Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM)
Vol. 8, No. 3, November 2025, pp. 622 – 634
p-ISSN: 2614-3372 | e-ISSN: 2614-6150

AI-Based Prediction of Fatalities in Flight Accidents: Insights from 75 Years of Aviation Accident Records

[1*]Muhammad Ridho Ramadhan, [2]Avelia Fairuz Faadhilah, [3]Roniyansyah, [4]Elmo Juanara
[1]Department of Electrical and Information Engineering, Universitas Gadjah Mada, Indonesia
[2]Radio Telecomm. and Microwave Lab, School of Electrical Eng. & Informatics, Institut Teknologi Bandung, Indonesia
[3]Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Lambung Mangkurat, Indonesia
[4]School of Knowledge Science, Japan Advanced Institute of Science and Technology, Japan
Email: [1]mridhoramadhan@mail.ugm.ac.id, [2]23222326@std.stei.itb.ac.id,
[3]1911012110014@mhs.ulm.ac.id, [4]elmo.juanara@jaist.ac.jp

Article Info

Article history:
Received Aug 06th, 2025
Revised Sep 15th, 2025
Accepted Oct 19th, 2025

Keyword:
Aircraft Accidents
Aviation Safety
Exploratory Data Analysis
Fatality Prediction
Machine Learning

ABSTRACT

Aviation accidents have plateaued in terms of safety improvements since the late 1990s, underscoring the need for advanced analytical approaches. This study employs a data-driven framework utilizing Artificial Intelligence (AI) on a comprehensive dataset spanning 75 years of global aviation accidents. This enables the identification of long-term safety patterns that are often overlooked in studies limited to specific regions or flight phases. The study aims to analyze long-term trends and predict future aviation accidents using Machine Learning (ML) classification models. This study involved web scraping the Aviation Safety Network (ASN) database to compile the dataset, followed by Exploratory Data Analysis (EDA) to obtain insights. Support Vector Machine (SVM), Random Forest (RF), and Categorical Naive Bayes were employed for fatality prediction. EDA results show that while the number of fatal accidents has declined, scheduled passenger service and the en-route flight phase show the highest proportion of occurrences. Furthermore, the maneuvering flight phase and military service have a maximum likelihood of a fatal outcome. The predictive models achieved accuracies of approximately 79-80%. The SVM model, with the highest F1-score (79.85%), proved to be the most balanced in terms of specificity for non-fatal incidents and sensitivity for fatal ones. This result provides safety practitioners with a reliable framework for evidence-based decision-making.

Corresponding Author:
Muhammad Ridho Ramadhan,
Department of Electrical and Information Engineering,
Universitas Gadjah Mada,
Jl. Grafika No. 2 Kampus UGM, Yogyakarta 55281.
Email: mridhoramadhan@mail.ugm.ac.id

DOI: http://dx.doi.org/10.24014/ijaidm.v8i3.37948

1. INTRODUCTION

Aviation safety has prioritized its continuous effort to reduce the risk of accidents. Accident rates have been consistently declining since the 1960s, and the industry has experienced substantial safety enhancements for decades. However, this positive trend has slowed, reaching a plateau in recent decades, suggesting that conventional safety enhancement methods may be approaching their limits of effectiveness [1]. This stagnation emphasizes the urgent need for any stakeholders to create and explore innovative approaches to enhance the effectiveness of existing accident prevention and mitigation strategies [2]. Although most incidents are minor, the accumulation of these events can result in severe complications if they are not properly detected and managed [3]. Artificial Intelligence (AI) and Machine Learning (ML) have emerged as a promising solution which provides the ability to process and analyze massive volumes of historical data to find patterns, trends, and non-linear correlations that traditional statistical analysis methods often miss [4]. By employing predictive models, the aviation sector can shift from a reactive to a proactive approach to addressing safety hazards [5].

Recent research has further expanded the range of algorithms considered. Zhang and Mahadevan [6] applied ensemble ML models to predict aviation incident risk and reported improved accuracy, although their work relied on limited operational data. Omrani et al. [7] compared neural networks, decision trees, and Support Vector Machine (SVM) for accident severity prediction in civil aviation, showing both the potential and the limitations of classical ML approaches. A study published by Bilgic et al. [8] explored several algorithms for predicting injuries and fatalities in aviation accidents and confirmed the need for models that can reliably separate fatal from non-fatal outcomes. Xia et al. [9] used ADS-B data to detect anomalies in commercial aviation and predict accident precursors, highlighting the importance of real-time monitoring. Demir et al. [5] presented a systematic review that underscored the growing role of AI in aviation safety but noted that most studies still rely on regional or short. However, these studies frequently face limitations due to their shorter time frames, inability to cover all flight phases, or lack of access to comprehensive global accident datasets. These limitations highlight the need for a more comprehensive analysis of accident risk.

To address this gap, this study suggests an AI-based approach to analyze 75 years of global aviation accident records. The main objective is to create a reliable model for predicting fatalities, which are defined as deaths that happen within 30 days of an incident, according to ICAO standards [10]. This study focused on the key features, such as the type of aircraft, the type of service, and the phase of flight. Empirical studies have demonstrated that the type and age of an aircraft are important factors that affect risk profiles and accident patterns [11]. Takeoff and landing are also statistically the most hazardous phases for accidents [12].

Thus, this study has two main goals. The first objective is to conduct a comprehensive Exploratory Data Analysis (EDA) on 75 years of global aviation accident records to identify significant patterns and high-risk factors that contribute to fatalities. Secondly, we aim to develop and evaluate an AI-based predictive model that may employ ML algorithms trained on 75 years of global aviation accident records to classify the results of aviation accidents into two categories, which are fatal and non-fatal. A similar approach was also applied in a study by Juanara and Lam [13] for the classification of early warning levels in disaster scenarios. This study makes two main contributions. First, it provides a validated predictive framework that demonstrates a balanced approach to predict fatalities. Second, it presents crucial data-driven insights derived from the historical analysis, which indicate specific high-risk scenarios such as very high fatality rates in military operations and during the maneuvering phase of flight. The contribution of this study lies in its global and long-term scope, which bridges gaps left by prior works limited in geography, timeframe, or operational phases. Beyond methodological contributions, the findings offer practical value by equipping aviation regulators and operators with a reliable predictive framework to strengthen proactive safety management strategies. Simultaneously, these contributions provide a foundation for developing more effective and targeted safety strategies.

This paper is structured as follows: Section 2 outlines the research methodology, Section 3 presents the results of the experiments and EDA, then discusses the analysis of the findings, and Section 4 concludes the study and suggests directions for future research.

2. RESEARCH METHOD

This study will discuss the use of EDA as a method to explore the aviation accident dataset from 75 years of records and create a model to predict future aviation accidents using several types of ML algorithms.

