Anomaly Detection in Flight Data Using the Naïve Bayes Classifier

Murtaja S. Jalawkhan
College of Science, Computer Dept.
University of Baghdad
Baghdad, Iraq
murtagaa95@gmail.com

Tareef K.Mustafa
College of Science, Computer Dept.
University of Baghdad
Baghdad, Iraq
Tareef_alshaibi@yahoo.com

2021 7th International Conference on Contemporary Information Technology and Mathematics (ICCITM) | 978-1-6654-2032-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCITM53167.2021.9677655

***Abstract*—Safety is the key to reliable civil aviation. In the airline industry, there is a growing emphasis on proactive safety management systems in order to improve the safety of current aviation operations. These systems utilize anomaly detection techniques to recognize and reduce the risk of accidents occurring. This work develops a new anomaly detection approach for commercial flight operations using routine operational data to enhance proactive safety management systems and utilizes data mining techniques to identify abnormal situations instantaneously during flights using real-life FDR (Flight Data Recorder) data. The Naïve Bayes classifier was used to detect normal and abnormal situations. This classifier was applied to a dataset of 100 flights and new abnormal situations could be recognized with a high probability of detection and a low probability of false alarm. The results strongly suggest that anomalies detected in a variety of flights can be recognized, which can help airlines with many different approaches, such as the deployment of predictive maintenance, the detection of early signs of performance divergence, safety support, and the training of staff accordingly.**

Keywords— Anomaly detection, Flight data recorder, Data Mining, Naïve Bayes.

I. INTRODUCTION

In the aviation field, fatal accidents have decreased every decade since the 1950s, which is considered to be a significant achievement given the massive growth in air travel since then. Aircraft have become more reliable due to a number of factors, including flight operations with more accurate and detailed weather information, the improvement of navigational aids, and the enhancement of pilot training with the help of flight simulators having become progressively more sophisticated. This has been achieved through the collaborative efforts of governments, airlines, manufacturers and other parties. However, more recent documentation suggests that there has been a decrease in the rate of progress in aviation safety [1].

In recent years, data analytics has gained significant attention in many different industries due to the increasing value of the information and data it provides [2]. This is due to the generation of large amounts of data on each flight as modern aircraft are usually provided with digital FDRs (Flight Data Recorders) [3].

With the rise in the number of sensors and actuators on aircraft, it is logical to take advantage of those FDR measurements with anomaly detection techniques to enhance the safety levels in aviation systems so as to identify early indicators of safety degradation, reveal hidden threats, decide when maintenance should be performed, prevent any potential accidents, and instruct any relevant personnel accordingly [4].

Stakeholders in the aviation industry have endeavored to use different approaches to identify anomalies and evaluate flight safety concerns. One approach in particular is the Event Exceedance Analysis methodology, which focuses on discovering whether a predefined safety threshold has been exceeded and the data is implemented using the Flight Operations Quality Assurance (FOQA) program. These data are captured and recorded from the aircraft either by the Quick Access Recorders (QARs) or the Flight Data Recorders (FDRs). Then the data are downloaded to a computer or another electronic device to be aggregated from the fleet. Finally, the data are analyzed to find abnormal situations in the flight parameters. These parameters involve continuous features (i.e., ground clearance), discrete features (i.e., flap position), binary features (i.e., switch settings), and category features (takeoff and landing airport). A typical FOQA file has several hundred to thousands of parameters recorded at a frequency of at least 1 Hz [5].

It is important to recognize that aviation data present challenges for a variety of reasons, including their large volume, high dimensionality, heterogeneity (mixed categorical and continuous attributes), multi-modality (multiple modes of nominal and non-nominal operations involving a variety of aircraft, airports and airspace), and temporality (long time series) [6]. Additionally, the challenge is expected to become even greater in the future as the number of sensor-equipped aviation systems continues to grow. The findings of the study indicate that a relatively high anomaly score typically corresponds to periods of expected failure. Automatically injecting faults into flight data using recommended software has resulted in abnormal occurrences being recognized as such. This paper proposes a novel strategy in the aerospace field that utilizes a Naïve Bayes classifier and produces very promising results even with massive amounts of data.

The paper is organized as follows: The second section covers the literature review. The third section presents a description of the proposed method in the context of current concepts used in its construction with an emphasis on the unique aspects of this method. In the fourth section, we discuss our experiment performed on NASA's Aviation Safety Information Analysis & Sharing (ASIAS) flight recorder data and examine whether our new classifier produces a superior outcome. Finally, in the fifth section, we summarize our conclusions and future work.

