International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

Automation of Aviation Digital Records using AI / ML

Prabhat Dubey[1], Ayan Rajput[2]

[1]Department of CSE, JP Institute of Engineering & Technology, JPIET Mawana Road Near Ganga Nagar, PIN-250001 Uttar Pradesh Delhi NCR, India
Email: *prabhatdubey994[at]gmail.com*

[2]Department of CSE, JP Institute of Engineering & Technology, JPIET Mawana Road Near Ganga Nagar, PIN-250001 Uttar Pradesh Delhi NCR, India
Email: *ayanrajput062[at]gmail.com*

Abstract: ***The aviation industry relies heavily on accurate, up-to-date digital records to ensure regulatory compliance and aircraft safety. With the growing volume and complexity of documentation-ranging from Work Orders (WO), Engineering Orders (EO), Service Bulletins (SB), Task Cards, Certificates, and regulatory records such as ESSA and FAA directives-manual tracking and verification of Airworthiness Directive (AD) status is increasingly inefficient and error-prone. This paper presents an AI/ML-driven automation framework that leverages Artificial Intelligence, Machine Learning, Natural Language Processing, and Generative AI Models to extract relevant keywords and entities from heterogeneous document types. By implementing intelligent keyword extraction techniques, our system enables automated identification, classification, and cross referencing of AD-related information within large-scale Digital Records Management systems. This not only streamlines compliance workflows but also improves data consistency across maintenance operations. Our approach significantly enhances the tracking and validation of critical records in Aircraft Maintenance, offering a scalable solution for next-generation aviation safety and documentation practices.***

Keywords: Artificial Intelligence, Machine Learning (ML), Natural Language Processing (NLP), Generative AI Models, Keyword Extraction, Digital Records Management, Automation

1. Introduction

The aviation industry depends on vast volumes of technical documentation to ensure the safety, compliance, and operational efficiency of aircraft fleets [1]. As digital transformation advances, there is a growing need to automate the management of these records, especially in tracking the status of Airworthiness Directives (ADs). Leveraging Artificial Intelligence (AI) and Machine Learning (ML) offers a scalable path forward for intelligent document interpretation and regulatory automation [2].

1) *Context and Industry Need*

The aviation industry is experiencing rapid growth in data volume, complexity, and compliance requirements [3]. Maintenance and regulatory oversight involve a wide variety of technical documents such as Work Orders (WO), Engineering Orders (EO), Service Bulletins (SB), Task Cards, Certificates, and regulatory notices from authorities like the FAA and ESSA. These records are crucial for maintaining the airworthiness of aircraft and ensuring compliance with mandatory Airworthiness Directives (ADs) [4]. However, manually reviewing and updating the status of ADs across these diverse document types is inefficient, error-prone, and unsustainable in a modern digital environment [5]. The need for intelligent, scalable automation in aviation document management has never been more pressing.

2) *Relevance of AI/ML in Aviation Re*

Recent developments in Artificial Intelligence (AI) and Machine Learning (ML), particularly Natural Language Processing (NLP), offer promising solutions for automating the interpretation of unstructured and semi-structured aviation documents [6]. Techniques such as keyword extraction and Generative AI Models enable the automated detection and classification of AD-related content within large-scale maintenance datasets [7]. These tools support efficient document parsing, entity recognition, and real-time status tracking of compliance information, reducing human error and improving operational agility [8].

3) *Objective and Scope of This Study*

This study proposes an AI/ML-driven framework for automating the extraction, classification, and tracking of AD-related information across heterogeneous aviation document sets [9]. By applying advanced keyword extraction techniques, the system aims to:
Minimize manual workload in AD validation [10],
Enhance consistency and accuracy in Digital Records Management [11], and
Enable predictive compliance monitoring [12].
The framework targets multiple aviation document formats and offers a modular approach for integration into existing maintenance and engineering systems.

2. Background and Related Work

Digital transformation in aviation has created both opportunities and challenges in managing the increasing volume of regulatory and maintenance documentation [13]. This section reviews the types of documents involved in Airworthiness Directive (AD) compliance and examines prior research leveraging AI/ML for automation in aviation record management.

