iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI:

Passenger Survival Probability Prediction in Aviation Accidents Using Machine Learning-Based Risk Analysis

Mrs. K.Lakshmi Viveka[1] , Kothapalli Mounika Satya Sri[2], Nukapayyi Harshita[3], Jalagadugula Puneeth Babu[4], Lakkimsetty Sri Sadhvika[5], Kommula Devi Veera Venkata Satya Sai Sanjay[6]

[1]*Associate Professor, Department of Computer Science and Engineering (AI), Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437*

[2,3,4,5,6] *B.Tech Students, Department of Computer Science and Engineering(AI), Pragati Engineering College, ADB Road, Surampalem, Near Kakinada, East Godavari District, Andhra Pradesh, India-533437*

Abstract: This project presents a machine learning-driven framework for predicting passenger survival probability in aviation accidents through structured risk analysis. The system utilizes historical aviation accident datasets containing attributes such as aircraft specifications, phase of flight, weather conditions, engine configuration, and injury severity distributions. Advanced preprocessing techniques are applied to handle missing values, normalize categorical variables, and compute survival metrics from injury data.A supervised ensemble learning model based on the Random Forest algorithm is developed to classify survival probability into three categories: Critical Risk, Moderate Risk, and High Survivability. The model generates probabilistic outputs that are further transformed into a dynamic safety score, improving interpretability and decision-making. The system is deployed using a Flask-based web application integrated with visualization tools such as probability charts and geospatial heatmaps. The framework demonstrates strong predictive capability, efficient handling of heterogeneous data, and scalability for real-world deployment. This work contributes to aviation safety analytics by enabling proactive risk estimation instead of traditional retrospective analysis.

Keywords: Machine Learning, Aviation Safety, Survival Prediction, Random Forest, Risk Analysis, Predictive Modeling, Data Preprocessing, Feature Engineering, Flask Web Application, Data Visualization.

I. INTRODUCTION

Aviation is one of the safest modes of transportation, yet accidents—though rare—can lead to severe consequences. Traditional aviation safety systems focus mainly on post-accident investigation, analyzing causes after incidents occur. However, these approaches lack predictive capability for estimating survival probabilities beforehand. With the growth of data analytics and machine learning, it is now possible to analyze historical aviation accident datasets and identify patterns that influence passenger survival. These datasets include parameters such as aircraft type, weather conditions, engine configuration, and phase of flight, which significantly impact survival outcomes. This project introduces a machine learning-based predictive framework that transforms historical aviation data into actionable intelligence. By leveraging Random Forest classification and structured feature engineering, the system estimates survival probabilities and categorizes risk levels. The integration of visualization tools such as heatmaps and probability charts further enhances interpretability. This approach shifts aviation safety from descriptive analysis to predictive intelligence, enabling proactive decision-making and improved safety strategies

A. Problem Statement

Current aviation safety analysis systems are primarily retrospective and lack predictive modeling capabilities. They analyze accident causes but do not estimate survival probability under different conditions.

There is no integrated system that:

- Uses machine learning to predict survival probability
- Handles complex aviation datasets effectively
- Provides probabilistic risk classification
- Offers intuitive visualization for decision-making

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Hence, there is a need for an intelligent predictive system that can analyze aviation accident data and estimate passenger survival probability in real time.

B. Motivation

The availability of large-scale aviation datasets and advancements in machine learning motivate the development of predictive safety systems. Traditional statistical methods fail to capture complex relationships between multiple influencing factors such as environment, aircraft design, and operational conditions.

The motivation of this project is to:

- Apply AI for proactive aviation safety analysis
- Improve interpretability using probability-based predictions
- Enable better decision-making through visualization
- Transform raw accident data into predictive insights

C. Key Objectives of this Research Include

- To design a machine learning-based survival prediction system
- To preprocess and analyze aviation accident datasets
- To classify survival probability into risk categories
- To generate interpretable safety scores
- To develop an interactive web-based visualization system

II. LITERATURE SURVEY

Recent research in aviation safety has focused on applying machine learning techniques such as Decision Trees, SVM, Logistic Regression, and Random Forest for accident analysis. Ensemble models like Random Forest have shown higher accuracy and robustness in handling structured datasets.

