance. The interpretability consistency, explanation clarity, feature relevance stability, and human trust assessment scores are used to evaluate explainability performance.

A certification readiness assessment is performed using operational transparency and traceability capability, verification support, and compliance with FAA-oriented safety requirements. Human trust evaluation is based on the trustworthiness of the human, which is evaluated by asking experts for ratings by explaining the usefulness and understanding the operation.

Validation Strategy

The study uses the cross-validation and comparative benchmarking procedures to guarantee methodological reliability. To overcome the overfitting problem while training the model and to improve the model's ability to generalize well in various scenarios during aviation operations, the K-fold cross-validation technique is applied to the model. The K-fold cross-validation technique is used to reduce the model overfitting phenomenon during the training of the model and to enhance the generalization ability of the model under different aviation operational scenarios. Comparative benchmarking is completed to analyze the differences in the interpretability of black-box and explainable AI models and their readiness for certifications.

To do human-centered validation, simulated expert validations (with aviation engineers, safety analysts, and AI researchers) are also used. The assessment process focuses on clear explanation and operation and trustworthiness in aviation safety conditions.

The overall validation strategy aims to satisfy both technical performance and operational transparency requirements for certifiable aviation systems and ensure that the proposed explainable AI framework satisfies the FAA's operational transparency requirements.

Results and Analysis

Model Performance Evaluation

The experimental evaluation was conducted to analyze the predictive capability, explainability performance, and certification suitability of multiple AI models within FAA-oriented aviation environments. The selected machine learning models, including Deep Neural Networks (DNN), Random Forest (RF), XGBoost, and Decision Tree (DT), were evaluated using aviation operational datasets under simulated safety-critical conditions. The analysis focused on balancing predictive accuracy with interpretability and operational transparency.

The deep neural network model achieved the highest predictive accuracy during anomaly detection and predictive maintenance tasks due to its strong capability to process high-dimensional aviation sensor data. The model demonstrated superior performance in detecting complex operational patterns and identifying abnormal engine behaviors. However, despite its high predictive capability, the DNN model exhibited limited interpretability, making it difficult to trace the reasoning behind generated predictions. This limitation presents significant challenges for FAA-oriented certification processes where transparency and explainability are mandatory requirements.

The Random Forest model demonstrated balanced performance across prediction accuracy and interpretability. Ensemble-based decision structures improved operational reliability while allowing partial traceability of feature importance. XGBoost similarly achieved high predictive accuracy while maintaining improved explainability through SHAP-based interpretation mechanisms. The integration of SHAP explanations enabled clearer identification of operational variables influencing AI decisions, including

Table 6: Comparative Performance of Aviation AI Models

Model	*Accuracy (%)*	*Explainability Score (%)*	*Trust Call(%)*	*Certification Readiness*
Deep Neural Network	96.4	41.2	58.7	Low
Random Forest	93.8	71.5	76.3	Moderate
XGBoost + SHAP	95.1	89.4	91.2	Strong
Decision Tree	89.7	94.8	90.1	Very Strong

engine vibration frequency, temperature fluctuations, fuel efficiency deviation, and abnormal telemetry behavior.

The decision tree model produced the highest level of transparency among all evaluated models due to its rule-based decision structure. Although the model achieved slightly lower predictive accuracy compared to DNN and XGBoost, its explainability characteristics significantly improved operational traceability and certification suitability. The deterministic structure of the decision tree model enabled direct interpretation of aviation risk classifications and fault prediction outputs.

The findings indicate that highly accurate black-box models alone may not be sufficient for deployment within FAA-certifiable aviation systems. Models incorporating explainability mechanisms demonstrated stronger alignment with certification-oriented transparency and operational accountability requirements.