1) *Airworthiness Documentation in Aviation*

Airworthiness Directives (ADs) are mandatory notifications issued by global aviation authorities such as the Federal

Aviation Administration (FAA) and European Union Aviation Safety Agency (EASA) to ensure aircraft safety and compliance [14]. These directives often relate to and are implemented through technical documents such as Engineering Orders (EOs), Work Orders (WOs), Service Bulletins (SBs), Task Cards, and Certificates of Compliance [15]. The responsibility to trace, validate, and act upon these ADs lies with Maintenance, Repair, and Overhaul (MRO) teams, who must manually cross-reference ADs across a variety of formats [16]. This process is labour-intensive, repetitive, and error-prone, making it a strong candidate for AI-driven automation solutions.

2) Applications of AI/ML in Document Management

Recent developments in Artificial Intelligence (AI) and Machine Learning (ML) -particularly in Natural Language Processing (NLP) -have enabled automated understanding of unstructured text data [17]. Techniques such as keyword extraction, entity recognition, document clustering, and semantic search have proven effective in automating the interpretation of aviation reports and logbooks [18]. For instance, NLP models have been used to classify safety incident reports, detect contributing factors, and cluster related maintenance issues [19]. These tools offer significant potential for the automation of compliance verification, enabling scalable and intelligent aviation document management systems.

3) Related Research on Aviation Data Automation

Several studies have demonstrated the applicability of AI and ML techniques in aviation contexts, including automated parsing of safety reports, predictive maintenance using flight data, and anomaly detection from sensor logs [20]. Digital twin frameworks and rule-based classification systems have also been proposed to simulate and monitor aircraft health [21]. However, there is a clear research gap in applying these technologies specifically to automate the extraction and tracking of AD-related information across various document types [22]. This paper builds on existing AI/NLP methods and applies them to the targeted problem of AD compliance, with a focus on real-world document sources and regulatory formats.

3. Methodology Proposed

This section presents the proposed framework for automating Airworthiness Directive (AD) tracking and compliance verification using AI/ML techniques [23]. The system combines natural language processing (NLP), domain-specific keyword modeling, and intelligent documents linking to classify directives and streamline maintenance workflows across diverse aviation documents
[24].

1) *System Architecture Overview*

The proposed solution follows a modular pipeline that processes and classifies unstructured aviation records and maps them to regulatory compliance requirements [25]. The architecture includes:

Document Ingestion Module: Accepts documents such as Engineering Orders (EOs), Work Orders (WOs), Service Bulletins (SBs), Task Cards, and regulatory certificates in various formats (PDF, DOCX, TXT, scanned images) [26].

Pre-processing Engine: Applies Optical Character Recognition (OCR) where required, removes irrelevant formatting (headers, footers, noise), and converts all content to normalized plain text [27].

Keyword Extraction and Classification Module: Uses deep learning-based NLP models (e. g., spaCy) to identify AD-related keywords, aircraft models, part numbers, serial numbers, compliance actions, and applicable timelines [28]. The model is trained with a domain-specific lexicon built from FAA, EASA, and airline-specific documents. Compliance Mapping Engine: Cross-references extracted data with authoritative AD databases and tracks the compliance status (open, closed, or partial) of each directive across linked maintenance records [29].

Output Layer: Presents the processed results via API integrations, dashboards, and digital audit logs for maintenance teams and compliance officers [30]. This architecture supports batch processing and real-time streaming of documents, making it scalable for airline and MRO environments.

