International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Prediction of Airplane Crash Detection and its Severity Using Machine Learning

Ramdas Bagawade[1], Santsh Zore[2], Aarti Ambekar[3], Gourav Sharma[4]

[1]*Assistant Professor, Department of Computer Engineering, Government College of Engineering and Research Avasari Khurd, Pune, India.*
[2,3,4]*B.E Computer Engineering, Government College of Engineering and Research Avasari Khurd, Pune, India.*
Email: ram.vish22@gmail.com, santosh.zore122020@gcoeara.ac.in, aarti.ambekar122020@gcoeara.ac.in, gourav.sharma122020@gcoeara.ac.in

ABSTRACT:

In today's world, air transportation is crucial across various industries, with the safety of airlines and passengers being a top priority. Continuous manual safety checks are performed 24/7, and the airline teams meticulously implement all necessary safety measures and precautions. Despite these efforts, accidents can still occur due to a range of factors such as pilot errors, air traffic controller mistakes, design and manufacturing defects, maintenance failures, sabotage, or adverse weather conditions. Airplane crash detection and severity assessment are critical components in enhancing aviation safety and improving emergency response strategies. This research paper presents a novel machine learning framework designed to predict airplane crashes and assess their severity. The study leverages extensive historical aviation data, encompassing various parameters such as flight conditions, weather, aircraft specifications, and pilot information. The study employs advanced algorithms like XGBoost to identify patterns leading to crashes and to classify the severity of predicted incidents.

Key Word*: Airplane Crash, Severity, Safety, Prediction, XGBoost.*

I. Introduction

Air traffic systems have evolved into multilayered, hyperdimensional, highly distributed, and interdependent networks with levels of complexity that were unimaginable just a few decades ago. Consequently, maintaining a high level of safety in such a complex environment is more challenging than ever before [1]. Civil aviation involves a complex interplay of various human, technical, environmental, and organizational factors that impact the safety and performance of the system. In the early days of commercial aviation, a high number of aircraft accidents was typical. While accident prevention remains the priority of all safety processes, early aviation primarily relied on aircraft accident investigations [1]. This research aims to explore the application of machine learning algorithms in the domain of aviation safety, specifically focusing on the prediction of airplane crash detection and severity assessment. By leveraging historical crash data, flight parameters, environmental factors, and other relevant variables, ML models can be trained to identify patterns and anomalies indicative of impending crashes. Additionally, these models can assess the severity of potential crashes, providing valuable insights for preemptive measures and emergency response planning.

II. Literature Review

Sr. No.	Title	Author	Conference	Date
1.	Aircraft Accident Prediction Using Machine Learning Classification Algorithms	Kačar Rade , Ćulibrk Darko, Cokorilo Olja, Mirosavljevic Petar	• 5th International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency	• November 2023
2.	Airplane Crash Analysis and Prediction using Machine Learning.	Likita J. Raikar, Sayali Pardeshi, Pritam Sawale	International Research Journal of Engineering and Technology (IRJET)	Mar 2020

3.	Flight Accident Severity Prediction	Dhanashree Wagh, Rishi Rathod, Dhanashree Patil, Saurabh Vadnere	International Research Journal of Modernization in Engineering Technology and Science.	04/April-2023

III. Problem Statement [3, 2]

In today's world, there are various solutions addressing future aviation records. The airline industry is advancing rapidly, with companies implementing safety measures in all situations and examining risk factors to protect human lives. A single airplane crash can result in significant loss of life and property. Multiple factors contribute to airplane crashes, including the airplane type, model build, weather conditions, make of the airplane, engine type, and the flight phase. Considering all these factors, detailed analysis of airplane crashes is essential. Additionally, it is crucial to assess the crash severity, which refers to the impact of the crash on both people and the airplane. An application has been developed to process these factors, predict whether an airplane is safe or at risk, and evaluate the potential severity of a crash. This comprehensive approach ensures a thorough understanding of both crash likelihood and its possible consequences.