2.1. Data Collection

The dataset is obtained from the Aviation Safety Network (ASN), a worldwide repository of information that provides up-to-date, complete, and reliable authoritative information on airline accidents and safety issues [14]. The data used in this paper corresponds to all the 18,652 accidents and incidents reported during a 75-year period spanning from January 1951 to December 2024. We distribute all the data into three categories: aircraft manufacturers, flight phase, and aircraft type of service. This data is obtained by using a web scraping method, which is a technique used to automatically extract data from websites. Web scraping is systematically implemented based on the flowchart, as shown in Figure 1.

The implementation of the web scraping method in this study is slightly similar to [15], which relied on the BeautifulSoup library to parse and extract relevant data from the HyperText Markup Language (HTML) content of the ASN website. The program begins by defining the range of years from which accident data will be collected, which spans from 1951 to 2025. The next stage is sending HyperText Transfer Protocol (HTTP) requests to the ASN website. The scraper accesses the ASN database by sending HTTP requests to specific year-based Uniform Resource Locators (URL). After sending the request, the scraper verifies the HTTP response status code. The next stage is identifying the number of pages for each year. Several years may accommodate multiple pages of accident records. To ensure all available accident data is retrieved, the scraper first detects pagination links on the ASN website. These links allow users to navigate through additional pages containing accident reports for a given year. The next step is to get the data on aviation accidents. The scraper goes through each page iteratively to obtain any details from the available accident records table. The scraper uses BeautifulSoup to locate and retrieve information from HTML tables representing individual accidents. The retrieved data contains the accident date, aircraft type, location, number of fatalities, and a direct link to the detail accident reports. After acquiring the URL to the detailed accident page, the scraper extracts further information, such as the flight phase and type of aircraft service. The collected data is then compiled into a Comma-Separated Values (CSV) file after it has been successfully extracted. Once all available pages for a given year have been processed, the scraper moves to the next year and repeats the entire process. The scraper continues this iterative approach until it successfully collects data from 1951 to 2025.

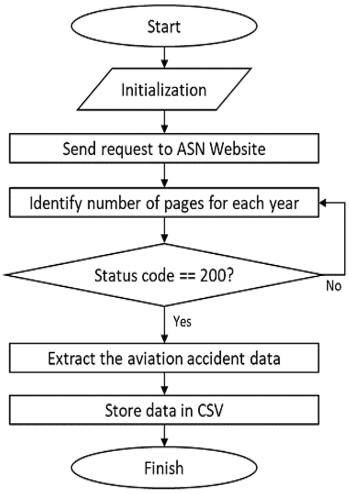


Figure 1. Flowchart of Web Scraping Method

2.2. Data Pre-processing

The data processing phase is essential to prepare the dataset for accurate and reliable analysis. This phase involves five key steps: flight data selection, data cleaning, feature selection, dataset standardization, and handling the imbalanced data.

2.2.1. Flight Data Selection

In this step, the dataset is filtered to retain only relevant and meaningful features necessary for analysis. This process helps eliminate unnecessary data to ensure that only useful attributes are considered for further processing.

2.2.2. Feature Extraction

To enhance the analytical value of the dataset, additional features are derived from existing columns. Specifically, the Manufacturer feature is extracted from the Aircraft column to categorize accident data based on aircraft manufacturers. Then, the year feature is extracted from the Date column to facilitate time-based trend analysis.

2.2.3. Data Cleaning

Data quality and integrity are ensured by handling missing or inconsistent values. Based on the data obtained from web scraping, the Type of Service column contains several records with unknown values. Since these records lack crucial information, they are removed from the dataset to maintain data reliability.

2.2.4. Data Standardization

After filtering and cleaning the data, standardization techniques are applied to ensure consistency across all attributes. One crucial aspect of this process is normalizing date formats, which ensures uniformity in time-based data, making it easier to analyze trends over different periods. Additionally, the number of Fatalities column is standardized by categorizing it into two distinct classes: Fatal and Non-Fatal. The fatal class includes accidents that resulted in at least one fatality, while the non-fatal class includes incidents that did not result in any fatalities. This classification provides the analysis easier by separating severe and non-severe aviation accidents, which allows it to be easier to recognize patterns and evaluate trends.

2.2.5. Handling Imbalanced Data

After completing data cleaning and standardization, class imbalance was addressed using the Synthetic Minority Oversampling Technique (SMOTE). This approach was applied before model training to ensure that the minority class (fatal accidents) was adequately represented, thereby enhancing the models' ability to generalize.

2.3. Exploratory Data Analysis

EDA is a procedure for analyzing data, a way to make interpretations or interpretations, a plan for obtaining data with the aim of facilitating analysis [16]. In this study, we conducted EDA to gain insights into the aviation accident dataset in the last 75 years period. In this study, this analysis consists of data visualization providing a clear and intuitive way to interpret the data. This stage is carried out to find the patterns and trends in aviation accident fatalities by examining key categorical variables.

2.3.1. Proportional and Temporal Analysis of Accident Fatalities

There were two main steps in the initial analysis of accident outcomes. In order to provide an overview of the severity of accidents are in general, we observe the overall distribution of accident severity by categorizing all incidents in the 75-year dataset into two main types: Fatal (where there was one or more fatalities) and Non-Fatal (where there were no fatalities). Next, we did a temporal trend analysis to examine how aviation safety has changed over time. At this stage, we collected the annual data on fatal accidents from 1950 to 2025. The time-series data was then plotted to demonstrate the long-term trends, determine when high volatility data occurs, and evaluate how safety performance has changed over the years.

2.3.2. Analysis of Fatalities by Aircraft Manufacturer

The analysis of fatalities by aircraft manufacturer comprised two steps. The accident data was organized by manufacturer to determine the number of fatal accidents attributed to each. This enabled the identification of manufacturers with the highest number of incidents overall. A proportional analysis was conducted to evaluate the overall inherent safety performance. This analysis then needs the calculation of the proportion of non-fatal accidents according to the total number of incidents for each manufacturer. This normalization provided a more precise understanding of risk which might take into account the different operational scales of various producers. This analysis supports safety audits and risk assessments by highlighting manufacturers whose designs warrant further attention.

2.3.3. Analysis of Fatalities by Phase of Flight

The analysis of fatalities by phase of flight was conducted to pinpoint operational vulnerabilities. Initially, accident records were categorized by flight phase and the absolute number of fatal accidents was aggregated for each. To understand the intrinsic risk of each phase independent of its duration or frequency, a proportional analysis was also performed. The ratio of fatal to non-fatal occurrences was determined for each phase. This important stage provided the differentiation between phases characterized by high incident frequency and those with the greatest likelihood of a fatal outcome per incident. These insights are crucial for flight safety programs, pilot training priorities, and regulatory oversight due to the highlighted areas where interventions could yield significant benefits.