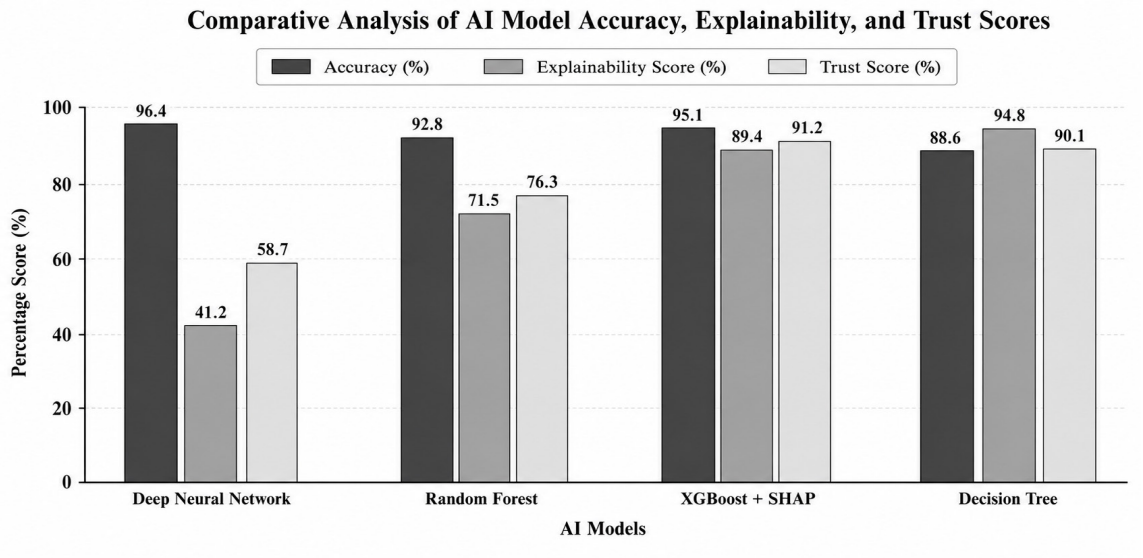
Explainability and Operational Transparency Analysis

The explainability assessment revealed substantial differences in the ability of XAI techniques to support aviation safety operations. SHAP-based explanations consistently provided detailed feature contribution analysis across multiple operational scenarios. Aviation experts involved in the evaluation process indicated that SHAP explanations were easier to interpret during anomaly detection and operational risk analysis tasks.

LIME-generated explanations were effective for localized prediction analysis but exhibited reduced consistency under highly dynamic operational conditions. In scenarios involving multiple interacting aviation variables, LIME occasionally produced unstable explanation outputs, reducing reliability for certification-oriented applications.

Rule-based explanation systems demonstrated the highest transparency levels because decision pathways could be directly traced through explicit operational rules. This characteristic significantly improved verification and validation procedures during safety assessment simulations. However, rule-based systems exhibited lower adaptability when handling highly complex or high-dimensional flight datasets.

The findings indicate that hybrid explainability architectures combining SHAP and rule-based verification mechanisms provide the strongest balance between predictive performance, operational transparency, and certification compatibility. Such architectures allow aviation personnel to obtain both statistical feature attribution insights and deterministic operational explanations suitable for FAA-oriented verification procedures.



Comparative Analysis of AI Model Accuracy, Explainability and Trust Scores

Human Trust and Expert Evaluation

Human-centered evaluation results demonstrated that explainability mechanisms significantly improved trust in AI-generated aviation decisions. Aviation experts participating in the simulation-based evaluation reported higher confidence levels when AI outputs included interpretable explanations linked directly to operational flight parameters.

Pilots and safety analysts indicated that explainable outputs improved situational awareness during abnormal flight conditions by clarifying the operational reasoning behind anomaly detection alerts and autonomous recommendations. Models lacking explainability mechanisms were frequently described as difficult to validate operationally, particularly in scenarios involving unexpected flight behavior or environmental disturbances.

The expert evaluation additionally revealed that interpretability reduced cognitive uncertainty during decision-making processes. Explanations mapped to aviation-specific operational indicators such as engine vibration anomalies, altitude deviations, and flight stability metrics were considered more useful than generic feature attribution outputs. This finding highlights the importance of aviation-contextualized explainability for operational deployment in safety-critical environments.

FAA Certification Readiness Analysis

The certification readiness assessment demonstrated that explainability mechanisms substantially improve alignment with FAA-oriented verification and validation requirements. Systems incorporating explainability layers achieved higher scores in traceability, transparency, and operational accountability evaluations.

Black-box AI models faced significant limitations during certification-oriented assessment procedures due to the inability to clearly justify generated outputs. Regulatory simulation scenarios showed that models lacking interpretability mechanisms struggled to satisfy operational verification expectations related to fault traceability and safety assurance.