Feature engineering techniques such as encoding categorical variables and deriving survival metrics improve model performance. Additionally, probabilistic prediction models provide better interpretability compared to binary classification.

Visualization tools such as heatmaps and dashboards enhance understanding of accident patterns. However, many existing systems lack integration of prediction, visualization, and deployment in a unified framework.

The proposed AeroSafe system addresses these gaps by combining machine learning, preprocessing, and visualization into a single platform

S.No	Citation	Research Focus	Methodology	Key Findings
1	Li et al., 2020	Aviation accident severity prediction	Logistic Regression, Decision Trees	Identified key factors influencing accident severity but lacked probabilistic prediction
2	Chen & Zhang, 2021	Risk modeling in aviation safety	Support Vector Machine (SVM)	Improved classification accuracy but struggled with large datasets
3	Breiman, 2001	Ensemble learning techniques	Random Forest Algorithm	Demonstrated high accuracy and robustness in classification tasks
4	Kotsiantis, 2007	Supervised machine learning overview	Comparative study of ML models	Highlighted advantages of ensemble methods over single classifiers
5	Sun et al., 2019	Aviation safety data analysis	Decision Trees and clustering	Provided insights into accident patterns but lacked predictive deployment
6	Zhang et al., 2022	Survival prediction in transport systems	Random Forest, Gradient Boosting	Showed ensemble models outperform traditional statistical methods
7	Goodfellow et al., 2016	Deep learning applications	Neural Networks	Demonstrated capability for complex pattern recognition
8	Géron, 2019	Practical machine learning systems	Scikit-learn implementation	Emphasized importance of preprocessing and feature engineering
9	Pedregosa et al., 2011	Machine learning libraries	Scikit-learn framework	Provided tools for efficient model development and evaluation
10	Hunter, 2007	Data visualization techniques	Matplotlib visualization	Improved interpretation of analytical results through graphs

III. BACKGROUND WORK

The field of aviation safety analysis has traditionally relied on statistical investigations and post-accident reporting mechanisms. These conventional approaches focus on identifying causes such as mechanical failures, environmental conditions, and human errors after incidents occur. While such methods provide valuable insights, they lack predictive capability and are not suitable for proactive risk assessment.

With the advancement of data science and machine learning, researchers have begun exploring predictive analytics to estimate accident severity and survival outcomes. Historical aviation accident datasets contain structured and semi-structured information such as aircraft specifications, number of engines, phase of flight, weather conditions, geographical location, and injury severity distribution. These datasets form the foundation for developing intelligent predictive models.

One of the key aspects of background work involves **data preprocessing**, which includes:

1) Handling missing and inconsistent values
2) Standardizing categorical attributes
3) Encoding features into numerical representations
4) Deriving survival metrics from injury data

Feature engineering plays a crucial role in transforming raw aviation data into meaningful predictive inputs. For example, survival rate can be calculated using the ratio of uninjured passengers to total passengers, which helps in categorizing survival risk levels.Machine learning models such as Decision Trees, Logistic Regression, and Support Vector Machines have been used in earlier studies. However, these models often struggle with complex nonlinear relationships present in aviation datasets. Ensemble learning methods, particularly **Random Forest**, have shown better performance due to:

- Ability to handle mixed data types
- Reduction of overfitting
- High classification accuracy

Additionally, modern systems incorporate **probabilistic prediction**, where models output probability distributions instead of fixed labels. This enables better interpretation of risk levels and supports decision-making.

Visualization techniques such as heatmaps, probability charts, and dashboards further enhance analytical understanding by presenting complex results in an intuitive manner. The combination of machine learning and visualization forms the foundation for intelligent aviation safety systems.

IV. PROPOSED MODEL

The proposed AeroSafe system is a **machine learning-based predictive framework** designed to estimate passenger survival probability using structured aviation accident data. The system follows a modular and layered architecture integrating data preprocessing, predictive modeling, and visualization.