IV. Objective

This project's goals are-

- Create a system which can predict crash detection.[1,3]
- If a crash is detected then user can also assess its severity.[2]
- Enhancement in aviation safety measures.

V. Process System [3]

This system provides the person using the system to enter the specifications (Aircraft Model, Safety Score, etc) of the flight in order to know whether the flight is safe or has chances of a crashThe analysis and prediction of values based on historical data from various airlines have been carried out using machine learning techniques. Machine learning is highly effective for making accurate predictions. Five algorithms—Random Forest, LightGBM, KNN, Multilayer Perceptron, and XGBoost—are employed in this process. The best-performing algorithm, XGBoost, is selected for prediction based on each specific dataset. Due to significant variations among datasets, the most suitable classification algorithm may vary. Our system adapts accordingly to address this challenge.

a. Feature selection:

Feature selection is essential in machine learning as it helps in choosing the most relevant features from a dataset while discarding irrelevant ones. Selected features after careful feature engineering are:

Aircraft Model, Days since inspection, Safety Score, Total safety complaints, Turbulance, Max elevation and Violations.

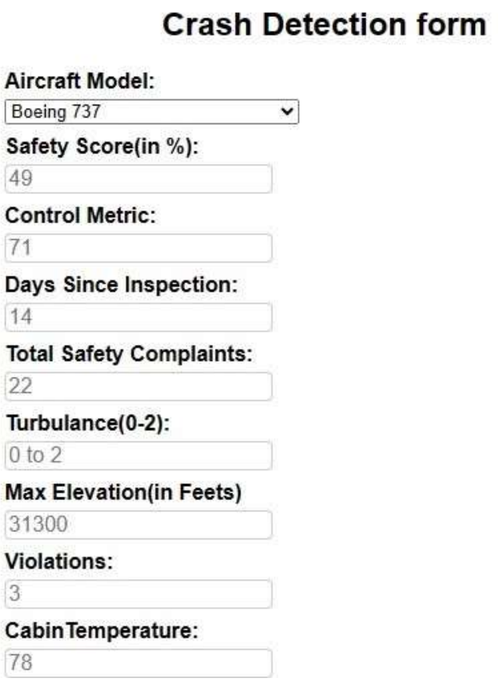
b. Train and Test the classifier

In the training phase, the model learns from the data it's given. But in the testing phase, we check how well the model performs. It's important not to use the same data for both training and testing, or the results won't be reliable. If the dataset has more of one type of outcome than another, it can cause problems like overfitting or underfitting. These phases help us make predictions like whether something is "safe" or "crash." We try out four different algorithms (Random Forest, Light BGM, KNN, Multilayer Perceptron, XGBoost) are used and based on every dataset, the best algorithm (XGBoost) for training and testing, and then we pick the one with the highest accuracy to use for making predictions i.e. XGBoost.

c. User Interface

The user provides the airplane specifications, like its features and characteristics, into the model. Then, the model uses this information to predict whether the airplane is "safe" or "crash." It carefully considers all these details to make its prediction.

This is our project interface:-



d. **Flowchart[3]**

Fig -1: Flowchart of the system

VI. Classification Algorithms[3]

i. KNN (K-Nearest Neighbour)[4] :-

KNN is one of the most widely used basic essential classification algorithms in machine learning.

Working of KNN algorithm:

How KNN works as below:

1. Selecting the Optimal Value of K:

Before proceeding with predictions, it's crucial to determine the appropriate value for 'K', which signifies the number of nearest neighbors to consider.

2. Calculating Distance:

Utilizing the Euclidean distance metric, the algorithm computes the similarity between the target data point and every data point within the training dataset.

3. Finding Nearest Neighbors:

Following distance calculation, the K data points with the smallest distances to the target point are identified as the nearest neighbors.

