A Comparative Overview of Accident Forecasting Approaches for Aviation Safety

Monika¹, SeemaVerma², and Pardeep Kumar³
¹ Assistant Professor, Department of Computer Science, Banasthali Vidyapith, Jaipur, Rajasthan
² Professor, Department of Electronics, Banasthali Vidyapith, Jaipur, Rajasthan
³ Associate Professor, Department of Electronics and Communication, SGT University, Gurugram, Haryana

E-mail: mor.monika@gmail.com

Abstract. The demand of air transportation is expected to be doubled over the next two decades as per the recommendations of the International Air Transport Association. This would prompt more aviation safety issues with increased air traffic congestion and load on the air transportation system. To estimate the level of risk and improve the forecasting ability, various methodologies have been proposed by the research community. As each methodology has its pros and cons, this manuscript provides a comparative study of various Data mining, Time series, Artificial Neural Networks, and ensemble Techniques on the aviation safety and forecasting complication. This paper concludes that different methods dealing with different information may be combined to have an outstanding prospective in aviation accident forecasting and to come up with a number of ways of enhancement and their assistance in decision making.

1. Introduction
The continuous growth in demand for air transportation and scarcity of infrastructure results in increased pressure on the aviation industry. As per the recommendations of International Air Transport Association (IATA), the demand for air transport is continuously increasing around the world which is expected to be doubled over the next two decades [1]. This rapid increase will prompt more safety issues with increased air traffic congestion and load on Air Transportation Management (ATM) system. Due to the severe consequences of aviation accidents, much more attention has been given to the safety of the airspace system. Figure 1a. shows the accident rates and Figure 1b. shows the fatalities caused from 2008 to 2019 on scheduled commercial flights. The rate of growth of accidents and fatalities indicates that flight safety is a major concern for the aviation industry and must be taken seriously. During the last 10 years, lots of research and efforts have been done exclusively in this direction to encourage and improve forecasting methodologies and risk estimation. Safety is always the primary goal of the International Civil Aviation Organization (ICAO).
To further improve safety and standards, ICAO has collaborated with many entities related to air transportation. The National Aeronautics and Space Administration (NASA) are running the Aviation Safety Reporting System (ASRS) program to achieve a higher degree of aviation system safety. ASRS fetches processes and interprets deliberately submitted incident reports from air traffic controllers, pilots, cabin crew, flight attendants, technicians, and other staff members involved in various aviation operations [3]. The Federal Aviation Administration (FAA) responsible for implementing the Federal Aviation regulations, providing the safest airspace in the world. It is working under the US Department of Transportation. National Transportation Safety Board (NTSB) is an independent agency responsible for investigating transportation accidents and issuing safety recommendations to the FAA. It provides comprehensive aircraft incident and accident databases. This includes the place of accident, aircraft types, date, air carrier information, etc. NASA in joint partnership with FAA and US national Airspace System (NAS) working to identify the risk factors and improve aviation safety by developing new techniques for improving the future aviation safety issues [4].

Safety analysis is basically a way to improve safety. The scope of analysis lies from the investigative approach to the predictive approach. The investigative approaches depend on data of accidents, incidents, and near misses. Predictive approaches can also be implemented to analyze the data of any incident occurrence and then identifying the risk factors which may result in an accident [5]. A huge amount of data repositories are maintained by the various regulatory bodies like NTSB, FAA, NASA, ICAO, airlines, etc. had been analyzed by the research community to facilitate the officers for taking precautionary actions against loss and risk of occurrences of accidents. A precise and sustainable accident forecast will not only lessen the fatality and financial loss but also augment the evolution of aviation safety administration [6].