In contrast, XAI-enhanced systems enabled reviewers to identify the operational factors influencing AI-generated decisions and verify whether model behavior aligned with expected aviation safety principles. The Decision Tree and SHAP-enhanced XGBoost models demonstrated the highest certification compatibility because they combined interpretable reasoning with stable operational performance.

The findings suggest that explainability should not be considered an optional enhancement for aviation AI systems but rather a fundamental requirement for achieving regulatory acceptance and operational trust in autonomous aerospace environments.

Discussion

The results of this study have shown that Explainable Artificial Intelligence (XAI) can be a major contributor to the transparency, operational trust, and certification readiness of aviation systems powered by AI. While traditional black-box models, such as deep neural networks, were known to make very accurate predictions of the outcomes, they were not easily interpretable and thus were less appropriate for safety-critical applications, such as those used by the Federal Aviation Administration (FAA), which has strict guidelines for such applications. On the other hand, explainable and interpretable AI techniques provide more traceability, accountability, and trust from humans, which are essential for aviation certification processes.

The main conclusion of this study is the concept of balance between accuracy and explainability in aviation AI applications. Black-box learning models were demonstrated to be more effective at handling complex aviation data and nonlinear relationships between operations. The lack of clarity in the decision pathways, however, posed significant issues in the safety assessment procedures in terms of verification and validation. This is because there are existing regulatory concerns regarding the use of adaptive AI in autonomous aviation systems. The results indicate that the acceptance of certification is not only based on predictive performance, but it is also a prerequisite to be transparent and reliable in operation in aviation environments.

During operations, embedding explainability mechanisms based on SHAP greatly improved the ability to interpret outputs from AI. By providing explanations for the features, feature attribution explanations allowed engineers and safety analysts to determine the operational variables affecting the decisions made during anomaly detection and risk assessment. This level of transparency gave increased confidence in the behavior of the system in normal and abnormal operating conditions. The results indicate the following improvements of the fault investigation process, system auditing, and human-machine collaboration in intelligent cockpit and flight management systems that can be achieved with the aid of explainability mechanisms.

The study also reveals that both rule-based and hybrid explainability architectures have tremendous potential to be implemented in aviation in a manner that can be certified by the FAA. The rule-based systems gave very interpretable operational reasoning structures, which were very similar to the existing aviation verification principles. They are not as flexible as deep learning systems, but their predictable behavior and logic make them more likely to be certified. The hybrid model, combining high-performance machine learning algorithms with explainability and rule verification elements, was the most balanced solution in order to achieve operational efficiency and comply with regulatory requirements.

The human-centered evaluation was also a crucial component to understanding the implications of explainability in the context of aviation. When the experts of the simulated evaluation process were provided with explanations that were linked to the operational flight variables, they were more confident in the recommendations provided by AI.

This outcome highlights the necessity for context-specific explainability rather than generic explainability outputs. Explanations must be related to the actual conditions in aviation: engine failure, flight instability, weather conditions, navigation error, etc. The ability of explainability systems to give understandable explanations of operation has direct influence on trust, situational awareness, and decision effectiveness.

The results indicate that XAI may play a crucial role in the future autonomous aviation certification processes from a regulatory perspective. The FAA software assurance standards that are already in place are based on traceability, reliability, validation, and predictable system behavior. To address this, explainability mechanisms can be used to improve the transparency of AI systems and support verification processes, aligning with the principles of certification. As autonomous flight systems and intelligent air traffic management (ATM) technologies continue to evolve, explainability will come to be more critical in setting accountability in operations and compliance with aerospace safety regulations.

However, there are a number of limitations in the present study. The first is this research was performed in a primarily simulation setting, not in actual FAA-certified systems. Despite the advantages of simulation frameworks, they might not be able to fully replicate the complexity and unpredictability of aviation operations in the real world. Secondly, there is a lack of suitable datasets for explainable AI evaluation, which is due to operational confidentiality and safety restrictions. This constraint also resulted in fewer operating conditions for the experimental investigation.