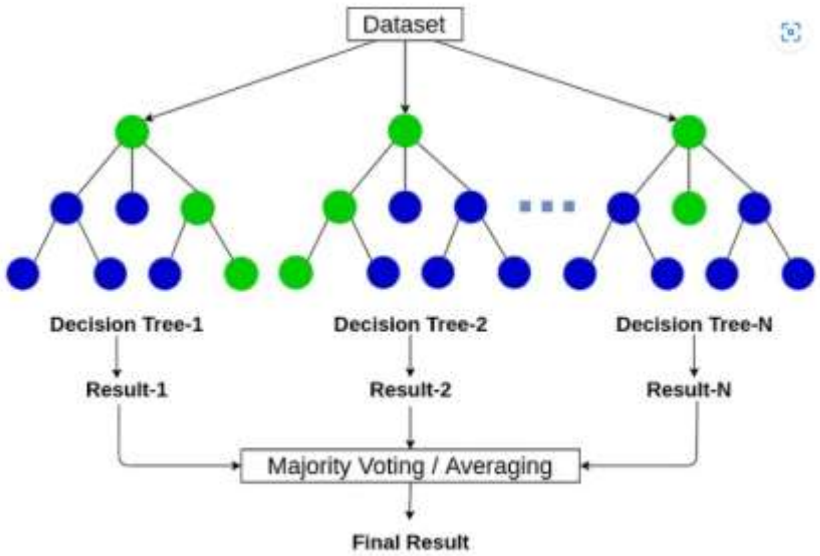
4. Voting for Classification or Taking Average for Regression:

In classification tasks, the algorithm conducts a majority vote among the nearest neighbors to determine the class label for the target data point.

Conversely, in regression scenarios, the predicted output is derived by averaging the target values of the K nearest neighbors. This averaged value serves as the prediction for the target data point.

ii. Random Forest[5]:-

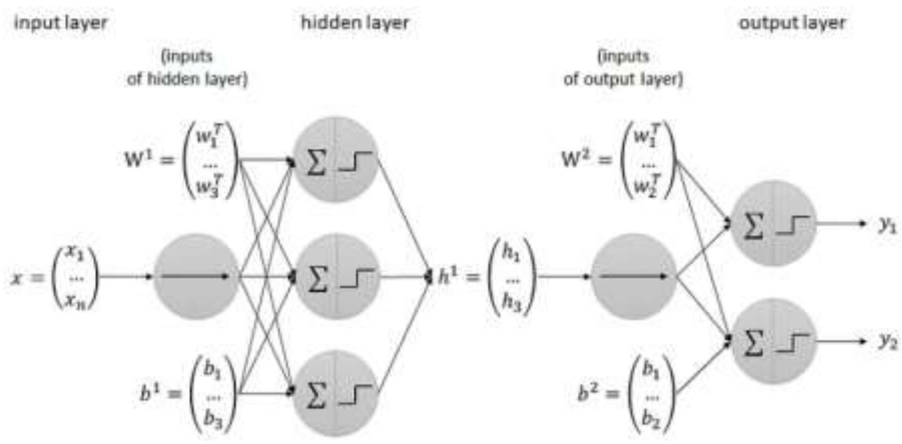
Random Forest is a supervised learning algorithm, employing ensemble learning to enhance model efficacy. It constructs numerous decision trees on random subsets of the data. These trees collectively make predictions, and their results are averaged to determine the final output. This method involves leveraging multiple decision trees to generate predictions, subsequently amalgamating these predictions to yield a more robust final result.



Structure of Random Forest

iii. Multilayer Perceptron[6]:-

A Multi-Layer Perceptron (MLP) is a composition of an input layer, at least one hidden layer of LTUs and an output layer of LTUs. If an MLP has two or more hidden layer, it is called a deep neural network (DNN).