There are many factors responsible for aviation accidents, together with certain and unforeseen factors. Few unanticipated factors involve harsh weather like rain, low cloud, thunderstorms, hails, winds shear, lightning, etc. and bird ingestion [7] and certain factors cover pilot error, mechanical defects, air traffic control errors, ground support faults, inexperienced crew, workshops and factories maintenance errors, etc. [8]. Owing to the participation of the real-time dynamic systems it is impossible to predict with a simple data model. A lot of researches have been carried out for aviation accident forecasting using various advanced techniques based on statistical methods, data mining, artificial intelligence, and machine learning [9]. These techniques can be further categorized into prominent groups like Bayesian Methods, Times Series methods, mining methods, Artificial Neural
Networks (ANN), ensemble methods, etc. [10]. This paper provides a comparative analysis of all these techniques in concern to their applicability to aviation safety. This paper focuses on various models used for aviation accident prediction to come up with several ways of enhancement and their assistance in decision making.

The rest of the article is organized as follows: Section 2 describes the various techniques applied for safety prediction in aviation. Section 3 provides a comparative discussion on data analysis approaches used in an aviation accident and incident data. The concluding remarks and future scopes wrapped up in section 4.

2. Literature Review

Over the past few years, several studies have ventured to pay attention to aviation accidents due to an increase in global aviation transportation. It is necessary to perform detailed research for enhancing the safety of passengers and reducing the accident rate. Here, various methods those are responsible for valid and accurate prediction of accidents are identified and are discussed to improve forecasting capability for aviation and to minimize the possibility of occurrence of incidents or accidents. These methods are constructed over historical data provided by various agencies like ASRS, FAA, NTSB and NASA etc.

2.1. Bayesian Networks (BNs)

To build a model from existing accidental data, Bayesian Networks (BN) can be employed. These networks are a probabilistic graphical model that may be used for prediction, diagnostic, anomaly detection. Due to complicity and uncertainty of accident data, BN’s are one of the better options for prediction or identification of the accidents. Bayes theorem states that posterior knowledge can be deducted getting prior knowledge and data under observation. After the availability of inspection data for one node and by updating its probabilities, then the probabilities of associated nodes in the network scenario can also be updated easily. That’s why BN are widely used for solving data scarcity problem. The noisy-OR gates and recursive Noisy-OR rule are used to generate conditional probabilities table in BNs [11]. Figure 2. shows the Bayesian process.

![Bayesian process](image)

Figure 2: Bayesian process [11]

Flight delay is one among the prominent factors of aviation safety hazard. Huawei Wnag et al. [12] have used BNs to build the safety assessment model for aviation-based flight delays. It is found that
the civil aviation safety level reduces with the increase in flight delays. One of the major reasons for Mid-Air collision accident is Loss of Separation (LOS). Schon Z. Y Liang Chen et al. [13] model the possibility of these precursors in combination with BN and results show that factors responsible for LOS could forecast the severity and in turn can forecasting. Eduardo S. Arya et al. studies the factors responsible for run-way overrun and their classification with the help of BNs.

Lots of studies investigate the fitness of various crash prediction frameworks to calculate the severity and frequency of crashes. Generalized linear regression models like Logit or Probit models and passion models have been applied to investigate the relationship between accident occurrences and the associated risk factors Huang et al. [14] have proposed 5XT level hierarchies for studying the multilevel data structures available in crash data. This framework based on Bayesian modeling has the potential to model the heterogeneities existing in the multilevel data structures.

A methodology for the prediction of road accidents is presented by Markus Deublien et al. [15]. This is a combination of hierarchical multivariate passion-log-normal regression and Bayesian probabilistic networks. First, the response variables (i.e. the number of fatalities) and risk determining factors are observed. Then, gamma updating of the response variables is done. After this, the Bayesian inference algorithm is applied to model the non-linear relationship between risk indicating response variables and other uncertainties.