2.3.4. Analysis of Fatalities by Type of Service

The analysis of fatalities categorized by the type of service was designed to emphasize risk profiles specific to each purpose. Each record of a fatal accident was initially classified according to its operational service, and the data was compiled to figure out the total number of fatal incidents for each category. To provide a more insightful risk assessment, a proportional analysis was then conducted by calculating the percentage of fatal accidents within each service category. This normalization was critical to identify which operational contexts, such as military flights, carry the highest intrinsic risk of a fatal outcome, independent of their overall frequency of operations. The objective was to identify the operating contexts that are most and least susceptible to fatal accidents and to understand the correlation between the type of service and fatality rates could help in addressing the systemic weaknesses within the particular operational categories.

2.4. ML Algorithms

This study investigates several widely used ML algorithms to identify hidden patterns in data and analyze the impact of input variables on outcomes. Three different algorithms used in this study are SVM, Random Forest (RF), and Categorical Naive Bayes to predict future aviation accidents based on the given dataset. The selection of ML algorithms in this study was based on both theoretical considerations and their complementary strengths. SVM was chosen for its effectiveness in producing optimal separating hyperplanes in high-dimensional feature spaces, making it suitable for complex classification tasks with overlapping classes. RF was included for its robustness in capturing nonlinear relationships among features and its ability to provide interpretability through feature importance scores. Finally, Categorical Naïve Bayes was employed due to its probabilistic simplicity and efficiency in handling categorical predictors, serving as a strong baseline for comparison. This diverse combination of algorithms enables a comprehensive evaluation of predictive performance from linear margin-based, ensemble non-linear, and probabilistic perspectives. Each algorithm employs the distinct approach based on different assumptions and statistical techniques, which results to different strengths and weaknesses. Consequently, it is crucial to implement numerous algorithms on a specific issue in order to identify the optimal solution. In the end, the models were evaluated using accuracy, F1-Score, sensitivity and specificity [17].

2.4.1. Support Vector Machine

SVM which was introduced by Vapnik, is a supervised learning method widely employed for many cases of classification and regression tasks. This algorithm aims to identify a single hyperplane that can maximize the margin for linear separation of the classes. This method demonstrates particular effectiveness in scenarios with limited training data, where traditional statistical methods dependent on large datasets may fail to ensure an optimal solution. The SVM algorithm can make a more stable model by maximizing the distance between classes with the hyperplane and has a good generalization ability for any new data testing. The data points that most determine the position of this hyperplane are known as support vectors, which play an important role in forming the boundary separating the classes in a high-dimensional feature space [18]. To determine the optimal hyperplane that separates data into two or more categories, support vector machine identifies the decision boundary that maximizes the margin between classes. The decision function of SVM can be represented by equation (1).

$$f(x) = w^T \Phi(x) + b \tag{1}$$

where w is the normal vector to the hyperplane, b denotes the bias term while $\Phi(x)$ represents a nonlinear mapping of the input vector x into a higher-dimensional feature space [19] .

2.4.2. Random Forest

The random forest algorithm is a machine learning model that is built from a tree-based block arrangement. Tree-based models process the dataset recursively and divide it by certain criteria until a stopping condition is met. At the bottom of the decision tree is something called a leaf node. The variation of the partition criteria and the stopping conditions set becomes the rules in designing decision trees for classification tasks (categorical outcomes, for example, logistic regression) and regression tasks (continuous outcomes) [20]. One of the important features of random forests is the ability to determine the variables that influence the prediction. This model can accept raw data and also model nonlinear relationships and accept regression and classification problems at the same time [21]. As machine learning technology advances, random forests continue to be used as a comparison to newer techniques in classification and regression.

The Classification and Regression Tree (CART) algorithm in Random Forest will produce tree nodes that can represent binary decision rules. This division is done on features that maximize information gain (IG) recursively on the data. IG is defined in equation (2).

$$IG\left(D_p, f\right) = I\left(D_p\right) - \sum_{j=1}^{m} \left(\frac{N_j}{N_p} I\left(D_j\right)\right) \tag{2}$$

where D_p and D_j are the data sets at the parent and the j-th child node, f is the variable to perform the split of the feature space, j is the index of the child node ranging from 1 to m and m indicates the total number of child nodes produced from the split. I is Gini impurity function whereas N_p and N_j are the number of samples at the parent and child nodes. The Gini impurity reflects the probability of misclassifying an observation. Formally, it can be described in the equation (3).

$$GINI\ (D) = 1 - \sum_{i=1}^{k} p_i^2 \tag{3}$$

where D is the dataset containing samples of k classes and p_i is the proportion of the samples that belong to class i for a particular node [22].

2.4.3. Categorical Naïve Bayes

There are three Naive Bayes classification methods, namely Bernoulli Naive Bayes, Gaussian Naive Bayes and Categorical Naive Bayes. Bernoulli Naive Bayes classifies data distributed according to the Bernoulli distribution, namely binary values. As for Gaussian Naive Bayes, it is the most commonly used classifier method and its decision function is derived from the Gaussian distribution. Meanwhile, in this study, the third method is used, namely Categorical Naive Bayes, which is specifically intended for data that is categorically not trusted [23]. To avoid the problem of zero frequency, the smoothing parameter (α > 0) is used in the decision function in equation (4).

$$P(v_i = t|y = c, \alpha) = \frac{N_{tc}+\alpha}{N_c+\alpha n_i} \quad (4)$$

Where v_i is the i^{th} feature, t is a specific category of feature v_i, N_{tc} is the number of times the category t appears in the sample from class c, N_c is the total number of samples in class c, and n_i is the number of categories that exist in the i^{th} feature.

2.5. Model Evaluation

The performance model can be assessed through insight into the correctness of several categorization model components through the usage of a confusion matrix. The confusion matrix for the two-path clustering is shown in Table 1. True positive (TP) denotes that the system accurately identified a positive prediction, which is a fatal class, while True Negative (TN) means that the system successfully identified a negative prediction, which is a non-fatal class. False Positive (FP) describes how the system incorrectly classifies a negative prediction as positive, while False Negative (FN) describes how the system incorrectly labels a positive prediction as negative. Accuracy refers to how closely the value predicted by the system matches the actual value, recall assesses the ability of the model to identify all positive cases in the dataset, precision provides with information about level of certainty within the model's prediction of the information being positive, f1-score is a metric that compares the average level of recall and precision by setting the highest value to 1 and the lowest to 0 [24].

Table 1. Confusion Matrix

Actual Value	Predicted Value	
	Fatal	Non-fatal
Fatal	TP	FP
Non-Fatal	FP	TN

3. RESULTS AND ANALYSIS

This section will delve into a detailed description of the results obtained from the EDA and the classification using three ML models. In addition, we will present the evaluation results obtained from comparing the actual data labelling with the predictions made by the model. Furthermore, we also describe the characteristics of the data by carrying out EDA. All the model training and data analysis of experiments were carried out using Google Colab.