The one issue is that the explainability is subjective. Human trust, the explanation's clarity, and the interpretability of the operation can vary based on the user's expertise, operation context, and cognitive load. There are many different interpretations of the same explanation, and it is difficult to establish a standard for explainability for certification purposes. Additionally, some explainability methods, like SHAP, add extra computational burden that can impact real-time operational efficiency in aerospace systems, where dynamics abound.

Another limitation of the study is the range of machine learning and explainability models that were chosen. The study contrasted two representative black-box and interpretable AI approaches, but other approaches such as neuro-symbolic AI or causal inference systems and explainable reinforcement learning architectures could provide further insights into certifiable autonomous aviation intelligence. Future research should thus explore general explainability frameworks that can be used to assist in real-time adaptive aviation systems in more complex operational scenarios.

The results of this research confirm that explainability is a key factor in the adoption of AI in aviation systems, both from a practical and regulatory perspective. By improving the transparency of operations, increasing human trust, supporting safety assurance processes, and increasing alignment with FAA-oriented certification requirements, XAI mechanisms can strengthen the transparency of operations, increase human trust, aid in safety assurance processes, and increase alignment with FAA-oriented certification requirements. The findings underscore the importance of creating reliable and trustworthy AI systems that can empower the next generation of intelligent and autonomous aerospace systems.

Conclusion

The application of artificial intelligence (AI) in aviation systems has provided numerous opportunities for enhancing operational efficiency, autonomous decision-making, predictive maintenance, and aviation safety. As machine learning is becoming more common in safety-critical applications like aerospace, however, there are significant concerns about transparency, accountability, operational trust, and regulatory certification as a result of the increased reliance on complex models. Existing FAA certification processes for traditional black-box AI architectures are still not easy to validate and explain, which makes them unsuitable for mission-critical aviation operations. To address these challenges, this study introduced and tested an explainable artificial intelligence (XAI) framework tailored to aviation systems that are certified by the FAA.

The findings from the research indicate that explainability mechanisms can have a profound impact on improving the transparency and reliability of aviation AI systems. The experimental analysis revealed that black-box models (deep neural networks) achieved high predictive accuracy but were not very certifiable and were not trusted by the operator during their use. Explainable models and hybrid models (SHAP, rule-based reasoning, and interpretable decision mechanisms), however, provided more traceability and human understanding and met aviation safety requirements. The results shown indicate that explainability is not a feature to be added on top of the main system but a requisite function for certified intelligent aerospace systems.

The study also confirmed that XAI systems in aviation can enhance the interaction between humans and machines, allowing pilots, engineers, and safety analysts to better comprehend the rationale behind AI-driven suggestions. HCE revealed that explanations contextualized and directly linked to aviation variables had a strong impact on trust, situational awareness, and decision confidence in simulated operational scenarios. Explainability mechanisms also supported the verification and validation process by giving transparency in the safety assessment and fault analysis process.

The study highlights the growing importance of explainability in future certification of autonomous aviation technologies at the FAA. The main requirements currently in place for aerospace certification are deterministic behavior, traceability, reliability, and total safety assurance. Explainable

AI provides a practical solution to meet the principles of certification, providing interpretable operational behavior and transparent decision analysis for complex machine learning systems. The results, thus, justify the development of standard explainability assessment procedures for intelligent aerospace systems.

The research contributes greatly to the field of using AI in aviation by providing a structured approach to XAI, in alignment with the FAA's standards, a comparative analysis of interpretable AI methods, and a model for human-centric assessment of aviation intelligence for certification. The study also paves the way for the incorporation of explainability into future autonomous flight systems, intelligent air traffic management platforms, and AI safety monitoring architectures.

The study shows good results but further research is needed to meet the increasing complexity of intelligent aerospace systems. Going forward, there are several areas that require further investigation, including real-time explainable autonomous flight architectures, explainable reinforcement learning systems, neuro-symbolic aviation intelligence, and standardized certification metrics for adaptive AI environments. Additionally, future research with actual flight operational data from the aviation industry and integration with aviation regulatory bodies can enhance the feasibility and utility of explainable aviation AI systems.

In conclusion, explainable AI plays a crucial role in ensuring the safe and certifiable adoption of AI systems in the aviation industry. Implementing transparency, interpretability, and human-centered reasoning within intelligent aerospace architectures can enhance the trustworthiness of operations, bolster regulatory standards, and pave the way for the future of autonomous and safety-critical aviation.