II. LITERATURE REVIEW

A. Related Works

The study of flight anomaly began 21 years ago [7]. Since then, a number of studies on the subject of anomaly detection based on FDR data have already been conducted. Paper [8] proposes a method to extract an advantage in the analysis subject of flight data. The basic algorithm uses a feature-extraction algorithm known as Symbolic Dynamic Filtering (SDF), which was the first attempt to explore the applicability of its current features in real-world flight data analysis. The technique of extracting SDF-based features is easily affected by signal distortions because SDF is able to detect changes in the statistical characteristics of a signal. This work uses an extraction tool with its advantage of detecting anomalies.

A number of researchers [9] have endeavored to detect anomalies in flight trajectories and predict diversions in flight transportation. However, this paper introduces a prediction model that only requires specific information about an airplane (its velocity, position and intended destination). This information is utilized to identify the normal and abnormal behavior of an airplane. The model predicts a diversion when an airplane displays abnormal behavior over a long period of time. It was concluded that this approach, with excellent precision, would be capable of identifying potentially abnormal airplanes and recall them even without any knowledge of their planned trajectories.

Another study [10] applies a GMM (Gaussian Mixture Model) based clustering to digital flight data to identify flights with abnormal data patterns. These abnormal flights may cause an increase in the level of risk under the assumption that normal flights have patterns in common. The results revealed that Cluster AD-Data Sample successfully identified unusual flights with higher threats, making it a helpful tool for airline operators to identify early indicators of safety degradation. In another paper [11], a new approach to anomaly detection is proposed that does not require the hand-crafting of features needed for state-of-the-art PHM algorithms to perform system diagnoses effectively on large volumes of data from arrays of sensors on aircraft. The authors therefore proposed leveraging existing unsupervised learning methods based on Deep Auto-Encoders (DAEs) on raw time-series data from multiple sensors to build a robust model for anomaly detection. The paper demonstrates good fault detection rates (97.8%) with no false alarms with robust fault disambiguation. The RNN (Recurrent Neural Networks) method has been proposed to detect anomalies in in-flight data [12]. RNN with Long Short-Term Memory cells (LSTM) and RNN with Gated Recurrent Units (GRUs) have the ability to handle multivariate sequential time-series data without dimensionality reduction and they can detect anomalies in latent features. The RNNs were able to unmistakably identify 8 of 11 abnormalities in addition to one extra abnormality that was close to being identified. The authors of [13] developed a method that integrates the recent advancements in deep generative models to create an unsupervised machine learning algorithm known as the Convolutional Variational Auto-Encoder (CVAE), which takes observed operations and identifies anomalies for important flight operations in high-dimensional, heterogeneous aviation time series. This algorithm has been tested in the aviation domain and has been proven to surpass existing anomaly detection techniques by a margin of 48 pp in precision and 29 pp in recall. The increasing visibility of previously unknown vulnerabilities translates into an increased risk of an incident or accident occurring if these issues are not monitored and remedied.

B. Naive Bayes Classifier

The Naïve Bayes model is an extremely simplified Bayesian probability model [14]. The Naïve Bayes classifier functions on a powerful independence assumption, which means that the probabilities of the attributes do not affect each other. The Naïve Bayes classifier makes $2n!$ independent assumptions for a given series of n attributes and the results are often correct. As revealed in [14], it determines which of the circumstances affect the Naïve Bayes classifier and concludes that the error is an outcome of three main elements, namely training data noise, bias and variance. Training data noise may be reduced only by selecting appropriate data for training, which can be done by dividing the data into many different groups using a machine learning algorithm. While bias is the error caused by the very large groupings of the training data, variance is the error caused by the very small groupings of these data.

C. Anomaly Detection

Anomaly detection can be defined as the detection of abnormal data points or observations that differ from the normal behavior in a dataset. It is also named fault detection, deviation detection, and outlier detection depending on various application fields. A very large number of abnormal identification methods have been developed. Some are general and are able to solve diverse application problems, and many are specific to solve specific problems in a specific application field [4]. To determine the most suitable anomaly detection technique, a number of items need to be taken into consideration as there is no foreknowledge of the abnormalities in the flight data (if there are any) and the obtained data from routine operations are not labeled. Accordingly, only semi-supervised (or unsupervised) learning methods can be utilized although, even with the use of these methods, the number of abnormalities is presumed to be a small fragment of all the data in the FDRs [15].