3.1. Exploratory Data Analysis

The analysis begins by examining the overall distribution and temporal trends of accidents. Then, this is followed by a detailed breakdown of accident fatalities based on three key categorical variables: aircraft manufacturer, type of service, and phase of flight, to identify specific high-risk scenarios.

3.1.1. Proportional Analysis of Accident Fatalities

The distribution of accident fatalities throughout 75 years of aviation history is shown in Figure 2. Accidents are classified into two categories: fatal and non-fatal incidents. The total fatalities amount to 6,237, whereas non-fatal occurrences total 10,996 of all flight accidents. This visualization illustrates the notable distinction between fatal and non-fatal occurrences, which underscores the severity and prevalence of each category. The data demonstrates that non-fatal occurrences are approximately twice as frequent as the fatal ones. The significant number of fatal accidents highlights the importance of improving the safety protocols and technologies to further decrease mortality rates.

3.1.2. Temporal Trends of Fatal Accidents (1950-2025)

The historical trend of fatal aviation accidents from 1950 through the early 2025 is illustrated in Figure 3. This decline in number has occurred despite periods of any significant volatility, particularly in the early decades of the jet age [1]. From the 1950s to the late 1970s, the number of fatal accidents increased each year, with the most happening in 1972 with 158 recorded incidents. The rapid growth of commercial aviation and the introduction of new technologies during that time are likely what caused this instability. Thereafter, the trend gradually declined and became more consistent in the 21st century, with annual cases generally <100 after 2000. This decline has been attributed to advances in aircraft technology, such as flight management systems (FMS) and fly-by-wire [25]. The data representing the year of 2025 only cover the first two months (through February), and therefore cannot be used to estimate annual totals.

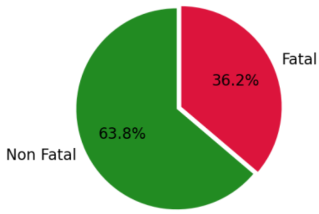


Figure 2. Distribution of Aviation Accidents By Fatality Type (1950-2025)

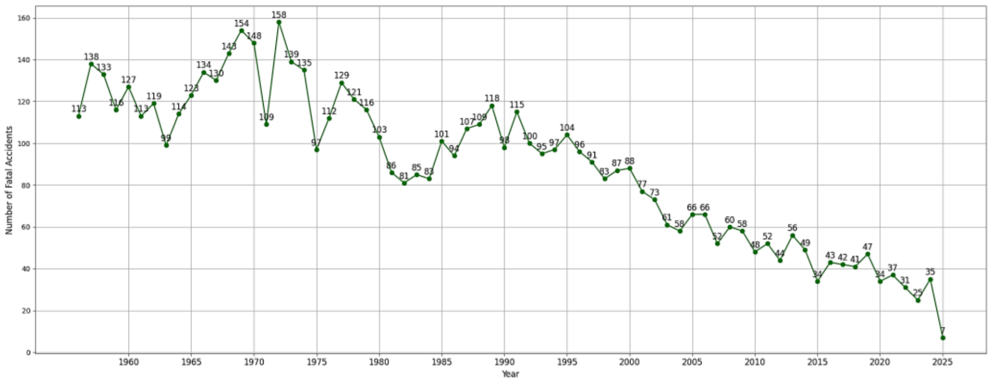
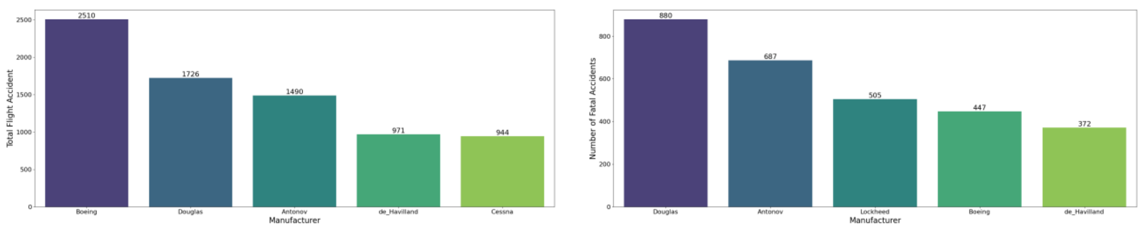


Figure 3. Annual Number of Fatal Aviation Accidents (1950-2025)

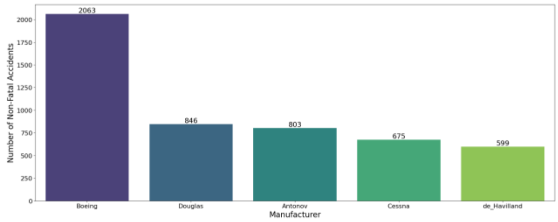
3.1.3. Analysis of Fatalities by Aircraft Manufacturer

Further analysis of the 75-year aviation accident dataset examines the distribution of aviation occurrences among aircraft manufacturers which is a crucial factor for advanced risk assessment and predictive modelling. Figure 4a illustrates the total number of recorded aviation occurrences by manufacturers, including fatal and non-fatal accidents. It can be seen that Boeing has the highest number of aviation accidents in the latest 75 years period. In Figure 4b, legacy manufacturers account for the largest number of fatal accidents reflect the historical context related to the limitations of past technology and safety systems [26] and extensive operational exposure, rather than current safety standards. Figure 4c demonstrated all manufacturers with the number of non-fatal occurrences. The graph implies that the number of non-fatal occurrences is proportional with the number of total activities, making Boeing and Douglas becomes the first and the second highest in number of non-fatal occurrences.

To obtain a more normalized measure of safety performance, Figure 5 presents the percentage of non-fatal occurrences relative to total accident history from each manufacturer. This proportional view yields a crucial insight where a large group of manufacturers exhibit a 100% non-fatal accident rate which indicates zero fatal incidents in the dataset used in this study. However, this analysis should be interpreted with statistical caution due to the denominator effect. This means that a perfect rate for a manufacturer with very few total incidents (like Aerospace and Antonon, only had one flight during the period) is less statistically significant than a near-perfect rate for a major manufacturer with thousands of incidents. This difference is very important for getting a more comprehensive view of safety performance at different levels of operation.