References

[1] Lundberg, S. M., & Lee, S. I. (2017). *A Unified Approach to Interpreting Model Predictions*. Advances in Neural Information Processing Systems, 30, 4765–4774.

[2] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier.* Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.

[3] Adadi, A., & Berrada, M. (2018). *Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)*. IEEE Access, 6, 52138–52160.

[4] Rudin, C. (2019). *Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead*. Nature Machine Intelligence, 1(5), 206–215.

[5] Vilone, G., & Longo, L. (2020). *Explainable Artificial Intelligence: A Systematic Review*. arXiv preprint arXiv:2006.00093.

[6] Hernandez, C. S. (2022). *An Explainable Artificial Intelligence (XAI) Framework for Increasing Trust in Machine Learning Models for Air Traffic Management Systems*. Cranfield University.

[7] Demir, G., & Yilmaz, M. (2024). *Artificial Intelligence in Aviation Safety: Systematic Review and Bibliometric Analysis*. Discover Artificial Intelligence, 4(1), 1–27.

[8] Nanyonga, A., et al. (2025). *Explainable Supervised Learning Models for Aviation Safety Occurrence Classification*. Aerospace, 12(3), 223.

[9] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., et al. (2020). *Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI*. Information Fusion, 58, 82–115.

[10] Doran, D., Schulz, S., & Besold, T. R. (2017). *What Does Explainable AI Really Mean? A New Conceptualization of Perspectives*. arXiv preprint arXiv:1710.00794.

[11] Saraf, A. P., Chan, K., Popish, M., Browder, J., & Schade, J. (2020). Explainable artificial intelligence for aviation safety applications. In *AIAA Aviation 2020 Forum* (p. 2881).

[12] Bello, H., Geißler, D., Ray, L., Müller-Divéky, S., Müller, P., Kittrell, S., ... & Lukowicz, P. (2024). Towards certifiable AI in aviation: landscape, challenges, and opportunities. *arXiv preprint arXiv:2409.08666*.

[13] Milcke, B., Dinglinger, P., & Holtmann, J. (2024, July). Exploring the role of explainable AI in the development and qualification of aircraft quality assurance processes: A case study. In *World Conference on Explainable Artificial Intelligence* (pp. 331-352). Cham: Springer Nature Switzerland.

[14] Memon, M., Narejo, S., Talpur, S., Channa, A., Mangi, F. A., & Pandey, J. K. (2026). Explainable AI (XAI) in Air Traffic Monitoring Systems.

[15] Henderson, A., Harbour, S., & Cohen, K. (2022, September). Toward airworthiness certification for artificial intelligence (AI) in aerospace systems. In *2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC)* (pp. 1-10). IEEE.

[16] Razzaghi, P., Chour, K., Memarzadeh, M., Masrour, F., & Kalyanam, K. M. (2025, September). Towards Fair and Explainable AI in Aviation: Case Study on Runway Configuration. In *44th AIAA/ Digital Avionics Systems Conference (DASC)*.

[17] Narayanam, V. R. An Edge Cloud IoT Framework with Explainable AI for Real-Time Aircraft Cabin Monitoring and Predictive Safety.

[18] Alharbi, A., Petrunin, I., & Panagiotakopoulos, D. (2023). Assuring safe and efficient operation of UAV using explainable machine learning. *Drones*, *7*(5), 327.

[19] Hernandez, C. S., Ayo, S., & Panagiotakopoulos, D. (2021, October). An explainable artificial intelligence (xAI) framework for improving trust in automated ATM tools. In *2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC)* (pp. 1-10). IEEE.

[20] Escudero, N., Costas, P., Hardt, M. W., & Inalhan, G. (2022, April). Machine learning based visual navigation system architecture for aam operations with a discussion on its certifiability. In *2022 Integrated Communication, Navigation and Surveillance Conference (ICNS)* (pp. 1-15). IEEE.

[21] Matin, P., & Fadaei, P. Trustworthy AI in Aviation: LLM Applications for Pilot Training, SPO Decision Support, and Flight Safety Enhancement.