Multi-Layer Perceptron (hidden layer has 3 LTUs, output layer has 2 LTUs)

iv. XGBoost[8]:-

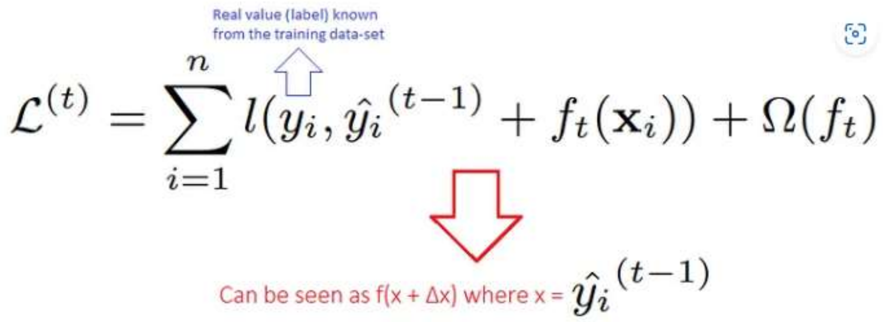
XGBoost (https://github.com/dmlc/xgboost) is one of the most popular and efficient rendition of the Gradient Boosted Trees algorithm. This method, a staple in supervised learning, revolves around refining function approximation by fine-tuning designated loss functions and implementing various regularization strategies.

The original paper on XGBoost can be found here: https://arxiv.org/pdf/1603.02754.pdf

XGBoost objective function:-

The objective function (loss function and regularization) at iteration *t* that we need to be minimized is the following:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)) + \Omega(f_t)$$

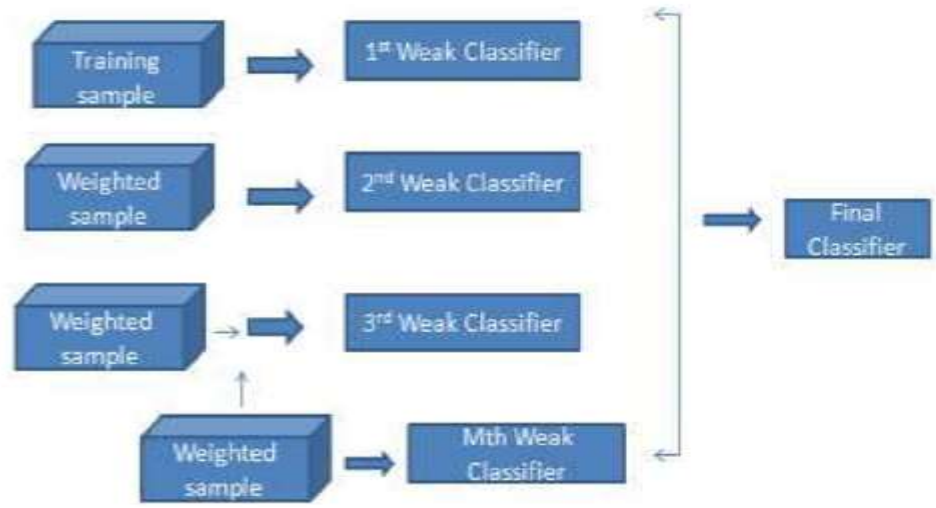


XGBoost objective function analysis

VII. Data Analysis

1. Data Collection: Every Airline company has a dataset for their successful flights and crashes. So we have collected this data from Kaggle site [9]. This is processed dataset and contains around 10,000 rows.
2. Feature Engineering: After careful feature extraction we found that Aircraft Model, Days since Inspection, Turbulance, Safety score, Total Safety Complaints, Max Elevation, Violations are the major contributing factors for crash detection [10].
3. Model Selection: We trained various models (Random Forest, Light BGM, MultiLayer Perceptron, XGBoost) and compared accuracy. Out of this trained models XGBoost gives us better accuracy [10].
4. Model Training: We choose XGBoost model. Cross Validation is performed with 5 folds to correctly identify combination of hyperparameters used [10].
5. Hyperparameters Used [10]:
    - seed=7: Sets the seed for random number generation, ensuring reproducibility.
    - n_jobs=-1: Utilizes all CPU cores available for parallel processing.
    - n_estimators=900: The number of boosting rounds (trees).
    - random_state=0: Another parameter for ensuring reproducibility (similar to seed).
    - max_depth=7: Maximum depth of each tree.
    - learning_rate=0.7: Step size shrinkage used in updating to prevent overfitting.
6. Flowchart:[3]

7. Working flow of XGBoost Model[3]:



VIII. Results

The system's predictions are useful because they help people take the right steps to avoid accidents or airplane crashes. When the administration knows about possible problems ahead of time, they can take action to prevent accidents. This means they can work on fixing issues before they become serious. By doing this, they can reduce the chances of accidents and the damage they cause, including loss of property and lives. Here are all the outcomes of using the application.