(a) (b)



(c)

Figure 4. Top Five Manufacturers Ranked by Different Category
(a) By The Total Number of Flight Accidents, (b) By The Total Number of Fatal Accidents,
(c) By The Total Number of Non-Fatal Occurrences

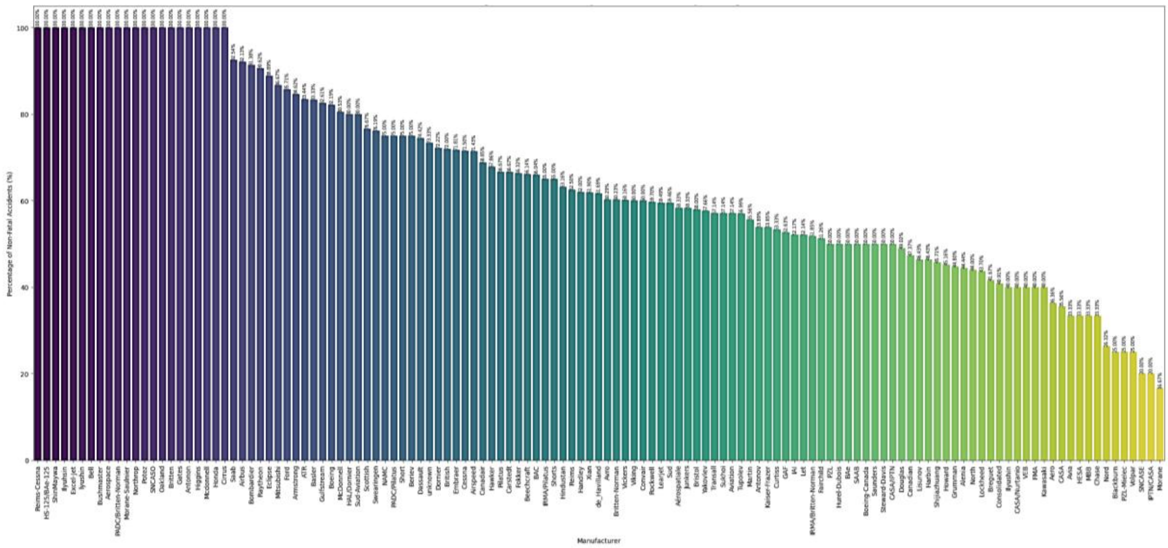
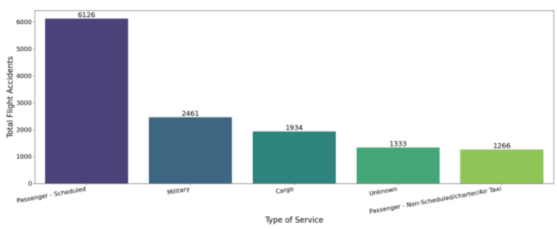


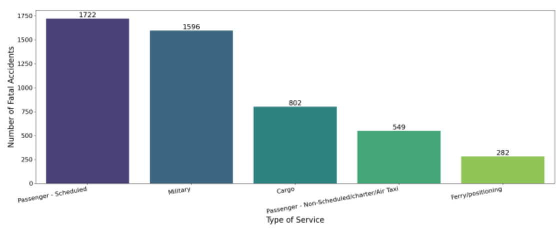
Figure 5. Percentage of Non-Fatal Occurrences by Manufacturers

3.1.4. Analysis of Fatalities by Type of Service

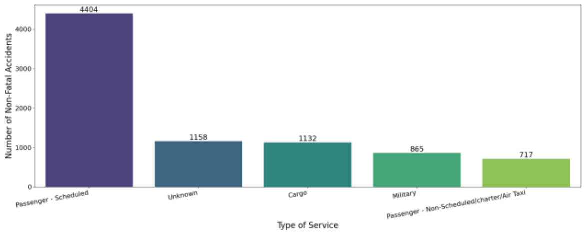
The analysis was expanded to include breaking down accident data by the operational of flights based on the Type of Service, which is a key factor that provides a lot of information about the risks involved in each mission. Figure 6a presents the baseline distribution of the total number of aviation occurrences, segmented by the type of service. Scheduled passenger aviation also recorded the highest number of fatal (1,722) and non-fatal (4,404) accidents (Figures 6b and 6c), in line with its large operational exposure, as this service accounts for the majority of global flight operations [27]. Military aviation had the second highest number of fatalities (1,596), but it had fewer non-fatal incidents (1,132), indicating differences in risk and severity compared to civil aviation. Proportional analysis (Figure 6d) confirms this difference: military aviation had the lowest non-fatal incident rate, at only 30.15%, meaning nearly 70% of accidents were fatal. These high risks are related to the nature of military operations such as involving combat scenarios, high-risk training, extreme performance maneuvers, and flights in unpredictable environments or against conditions that inherently involve greater risks and more work than routine civil transportation [28]. In contrast, Passenger - Scheduled flights have a non-fatal incident rate of 71.89%, demonstrating that while incidents occur, the overwhelming majority do not result in fatalities, reflecting the robust safety systems and protocols governing commercial aviation.



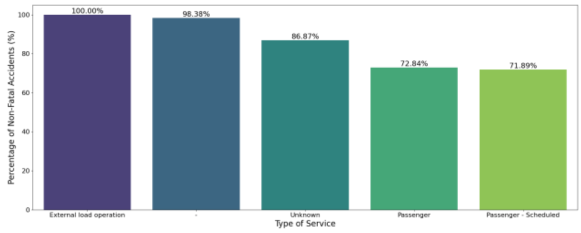
(a)



(b)



(c)

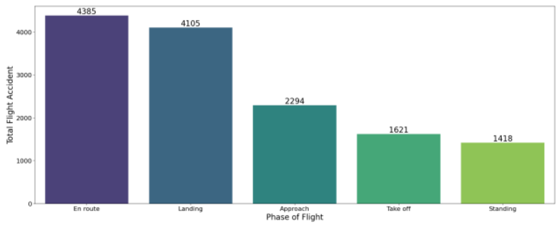


(d)

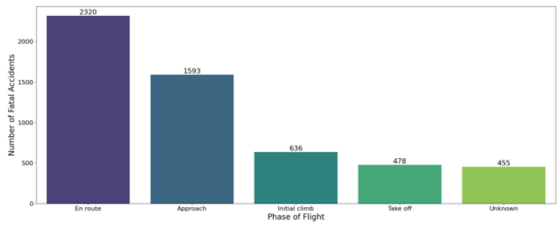
Figure 6. Top Five Type of Service Ranked By Different Categories
(a) By The Total Number of Flight Accidents, (b) By The Total Number of Fatal Accidents,
(c) By the Total Number of Non-Fatal Occurrences, (d) By Percentage of Non-Fatal Occurrences