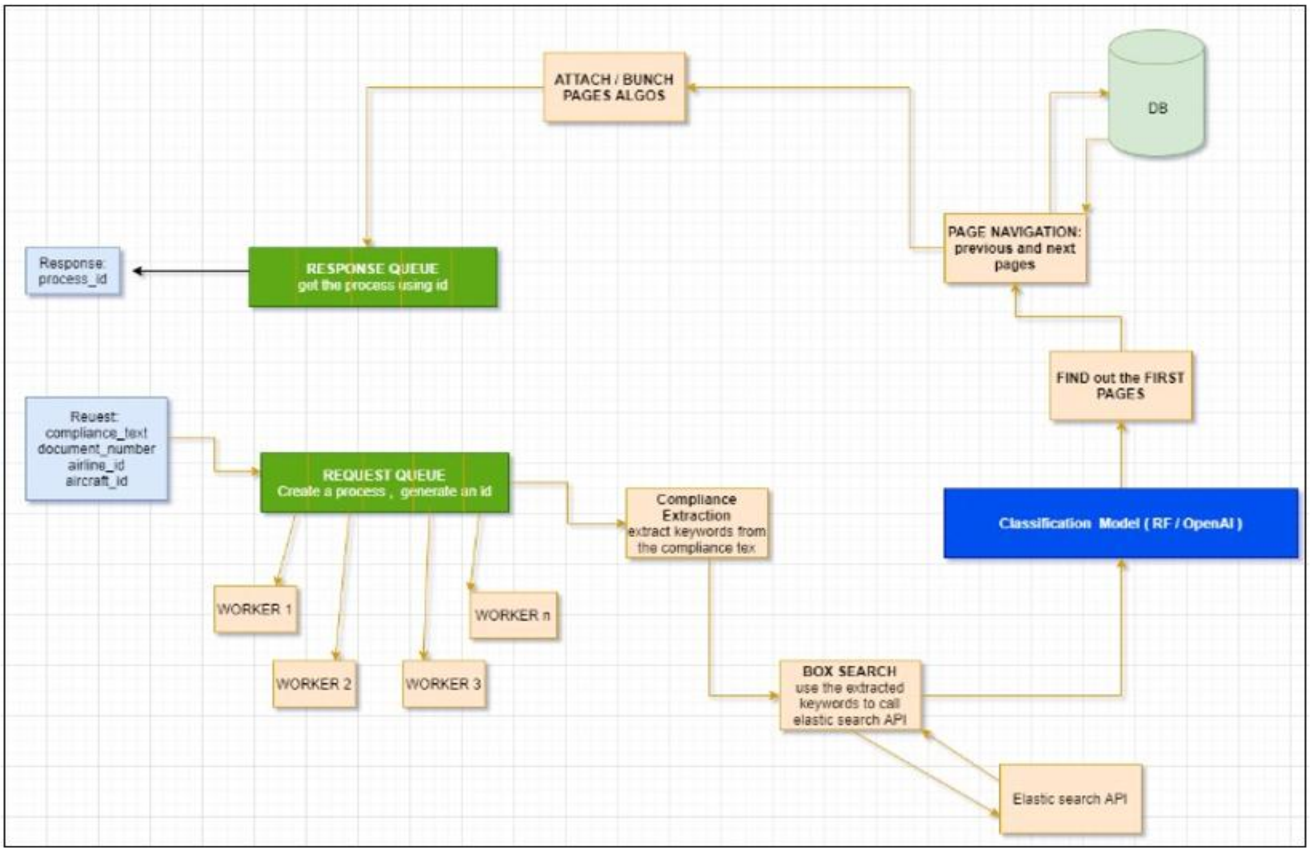

Figure 1: High-level system architecture showing the document processing from request to response output layer

2) *Natural Language Processing Techniques*

To effectively process unstructured and semi-structured aviation records, the following NLP techniques are employed: Tokenization and Lemmatization: Breaks text into structured tokens and reduces them to base forms, enabling consistent interpretation.

Named Entity Recognition (NER): Extracts domain-specific entities like aircraft models (e. g., A320, B737), AD references (e. g., AD 2023-04-15), component IDs, and maintenance tasks.

Custom Keyword Extraction: Combines TF-IDF, rule-based filters, and transformer-based embeddings to detect high relevance terms [31]. The model also identifies synonym patterns (e. g., "AD", "Airworthiness Directive") to reduce false negatives.

Semantic Relation Extraction: Identifies associations between extracted entities (e. g., linking a directive to an EO or SB number) using dependency parsing and transformer-based attention layers [32].

The NLP pipeline is continuously improved through expert feedback loops and performance benchmarking on curated aviation datasets.