A. System Architecture Overview

The system consists of the following major components:

1. Data Preprocessing Module
   - Cleans and filters aviation accident datasets
   - Handles missing values using imputation techniques
   - Encodes categorical features using Label Encoding
   - Computes survival rate using injury statistics
   - Selects relevant features for model training
2. Machine Learning Prediction Engine
   - Uses Random Forest Classifier
   - Trains model on structured dataset
   - Generates probability outputs for three classes:
     - Critical Risk
     - Moderate Risk
     - High Survivability
   - Converts probabilities into a dynamic safety score

3. Backend System (Flask Framework)
   - Handles API requests and routing
   - Loads trained model using Joblib
   - Processes user input and returns predictions
   - Provides JSON responses for frontend integration
4. Visualization Module
   - Displays probability distribution using charts
   - Shows safety score indicator
   - Generates global aviation accident heatmaps
   - Enhances interpretability of prediction results
5. User Interface Module
   - Accepts input parameters such as:
     - Aircraft type
     - Weather condition
     - Phase of flight
     - Engine count
   - Displays prediction results in graphical format

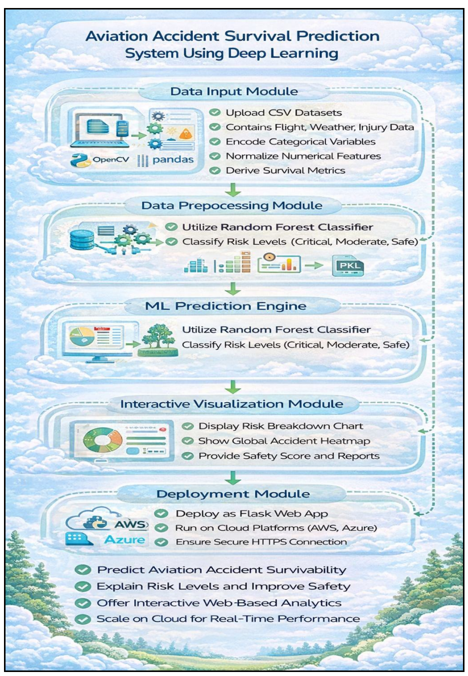


Figure 1: Architecture of Aviation Accident Survival Prediction System

Figure 1 illustrates the overall architecture of the AeroSafe system, designed to predict passenger survival probability in aviation accidents using machine learning techniques. The architecture follows a modular and layered pipeline, where each component performs a specific function, ensuring efficient data processing, prediction, and visualization.

V. IMPLEMENTATION RESULTS

The system design of the AeroSafe framework focuses on building a scalable, modular, and intelligent architecture for predicting passenger survival probability in aviation accidents. The design integrates machine learning, data preprocessing, and web-based visualization into a unified pipeline.

The architecture follows a layered approach, where each component performs a specific function, ensuring maintainability, flexibility, and efficient data flow. The system is designed to handle structured aviation datasets and provide real-time prediction outputs through a user-friendly interface

1) Home Page

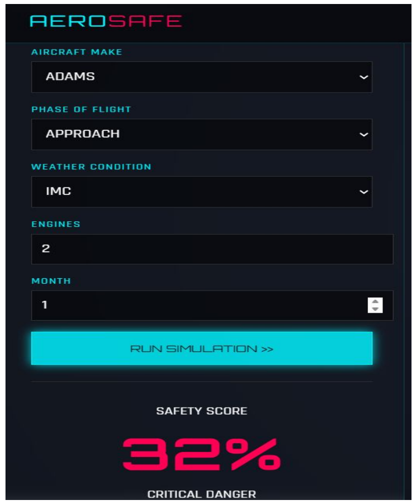


Figure 2: AeroSafe Prediction Interface and Safety Score Visualization

Figure 2 illustrates the interactive user interface of the AeroSafe system, which enables users to input aviation-related parameters and obtain survival probability predictions in real time. This interface represents the frontend layer of the system, designed with a modern and user-friendly layout to ensure ease of use and efficient interaction.