3.1.5. Analysis of Fatalities by Phase of Flights

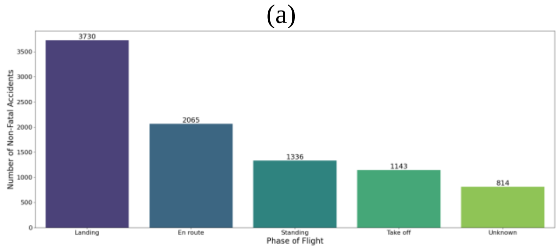
The investigation next segments accident data according to the phase of flight, a critical variable that pinpoints when an aircraft is most vulnerable. This study demonstrates a considerable difference between the phases with the highest number of fatal accidents and those with the highest proportional risk of fatality. Figure 7a provides a baseline overview of all recorded aviation occurrences, segmented by the phase of flight over the 75-year period. This combined dataset, including both fatal and non-fatal events, shows that the en route phase has the highest frequency of occurrences. As illustrated in Figure 7b, the En route phase accounts for the highest absolute number of fatal accidents, followed by the Approach phase. The high number for the En route phase is largely a function of duration-based exposure [1]. This phase constitutes the vast majority of a flight's time, naturally leading to a higher cumulative count of events over 75 years. This conclusion is further supported by Figure 7c, which shows the distribution of non-fatal occurrences. The landing and en route phases had the greatest counts, which demonstrate that phases with long durations or high frequency naturally accumulate more events of all categories. However, a proportional analysis provides a more insightful perspective of the intrinsic risks from each phase. Figure 7d depicts a proportional analysis that normalizes for exposure and displays the true, underlying risk of a fatal accident occurring throughout each period. The findings demonstrate that the maneuvering phase is the most dangerous, with a non-fatal percentage of only 25.80%, which implies a roughly 75% probability of a fatal accident. This is followed by the Approach (30.56% non-fatal) and Initial climb (35.04% non-fatal) phases. Conversely, ground operations such as Pushback/towing and Taxi are demonstrably the safest. This analysis quantifies the well-known critical phases of flight where the most safety-related improvements are concerned such as taxi, climb, approach, and landing [1] where the aircraft is in a high-energy state at low altitude with minimal margin for error. The En route phase, with a 47.09% non-fatal rate, is proportionally safer per-incident than the approach or climb phases, highlighting the importance of this normalized view.



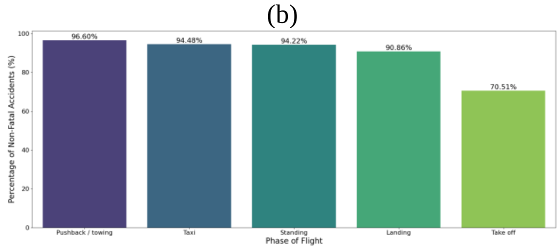
(a)



(b)



(c)

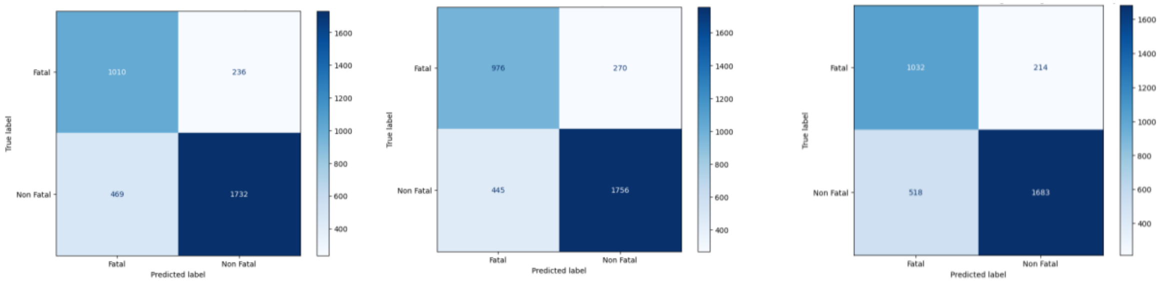


(d)

Figure 7. Top Five Phases of Flight Ranked By Different Categories
(a) By The Total Number of Flight Accidents, (b) By The Total Number of Fatal Accidents,
(c) By the Total Number of Non-Fatal Occurrences, (d) By Percentage of Non-Fatal Occurrences

3.2. Fatalities Prediction using Various ML

The final stage of this study involved the implementation and evaluation of three different ML models: SVM, RF, and Categorical Naive Bayes. The performances of each model were evaluated using several metrics which are commonly used in the classification task with key results in detail presented in Table 2 and further detailed by the confusion matrices in Figure 8. All three models exhibited robust predictive ability with accuracy and F1-scores between 78% and 80%.



(a) (b) (c)

Figure 8. Confusion Matrix of Prediction Performance
(a) Support Vector Machine, (b) Random Forest, (c) Categorical Naïve Bayes

Table 2. Performance of Each ML Model

ML Model	Evaluation Metrics			
	Accuracy	F1-score	Sensitivity	Specificity
Support Vector Machine (SVM)	79.55%	79.85%	78.69%	81.06%
Random Forest (RF)	79.25%	79.51%	79.78%	78.33%
Categorical Naive Bayes	78.76%	79.13%	76.46%	82.83%

A deeper analysis reveals a critical trade-off between sensitivity (the ability to correctly identify fatal accidents) and specificity (the ability to correctly identify non-fatal occurrences). This trade-off is important to understand how useful each model will be in real-world applications. Table 2 indicates that RF model achieved the highest sensitivity at 79.78%, whereas the Categorical Naive Bayes model demonstrated the highest specificity at 82.83%. SVM demonstrated the most balanced performance by achieving the highest overall accuracy of 79.55% and the best F1-score of 79.85% compared to other models.

The evaluation indicates a significant trade-off between sensitivity (the ability to accurately identify fatal occurrences) and specificity (the ability to accurately identify non-fatal occurrences). This feature is very necessary for predicting fatalities. It is more important to look at the details of prediction errors compared to only focusing on the overall accuracy. In the prediction of aviation fatalities, the most significant error is a false negative, which occurs when a fatal accident is inaccurately categorized as non-fatal accident. The sensitivity metric quantitatively assesses the ability of the model to avoid this kind of inaccurate prediction. The performance of the RF model demonstrated the highest sensitivity at 79.78% which indicates the effectiveness in capturing complex and non-linear relationships among diverse risk factors, such as the aircraft type, phase of flight, and type of service. From a safety management perspective, such a model could be particularly valuable for regulators or accident investigation authorities who must prioritize minimizing overlooked fatal risks, even if it results in more frequent alerts. In contrast, the false positive one occurs when a non-fatal accident is inaccurately categorized as fatal. This error, although less severe, might impact to resource allocation and data analysis. The categorical Naive Bayes model demonstrated the highest specificity at 82.83% thereby being the most effective in reducing false alarms. In operational terms, this could reduce unnecessary resource diversion and help safety teams focus on truly critical cases. The assumption of feature independence may enable the model to effectively identify significant indicators of non-fatal outcomes, such as incidents occurring during the Taxi phase, which improves its confidence in negative predictions. However, this simplicity likely results in an inadequate understanding of the complicated relationship of factors contributing to fatalities, which leads to reduced sensitivity.