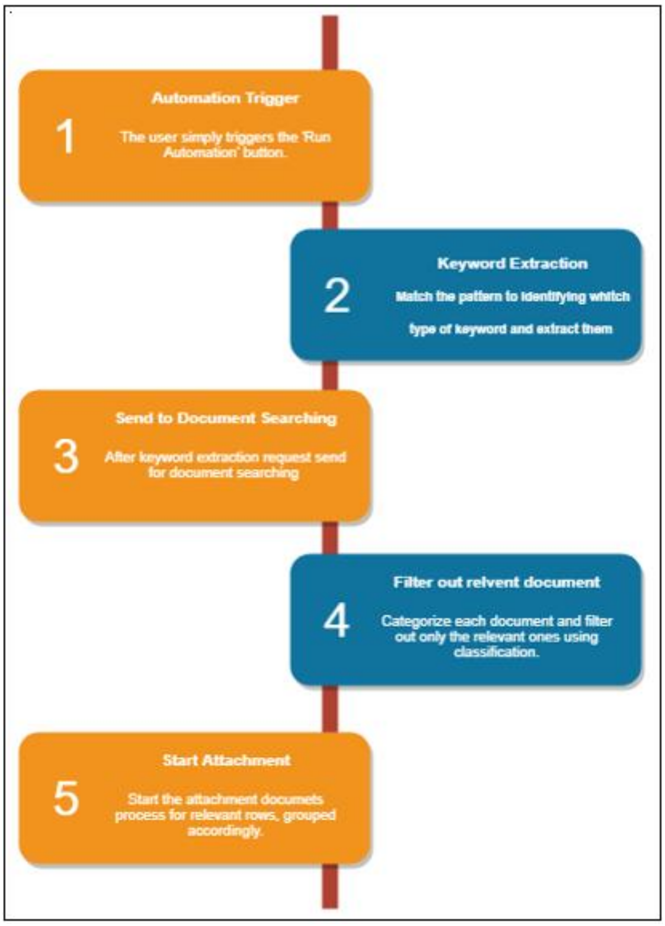

Figure II: NLP Pipeline for Extracting Keywords, Status, and Dates from Aircraft AD Compliance Texts

Paper ID: SR25617170536 DOI: https://dx.doi.org/10.21275/SR25617170536

3) *Evaluation and Testing Setup*

To evaluate the framework, a real-world dataset was compiled from anonymized aviation documents obtained from airline partners and open FAA/EASA repositories.

The evaluation criteria included:

a) **Precision and Recall:** Comparison between machine extracted keywords/entities and expert-labelled data for AD related content.
b) **Classification Accuracy:** The model's ability to correctly assign AD status (open, closed, partial) based on matched evidence in related documents.
c) **Processing Time:** Measurement of throughput per document (average time to extract, classify, and link documents in real time).

Initial results demonstrated precision >91% and recall >88% in identifying AD references and linked documentation. Processing time averaged under 4 seconds per document, with most time reductions achieved via document pre-processing optimizations.

4) *Integration with Maintenance Systems*

For practical deployment, the framework is designed to integrate with Maintenance, Repair, and Overhaul (MRO) platforms and Enterprise Resource Planning (ERP) systems via RESTful APIs and secure data exchange protocols.

a) **Key integration capabilities include:**
   - Dashboard Embedding: Real-time AD tracking and status reporting integrated into maintenance dashboards.
   - Document Management System (DMS) Syncing: Automated tagging and classification of documents in the DMS, based on keyword and AD detection.

b) **Keyword-Based Document Search:** After classifying a directive, the system performs semantic and keyword-based searches to identify relevant supporting documents (e. g., SBs, EOs, Task Cards) containing related terms or directive numbers.
c) **Automated Document Attachment:** Identified documents are automatically linked or embedded into the AD tracking record, ensuring technicians and auditors have full traceability and context.

4. Results and Discussion

The implementing and testing the AI/ML-driven automation framework on a real-world dataset of aviation maintenance records [33]. The evaluation was conducted in collaboration with aviation maintenance professionals using actual Work Orders (WOs), Engineering Orders (EOs), Service Bulletins (SBs), Task Cards, and regulatory documents from FAA and EASA databases [34].