D. Flight Data

Flight data can be explained as being digital FDR data having originated through aircraft operations. Each airplane has sensors that measure the parameters of a flight which usually include airspeed, altitude, accelerations, thrust. The flight parameters numbers may reach 2,000 relying on the recording abilities of modern airplanes (with an average of more than 300 flight parameters). Then these parameters are recorded on-board during every commercial flight by the digital FDR or by the QAR (Quick Access Recorder). Finally, the airlines collect the recorded data regularly to detect risks and evaluate daily operations by analyzing this data [16]. The recorded data on the digital FDR or QAR are sampled at individual rates. The identification of these data is recorded and the sampling rates differ by two factors, namely by the kind of recorder being used and by the configuration needs of the airline [16]. The complexity and size of the flight data creates challenges in the data analysis field.

III. METHOD

In Bayesian classification, there is an assumption that given data are a part of a particular class. Then the probability of that assumption is calculated to determine whether or not it is true [17]. The method requires one scan only for all the data and if there are extra training data at some point, then each training example can gradually increase or decrease the probability of an assumption to be true. This is considered to be among the most functional methods for specific kinds of problems [17].

In this section, the proposed method has been detailed in the following steps Fig. 1:

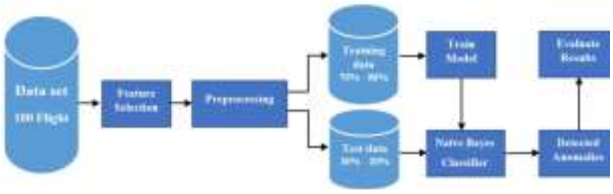

Fig. 1. Block diagram of the proposed system.

A. Pre-Processing

To make data more suitable for analysis by data mining algorithms, we simplify the disruption of data. Firstly, each sensor on the aircraft has a different recording rate, between 0.25 to 16 records/sec., which means that each parameter has a different number of records for the same flight. To unify the record numbers, we take one record each second and disregard the remainder. In the case of a recording rate of less than one, we duplicate records until we have one record per second. Secondly, the data are normalized to remove the distribution of numbers as a result of the differentiation of sensors. We transform the data into an ordinal form to make it more convenient for its implementation into the classifier. To perform a linear transformation on the original data, we use the Min-Max normalization.

B. Training

In the training phase, we take 80% of the dataset to train our model and apply the Naïve Bayes algorithm. Assuming that our training set includes n samples in xi, i = 1…n, which contains p attributes, i.e., xi = (xi1,xi2,…,xip). Each sample is supposed to be associated with only one class, y ∈ {y1,y2,…,yc}.

Naïve Bayes learning represents the construction of a Bayesian probabilistic model that assigns a posterior class probability to a sample: $P(Y = y_j | X = x_i)$. Then, we calculate the prior probability for each given class label, $P(y_i)$ = number of y_i / n. The simple Naïve Bayes classifier uses these probabilities to assign a sample to a class, as in "(1)" [18]:

$$P(y_j \mid x_i) = \frac{P(x_i \mid y_j)\, P(y_j)}{P(x_i)} \tag{1}$$

C. Test

In this approach, assuming that we have one record per second, we test each flight record to determine any anomalies. We assume that the individuals x_i are unconnected to each other. This is a powerful belief that is disregarded in most experimental applications and is accordingly naïve (thus its name). This belief suggests that $P(x_1 | x_2, x_3, \ldots, x_p, y_j) = P(x_1 \mid y_j)$, for example. Hence, the joint probability of x and y_j is " (2)" [18]:

$$\begin{aligned} &P(X|y_j)P(y_j) \\ &= P(x_1|y_j)\,.\,P(x_2|y_j)\,.\,.\,.\,P(x_p|y_j)\,P(y_i) \\ &= \prod_{k=1}^{p} P(x_k|y_j)\,P(y_i) \end{aligned} \tag{2}$$

In our study, we have only two classes, either True, which means that the record is anomalous, or False, which means we have a normal instant.

D. Performance Testing

This phase uses a confusion matrix to compute accuracy measurements. The result is an indicator that has the value of accuracy, precision, and recall. This measure is calculated from different variables, namely TN (True Negative), TP (True Positive), FN (False Negative) and FP (False Positive). In "(3)" to "(6)" are utilized to calculate system performance based on the previous variables.

$$\text{accuracy} = \frac{TN + TP}{FN + TN + FP + TP} \tag{3}$$

$$\text{precision} = \frac{TP}{(FP + TP)} \tag{4}$$

$$\text{recall} = \frac{TP}{(TP + FN)} \tag{5}$$

$$\begin{aligned} \text{F-score} &= 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \\ &= \frac{TP}{TP + \frac{1}{2}(FP + FN)} \end{aligned} \tag{6}$$

Accuracy can be defined as the most instinctive measure of performance and the ratio of accurately predicted observations to the total observations, while precision indicates the ratio of accurately predicted anomaly observations to the total predicted anomaly observations. Recall is the ratio of correctly predicted anomaly observations to all observations in the actual class. Finally, F1-Score is the harmonic mean of precision and recall [9].

