LITERATURE REVIEW OF MACHINE LEARNING TECHNIQUES TO ANALYSE FLIGHT DATA

S Kumar Jasra[1], J Gauci[1], A Muscat[2], G Valentino[3], D Zammit-Mangion[1] and R Camilleri[1]

(1) Institute of Aerospace Technologies, University of Malta, Msida, MSD 2080, Malta
(2) QuAero Ltd., 65/67 Triq It-Torri, Mosta MST 3503, Malta
(3) Department of Communications and Computer Engineering, University of Malta, Msida, MSD 2080, Malta

KEYWORDS: Flight Data Analysis, Machine Learning, Clustering, Anomaly Detection, Feature Selection

ABSTRACT:

This paper analyses the increasing trend of using modern machine learning technologies to analyze flight data efficiently. Flight data offers an important insight into the operations of an aircraft. This paper reviews the research undertaken so far on the use of Machine Learning techniques for the analyses of flight data by evaluating various anomaly detection algorithms and the significance of feature selection in Flight Data Monitoring. These algorithms are compared to determine the best class of algorithms for highlighting significant flight anomalies. Furthermore, these algorithms are analyzed for various flight data parameters to determine which class of algorithms is sensitive to continuous parameters and which is sensitive to discrete parameters of flight data. The paper also addresses the ability of each anomaly detection algorithm to be easily adaptable to different datasets and different phases of flight, including take-off and landing.

1. INTRODUCTION

1.1. Flight Data Monitoring

Flight Data Monitoring (FDM) is an activity carried out by airlines primarily as a means of monitoring and improving the safety and operation of their aircraft [1]. The data recorded by the flight data recorder on-board an aircraft is downloaded and analyzed through various tools and techniques, with the ultimate objective of using that analysis to improve civil aviation operations, establishing maintenance schedules, training pilots and modifying operational procedures amongst others, without compromising safety [2]. In this way FDM has two key objectives:

1. It detects technical flaws, unsafe practices or conditions outside desired operating procedures at an early stage, thus preventing potential incidents or accidents [3].

2. It provides an objective means of following-up on corrective actions by increasing training or altering Standard Operating Procedures [4].

The benefits of FDM are often highlighted by accident investigations. One such classical example is of Gulf Air's A320 flight GF072 accident on the 23rd. of August 2000 near Bahrain International Airport. The final report of the Accident Investigation Board (AIB) in summary stated that at the time of the accident, the flight data analysis system was not functioning satisfactorily. Non-availability of flight data analysis deprived the airline of a valuable safety analysis tool [5].

Once the flight data is downloaded from the aircraft, the entire flight duration known as a timeline is analyzed. Traditionally, FDM involves the use of statistical techniques to analyze data. An airline analyses flight data through dedicated software which highlights if certain parameters of flight data have exceeded pre-determined limits. This approach is based on known 'exceedances' i.e. predefined issues [6]. Such exceedances in the timeline are known as events. Exceedance or event detection is the standard FDM algorithmic methodology that checks the data for deviations from flight manual limits, SOP's and good airmanship. This relies on human experts to create a rule-based system that detects known safety issues, based on whether a small set of parameters exceed some predefined set of thresholds. A single event is a collection of multiple snapshots. A snapshot is a sample of a flight data parameter taken at a point during the flight. A collection of snapshots taken at the same point in time will create a vector which describes the condition of the aircraft at that point in time. Examples of events include:

1. High descent rate below an altitude of 400ft
2. Low approach speed below an altitude of 500ft
3. High unstick speed
4. Maximum operating altitude exceedance
5. Late selection of flaps to landing configuration
6. Exceedance of maximum operating altitude
7. Engine over-temperature
8. Excessive bank angles

9. Deviation from glidepath
10. High pitch of the aircraft during take-off

There can be over 100 such events defined by an airline and these must be checked for every flight of every aircraft.

Detecting an event is a laborious process. However, with developments in computational techniques, other avenues could be explored for flight data analysis. For example, in finance and online gaming, Machine Learning (ML) techniques have been major drivers in the development of these industries. ML is an application of Artificial Intelligence (AI) by which computer programs are designed to access data, discover patterns, learn and improve from experience, without further direct human intervention [7]. Such computational techniques would be very effective for the aviation industry.