The upper section of the interface includes multiple input fields that allow users to provide critical parameters influencing survival probability. These inputs include:

- Aircraft Make (e.g., ADAMS) – represents the manufacturer of the aircraft
- Phase of Flight (e.g., APPROACH) – indicates the operational stage during which the accident occurs
- Weather Condition (e.g., IMC – Instrument Meteorological Conditions) – reflects environmental conditions affecting flight safety
- Number of Engines – specifies aircraft configuration
- Month – captures temporal patterns in aviation accidents

2) Support Skin Types

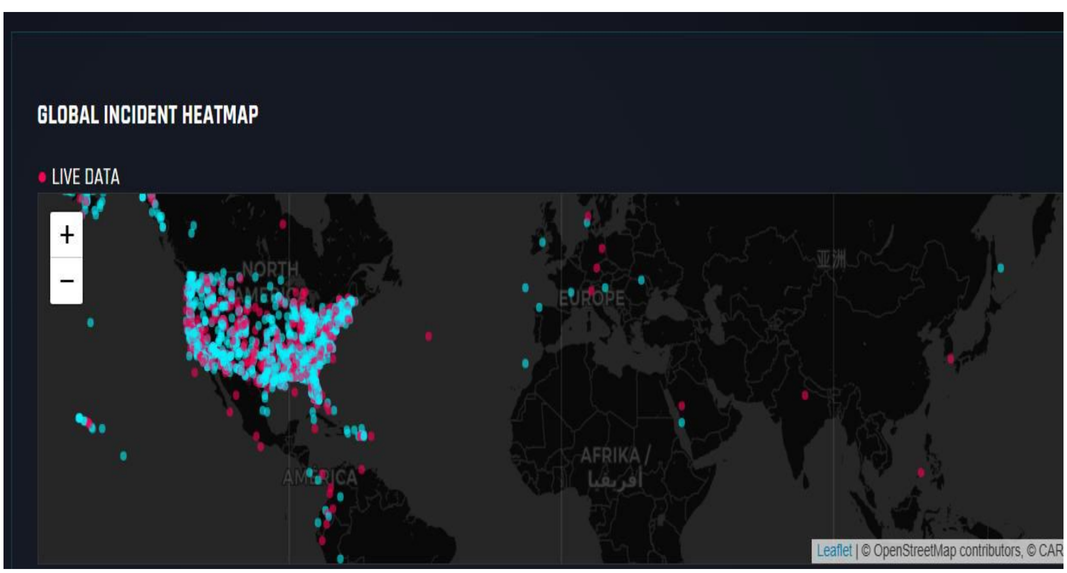


Figure 3: Global Aviation Accident Heatmap Visualization

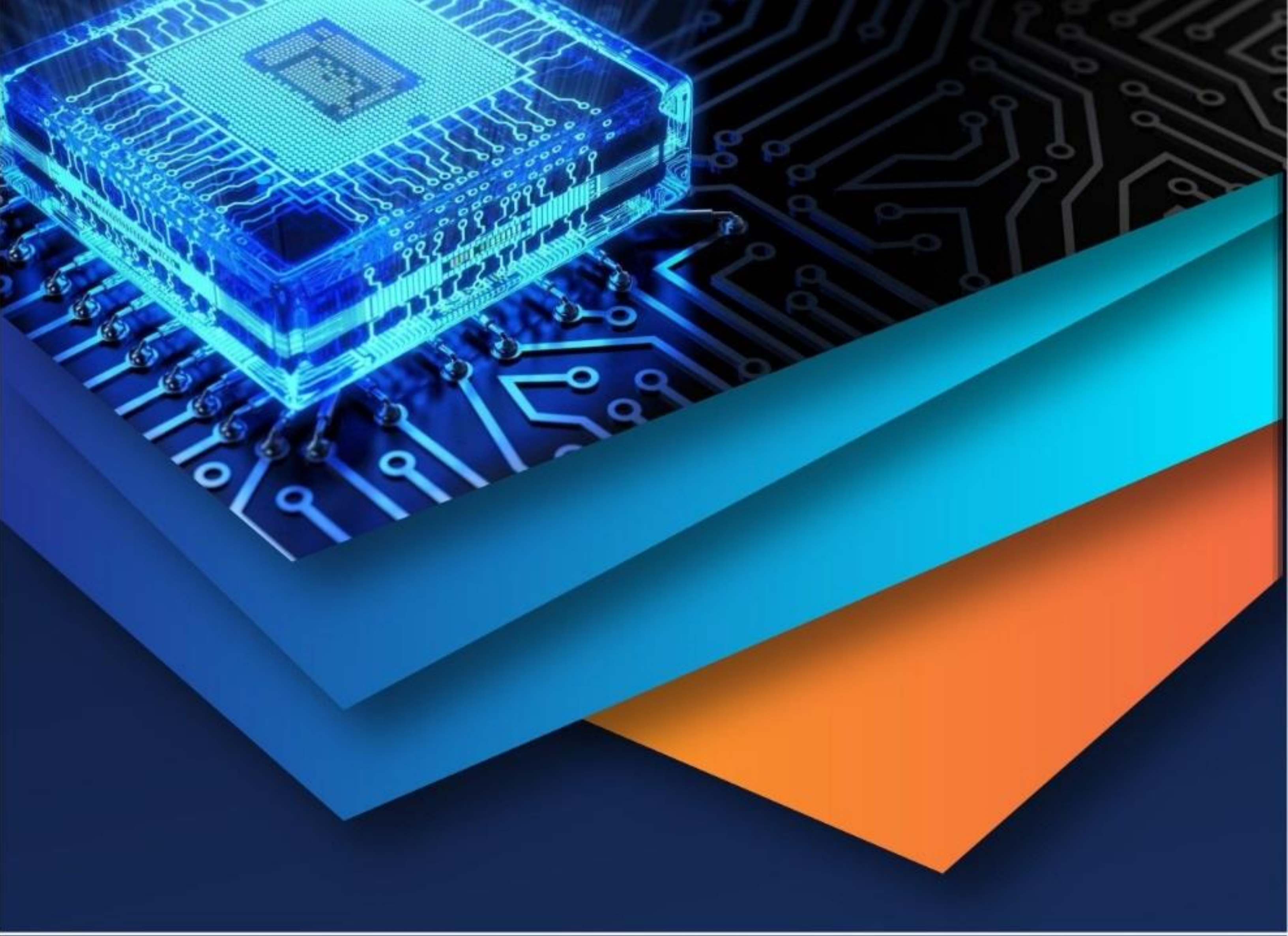
Figure 3 presents the global aviation accident heatmap, which is a key component of the AeroSafe system's visualization module. This figure illustrates the geographical distribution of aviation incidents across different regions of the world using an interactive map interface.The heatmap is generated using geospatial visualization techniques (implemented through libraries such as Leaflet.js), where accident data points are plotted based on their latitude and longitude coordinates. Each point on the map represents an aviation incident, and the clustering of points indicates regions with higher accident frequency.

VI. CONCLUSION

The AeroSafe system successfully demonstrates the application of machine learning in aviation safety by predicting passenger survival probability using structured risk analysis. The integration of preprocessing techniques, Random Forest classification, and probabilistic modeling enables accurate and reliable prediction of survival outcomes. The system transforms traditional aviation safety analysis from a reactive approach into a proactive predictive framework. The introduction of a dynamic safety