An innovative statistic method based on Bayesian inference and hierarchical structures in developed by Rosa Maria et al. [10]. This will help in forecasting future safety events and risks. Many efforts have been put to find more perspectives on preventing accidents and minimizing the risks to aircraft travelers. The operational and research staffs are continuously trying to find techniques for a more accurate prediction of risks. Peter Broober [16] has presented a solution to this based on Bayesian Belief Networks (BBN). The conditional probabilities of event occurrence are estimated to presents a model of expertise. If one can embed the real-world information in BN then it results in eliminating model complexity. But, the traditional BNs have the limitation of temporal dependency.

2.2. Time series
Time series analysis is a prediction strategy of statistical analysis that explains the statistical attributes of a variable and discloses the rule of small change in data pattern as stated by a statistical association between data. Time series analysis can be applied to aviation accident data to explore the variation law in long term historical data. This helps in predicting the probable transition of future aviation accidents. By providing a reference for aviation safety research Yafei Li [6] has provided an analysis of civil aviation accidents based on Mann-Kendall trend analysis and mutation analysis. They built Time-series based Autoregressive Moving Average (ARMA) which can predict the fluctuations in accidents and casualties. The author made a comparison of different forecasting phases to explore the trend of civil aviation accidents worldwide. A multi-variate time series search algorithm for anomaly detection has been proposed by Brayn Matthews et al. [17].

Andrej Lali’s [18] performed analysis on safety performance tool index and specifies new features, which can use time series analysis for improving recognition index by industry and future research on aviation safety domain. X. Y. Huang, et al. applies Grey Model to aviation safety prediction for the data that follows the exponential law [19]. Weiwei Zhang et al. perform time series forecasting in high dimensional parameter space for quantitative prediction and found that more accurate and macroscopic prediction is possible with the help as time series [20].

2.3. Artificial Neural Network (ANN)
This is the major technique used for Machine Learning (ML) algorithms. These are brain-stimulated systems as indicated by their name containing neural which resembles neurons in the human brain. The main idea is to create a statistical configuration similar to that of interrelated neurons in the human brain hooked up to a fire and generate an outcome when provided with any input stimulus
signal as shown in Figure 3. They are superb tools to discover such patterns that are numerous or complex to extricate by humans and train the machine to perceive them [21]. ANN can be easily fit for complicated non-linear associations among the input and output. So these are evidenced as a strong tool for predicting and forecasting aviation accidents. Jeevith Hegde and Borge Rokseth [22] specify that ANN is the best performing algorithm for risk evaluation in combination with SVM. ANN is the future requirements for real-time estimation and prediction of risk to cater to safety requirements by various regulatory bodies.

![Artificial Neural Networks](image)

**Figure 3:** Artificial Neural Networks

Shreyas Vathul Subramanian et al. [23] have used deep learning networks to improve safety in national the national airspace system by training the Long Short-Term Memory (LSTM) networks to analyze Go-Around (GAR) and Missed Approach (MA) data from the ASRS database. This forecast helps in identifying the factors which are responsible for the incidents. At the same time, it also provides perception about those sort of incidents whose probability of occurrence are less (or more) during the forecast period. The authors perform the data transformation, learning, and the prediction using by following multiple parameter combinations, a grid search was performed to build an LSTM network.

Ayca Altay et al. use aircraft age and the type of aircraft for prediction with the help of Genetic Algorithm and artificial neural networks. The results of this study provide more accurate forecasting having good correlation degree [24]. Xiaojing Yan [25] focuses on aviation accidents due to engine factors and forms a fault-tree model to be mapped with ANN. The accuracy of the proposed model is superior to the traditional learning methods.

Prabher Srinivasan et al. [26] uses word embedding techniques of NLP which forms a vector of words, where words which have same meaning appear closer to each other and use this vector to develop deep learning model with LSTM algorithm predicting whether a series of events may results in an aircraft accident or damage. This ANN helps the operator to work in a what-if situation and take a precautionary measure in advance.