1.2. Role of ML in FDM

The aviation industry operates at a tight financial margin [8,9]. Despite the fact that air travel continues to grow, the aviation industry remains susceptible to external factors such as oil prices [10]. Hence, the efficient analyses of flight data can improve flight operations, reducing fuel consumption as well as maintenance and insurance costs while increasing the level of safety [11]. However, the traditional statistical approaches for the analyses of flight data, which rely on pre-defined thresholds, do not provide enough information and are tedious. This limitation can be addressed by adopting ML techniques. These techniques can compare flight data parameters from a large number of flights and identify new or unknown patterns [12]. These patterns may show abnormal or inconsistent behavior with respect to most of the flights. The outliers are of interest and require further investigation. One example of an abnormal flight pattern is during the descent phase where the aircraft is not following the standard procedure for a stable approach where by the landing gear is not down by 1000 feet of altitude. Another example of an abnormal flight could be excessive pitch of the aircraft during take-off. With traditional methods of flight data analysis, the flight is flagged red if the aircraft's pitch crosses a specific threshold but with ML algorithms the abnormal pitch could also be detected as shown in Fig. 1.

Modern ML algorithms are best suited not just to classify flights as safe or unsafe, but also to further analyze flight details and explain in more detail the reasons for flights being unsafe. With the development of ML algorithms, the analysis of flight data is becoming more efficient and more suitable to make predictions from flight data [14, 15]. The following sections describe the ML techniques that have been used for FDM. In the last part, the conclusion is presented, and future work is discussed.

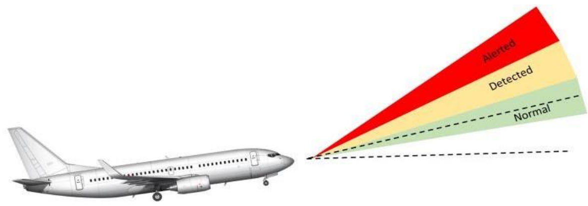

Figure 1. Aircraft pitch at take-off [13]

2. FEATURE SELECTION

Flight data consist of a large number of flight parameters, for example, altitude, computed air speed, etc. These parameters are crucial to analyze any flight data or to detect an anomaly. However, not all these parameters are equally important, and some parameters are more important and are thus recorded at high frequency. Currently in the aviation industry, experts decide on parameters which are important for FDM. Therefore, selecting the right kind of parameters has always been considered as a matter of a judicious choice.

A poor prior selection of parameters would lead to an inefficient FDM analysis. A better method is to have a learning algorithm with a capability to choose which set of parameters are the best for a particular scenario. This will help to create an FDM analysis which is capable of coping with (and adapting to) different types of data. Prior selection of parameters may bias the learning process and lead to worse anomaly detection performance when compared to leaving the choice to the machine i.e. the learning algorithm.

The process of automatic selection of attributes from the given data by the learning algorithm itself is known as Feature Selection (FS). Theodoridis and Koutroubas [16] define feature selection or feature reduction as the procedure in which, given a number of features, one selects the most important of them. The goal is to reduce the number of parameters, and at the same time, retain as much of their class discriminatory information as possible. More general methods that create new features based on transformations or combinations of the original feature set are termed Feature Extraction (FE) algorithms [17]. This procedure can reduce not only the cost of recognition by reducing the number of features that needs to be collected, but in some cases, it can also provide a better classification accuracy due to finite sample size effects [17]. FS thus helps in generalizing performance with more computationally efficient methods and identifying critical features.

The field of FS has a lot of potential for FDM, but it has not been explored much. A literature research uncovered only two major prior works [18, 19] which use this technique for FDM. In [18] the FE algorithm called Symbolic Dynamic Filtering (SDF) [20] is used, whereas [19] uses scalar feature selection, vector feature selection and Kalman filtering.

In [18], SDF, a robust time-series feature extraction tool for enhancement of performance of pattern classification is used. It involves four major steps:

1. Generating symbol sequences via portioning of the time series data sets.
2. Constructing Probabilistic Finite State Automata (PFSA) from the respective symbol sequences.
3. Extracting of features as probability matrices or as state probability vectors from PFSA.
4. Pattern classification based on the extracted features.

Algorithms, constructed in the SDF setting yielded superior performance in terms of early detection of anomalies and robustness to measurement noise in comparison with other existing techniques such as Principal Component Analysis (PCA), Neural Networks (NN) and Bayesian techniques [21]. SDF based feature extraction technique is sensitive to signal distortions and at the same time robust to measurement noise and spurious signals. It is also adaptable to low-resolution sensing due to the coarse graining in space partitions.