IV. EXPERIMENTAL RESULTS

The experimental tests were applied to FDR data obtained from NASA Aviation Safety Information Analysis and Sharing. The data contains routine operational data where faults can be injected manually or automatically (using software) [19]. First, we utilized 30% of all the normal flights and we randomly injected faults into 10% of the data record of each flight. In the training stage, we employed the Naïve Bayes algorithm for 80% of the whole dataset. While in the testing stage, we used another 20% of the flights which have 10% of anomalous records for testing. Finally, we measured the system performance as explained above.

The entire dataset has 500,812 data records from 100 individual flights, among which 488,611 data records are normal and the other 12,201 had been fault-injected to be labeled as anomalous. For each record, there are 186 sensors and some of the outputs of these sensors are continuous and others are Boolean, with the average flight length being around 2.5 hours. Nevertheless, we adopted a subset of the flight parameters selected by the expert domain, such as id: time interval, LONP: LONGITUDE POSITION LSP, LATP: LATITUDE POSITION LSP, GS: GROUND SPEED LSP, TAS: TRUE AIRSPEED LSP, ALT: PRESSURE ALTITUDE LSP, PS: STATIC PRESSURE LSP, HDGS: SELECTED HEADING, PTCH: PITCH ANGLE LSP.

When the fault injection process is used, some of the sensors will have an abnormal output compared to their normal output. Fig. 2 and Fig. 3 show the fault injection process at around 33 min. to 50 min. followed by the output of the sensors indicating an abnormal change in the Altitude (ALT) and Ground Speed (GS).

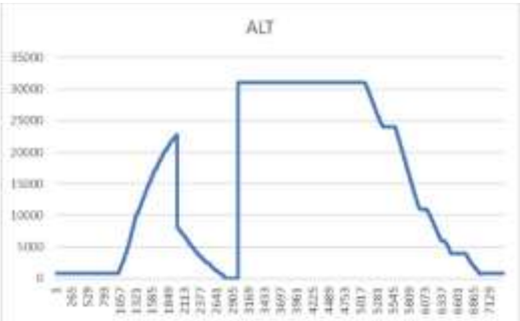

Fig. 2. Airplane Altitude.

Fig. 3. Ground Speed.

The model was coded in the C#.Net environment and trained on a laptop computer with an Intel i7-5500U processor (2.40 GHz base core clock) and 8 GB DDR4 memory running on a Windows 10 (64-bit) operating system. It took 7.13 minutes to train the model with 80 flights as training data and 3.15 minutes to test all 20 test flights (on average 9.4 seconds for each flight) using the aforementioned hardware. Table I shows the results obtained for the measurement of accuracy, precision, recall, and F1-score, where we aimed to achieve the best F1-score.

TABLE I. PERFORMANCE MEASURE WHEN THE FAULT INJECTION WAS APPLIED ON 25%, 50%, AND 100% OF THE SELECTED FEATURES

Performance Metrics	Ratio of Fault Injected Feature		
	25%	*50%*	*100%*
Accuracy	95.9%	99.06%	99.39%
Precision	99.78%	100%	100%
Recall	59.12%	90.61%	93.97%
F1-Score	68.09%	93.79%	96.36%

The results clearly show that the higher the percentage of parameters that are fault injected, the higher the performance metrics obtained. All indicators were great within 50% and 100% of the fault injection ratio. We obtained good precision and accuracy even with 25% and we had excellent precision in all cases that was not less than 99.78% with very high accuracy reaching 99.96%. The minimal result obtained was within the 25% fault injection ratio for recall and F1-score, which at minimum amounted to a 68.09% F1-score and 59.12% recall. Compared to other classifiers [6], [11], the Naïve Bayes classifier has a significant impact on enhancing detection accuracy in aircraft operations.