X Zheng, and M Liu [9] have identified seven approaches which are divided into two categories: time series forecasting and causality forecasting. Time series forecasting involves time series, gray model, Markov chain method, and neural network. Causality forecasting includes the Bayesian network, regression method, and scenario analysis. Here, the neural network and Grey model are combined which proves to be better in non-linear problems and helps in accident forecasting.
2.4. **Ensemble Methods**  
Ensemble methods are an extension of machine learning techniques which includes consideration of different prediction models to have more optimized prediction capability as shown in Figure 4. Here, the better-predicted outcome can be obtained by reducing the variance and bias. Ensemble methods generate a model that has the competency to different regions. The projected consequences can be obtained based on the weighted results of different models.

![Figure 4: Artificial Neural Networks](image)

Ziaoge et al. [1] develop a hybrid model that handles both structured and unstructured data to learn the relationship of abnormal events with their consequences. Textual or unstructured data after preprocessing is fed to SVM and categorical or structured data is trained using deep neural networks. In the next stage results predicted from DNN and SVM were fed to probabilistic fusion rules for prediction results and lastly expanded to event outcome analysis.

R. A. Burnett and D. Si [27] consider ML algorithms such as SVM, k-NN, ANN, and apply them on the FAA part-91 dataset that focuses on general aviation accidents. The authors created six datasets and used them for testing and generate confusion matrix for each dataset to observe the algorithm performance. Results show that better prediction is possible with ANN in comparison with traditional methods of statistical analysis but the data should not be ill-behaved.

Prediction technique based on deep learning neural networks has been used for civil aviation safety evaluation by Xiaomei Ni et al. [28]. This is a combination of Deep Belief Network (DBN) and Principal Component Analysis (PCA). DBN can predict the flight incident rate based on the outcome of PCA. The proposed technique is capable of extracting the main influencing elements to lower the flight incident rate.

Apoorv Maheshwary et al. [21] have performed a comparative study of various machine learning approaches for solving the problem of air travel demand and city pair estimation using neural network, SVM, regression and classification, etc. This comparative study selects a suitable algorithm for given problems. The forecasting model tries to map future demand relying on time ordered sequencing and to estimate the linkage between demand levels with its determining variables.

A greedy approach is designed by M Salama [29] using DBN and RBM, to extract useful features for forecasting and prediction and scaled the continuous data to binary. RBM has hidden layers to model the distribution. DBN classification uses three layers of RBMs. Initially, RBM accepts data from the first layer of visible nodes which is passed further to hidden nodes in the second layer. Finally, it is passed to the last layer having one visible node. The experimental results show that the proposed approach is better in terms of feature reduction or reducing the dimensions, clustering, and further classification of continuous data.

2.5. **Mining Methods**  
Data mining is generally based on pattern identification from the collected raw data. It helps in finding out hidden patterns from larger data sets. Mining helps in extracting knowledge from big data. This
process is termed as knowledge discovery from data (KDD). Many data mining approaches have been applied to the aviation accident dataset for analysis of aviation safety. S. Koteeswaran et al. [30] proposed an aviation accident prediction approach by unifying k-NN and co-relation based features selection technique with search based on oscillation search and it is named as "Improved Oscillated correlation feature selection (IOCFS)". The novel approach could help in making the aviation management system even better by predicting the cause of the accident and finding out the risk.

The main purpose of accident data is used to learn the cause of accidents and how to prevent them in the future. A. H. Rao and K. Marais [31] have proposed a state-based approach by defining a grammar that depicts the sequence of states and triggers. The performance of this model shows that the proposed approach gives a better count for accident causes using rule-based logic.

Arockia Christopher et al. [32] applying traditional classifier algorithms, like Bayes Classifiers, lazy classifiers, rules classifiers, decision tree classifiers in large-scale datasets in the aviation domain. The performance of different classifier algorithms has been evaluated based on feature selection methods like gain ratio, relief F, information gain, and One R attribute evaluation. It is found that the decision tree classifier has lower misclassification rates with better classification accuracy.