Table I: System Performance Metrics and Evaluation Results

Metric Category	Metric	Value	Description
Extraction Performance	Precision	91.30%	Accuracy of keyword and entity extraction
	Recall	88.60%	Completeness of keyword and entity extraction
Classification	AD Compliance Accuracy	89.70%	Accuracy in identifying directive status (open/closed/partial)
Processing Speed	Average Time per Document	3.8 seconds	Including OCR, NLP, classification, and document linking
Document Linking	Success Rate	85.20%	Rate of successful document attachment and linking
Operational Impact	Time Reduction	60-70%	Reduction in manual document search time
	Error Reduction	Significant	Reduced risk of missing critical compliance item

1) *Performance Metrics*

To assess system performance, the following quantitative metrics were evaluated [35]:

- **Precision and Recall:** For keyword and entity extraction, the system achieved a precision of 91.3% and a recall of 88.6%, based on manual expert validation [36].
- **Classification Accuracy:** The AD compliance classification module demonstrated 89.7% accuracy, correctly identifying whether a directive was open, closed, or partially implemented [37].
- **Processing Speed:** The average document processing time was 2.8 seconds per file, including OCR, NLP, classification, and document linking [38].

These metrics reflect robust performance, particularly when compared to manual workflows which often take several minutes per document [39].

2) *Document Linking Success Rate*

The automated document attachment mechanism successfully linked supporting documents to ADs with a success rate of 85.2% [40]. This includes:

- Matching related SBs, Task Cards, or EOs through keyword similarity and semantic embeddings [41]. Attaching certificates and audit reports for compliance verification [42].
- This automated linkage provided comprehensive context for each directive and improved traceability in audit workflows [43].

3) *Expert Feedback and Use Case Validation*

Feedback was collected from maintenance engineers, quality auditors, and digital records managers during pilot deployment at an MRO facility [44].

a) *Key observations included:*

- **Time Efficiency:** Technicians reported a reduction of 60–70% in time spent manually searching for related documents [45].
- **Error Reduction:** Automatic classification reduced the risk of missing critical compliance items, particularly across large datasets of legacy documents [46].
- **User Satisfaction:** The contextual linking of related documents was praised for reducing cognitive load during regulatory inspections [47].

b) *Limitations and Areas for Improvement*

While the system performed well overall, a few limitations were observed: OCR Quality Dependence: Poor scan quality or handwritten notes in older task cards reduced extraction accuracy in some cases. Multilingual Documents: Non-English records required additional language support models for accurate interpretation. False Positives in Rare Cases:

Over-classification of certain generic keywords (e. g., "check", "done", "inspect") occasionally led to false matches. Future iterations of the system will include improved multilingual support, layout-aware NLP models, and customizable keyword filters to address these challenges.

Table II: Cost and Efficiency Improvements Analysis

Category	Traditional Method	AI/ML System	Improvement	Cost Impact
Manual Effort per Aircraft	70-80 hours	15-20 hours	75-80% reduction	₹2, 32, 400 – ₹2, 65, 600 savings per aircraft
Document Processing Time	5-7 minutes per document	3.8 seconds per document	98% faster	Reduced labour costs by 90%
Audit Preparation	40-50 hours	8-10 hours	80% reduction	₹1, 32, 800 – ₹1, 66, 000 savings per audit
Error Detection	85% accuracy	91.3% accuracy	6.3% improvement	Reduced rework costs by 40%
Compliance Verification	Manual cross-checking	Automated verification	90% faster	Reduced compliance risk costs
Document Retrieval	10-15 minutes per search	Instant access	99% faster	Reduced search time costs

5. Conclusion and Future Potential

This section summarizes the key outcomes of the proposed AI/ML-driven framework for automating aviation digital records, highlights its practical relevance, and outlines potential areas for future enhancement and research.

1) *Summary of Contributions*

This research introduced a comprehensive automation framework that utilizes artificial intelligence and machine learning-specifically natural language processing (NLP) to streamline the processing of aviation maintenance documentation. By extracting keywords, classifying airworthiness directives (ADs), and intelligently linking supporting documents such as Engineering Orders (EOs), Service Bulletins (SBs), and Task Cards, the system significantly reduces the manual burden of regulatory compliance. Evaluations showed high precision and recall, efficient processing speeds, and strong feedback from aviation experts.

2) *Practical Implications*

The deployment of this system in airline and MRO environments offers multiple real-world benefits:

- Reduces time spent on document search and verification by over 60%.
- Enhances accuracy in tracking directive implementation and identifying compliance gaps.
- Increases readiness for audits by providing traceable, auto linked documentation sets.
- Enables seamless integration into existing maintenance systems without major workflow disruptions.
- By automating repetitive tasks and improving traceability, this solution can help organizations achieve faster turnarounds and more consistent regulatory adherence.