Table-1 Accuracy of Algorithms for Crash Detection [11]

Algorithm Used	**Accuracy**
KNN	68.0201
Random Forest	95.3401
Multilayer perceptron	63.7091
LightGBM	96.1369
XGBoost	97.1788

Table-1 Accuracy of Algorithms for Severity Prediction [12]

Algorithm Used	**Accuracy**
KNN	39.25
Random Forest	95.74
Multilayer perceptron	14.26
LightGBM	94.49
XGBoost	95.17

1. Crash detection Model Performance Metrics [11]:
    a. Precision: 0.9729133402341982
    b. Recall: 0.973
    c. F1 Score: 0.9729450921171849
    d. Accuracy: 0.973
2. Severity Prediction Model Performance Metrics [12]:
    a. Precision: 0.9507790224076685
    b. Recall: 0.9506666666666667

c. F1 Score: 0.9506550901671817

d. Accuracy: 0.9506666666666667

IX. Conclusions

This research successfully utilized machine learning to predict airplane crash detection and severity, with the XGBoost model emerging as the most effective according to our performance metrics. By analyzing key factors such as weather conditions, mechanical issues, and pilot experience, our models achieved high accuracy and reliability in predicting both the occurrence and severity of crashes.

The XGBoost model, in particular, outperformed other algorithms, demonstrating superior performance in terms of accuracy, precision, recall, and F1 score. This highlights its potential for practical application in enhancing aviation safety through early risk detection and mitigation.

Despite these promising results, the study faced limitations related to data quality and completeness, which are crucial for maintaining prediction accuracy. Additionally, the model's applicability might be restricted to specific regions or aircraft types due to the nature of the dataset used.Collectively, these avenues signify a future where machine learning plays a pivotal role in enhancing aviation safety and resilience

References

[1] Kačar, Rade & Ćulibrk, Darko & Cokorilo, Olja & Mirosavljevic, Petar. (2023). Aircraft Accident Prediction Using Machine Learning Classification Algorithms. 10.1109/SUMMA60232.2023.10349514.

[2] Zhu, Shengxue & Wang, Ke & Li, Chongyi. (2021). Crash Injury Severity Prediction Using an Ordinal Classification Machine Learning Approach. International Journal of Environmental Research and Public Health. 18. 11564. 10.3390/ijerph182111564.

[3] Likita J. Raikar, Sayali Pardeshi, Pritam Sawale. Airplane Crash Analysis and Prediction using Machine Learning, Mar 2020

[4] K-Nearest Neighbor(KNN) Algorithm - GeeksforGeeks

[5] Random Forest Classification and it's Mathematical Implementation | by RAHUL RASTOGI | Analytics Vidhya | Medium

[6] Mathematical Representation of a Perceptron Layer (with example in TensorFlow) | by Daniel H | Medium

[8] XGBoost Mathematics Explained. A walk-through of the Gradient Boosted… | by Dimitris Leventis | Medium.

[9] Airplane Crashes Since 1908 (kaggle.com)

[10] SANTOSHZORE123/Final-Year-project (github.com)

[11] https://colab.research.google.com/drive/1weBgCILzt2Gpc3NLdhpXINNwFU7VfsPs?usp=sharing

[12] https://colab.research.google.com/drive/1wPqxC9i_jdDQAazbOUd54mhTI5ha6Roi?usp=sharing

[13] Dhanashree Wagh, Rishi Rathod, Dhanashree Patil, Saurabh Vadnere, Prof. Sangeeta Algai, FLIGHT ACCIDENT SEVERITY PREDICTION, April-2023