Bryan Mathews et al. [17] have designed as aviation safety knowledge discovery (AVSKD) anomaly detection methods to scan multi-variate historically observed data. After cleaning and filtering the initial raw FOQA, AVSKD method has been applied for anomaly detection. Once anomalies are discovered, the severity and frequency of those events are calculated and validated to generate a report. Airspace operations occur due to cooperation between Air Traffic Control (ATC) and flights in the airspace. An analysis of stabilized approaches for airspace risk management is presented by Zhenming Wang et al. [33]. Here, a simulation-based and data analysis approach is applied to identify various events and their potential risk.

3. Comparative Analysis
The above-mentioned methods can be exercised, pro-actively for boosting the safety of aviation. All these approaches are used beneath some hypotheses and confined conditions, which have their intrinsic pros and cons.

The main superiority of ANN lies in the capability of non-linear modelling and self-learning and organizing. The model is suitably created based on features selected from the data set by training. Artificial neural networks have the drawback that they are not capable of semantic representation and symbolic reasoning. An ANN works just like a "black-box" approach i.e. the non-straight associations of causes and consequences are not easily explainable, which makes it difficult to interpret results. Also, ANN does not ensure its convergence to global minima and it is also tough to decide about its parameters like the number of nodes, hidden layers, activation function, etc. ANN performs well but they need large data set for higher accuracy results.

Compared with ANN, real-world reasoning is carried out by Bayesian rigorous conception of the approach involved instead of the interrelationship of data and hypothesis, that's why, forms the basis of giving the objective interpretation. The reasoning is possible because it involves evidence and observation to generate the posterior likelihood of any variable. BNs are more easily interpretable by humans and increase the certainty of the correct model for adoption. Due to these features, aviation accident data can be studied and different models can be trained using BNs.

Data mining techniques like SVM, regression, etc. can easily recognize predictable behaviour for future forecasting. Different patterns in data set and errors which may result in losses can be easily identified using data mining. Unlike ANN, data mining also performs poorly under small data set and requires high tuning of a huge amount of hyper parameters, failing to which can lead to the ineffective training process.

Time series analysis is capable of identifying the factors responsible for the fluctuation of series. It does the forecasting based on past behaviour, but the historical data may not provide the real picture of the underlying trend. Different time series methods may produce a big deviation in mid-term and long-term forecasting due to different information about aviation accidents. Also if data is non-linear and
random then we are not able to apply most of the time series methods, because it has to satisfy stationary condition i.e. exponential smoothing. So, there is a need to incorporate microeconomic variables into the time series equation for improving forecasting capability.

Which ensemble method proves to be good depends on the problem at hand but generally; they provide better performance in comparison to individual models with less noisy aggregated results. Ensemble methods may not work well in case of probabilistic reasoning as the Bayesian network does. In aviation accident data analysis, ensemble methods prove to be very efficient having the drawback of lesser inter-ability due to its complicity. They are more suitable for real-time data.

4. Conclusion and future Scope

This manuscript focus on various models used for aviation accident prediction to come up with several ways of enhancement and assist in decision making. Here, various knowledge discovery methods have been outlined, that can identify precursors to aviation safety accidents and incidents for emergency response. Based on these methodologies, we have developed various conclusions.

Model selection by deploying different methods may lead to different models, which makes it difficult in deciding which one is better. So the combination of various methods results in better prediction, performance, which in turn depends on the model parameters and selection process.

Each method has its pros and cons in its applicability to different types of accidents and it is not possible to account for all information of accident data set in forecasting techniques. So, different methods of dealing with different information may be combined to have an outstanding perspective in aviation accident forecasting.

Also, one more drawing indicates that in the future there is a need to focus more on the technique selection and its improvement rather than developing forecasting methods. Due to the heterogeneity of the factors and absence of real-time data of aviation accidents some newer techniques such as genetic algorithms, fuzzy-based techniques, and machine learning techniques, etc., should be implemented to develop an intelligent model which also consider the influencing factors of accidents to find the hidden patterns that may be used for flight training methods and better prediction accuracy heading to lessen the aviation fatalities and injuries.