Given this analysis, the selection of an optimal model depends on the strategic priority. A model prioritizing only sensitivity (like RF) would successfully flag more potential tragedies but would also generate more false alarms. This approach could be suitable for high-level safety oversight programs where the cost of missing a potentially fatal case is far greater than dealing with surplus alerts. A model that emphasizes the specificity metrics, such as Categorical Naive Bayes, demonstrates a strong ability to predict for non-fatal accidents. This type of model might be most appropriate for daily operational monitoring in airlines, where reducing false alarms ensures smoother workflows and more efficient allocation of safety resources. However, this type of algorithm might frequently miss a significant number of fatal incidents. A balance between these two objectives is important to ensure the development of a practical and reliable system. The SVM model is therefore selected as the optimal model for this study. The highest F1-score (79.85%) proves mathematically that it achieves the best balance between sensitivity and precision. The SVM model is great at finding the best decision boundary to separate the two classes which effectively navigates the trade-off between missing a fatal accident and creating a false alarm. This balanced, robust performance makes the SVM model the most suitable and reliable for this critical predictive task, providing a practical tool for proactive risk assessment that can complement existing safety management approaches of the aviation industry. In real-world applications, an SVM-based system could be integrated into Flight Data Monitoring (FDM) or safety management frameworks, allowing both regulators and operators to make data-driven interventions that strengthen proactive risk mitigation.

4. CONCLUSION

This study successfully developed and evaluated an AI-driven framework to predict aircraft accident fatalities using a 75-year historical dataset. The analysis result obtained from EDA suggested that while the total number of fatal accidents has declined over time, the risk ratio remains high in some operational conditions. Scheduled passenger service and the en route flight phase are the categories with the highest frequency of aviation accidents, whereas military service and the maneuvering flight phase exhibit the highest likelihood of a fatal outcome in the event of an incident, which emphasizes the necessity of contextual analysis beyond the raw incident counts. In the predictive modelling stage, three ML algorithms were assessed. RF achieved the highest sensitivity (79.78%), making it the most effective at correctly identifying fatal accidents. Conversely, Categorical Naive Bayes yielded the highest specificity (82.83%), proving most adept at identifying non-fatal incidents. However, the SVM emerged as the most superior model overall, securing the highest F1-Score (79.85%) and accuracy (79.55%). This confirms its optimal balance between detecting fatal outcomes and avoiding false alarms, making it the most reliable and well-rounded model for practical risk assessment.

This study has significant implications for the aviation industry. The results from EDA can help regulators and operators to figure out which risk mitigation strategies are the most crucial in which areas. This includes the enhancement of safety procedures for non-commercial operations and the development of more intensive pilot training for high-risk flight phases, such as approach and initial climb. The predictive model that has been developed is a proactive instrument for assessing risk that can improve the current reactive safety management approaches. From a theoretical perspective, this research also advances the literature on AI applications for global aviation safety by demonstrating how long-term, worldwide accident data can be systematically modeled using ML. Unlike prior studies constrained by regional or short-term datasets, this work shows that AI can uncover broader, structural patterns of risk across service types and flight phases, thereby strengthening the academic foundation for data-driven safety management.

However, this study has limitations. Its reliance on historical data may not fully capture the impact of the latest safety technologies. The binary classification of the type of fatalities (Fatal and Non-Fatal) represents a simplification of the range of accident severity, which actually can be categorized into more than these two types. Future research should integrate more comprehensive datasets that include more variables, such as weather conditions, human factors, and information about aircraft maintenance, in order to improve the model performance. In addition, exploring deep learning, which is known as the more advanced algorithms, could potentially provide better prediction models that have been implemented in other disaster risk fields of study such as Juanara and Lam [29]. Furthermore, future studies should also investigate the integration of predictive models into real-time monitoring systems, such as FDM and Safety Management Systems (SMS), to provide early warnings and enable proactive interventions. Such integration would bridge the gap between theoretical modeling and practical implementation, supporting a more robust and proactive approach to aviation safety management.

REFERENCES

[1] The Boeing Company, "Statistical Summary of Commercial Jet Airplane Accidents: Worldwide Operations 1959-2023," 2024, Accessed: Mar. 09, 2025. [Online]. Available: https://www.boeing.com/content/dam/boeing/boeingdotcom/company/about_bca/pdf/statsum.pdf.

[2] T. T. R. Koo, C. Caponecchia, and A. Williamson, "Measuring the effect of aviation safety risk reduction on flight choice in young travellers," *Saf Sci*, vol. 73, pp. 1–7, Mar. 2015, doi: 10.1016/j.ssci.2014.10.008.

[3] M. Rey, D. Aloise, F. Soumis, and R. Pieugueu, "A data-driven model for safety risk identification from flight data analysis," *Transportation Engineering*, vol. 5, p. 100087, Sep. 2021, doi: 10.1016/j.treng.2021.100087.