In [19], scalar feature selection processing helps to select features identified by the ambiguity function and the correlation coefficient method. The ambiguity function is used to examine the overlap of the features, thus helping to reduce the number of features and the correlation coefficient is used to identify highly correlated features. Features which are highly correlated are replaced with most significant feature. For two flights from different classes (e.g. normal and anomalous), a parameter preferably having a normal distribution is taken. Then it’s mean, and variance are taken for those two flights. Statistical tests such as the t- test are applied to this feature. After the test, if the two means are in the same interval of significance, they are considered equal and thus that feature does not provide any discriminatory information between the two flights and is dropped.

The reduced set of features is then passed to vector feature selection process. The objective of this processing is to select the best features that yield the maximum discrimination between the set of flights. All the combinations of features are tried using methods such as Backward & Forward selection, Floating Sequential search and Branch & Bound. A detailed discussion of these methods is beyond the scope of this paper. The Kalman filter is used to determine the importance of each selected feature to discriminate between flights by characterizing low probability density function valued samples as unusual.

Since the work done in implementing FE algorithms for the FDM is limited, the full potential of FE algorithms is yet to be explored by experimenting with PCA, NN and Independent Component Analysis (ICA) for FDM. The work in [18] has shown that SDF has significantly improved the anomalies detected from real-life flight data.

Having described the techniques to choose our feature set, the following section focuses on using the features to detect anomalies.

3. ANOMALY DETECTION

Anomaly detection or outlier detection refers to the task of identifying new or unknown patterns which, in many cases, are abnormal or inconsistent with the rest of the data set [18]. An outlier is an observation that deviates too much from other observations. As compared to the exceedance-based approach mentioned earlier, outlier or anomaly detection technique is more effective as it is not dependent on human generated rules and can even detect previously unknown issues.

Furthermore, the exact definition of an outlier depends on the context. Definitions fall roughly into five categories [22]:

1. Distribution-based, where outliers are observations which deviate from a standard distribution.
2. Depth-based which relies on the computation of different layers of k-d convex hulls.
3. Clustering-based methods, which define outliers as observations that do not fit in the overall clustering pattern.
4. Density-based methods, which detect outliers as objects that are in a less-dense region of the feature space than the rest of the dataset. Objects can be outliers relative to their neighborhoods, particularly with respect to densities of the neighborhoods.
5. Distance-based methods, which define outliers as an observation that is some mini mum distance away from a certain percentage of observations in the dataset.

Outliers can be detected using supervised, unsupervised, as well as semi-supervised ML techniques. Supervised techniques make use of a dataset to train the models. The learning algorithm produces an inferred function from the input and correct output. One example of this technique is any automatic speech-to-text system in a mobile phone, which trains itself to recognize the user's voice. On the other hand, unsupervised techniques allow learning algorithms to infer a function which describes a hidden structure from unlabeled data. The algorithm does not figure out the right output, but it uncovers patterns in the data. For example, customer segmentation based on customer purchasing history and behaviors helps to create a company's customer base. The customer data can be segmented into clusters using unsupervised learning that can then be further analyzed to identify certain associated patterns. This leads to a more customized customer approach and is core to customer satisfaction as well as retention.

Semi-supervised techniques generally use a small amount of labelled data and a large amount of unlabeled data. The algorithms based on these techniques help to considerably improve the learning accuracy of a system. One example of the application of a semi-supervised technique is the prediction of future stock values on the basis of a few past values.

For the analysis of flight data, the authors of this paper have studied- work done in the application of the above methods only in the field of FDM. The observations found are discussed in the following sections and the concluding remarks about the above methods can be found in Section 4.

3.1. Distribution-based & Depth based methods

In distribution-based methods, a standard distribution (e.g. Normal, Poisson, etc.) is used to fit the data set. Outliers are defined based on the probability distribution. The biggest drawback of this method is that most of the distributions used are univariate i.e. dependent on one variable [23]. In the case of FDM, the underlying distribution is unknown and flight data has a large number of parameters. Therefore, identifying the underlying distribution or fitting the flight data with a standard distribution is a costly as well as time-consuming process, and hence infeasible.

On the other hand, in depth-based methods, each object is represented as a point in k-d space and is assigned a spatial depth. However, when applied to outlier detection, this approach becomes inefficient for large datasets [23].