References
[1] Zhang X and Mahadevan S 2019 Decision Support Systems 11648–63
[2] Accident-statistics://URLhttps://www.icao.int/safety/iStars/Pages/Accident-Statisticse
[3] Asrs program briefing.
[4] Janakiraman V M and Nielsen D 2016 2016 International Joint Conference on Neural Networks (ICCNN) (IEEE) pp 1993–2000 .
[5] Oster Jr C V, Strong J S and Zorn C K 2013 Research in transportation economics 43148–164.
[6] Li Y 2019 Mathematical Problems in Engineering 2019
[7] MAGONO C 1980 Amsterdam, Elsevier Scientific Publishing Co.(Developments in Atmospheric Science. 12.
[8] Janic M 2000 Journal of Air Transport Management 643–50.
[9] Zheng X and Liu M 2009 Journal of Loss Prevention in the process Industries 22484–491
[10] Vald’es R M A, Comendador V F G, Sanz L P and Sanz A R 2018 Safety science 104216–230.
[11] Chen W and Huang S 2014 Transportation Research Record 2449105–113.
[12] Wang H and Gao J 2013 Mathematical Problems in Engineering 2013.
[13] Cheng S Z L, Vald’es R M A, Comendador V F G and Rom’an R 2019
[14] Huang H, Chin H C, Haque M et al. 2008 87th Annual Meeting of Transportation Research Board (TRB) (Transportation Research Board).
[15] DeubleinM, SchubertM, AdeyBT, K’ohlerJandFaberMH2013AccidentAnalysis&Prevention
51274–291.
[16] Brooker P 2011 Safety Science 491142–1155.
[17] MatthewsB, DasS, BhaduriK, DasK, MartinRandOzaN2013JournalofAerospaceInformation
[18] Lali’s A 2017 Transport Problems 12.
[19] GanXS, LiXY, ChenJL and ZhaoHT 2014 Advanced Materials Research vol 1049 (TransTech Publ) pp 1351–1354
[20] Zhang W, Li Q and Yu H 2019 2019 IEEE 1st International Conference on Civil Aviation Safety and Information Technology (ICCASIT) (IEEE) pp 54–57.
[21] Maheshwari A, Davendralingam N and DeLaurentis D A 2018 2018 Aviation Technology, Integration, and Operations Conference p 3980.
[22] Hegde J and Rokseth B 2020 Safety science 122104492
[23] SubramanianSV and RaoAH 2018 2018 AIAA Modeling and Simulation Technologies Conference p 0424
[24] Altay A, Ozkan O and Kayakutlu G 2014 Journal of Aircraft 5147–53
[25] Yan X, Wang H, Fu Q and Zhao J 2019 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) (IEEE) pp 898–906
[26] Srinivasan P, Nagarajan V and Mahadevan S 2019 AIAA Aviation 2019 Forum p 2938
[27] Burnett R A and Si D 2017 Proceedings of the International Conference on Compute and Data Analysis pp 60–68
[28] Ni X, Wang H, Che C, Hong J and Sun Z 2019 Safety science 11290–95
[29] Salama M A, Hassanien A E and Fahmy A A 2010 The 10th IEEE international symposium on signal processing and information technology (IEEE) pp 473–477
[30] Koteeswaran S, Malarvizhi N, Kannan E, Sasikala S and Geetha S 2019 Cluster computing 22 11379–11399
[31] Rao A H and Marais K 2020 Reliability Engineering & System Safety 193106670
[32] Christopher AA, Vivekanandam VS, Anderson AA, Markkaneyanan Sand Sivakumar V 2016 The Aeronautical Journal 1201849–1866
[33] Wang Z, Sherry L and Shortle J 2015 2015 Integrated Communication, Navigation and Surveillance Conference (ICNS) (IEEE) pp W3–1