[4] S. Athmaja, M. Hanumanthappa, and V. Kavitha, “A survey of ML algorithms for big data analytics,” in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, IEEE, Mar. 2017, pp. 1–4. doi: 10.1109/ICIIECS.2017.8276028.
[5] G. Demir, S. Moslem, and S. Duleba, “Artificial Intelligence in Aviation Safety: Systematic Review and Biometric Analysis,” *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 279, Nov. 2024, doi: 10.1007/s44196-024-00671-w.
[6] H. Zhang and S. Mahadevan, “Ensemble ML models for aviation incident risk prediction,” *Reliab Eng Syst Saf*, vol. 186, pp. 35–45, 2019, doi: 10.1016/j.ress.2018.11.011.
[7] F. Omrani, H. Etemadfard, and R. Shad, “Assessment of aviation accident datasets in severity prediction through ML,” *J Air Transp Manag*, vol. 115, p. 102531, Mar. 2024, doi: 10.1016/j.jairtraman.2023.102531.
[8] K. Bilgic, T. K. Dasaklis, and Y. M. G. Costa, “Prediction of injuries and fatalities in aviation accidents,” in *Proceedings of the 2017 ACM International Conference on Computing Frontiers (CF’17)*, Siena, Italy, May 2017, pp. 299–304. doi: 10.1145/3093241.3093288.
[9] T. Xia, H. Li, Y. Zhang, and H. Chen, “ML-based anomaly detection and prediction in commercial aviation,” *PLoS One*, vol. 20, no. 8, p. e0317914, Aug. 2025, doi: 10.1371/journal.pone.0317914.
[10] International Civil Aviation Organization (ICAO), “Annex 13 to the Convention on International Civil Aviation: Aircraft Accident and Incident Investigation. ,” Montreal, Canada, 2020.
[11] K. B. Marais and M. R. Robichaud, “Analysis of trends in aviation maintenance risk: An empirical approach,” *Reliab Eng Syst Saf*, vol. 106, pp. 104–118, Oct. 2012, doi: 10.1016/j.ress.2012.06.003.
[12] J. Clare and K. I. Kourousis, “Learning from Incidents in Aircraft Maintenance and Continuing Airworthiness Management: A Systematic Review,” *J Adv Transp*, vol. 2021, pp. 1–13, Feb. 2021, doi: 10.1155/2021/8852932.
[13] E. Juanara and C. Y. Lam, “Classification of Non-Seismic Tsunami Early Warning Level Using Decision Tree Algorithm,” *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 3, pp. 378–391, Oct. 2024, doi: 10.20473/jisebi.10.3.378-391.
[14] H. Ranter and F. I. Lujan, “Aviation Safety Network.” Accessed: Mar. 27, 2025. [Online]. Available: https://aviation-safety.net/about/
[15] A. Abodayeh, R. Hejazi, W. Najjar, L. Shihadeh, and R. Latif, “Web Scraping for Data Analytics: A BeautifulSoup Implementation,” in *2023 Sixth International Conference of Women in Data Science at Prince Sultan University (WiDS PSU)*, IEEE, Mar. 2023, pp. 65–69. doi: 10.1109/WiDS-PSU57071.2023.00025.
[16] F. B. Hermawan, T. E. Sutanto, and A. Santoso, “Exploratory Data Analysis of Indonesian Presidential Election Candidate Campaign in 2019 on Twitter,” *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 7, no. 2, p. 229, Apr. 2024, doi: 10.24014/ijaidm.v7i2.26308.
[17] Y. Guo, Y. Sun, Y. He, F. Du, S. Su, and C. Peng, “Deep-Learning-Based Model for Accident-Type Prediction During Approach and Landing,” *IEEE Trans Aerosp Electron Syst*, vol. 59, no. 1, pp. 472–482, Feb. 2023, doi: 10.1109/TAES.2022.3184282.
[18] D. Mustafa Abdullah and A. Mohsin Abdulazeez, “ML Applications based on SVM Classification A Review,” *Qubahan Academic Journal*, vol. 1, no. 2, pp. 81–90, Apr. 2021, doi: 10.48161/qaj.v1n2a50.
[19] J. Guo, J. Guo, Q. Zhang, and M. Huang, “Research on Rockburst Classification Prediction Based on BP-SVM Model,” *IEEE Access*, vol. 10, pp. 50427–50447, 2022, doi: 10.1109/ACCESS.2022.3173059.
[20] L. Yang *et al.*, “Study of cardiovascular disease prediction model based on random forest in eastern China,” *Sci Rep*, vol. 10, no. 1, p. 5245, Mar. 2020, doi: 10.1038/s41598-020-62133-5.
[21] M. Schonlau and R. Y. Zou, “The random forest algorithm for statistical learning,” *The Stata Journal: Promoting communications on statistics and Stata*, vol. 20, no. 1, pp. 3–29, Mar. 2020, doi: 10.1177/1536867X20909688.
[22] S. Nieland, R. Oostendorp, M. Heinrichs, and R. Cyganski, “Transferability analysis of user groups in travel behaviour surveys using a random forest classification model,” *Transportation Research Procedia*, vol. 76, pp. 81–95, 2024, doi: 10.1016/j.trpro.2023.12.040.
[23] O. Peretz, M. Koren, and O. Koren, “Naive Bayes classifier – An ensemble procedure for recall and precision enrichment,” *Eng Appl Artif Intell*, vol. 136, p. 108972, Oct. 2024, doi: 10.1016/j.engappai.2024.108972.
[24] F. H. M. Gandadipoera, R. Andrian, and R. Safei, “Tree Damage Type Classification Based on Forest Health Monitoring Using Mobile-Based Convolutional Neural Network,” *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 7, no. 2, p. 299, Apr. 2024, doi: 10.24014/ijaidm.v7i2.29421.
[25] Airbus, “A Statistical Analysis of Commercial Aviation Accidents 1958-2023,” Blagnac, France, 2024.
[26] E. A. Okine, E. Zarei, and B. J. Roggow, “Exploring the intellectual insights in aviation safety research: A systematic literature and bibliometric review,” *Saf Sci*, vol. 170, p. 106354, Feb. 2024, doi: 10.1016/j.ssci.2023.106354.
[27] Federal Aviation Administration (FAA), “Air Traffic by The Numbers,” Washington D.C. , United States., 2024.
[28] H. S. Eneterio, M. F. F. Ricco, and N. G. da Paixão Eneterio, “Risk management in military operations: An application of the Mosler method,” *Journal of the Americas*, vol. 2, pp. 226–239, 2020.
[29] E. Juanara and C. Y. Lam, “Deep learning-based waveform forecasting and physical simulations for volcano collapse tsunami early warning: The Anak Krakatau case,” *Ocean Engineering*, vol. 335, p. 121683, Aug. 2025, doi: 10.1016/j.oceaneng.2025.121683.

BIBLIOGRAPHY OF AUTHORS

M. Ridho Ramadhan received the bachelor's degree in electrical engineering with Cum Laude honors from Universitas Sriwijaya, Indonesia in 2021. During his undergraduate studies, he was ranked as an outstanding student in Faculty of Engineering at the Universitas Sriwijaya in 2020, with over 30 achievements, more than 30 event speakers, three co-authored books, and four published scientific papers. His research interests include deep learning, medical image processing, infrared thermography, healthcare device innovation, and data visualization. Currently, he is pursuing a Master's degree in Electrical Engineering at Universitas Gadjah Mada, supported by LPDP scholarship.

Avelia Fairuz Faadhilah is a graduate student in Electrical Engineering at the Bandung Institute of Technology (ITB), focusing on signal processing and RF antenna design. She holds a B.Eng from Andalas University with honors. Her research spans full-duplex antenna development using 3D printing, and audio quality assessment using ITU-R standards. She has published in international conferences such as COMNETSAT, ISAP, ICEAA, and AES. Avelia has hands-on experience with MATLAB, SDR systems, and RF planning tools like Atoll. She has served as a teaching and lab assistant in wireless communication and electric circuits. Her technical interests include antenna design, audio processing, and signal filtering. Avelia was awarded the Indonesia Education Scholarship and participated in the International Leadership in Action Summit. She aims to contribute to innovative and practical solutions in wireless systems.

Roni is a novice researcher and a final-semester chemistry student. His final project focuses on environmental chemistry, natural materials, and adsorption. He currently works as a laboratory assistant at a madrasah (Islamic boarding school) in Banjar Regency, South Kalimantan. He enjoys participating in various research projects, discussions, and youth activities to support his development as a trainer for the school's youth science group.

Elmo is a Ph.D. researcher specializing in disaster prediction, particularly tsunami forecasting caused by volcanic activities. His work integrates numerical simulation, ML, and deep learning to improve early warning systems. He's passionate about applying AI and data science for resilience and disaster risk reduction in coastal and vulnerable regions.