These statistical methods are becoming outdated with the onset of ML techniques. The major reason for this is the volume of flight data the vast number of features and the computing power required. Therefore, the authors of this paper advocate the use of ML techniques for FDM as they are data-driven and require no prior assumptions about the underlying relationships between the variables as in the case of statistical modelling. ML is more efficient on high volume data sets as more data makes the prediction more accurate.

3.2. Clustering-based & Density based methods

Clustering is an unsupervised ML technique in which the data points are assigned to subsets or groups (called clusters) based on similarities between parameters. It has been commonly used for statistical data analysis where the parameters are predefined [24]. In clustering, the algorithm itself groups the data points. Objects with common features are grouped together into clusters. When applying clustering techniques to flight data, one can assume that the majority of flights are safe and can be classified as such. It can therefore be determined that those flights which do not fit the normal flight pattern would be unsafe and further investigation is required to determine the reasons for these abnormalities [25].

Expert domain knowledge has shown that most incidents occur, or have precursors, during the descent phase of the flight. From the statistics presented in Fig 2., it can be seen that Approach and Landing Accidents (ALAs) account for more than 50% of all the accidents [26]. This is the major reason behind the focus of research work on approach and landing.

In a case study presented in [25], partitional and density-based clustering technique are used to identify and compare anomalies during the approach phase of several flights. Flights that differ from the norm follow a different approach pattern than usual and need to be flagged for further investigation. Clustering techniques were applied to 100 data points from six aircraft of the same type with approaches to the same airport runway. The data was de-identified to remove any implications to any pilot, airline, airport or aircraft manufacturer. To validate the usage of clustering, the author in [25] uses the unstable approach as the known safety incident because much work has been done to identify unstable approaches. The author defines a stable approach through a number of aircraft parameters which should stable at a predetermined setting by the time the aircraft is at an altitude of 1000 feet above the landing runway threshold (Instrument Flight Rules (IFR)) prior to landing. The aircraft should be on the glide path and with a proper air speed, with a stable descent rate and engine power setting, and configured correctly for landing at this predetermined point on the approach (1000 feet above the landing runway threshold). The parameters are shown in Tab 1. The author investigates flight data parameters from altitudes of 12,000 feet to touchdown during the descent phase of the flight.