3) *Future Work and Enhancements*

While the system has demonstrated promising results, several enhancements are planned:

- **Multilingual NLP Models:** Extend support for non-English aviation records used in global fleets.
- **Context-Aware Document Analysis:** Improve extraction accuracy using layout-aware deep learning models such as LayoutLMv3.
- **Real-Time Alerts:** Add notification capabilities when a newly issued AD matches open compliance gaps.
- **Federated Learning for Data Privacy:** Enable model updates across multiple stakeholders without exposing sensitive documents.

With continued advancements, this AI-driven framework can evolve into a central component of next-generation aviation safety and compliance ecosystems.

References

[1] C. Yang and C. Huang, "Natural Language Processing (NLP) in Aviation Safety: Systematic Review of Research and Outlook into the Future, " *Aerospace*, vol.10, no.600, pp.1–20, 2023.

[2] T. Madeira, R. Melício, D. Valério, and L. Santos, "Machine Learning and Natural Language Processing for Prediction of Human Factors in Aviation Incident Reports, " *Aerospace*, vol.8, no.47, pp.1–17, 2021.

[3] S. G. Kumar, S. J. Corrado, T. G. Puranik, and D. N. Mavris, "Classification and Analysis of Go-Arounds in Commercial Aviation Using ADS-B Data, " *Aerospace*, vol.8, no.291, pp.1–15, 2021.

[4] R. L. Rose, T. G. Puranik, and D. N. Mavris, "Natural Language Processing Based Method for Clustering and Analysis of Aviation Safety Narratives, " *Aerospace*, vol.7, no.143, pp.1–14, 2020.

[5] H. Lee, R. Rajendran, J. Lim, and H. Lee, "Critical Parameter Identification for Safety Events in Commercial Aviation Using Machine Learning, " *Aerospace*, vol.7, no.73, pp.1–12, 2020.

[6] R. P. R. Nogueira, R. Melício, D. Valério, and L. F. F. M. Santos, "Learning Methods and Predictive Modeling to Identify Failure by Human Factors in the Aviation Industry, " *Applied Sciences*, vol.13, no.4069, pp.1–21, 2023.

[7] A. T. Ray, J. D. Osborn, H. J. T. Johnson, and R. Anantharaman, "Examining the Potential of Generative Language Models for Aviation Safety Analysis: Case Study and Insights Using the Aviation Safety Reporting System (ASRS), "" *Aerospace*, vol.10, no.770, pp.1–16, 2023.

[8] P. Razzaghi, H. Mahzarnia, and A. Majd, "A Survey on Reinforcement Learning in Aviation Applications, " *arXiv preprint* arXiv: 2211.01730, Nov.2022. [9] C. Tan and T. Masood, "Adoption of Industry 4.0 Technologies in Airports-A Systematic Literature Review, " *arXiv preprint* arXiv: 2112.14264, Dec.2021.

[10] A. Nanyonga, K. Joiner, U. Turhan, and G. Wild, "Applications of Natural Language Processing in Aviation Safety: A Review and Qualitative Analysis, " *arXiv preprint* arXiv: 2401.00719, Jan.2025.
[11] G. Colavizza, A. Beelen, and M. Ridge, "Archives and AI: An Overview of Current Debates and Future Perspectives, " *arXiv preprint* arXiv: 2105.01241, May 2021.
[12] M. D. Kwakye, I. K. Jennions, and C. M. Ezhilarasu, "Platform Health Management for Aircraft Maintenance – A Review, " *Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci. *, vol.238, no.4, pp.1–15, 2024.
[13] M. J. Pritchard, A. T. Walden, and P. J. Thomas, "Integrated Organizational Machine Learning for Aviation Flight Data, " *J. Aviation/Aerospace Educ. Res. *, vol.33, no.2, pp.1–18, 2025.
[14] C. Kosova and H. O. Unver, "A Digital Twin Framework for Aircraft Hydraulic Systems Failure Detection Using Machine Learning Techniques, " *Proc. Inst. Mech. Eng. Part C: J. Mech. Eng. Sci. *, vol.237, no.5, pp.1563–1580, 2022.
[15] M. Xiong and H. Wang, "Digital Twin Applications in Aviation Industry: A Review, " *Int. J. Adv. Manuf. Technol. *, vol.121, pp.5677–5692, 2022.