TABLE II. COMPARISON BETWEEN OUR MODEL AND OTHER CLASSIFIERS

All classifiers		Precision	F1-score
DAE		97.8%	-
Fcae_30_20_10		97.1%	90.2%
Cae_30_20_10		93.4%	87.5%
Lstmae_pca_10		80.5%	79.4%
Lstmae_pca_30		85.8%	83.7%
Proposed method in 3 cases	100%	100%	96.36%
	50%	100%	93.79%
	25%	99.7%	68.09%

In Table II, we present a comparison between our model and other classifiers from papers [6] (DAE) and [11] (Fcae_30_20_10, Cae_30_20_10, Lstmae_pca_10, Lstmae_pca_30). The comparison shows clearly the improvement in the detection accuracy by the proposed model within a reasonable execution time.

V. CONCLUSION

In the current work, with a complete inspection of the historical flight sensor data, a classification-based abnormal identification model has been introduced to determine both anomalous and normal situations. After composing the abnormal identification as a classification problem, flight phase-based features on relevant flight sensors were selected based on domain expertise. There are many challenges we face when dealing with FDR data, such as the large data volumes, lack of labeled data, and the increasing numbers of sensors (multiple modalities) exacerbating the challenges of being able to hand-craft the features needed for the algorithm to perform system diagnoses effectively. The Naïve Bayes classifier was used to classify anomalous and normal situations. We demonstrated that Naïve Bayes functions best in two cases, the first having been (as expected) with completely independent features and the second (which was not expected) with functionally dependent features. The experimental results on NASA's datasets show that the proposed method for abnormal identification can successfully accomplish the main goal of our work with high scores in the evaluation measurements. Our results appear to be very promising, even for classes that are difficult to detect. For future work, we suggest applying the method to a real-time system with monitoring simulations and visualizing the results for users.

REFERENCES

[1] T. Balcerzak, and K.J.R.e.d.d.d.l.n.m.y.a. Kostur, Flight Simulation in Civil Aviation: advantages and disadvantages. 2018(35): p. 35-68.

[2] C.H. Lee, et al. Anomaly detection of aircraft engine in FDR (flight data recorder) data. in IET 3rd International Conference on Intelligent Signal Processing (ISP 2017). 2017. IET.

[3] ATR, Flight Data Monitoring: On ATR Aircraft. 2016.

[4] Li, L., Anomaly detection in airline routine operations using flight data recorder data. 2013, Massachusetts Institute of Technology.

[5] K. Sheridan, et al. An application of dbscan clustering for flight anomaly detection during the approach phase. in AIAA Scitech 2020 Forum. 2020.

[6] L. Basora, et al., Aircraft Fleet Health Monitoring with Anomaly Detection Techniques. 2021. 8(4): p. 103.

[7] M.Y. Pusadan, J.L. Buliali, and Ginardi, R.V.H.J.P.C.S., Cluster Phenomenon to Determine Anomaly Detection of Flight Route. 2019. 161: p. 516-526.

[8] S. Das, et al. Anomaly detection in flight recorder data: A dynamic data-driven approach. in 2013 American Control Conference. 2013. IEEE.

[9] Di Ciccio, C., et al., Detecting flight trajectory anomalies and predicting diversions in freight transportation. 2016. 88: p. 1-17.

[10] L. Li, et al., Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring. 2016. 64: p. 45-57.

[11] K.K. Reddy, et al. Anomaly detection and fault disambiguation in large flight data: a multi-modal deep auto-encoder approach. in Annual Conference of the PHM Society. 2016.

[12] A. Nanduri, and L. Sherry , Anomaly detection in aircraft data using Recurrent Neural Networks (RNN). in 2016 Integrated Communications Navigation and Surveillance (ICNS). 2016. Ieee.

[13] M. Memarzadeh , B. Matthews, and I.J. A. Avrekh, Unsupervised anomaly detection in flight data using convolutional variational auto-encoder. 2020. 7(8): p. 115.

[14] S. Mukherjee, and , N.J.P.T., SharmaIntrusion detection using naive Bayes classifier with feature reduction. 2012. 4: p. 119-128.

[15] T.G. Puranik, and D.N.J.J.o.A.I.S., Mavris, Anomaly detection in general-aviation operations using energy metrics and flight-data records. 2018. 15(1): p. 22-36.

[16] S. Das, S., et al. Comparison of algorithms for anomaly detection in flight recorder data of airline operations. in 12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. 2012.

[17] D.K. Ahirwar, et al., Anomaly detection by Naive Bayes & RBF network. 2012. 1(1): p. 14-18.

[18] S. Ranganathan, K. Nakai, C. and Schonbach, Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics. 2018: Elsevier.

[19] X. Zhou, Y. Zhong, L. and Cai, Anomaly detection from distributed flight record data for aircraft health management. in 2010 International Conference on Computational and Information Sciences. 2010. IEEE.