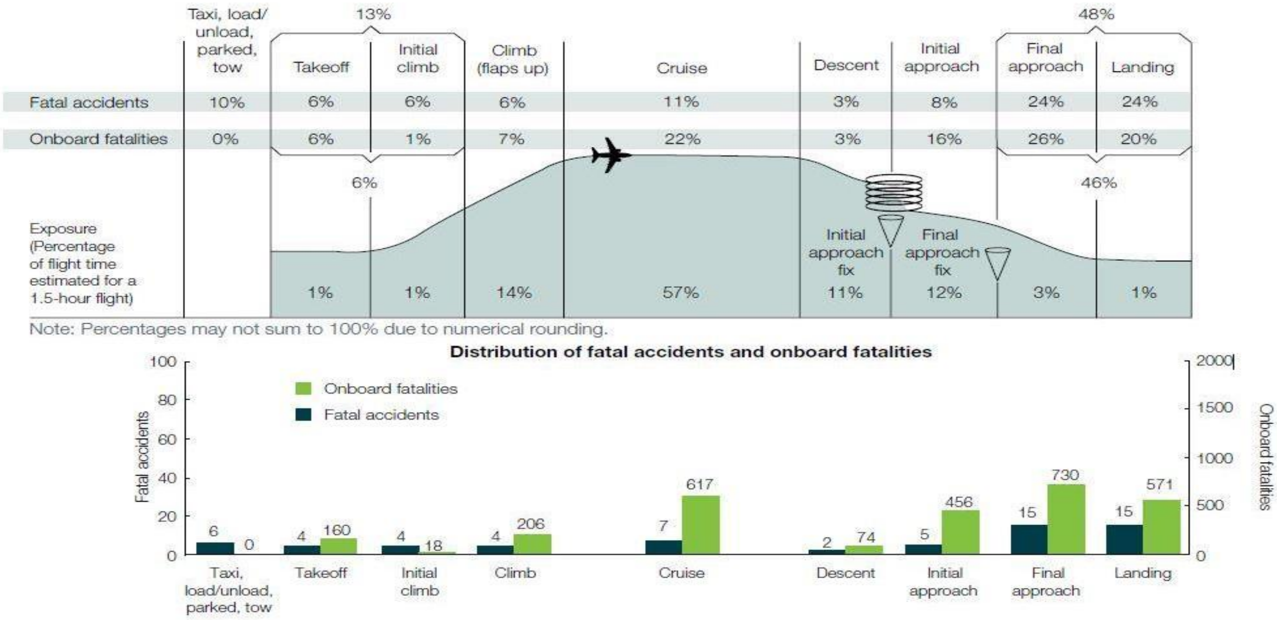

Figure 2. Percentage of fatal accidents and onboard fatalities [26]

The author in [25] uses the partitional method of clustering. This method applies a degree of membership to every object in the dataset and iteratively changes object memberships in order to solve an objective function. Partitional based clustering can be classified as Soft clustering (overlapping clustering) or Hard clustering (or exclusive clustering). With soft clustering, fuzzy sets are used to cluster data, so each object belongs to two or more clusters with different degree of membership. In contrast with hard clustering, objects are grouped in an exclusive way i.e. if an object belongs to a definite cluster then it cannot be included in another cluster [27].

Table 1. Unstable approach parameters [25]

Approach Rule	Recorded Parameter	Unit
Established on the glide path	Glideslope deviation	Dots
Proper air speed	Indicated Airspeed	Knots
Stable descent rate	Inertial Vertical Velocity	feet/minute
Stable engine power setting	Engine 1 N1 Speed	%
Proper landing configuration	Landing gear down and Flap Configuration	Degrees

In [25] k-means and k-medoid are used as hard partitional clustering algorithms, while the Fuzzy c-means (FCM) algorithm is used as a soft partitional clustering algorithm. In addition to this, Fuzzy clustering by Local Approximation of Memberships (FLAME) is also used as a soft portioning clustering algorithm. A short discussion on this work is presented in following section.

In the k-means clustering algorithm, each of the k clusters is represented by the mean or weighted average of its points. K-means is based on the centroid of the cluster whereas k-medoid is based on the medoid of the cluster. In [25], the author finds that in case of k-medoid as the value of k increases there is clear segregation between good and bad clusters. Also, FCM algorithm determines clusters almost similar to k-means algorithm. On the other hand, the FLAME algorithm shows more clear clusters as it involves two steps.

1. It first creates a k-nearest neighbors' graph to identify objects with the highest local density i.e. objects in the same neighborhood.
2. It then identifies objects with a local density lower than a threshold (outliers).

Fuzzy memberships are then assigned to objects with varying degree of memberships only to objects in regions of highest local density and not to objects in regions of lower local density. Therefore, in FLAME a partitional-based clustering technique is mixed with a density-based clustering technique i.e. only spatial proximate objects belonging to the same cluster. This mixed method of clustering produces better and more clearly segregated clusters. This method is also more suitable for real-life datasets as it can adapt to different types of data sets.

Cluster based Anomaly Detection (ClusterAD) [14, 15, 28] is based on the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. The first step of the algorithm is to transform data from the Flight Data Recorder (FDR) into high-dimensional vectors, which capture the multivariate and temporal characteristics of each flight. In a second step, the dimensions of the aforementioned vectors are then reduced to address issues related to data sparseness and multicollinearity. The third step is to apply cluster analysis on the aforementioned vectors of reduced dimensions. Groups of proximate vectors are clusters, or the common patterns; standalone vectors are anomalies, or abnormal flights. This algorithm was applied in [14] to analyses the flight phases of takeoff and final approach. It automatically determined the number of clusters, and found clusters based on a density criterion in the data and found outliers in the feature space.

3.3. Distance-based methods

In distance-based methods, outliers are the points farthest away from other points. In [29], a distance-based outlier detection technique is proposed based on the idea of nearest neighbors and is the most popular method in this category.

The work in [18] uses iOrca [30], a scalable version of the Orca developed by [31] on real-world data sets of a commercial passenger jet airline for the descent phase of the flight. Orca is a k-nearest neighbor based unsupervised anomaly-detection algorithm. Orca uses a nested loop structure to calculate pairwise distances between data points. In [30], the authors introduced a novel indexing strategy and an early termination criterion to make Orca scalable to extremely large data sets. The indexed version of Orca is known as iOrca.

The work done in [18] is very important for FDM as this is the first time that, a feature extraction technique-SDF was used, and for the first time, anomalies were ranked on the basis of a score. iOrca with SDF reported more anomalies than when compared to the run with temporal features, as the SDF-based features were more unique in representing anomalous states of individual flights. Furthermore, [18] shows that the distance-based anomaly detection technique is more efficient than the exceedance-based technique.

However, the distance-based methods have exponential time complexity since every data point needs to be compared to every other, in order to find the nearest neighbor [6]. Since the volume of flight data is huge and the number of parameters affecting the data set is large, this method in its crude form is not suitable for FDM due to its time complexity.

3.4. Other Unsupervised Methods

Multiple Kernel Anomaly detection (MKAD) [32], combines the strength of both vector space-based algorithms as well as sequential anomaly detectors to detect a variety of anomalies from heterogenous data sources. MKAD is a multiple kernel learning approach to detect anomalies in complex heterogenous systems such as FDM, that involves various data sources and data structures. MKAD is based on classical one-class Support Vector Machines (SVMs) [33] which is an unsupervised learning method that finds a set of outliers using a decision boundary. Given a dataset that has a probability distribution of P, one-class SVM aims to find a subset of the dataset S such that the probability that a test point from P lies outside S is bounded by some value. The limitation of MKAD is that it assumes one type of data pattern for normal operations, which is not always valid in real operations, since standards vary according to flight conditions.

The work in [14] compares the performance of ClusterAD with those of MKAD, as well as with exceedance-based detection. Results showed that both ClusterAD and MKAD were able to identify operationally significant anomalies, surpassing the capability of exceedance-based detection. ClusterAD performed better with continuous parameters and earlier known safety issues, whereas MKAD was more sensitive towards discrete parameters. Also, MKAD supported more heterogeneous data sets and both MKAD as well as ClusterAD were able to detect anomalies across a fleet of aircraft. In the future, a combination of these two methods can be a powerful algorithm for FDM.

The work in [6] compares the performance of iOrca with that of MKAD algorithm. Both these algorithms output anomaly scores and the location of the anomalies. MKAD characterizes whole flight as normal or abnormal, whereas iOrca points out at which point in time the abnormality has been found. Thus, MKAD is better for reporting anomalies at fleet level and iOrca performs better to detect anomalies within a flight.

Furthermore, in the work done in [34] a Self-Organizing Map Neural Network (SOM NN) model of anomaly detection in aircraft operation is compared with a one-class SVM model. SOM NN is an unsupervised, two-tier neural network consisting of input layers and output layers. The input layer consists of vectors of data which needs to be structured and the output layer represents the organized version of the data. SOM helps in transforming data points originally in high dimension into 2D space. This helps in further analysis and structure of the dataset can be viewed. Thus, SOM is used to cluster the data points. From the clusters, anomalous data points can be discovered.

The result from the work [34] shows that one-class SVM performs better as the threshold (boundary) used allows for more data to be classified as an anomaly. The work also highlights the role of available computing power while deciding upon the model. SOM NN improves its algorithm complexity with less computational resources. If good computing power resources are available (e.g. supercomputer) then one-class SVM is the better choice.

4. CONCLUSION AND FUTURE WORK

Until now, FDM has not exploited ML techniques fully. ML techniques have the potential to significantly improve the analysis of flight data by discovering unknown patterns. The application of ML to FDM has been very limited and mainly been concerned with the descent phase. Traditional FDM analysis is limited to detect only pre- programmed events, which require knowledge of a particular scenario (usually lessons learnt from historic flight safety incidents). However, ML also comes with its share of challenges. These include:

1. Adaptability to different datasets is an important consideration for selecting an ML algorithm for FDM [25]. As the volume of the flight data is very large, it is necessary that the algorithm is optimized to speed up the process of data analysis and outlier identification.
2. FDM parameters are stored at different rates and these need to be pre-processed to allow analysis.
3. Computational power of the resource available.
4. Choosing the right ML tools and techniques for the given problem, for example choosing between supervised or unsupervised method for FDM.

Unsupervised learning methods are more favourable as they support heterogenous data sets with a large number of parameters as is the case with the flight data. This is the major reason that the majority of research work done is based on unsupervised methods. However, supervised methods also need to be researched as methods like classification can be of great use, for example, classification of anomalies and parameters.

Amongst the unsupervised techniques, anomaly detection based on clustering was found to be most accurate. However, it has its own limitations; the major limitation being that the number of clusters needs to be known a priori. This is often difficult to implement in practice and an iterative process is normally used to determine the best fit. Also, anomaly detection methods require a detection threshold to be set in advance. This threshold decides percentage of flights to be detected. A tight detection threshold will miss true anomalies whereas a loose detection threshold will trigger false alarms. So, ML techniques in future can help to decide what optimum threshold level should be used.

It is also important to note that the algorithms which have been used for FDM focus mainly on segregating abnormal events from normal events. So, in future one of the other areas of interest could be exploring an algorithm which could optimize outlier detection for flight data. It would be ideal if the anomaly detection algorithm could also determine the exact parameters which made that flight anomalous and also provide a sort of score for each of those parameters as the algorithms output anomaly score in [6]. In addition, a flight would not be simply classified as normal or anomalous but would be assigned a degree of being an outlier. Currently, the major research work is focused on descent phase of the flight as it is the most sensitive part. In future, the ML algorithms should be tested for other phases of the flight as well. The major challenge in this case is to make the time series of different flights comparable as the temporal patterns in other flight phases are very diverse.

ML tools and techniques are reshaping the way in which data is being analyzed in almost every industry, with the aviation industry being no exception. For FDM analysis, ML has a bigger role to play in the future and this area requires further research.

5. REFERENCES

1. Fernandes, R. (2002). *An analysis of the potential benefits to airlines of flight data monitoring programs* (MSc). School of Engineering Air Transport Group, Cranfield University.

2. Chrysanthos, N. (2014). *Kernel Methods for Flight Data Monitoring* (PhD). University of Troyes Technology.

3. Flight Safety Foundation (1998). Aviation safety: US efforts to implement flight operational quality assurance programs. Flight Safety Digest, Vol. 17, No.7-9, p2.

4. Federal Aviation Administration. (2001). Flight operational quality assurance program; final rule. 14 CFR Part 13 [Docket No. FAA-2000-7554; Amendment No. 13-30]. Federal Register, Vol. 66, No. 211, p55042-55048.

5. Civil Aviation Affairs. (2018). Retrieved from https://www.bea.aero/docspa/2000/a40-ek000823a/htm/a40-ek000823a.html

6. Matthews, B., Das, S., Bhaduri, K., Das, K., Martin, R., & Oza, N. (2014). Discovering Anomalous Aviation Safety Events Using Scalable Data Mining Algorithms. Journal of Aerospace Information Systems, 11(7), 482-482. http://dx.doi.org/10.2514/1.i010211

7. What is Machine Learning? A definition - Expert System. (2018). Retrieved from http://www.expertsystem.com/machine-learning-definition/

8. Why airlines make such meagre profits. (2018). Retrieved from https://www.economist.com/the-economist- explains/2014/02/23/why-airlines-make- such-meagre-profits

9. EBIT margin of commercial airlines worldwide from 2010 to 2018, r. (2018). EBIT margin of airlines worldwide 2018 | Statistic. Retrieved from https://www.statista.com/statistics/225856/ebit-margin-of-commercial-airlines-worldwide/

10. Airlines Have an Insanely Small Profit Margin. (2018). Retrieved from http://www.businessinsider.com/airlines-have-a-small-profit-margin-2012-6

11. Smart, E. (2011). Detecting Abnormalities in Aircraft Flight Data and Ranking their Impact on the Flight (PhD). Institute of Industrial Research, University of Portsmouth.

12. Shah, A. (2018). Retrieved from https://www.edvancer.in/machine-learning-vs-statistics

13. EASA. (2018). Retrieved from https://www.easa.europa.eu/sites/default/files/dfu/16T0153_ATR_FDM_2016.pdf

14. Li, L., Das, S., John Hansman, R., Palacios, R., & Srivastava, A. (2015). Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations. Journal of Aerospace Information Systems, 12(9), 587-598. http://dx.doi.org/10.2514/1.i010329

15. Li, L., Hansman, R., Palacios, R., & Welsch, R. (2016). Anomaly detection via a Gaussian Mixture Model for flight operation and safety monitoring. Transportation Research Part C: Emerging Technologies, 64, 45-57. http://dx.doi.org/10.1016/j.trc.2016.01.007

16. Theodoridis S., & Koutroumbas K. (2001) Pattern Recognition and Neural Networks. In: Paliouras G., Karkaletsis V., Spyropoulos C.D. (eds) Machine Learning and Its Applications. ACAI 1999. Lecture Notes in Computer Science, vol 2049. Springer, Berlin, Heidelberg

17. Jain, A., & Zongker, D. (1997). Feature selection: evaluation, application, andsmall sample performance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(2), 153-158. doi: 10.1109/34.574797

18. Das, S., Sarkar, S., Ray, A., Srivastava, A., & Simon, D. (2013). Anomaly detection in flight recorder data: A dynamic data-driven approach. In American Control Conference (ACC) (p. 5132-5137). Washington DC: IEEE. Retrieved from https://pdfs.semanticscholar.org/6f9a/a72650108ee6f531b4a5eeef2139252c335a.pdf

19. Mugtussidis, I. (2000). Flight Data Processing Techniques To Identify Unusual Events (Ph.D). Virginia Polytechnic Institute and State University.

20. Ray, A. (2004). Symbolic dynamic analysis of complex systems for anomaly detection. Signal Processing, 84(7), 1115-1130. doi: 10.1016/j.sigpro.2004.03.011

21. Sarkar, S., Mukherjee, K., Jin, X., Singh, D., & Ray, A. (2012). Optimization of symbolic feature extraction for pattern classification. Signal Processing, 92(3), 625-635. doi: 10.1016/j.sigpro.2011.08.013

22. Jin, W., Tung, A. K. H., & Han, J. (2001). Mining top-n local outliers in large databases. In F. Provost, R. Srikant, M. Schkolnick, & D. Lee (Eds.), Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 293-298)

23. Breunig, M., Kriegel, H., Ng, R., & Sander, J. (2000). LOF: Identifying Density-Based Local Outliers ACM SIGMOD Record, 29(2),93-104. doi: 10.1145/335191.335388

24. Jain, A., Murty, M., & Flynn, P. (1999). Data clustering: a review. ACM Computing Surveys, 31(3), 264-323. doi: 10.1145/331499.331504

25. Jesse C., Liu H., Smart E., Brown D. (2008) Analysing Flight Data Using Clustering Methods. In: Lovrek I., Howlett R.J., Jain L.C. (eds) Knowledge-Based Intelligent Information and Engineering Systems. KES 2008. Lecture Notes in Computer Science, vol 5177. Springer, Berlin, Heidelberg

26. BOEING, 1959-2001, Statistical Summary of Commercial Jet Airplane Accidents (2018). Retrieved from https://www.boeing.com/resources/boeing dotcom/company/about_bca/pdf/statsum.pdf

27. Bora, D., & Gupta, D. (2014). A Comparative study Between Fuzzy Clustering Algorithm and Hard Clustering Algorithm. International Journal of Computer Trends and Technology, 10(2), 108-113. doi: 10.14445/22312803/ijctt-v10p119

28. Li, L., Gariel, M., Hansman, R.J., & Palacios, R.H. (2011). Anomaly detection in onboard-recorded flight data using cluster analysis. 2011 IEEE/AIAA 30th Digital Avionics Systems Conference, 4A4-1-4A4-11.

29. Knorr, E., Ng, R., & Tucakov, V. (2000). Distance-based outliers: algorithms and applications. The VLDB Journal the International Journal on Very Large Data Bases, 8(3-4), 237-253. doi: 10.1007/s007780050006

30. Bhaduri, K., Matthews, B. L., & Giannella, C. R. (2011). Algorithms for speeding up distance-based outlier detection. Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 11. doi:10.1145/2020408.2020554

31. Bay, S. D., & Schwabacher, M. (2003). Mining distance-based outliers in near linear time with randomization and a simple pruning rule. Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 03. doi:10.1145/956750.956758

32. Das, S., Matthews, B. L., Srivastava, A. N., & Oza, N. C. (2010). Multiple kernel learning for heterogeneous anomaly detection. Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 10. doi:10.1145/1835804.1835813

33. Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the Support of a High-Dimensional Distribution. Neural Computation, 13(7), 1443-1471.doi:10.1162/089976601750264965

34. Megatroika, A., Galinium, M., Mahendra, A., & Ruseno, N. (2015). Aircraft anomaly detection using algorithmic model and data model trained on FOQA data. 2015 International Conference on Data and Software Engineering (Icodse). doi: 10.1109/icodse.2015.7436969