score improves interpretability, allowing users to easily understand survival risk levels. Additionally, the integration of visualization tools such as charts and heatmaps enhances user interaction and analytical clarity. The use of open-source technologies ensures cost-effectiveness, scalability, and ease of deployment. The modular architecture allows future enhancements such as integration with real-time flight data, advanced deep learning models, and cloud-based deployment. Overall, this project highlights the potential of AI-driven predictive systems in improving aviation safety and enabling data-driven decision-making, thereby contributing to the development of intelligent aviation risk management systems.

REFERENCES

[1] Y. Li, X. Wang, and J. Liu, "Aviation accident severity prediction using machine learning techniques," Safety Science, vol. 129, pp. 104–112, 2020.
[2] H. Chen and Q. Zhang, "Support vector machine-based risk assessment in aviation safety," Journal of Air Transport Management, vol. 92, pp. 102–110, 2021.
[3] L. Breiman, "Random forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[4] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques," Informatica, vol. 31, no. 3, pp. 249–268, 2007.
[5] J. Sun, Y. Zhao, and L. Qin, "Data mining approaches for aviation accident analysis," IEEE Access, vol. 7, pp. 123–134, 2019.
[6] T. Zhang, H. Liu, and M. Chen, "Ensemble learning for transportation risk prediction," Transportation Research Part C, vol. 134, pp. 103–115, 2022.
[7] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
[8] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd ed. O'Reilly Media, 2019.
[9] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[10] J. D. Hunter, "Matplotlib: A 2D graphics environment," Computing in Science & Engineering, vol. 9, no. 3, pp. 90–95, 2007.



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY