Transportation Engineering 5 (2021) 100087

Contents lists available at ScienceDirect

Transportation Engineering

journal homepage: www.sciencedirect.com/journal/transportation-engineering

Full Length Article

A data-driven model for safety risk identification from flight data analysis

Mickael Rey [a], Daniel Aloise [*,b,c], François Soumis [b,c], Romanic Pieugueu [c]

[a] Ecole Polytechnique, 91128 Palaiseau Cedex, France
[b] Polytechnique Montréal, 2500, chemin de Polytechnique Montréal (Québec), Canada
[c] Groupe de Études et de Recherche en Analyse des Décisions (GERAD), 3000, ch. de la Côte-Sainte-Catherine Montréal (Québec), Canada

ARTICLE INFO

Keywords:
Air transportation
Decision-support systems
Risk prediction
Artificial intelligence,

ABSTRACT

Most aviation accidents take place in the final phase of a flight. One possible accident is the runway overrun - the fact that an aircraft leaves the runway unexpectedly on landing. Even though such accidents are well documented and studied in the aviation industry, this paper aims at identifying less direct links between data recorded by planes and the risk of runway overrun, or linked events. Indeed, a better understanding of these events using available flight data helps to reduce their number. Nonetheless, such analysis is not straightforward given the massive volume of data collected during the flights. For that purpose, we propose a data-driven approach with the use of data analysis methods and machine learning tools. After a quick correlation analysis, a boosted tree classifier was trained to classify flights as safe or at risk. The classifications were accurate enough to extract contributing factors, and a more extensive analysis was conducted on multiple airports. That analysis revealed the importance of particular factors, leading to new insights about potential approaches to aviation safety.

1. Introduction

In civil aviation, accidents often lead to substantial consequences. Thanks to strategies that were initially reactive and later proactive [1], players in the aviation industry have continuously succeeded in reducing the number of accidents and incidents. To illustrate, the number of deaths per passenger-hour was divided by 10 between 1996 and 2004.

Various types of incidents can happen during a flight. These are generally not problematic, but their accumulation can lead to complications, and even accidents, if the problem is not detected early enough and correctly managed. Moreover, these incidents are not evenly distributed over the duration of a flight. Between 1959 and 2008 [2], nearly 46% of fatal accidents occurred during the final approach or landing. These numbers are still high today (38% of serious accidents between 2015 and 2020 [3]).

When landing, there is a risk that the plane may not be able to stop within the paved surface resulting in an overrun at the end or the side of the runway. These events, which can seriously damage the aircraft and the surrounding areas, can be related to unstable approaches, poor weather conditions, or braking problems, among other factors. The early detection and prediction of such incidents can help to prevent flight accidents. Nonetheless, considering the enormous amount of data collected by operating planes, automatic data analysis methods seem to be the most appropriated tool to try and shed light on information regarding the origin of incidents in aviation.

Machine learning tools can help tackle multiple issues by means of a data-driven methodology approach. Whether it is to predict travel time on roads [4], to diagnostic cancer [5] or to help find the suitable amount of fuel for an engine [6], this quickly developing field often brings promising results. It regroups useful techniques for performing tasks such as classification or prediction, among other things. The advantage is that this is performed without an explicit programming of how to do so, so that machine learning can be used to make decisions based only on the data available. For instance, Kimera and Nangolo [7] used a machine learning approach to reassert the optimal time for maintenance of ballast pumps on floating docks, then predicting their expected time of failure.

Nonetheless, the use of machine learning requires reflection about potential misuses and consumed resources, which have been increasingly powered by steady and rapid advances in artificial intelligence [8]. In order to envisage the use of machine learning in a sustainable and trustworthy way, specially in aviation, its associated models and algorithms need to prioritize transparency so that ML-supported systems

* Corresponding author.
E-mail addresses: mickael.rey@polytechnique.edu (M. Rey), daniel.aloise@polymtl.ca (D. Aloise), francois.soumis@polymtl.ca (F. Soumis), romanic.pieugueu@gerad.ca (R. Pieugueu).

https://doi.org/10.1016/j.treng.2021.100087
Received 26 April 2021; Received in revised form 22 July 2021; Accepted 23 July 2021
Available online 25 July 2021

Table 1
Number of flights available for each airport.

Airport	Number of flights	Proportion of eventful flights (%)
North American airport 1	500	50
European airport 1	441	44
European airport 2	280	12
European airport 3	358	30
North American airport 2	500	50
Central American airport	297	16
South American airport 1	486	51
South American airport 2	384	35
Asian airport	495	50

could be investigated in detail in case they are hacked or invaded, or even if their guidance is arguable. Moreover, machine learning models tend to be complex to augment their classification performance, requiring the adjustment of several hundreds or thousand of parameters while demanding a massive amount of computational and energy resources (see e.g. [9]). These issues constitute major challenges for the use of AI by sustainable societies [10].

The goal of the work presented here is to find relationships between the risk that a particular event occurs and all information gathered by an aircraft. The events considered are the previously described runway overrun and unstable flight approach incidents. To that purpose, machine learning tools were used to detect potential links between the data recorded during a flight and the risk of occurrence of the selected events. The link between the data from a flight and a runway excursion incident is probably complex, and it is conceivable that some aspects may not be obvious. For this reason, data mining/machine learning models are employed to reveal correlations between pre-processed values and the incident in question in a interpretable way.

The International Air Transport Association (IATA) is the trade association for the worlds airlines, representing some 290 airlines. IATA has been analyzing the aggregated de-identified flight data to monitor eventful flights and highlight areas of flight safety concern, with benchmarking available at a global, regional and airport level. To better understand how the multiple features correlate to the certain types of events, IATA provided the data used in this work in the form of aggregated and de-identified time-series aircraft state parameters. These recordings are studied in our present paper and analysed by means of a machine learning approach so as to identify the features which are the most correlated to the studied events.

Our paper is organized as follows. Section 2 presents a review of the related works using similar methodological tools for the aviation field. Then, the methodology, composed by the pre-processing phase of the data, as well as the modelling part, is explained in Section 4. The results of the conducted study are presented in Section 5. Finally concluding remarks are given in Section 6.

2. Related works

Data recorded by aircrafts have been used for decades in order to assist pilots to limit flight risks. For example, runway overrun prevention systems have been created by manufacturers, such as Airbus and Boeing, by computing the distances required to stop a plane according to the state of the flight at a given moment (speed, approach angle, wet or dry runway, etc.). However, the data collected and recorded by aircrafts are vast, requiring the use of sophisticated data analysis methods to investigate less direct casual links for flight accidents. Machine learning can process a large amount of information from flights (both eventful and uneventful) making use of the multitude of available sensored data.

At first, most of the work concerning the study of air accidents using flight data seemed to focus on the evolution of the number of accidents (each year or according to the number of passenger-kilometers, e.g. [11], [12], [13]). These works use various methods, but the most efficient ones seem to use hybrid predictors. For example, the work in [11] proposed a prediction model by separating the linear and non-linear part of the time series of the number of incidents using an Auto Regressive Integrated Moving Average (ARIMA) and a SVM (Support Vector Machine).

Next, some researchers focused on classifying and studying incidents using information derived from flight information or from previous analysis of incidents. Zhang and Mahadevan [14] used neural networks and SVMs to classify incidents according to their severity, exploiting written reports. Drees and Holzapfelont [15] identified various contributing factors for incidents that are summed up in Contributing Factor Trees (CFT). This work is based on human choices and studies of anomalous values of certain variables measured during a flight. They considered runway overruns, the type of incident we investigate in our work, as an application example of their methods. However, their method differs from ours since it consists of a statistical analysis coming from a physical study, while ours is directly based on the recorded data, without prior consideration of the physical reasons for a runway overrun.

Finally, other works are centered on the risk analysis of individual flights, using flight recorded data. This data is massive corresponding to measurements made by the aircrafts sensors which typically collect several thousand values of hundreds of variables during a flight. This high scale data is both an advantage and a drawback since it offers great potential but is difficult to exploit. In particular, the available variables vary according to the type of the aircraft and the flight, which complicates their selection by a general approach. Nanduri and Sherry [16] used recurrent neural networks to analyze the evolution of time series, and to determine if a flight was abnormal, using artificially generated data with a hundred variables. This kind of model helped to deal with the large dimension situation, but their number of features was smaller than the total number of recorded ones. With a different approach, Memarzadeh *et. al.* [17] used a very interesting auto-encoding technique based on convolutional neural networks, that maps a flight into a smaller latent space and reconstruct it. Outliers, i.e., flights that present anomalous behaviour in their flight data, are recognized by their large reconstruction error.

Dimensionality reduction methods help to analyze complete flight data by generating more exploitable vectors. Li *et. al.* [18] used PCA [19] to project their dataset on a smaller space, on which a cluster analysis was performed to detect outliers and abnormal flight operation. Puranik *et. al.* [20] used a similar anomaly detection technique, by creating energy-based derived feature vectors on which a cluster analysis was conducted.

The listed works in this section differ from ours in the sense that their main goals are related to detecting abnormal events during flights, thereby identifying outliers in the flight data. Our main goal in this work is to identify contributing factors regarding the risk of runway overruns and unstable approaches.

3. Data preparation

Fortunately, runway overruns occur very rarely. However, most machine learning models require large amounts of data for training. Besides, if machine learning models can use safe flights for risk prediction, they also need to use data from eventful flights for comparison.

Due to the small amount of runway overrun events, the definition of eventful flights was extended to include events related to unstable approaches. With this new categorization of flights, we gathered recordings of planes heading towards nine different airports. For each airport, a few hundred flights were made available by IATA (see Table 1).

Since environmental conditions are different for each airport for the collected data, the main contributing factors for incidents might be potentially different at each location. Consequently, a separate but similar analysis was conducted for each airport. The extracted data contained time series of measurements made by each plane's sensors (*e.*

Algorithm 1 Wrapper algorithm ($\overline{S}$: whole set of data features)

$S \leftarrow$ random feature from $\overline{S}$
$c \leftarrow 0$
while $c < 0.9$ **do**
  $c \leftarrow \min_{f \in \overline{S} \setminus S} Cor(f, S)$
  $f_c \leftarrow \operatorname{argmin}_{f \in \overline{S} \setminus S} Cor(f, S)$
  $S \leftarrow S \cup \{f_c\}$
return S

Algorithm 1. Wrapper algorithm ($\overline{S}$: whole set of data features).

Table 2
Number of features before and after the wrapper algorithm and lasso selection.

Airport	Initial number of features	Final number of features
North American airport 1	906	217
European airport 1	882	133
European airport 2	842	110
European airport 3	858	129
North American airport 2	906	189
Central American airport	890	177
South American airport 1	890	174
South American airport 2	834	177
Asian airport	906	220

g. air temperature, acceleration), as well as other recorded flight feature values (*e.g.* selected airspeed), and parameters derived from the previous raw features (*e.g.* angle of attack). Our method made use of both raw and derived features as input for the trained machine leaning models.

3.1. Flight phase selection

First of all, we assume that the entire flight data is not necessary to identify the most crucial features to predict the risk of a runway overrun. Our hypothesis is that only the final phases of a flight should be considered. Consequently, only the measurements corresponding to the final approach and landing phases are kept and studied for each flight. We analysed data from the flights from the moment the aircrafts reached the altitude of 10,000 ft. during their descent procedures.

3.2. Colinearity removal and feature selection

Colinearity in features can harm the classification performance of machine learning models [21]. Hence, it is advisable to remove highly correlated features before training machine learning models, to get rid of redundant information. To perform this task, a wrapper algorithm was implemented based on the colinearity of the features during the landing phase. Algorithm 1 presents its pseudo-code. This method starts by selecting a random feature to the selected set of features S. Then, at each step, the least correlated feature to the features already in S is added to the latter. This is performed until that correlation is higher than 0.9. The correlation between a feature f and a set of features S is defined as $Cor(f, S) = \max_{g \in S} Cor(f, g)$.

Since the analysis was supposed to be as general as possible, the features that were available in less than 50% of flights (e.g. due to different aircraft models) were discarded beforehand, In addition to the feature selection expressed by Algorithm 1. This filtering process was performed regarding flights from each airport, separately.

Following the feature selection performed by Algorithm 1, a lasso-based feature selection [22] was performed as well to remove uninformative features. Finally, once automatic feature selection was completed, flight data analysts from IATA were consulted to validate it. In the end, only a few hundreds features were kept, depending on the airport (see Table 2).

3.3. Discretization of time series

Even after restricting the analysed flight data to the landing final phase and performing feature selection, the resulting time-series still contain a large amount of data for each flight, as each one of the plane's sensor can make measurements with a large frequency during several minutes. For example, the acceleration is collected at 8Hz during thousands of seconds. In addition, this information can be redundant as two measurements very close in time are likely to correspond to similar flight events.

In order to reduce the size of the analysed time series to a tractable size, we decided to take data snapshots. More explicitly, the value of

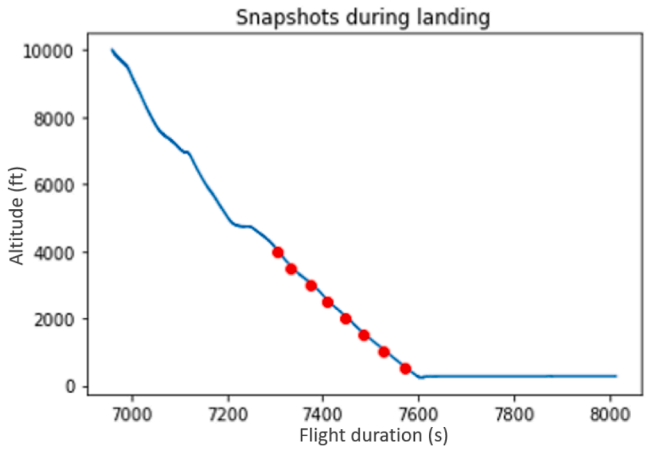


Fig. 1. Altitude (ft) during the approach (s). The red dots correspond to data snapshots. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

each feature is collected at given predefined frequencies, instead of keeping all the recorded information. In order to reduce the information loss incurred by such discretization, the snapshots were taken at multiple altitudes during the final phase of the flight. The values of the features were taken every 500 feet for altitudes between 4000 feet and 500 feet (see Fig. 1).

3.4. Data normalization

The final data preprocessing step consisted in the standardization of the dataset. Three types of standardization were applied.

The first one was specific to our dataset, which contains data collected from different aircrafts, each one of them with its own manufacturing characteristics. To get rid of this bias, the features specific to each aircraft were standardized accordingly. For example, the gross weight of a plane was adjusted according to the aircraft model – by dividing the measured weight by the mean weight value of the aircrafts of the same manufacturer.

Secondly, a normalization based on the months of the year was performed. We observed in our exploratory data analysis that eventful flights were concentrated in summer. That led to a bias, as some features such as the temperature and the pressure are highly correlated to the months, and could have been used by the models for directly detecting unstable flights. Hence, these features were standardized based on the months of the year to remove this bias. More explicitly, this means that a value that was measured in a given month is divided by the mean value of the corresponding feature for flights that took place in that month.

The final normalization was a classical one, which standardizes the features by adjusting them with a zero-mean and one-variance [23].

4. Methodology

The machine learning algorithm that was used in our conducted

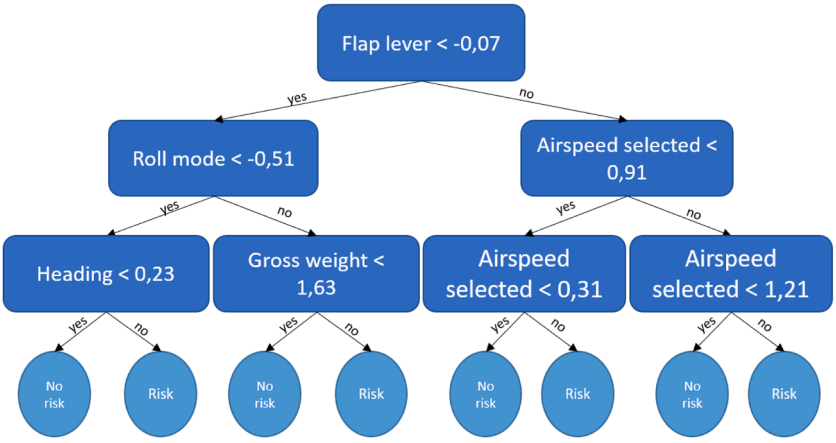


Fig. 2. Example of a simplified decision tree for flight risk analysis.

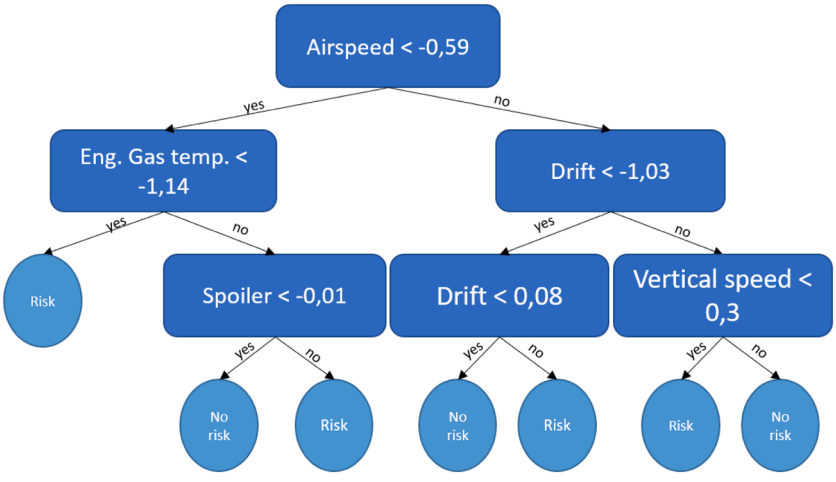


Fig. 3. Simplification of another decision tree used in this analysis.

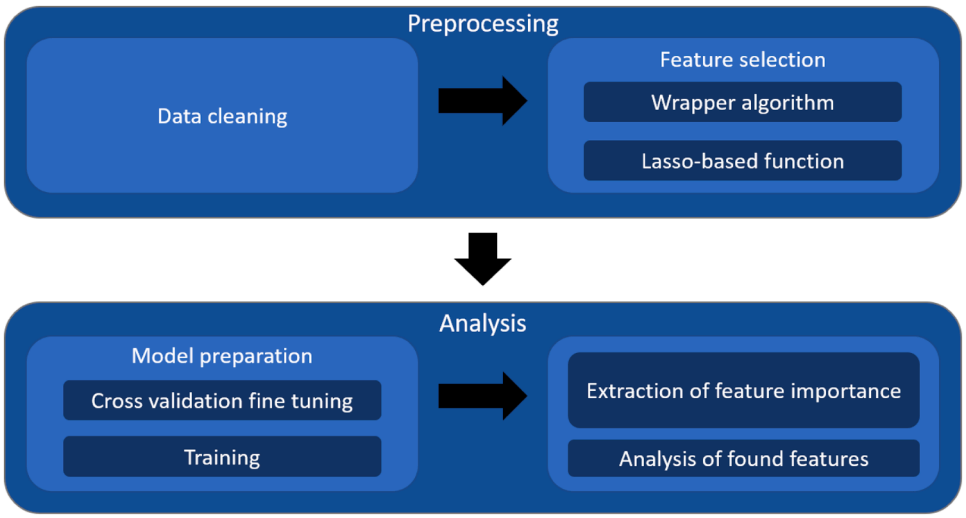


Fig. 4. Description of the complete analysis process.

Table 3
Classification metrics on the testing set for each airport obtained by the XGBoost predictor.

Airport	F1-score (%)	Recall (%)	Precision (%)
North American airport 1	94	93	96
European airport 1	91	89	93
European airport 2	55	46	76
European airport 3	90	89	93
North American airport 2	87	83	92
Central American airport	89	91	88
South American airport 1	92	96	89
South American airport 2	90	88	92
Asian airport	91	90	93

analysis and prediction is a boosted tree classifier, XGBoost [24]. This algorithm is a tree based classifier.

A decision tree corresponds to a series of successive nodes where feature values are evaluated for deciding the class of a data sample. An example of this intuitive process is presented in Fig. 2 where a flight is predicted as at risk or not at the leaf nodes of the tree depending on the values of the features consulted at the intermediate nodes.

Such a model is produced by training the decision tree to classify as well as possible the available labelled data, i.e., data for which the actual class (risk/no risk) is known. Thus, the tree nodes and their corresponding features used for class discrimination are thus chosen and hierarchically structured in order to maximize the accuracy of the decision tree.

In order to improve the generalization of such a model, it is possible to use multiple decision trees not necessarily trained on the same part of the dataset, e.g. via cross-validation (see Section 4.1). The combination of them yield a more robust classifier, and help reduce overfitting. Overfitting might happen when a model is trained too specifically on a dataset that captures only partially the observed phenomenon. Such a model obtains good results on the dataset it is trained on, but does not generalize well to new data samples. Various techniques exist to try to avoid overfitting, including combining multiple decision trees trained on different training sets. Each one of these trees is eventually different, implying different splits according to distinct features or feature values. Fig. 3 illustrates an example of another decision tree for flight risk prediction. Note that the features used for the splits are different from those used in Fig. 2.

Indeed, different decision trees might provide different predictions for the same input data. By combining them, one can achieve better classification performance in a process denoted *ensemble learning* [25].

Random forests [26] add a stochastic element to the model by randomly selecting subsets of the whole training dataset on which the decision trees are trained. Finally, boosted trees use a gradient boosting algorithm to weight the decision trees of a random forest according to their performance on correctly classifying the most critical data samples. XGBoost [24] is a popular exemplar of boosted tree, and it is the one used in this work.

Decision trees are less explicit about how the data features impact classification, but some metrics can extract such information. For those models, the importance of each feature may be determined using purity or the Gini index [27].

4.1. Cross-validation

An important part of machine learning is to verify that the trained models generalize well. The first idea that comes to mind is to split the dataset into a training set, on which the model is trained and fine-tuned (e.g. for the choice of hyper-parameters such as the number of decision trees in a forest), and a testing set on which the model is applied to verify its generalization to unseen (unlabelled) data. In our case, 20% of the dataset was dedicated to testing.

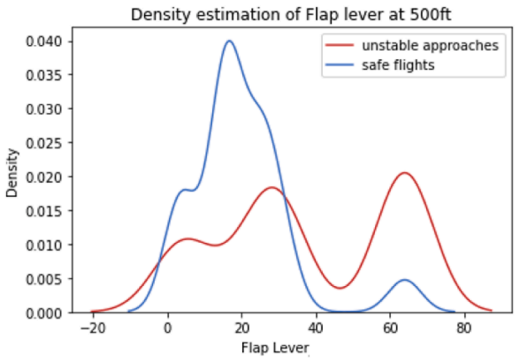
Cross validation [28] is a technique largely used for preventing overfitting. A k-fold cross validation consists in separating the dataset in k groups, and successively using one of the group as testing, while training on the $k-1$ remaining ones. Thus, cross-validation provides a better approximation of the real efficiency of a model and its variance. The selection of the hyper-parameters for our model was made using a 5-fold cross validation on the training set.

4.2. Complete workflow

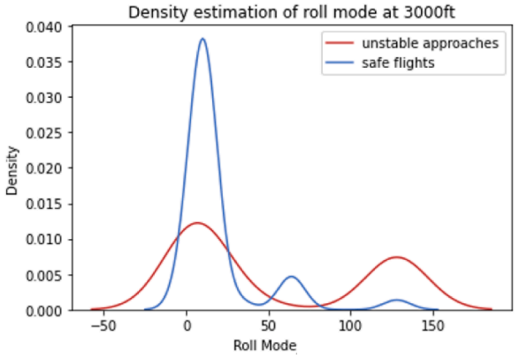
Fig. 4 presents the pipeline of our process to identify the most important features for predicting flight runaway overruns.

This process can be divided into two phases: preprocessing and analysis. The preprocessing phase, described in the first part of Section 3, consists of the data preparation (removal of unexploitable features, modification of time series into more usable structure, data standardization, etc.) and feature selection (reduction of the number of features to improve the model's efficiency).

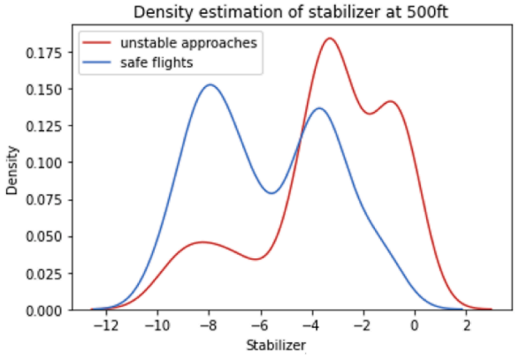
The second phase is data analysis. The chosen machine learning model, i.e. XGBoost, is fine-tuned by cross validation, and trained on the dataset for predicting the safety of a flight. Then, once it was trained, it was possible to verify that its results were satisfactory, and to extract the feature more commonly related to unstable flight approaches. Finally,



(a) Kde for flap lever at 500ft.

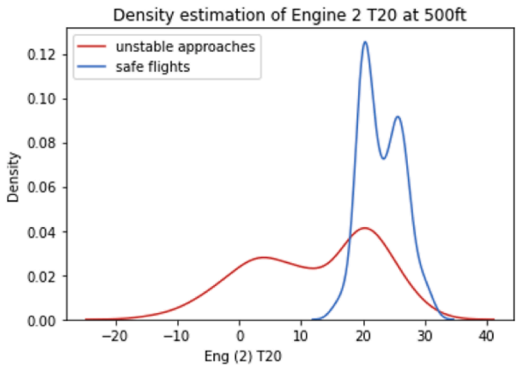


(b) Kde for roll mode at 3000ft.

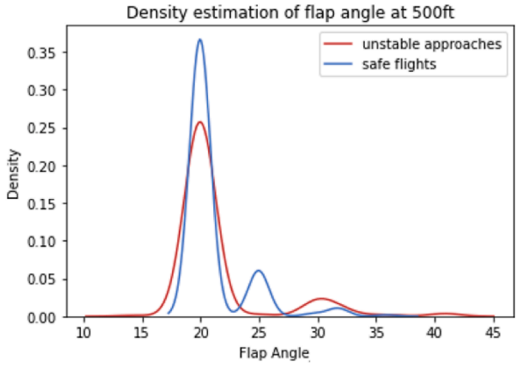


(c) Kde for stabilizer at 500ft.

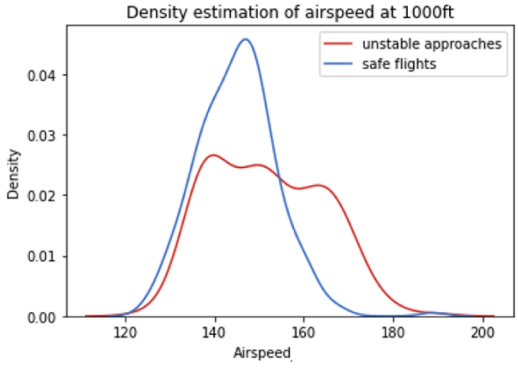
Fig. 5. Kernel density estimations of the top 3 features for both safe and risked flights at Asian airport.



(a) Kde for T20 selected at 500ft.



(b) Kde for the flap angle at 500ft.



(c) Kde for airspeed at 1000ft.

Fig. 6. Kernel density estimations of the top 3 features for both safe and risked flights at North American airport 1.

the extracted features are analysed. The results of this process are presented in the next section.

5. Results

To validate the assumption that the features importance for performing a prediction are indeed linked to flight safety issues, it is necessary for those predictions to be accurate. To assess the accuracy of the proposed model, we begin by computing the true positive, true negative, false positive and false negative counts across all the flights: an eventful flight predicted as unstable is a *true positive* (*TP*), an uneventful flight predicted as safe is a *true negative* (*TN*), an uneventful flight predicted as unstable is a *false positive* (*FP*), and an eventful flight predicted as safe is a *false negative* (*FN*). Using these numbers, we can evaluate the accuracy of the proposed model via the three following standard measures:

- Precision $= \frac{TP}{TP+FP}$;
- Recall $= \frac{TP}{TP+FN}$;
- F1-score $= 2\frac{\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$.

1000

Table 3 presents the F1-score as well as the precision and recall

metrics of our XGBoost for each test set used in our study. These metrics take into account the misclassified flights as well as the rightly classified ones [29], and a value close to one means a good overall prediction. Globally, the developed predictive model managed to classify the flights efficiently for all tested airports.

For 8 out of the 9 data instances, an F1-score of over 87% was reached, with 55% for the European Airport 2 which is the most imbalanced dataset of our study, with a minority of eventful flights. Similarly, the recall was larger than 82% for 8 out of 9 data instances, with a smaller value of 46% for the European Airport 2 as well.

Due to the fact that identified biases were removed in our primary analysis (see Section 3.4), we can assume that there exists a strong relationship between the selected data features and undesirable flight events during landing.

The important features identified by our model regarding each studied airport are presented in Appendix A. Here, we comment over the results obtained for the Asian airport and North American airport 1, for which the largest volume of data was available for our study.

For the Asian airport, the three most important features identified by our machine learning model were the flap lever (at 500ft.), the roll mode (at 3000ft.) and the stabilizer (at 500ft.). On the other hand, the top three contributing features identified for the North American airport 1 were the Engine T20 selected (at multiple altitudes), the flap angle (at 500ft.) and the airspeed (at 1000ft.). While the model successfully identified the importance of the features for each airport, such difference may result from the different characteristics of each airport, such as runway configurations, approach and landing procedures, weather, etc. However, considering that the aggregated dataset includes data from various aircraft types from multiple operators, which might vary with respect to which type of approach or landing procedures was used, it is impossible to conclude that these differences are solely based on the factors related to the airports, or from other external factors.

Now, in order to visualize the differences in values for these features according to the stability of the flight approach, we use a kernel density estimator. A kernel density estimator (kde) is designed to estimate the distribution of the values assumed by a variable. Given some values taken by a feature, this estimator approaches the probability distribution of a random variable from which the observed values would have been sampled. In Figs. 5 and 6, we plot the kde curves for the tree features found as the most important regarding the flights of the Asian and the North American 1 airports.

We can notice that even if similar values for these variables are taken by both uneventful and unstable flights, aircrafts with unstable approaches tend to report more extreme values than those with safe descent and landing. We observe that the plots in Figs. 5a, b, 6 a, b, c show less concentrated distributions for unstable flights. The remaining plot of Fig. 5c does not directly support this claim, but still shows a difference in the distribution of the two classes of flights.

6. Concluding remarks

Better understanding the reasons behind flight incidents is paramount to reduce their occurrence, as well as potential accidents. In this paper, a data-driven analysis was conducted in order to reveal links between recorded data collected during flights, and the risk of unstable approaches. Machine learning models were used for this purpose, in an attempt to detect non-linear connections among the flight data features.

After selecting the most relevant factor and preparing the data for analysis, a gradient boosted tree classifier was trained to predict the likelihood of unstable approaches in each flight, and the most important features for this task were extracted. This model performed very well in its classification task, and the features used were consequently considered as discriminating.

Although establishing a causality relation between the most contributing features and unstable approaches during landing is far from trivial, we demonstrate throughout our experiments the existence of links between the identified factors by our model and the considered events. We believe the identification of such links can be boosted by taking into consideration the expertise of flight specialists, particularly providing insights about the co-related dynamics between the features over time. Another research direction is to devise more advanced machine learning models able to grasp the dynamics of the task — remark that for a particular snapshot our data features are treated as independent, and snapshots are independent from each other in our study.

In the field of aviation safety, considering multiple approaches to the same issue can be beneficial to not miss any possibility. The data driven analysis considered here was made trying to avoid prior assumptions as much as possible thereby obtaining results directly from the flight measurements. Finally, our proposed methodology can be transposed to any other dataset, though it is recommended that the results are validated by subject matter experts to figure out whether causality relations can be actually established for risky flights.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was partially funded by MITACS-Canada.

Appendix A

Results for all airports Tables A.4 and A.5

Table A.4
Ten most important features found for each airport (part 1).

North American airport 1 Features	Importance	European airport 1 Features	Importance
Eng (2) T20 Selected 500ft	12.24	Eng (1) Oil Temp 2500.0ft	8.11
Eng (1) T20 Selected 500ft	6.83	Flap Angle 1000.0ft	4.76
Eng (2) T20 Selected 2000.0ft	4.97	Eng (2) T20 4000.0ft	4.06
Flap Angle 500ft	3.56	Eng (1) Oil Temp 3000.0ft	3.5
Airspeed 1000.0ft	3.39	Wind Speed 4000.0ft	3.42
Eng (2) T20 Selected 1000.0ft	3.21	Eng (1) Oil Temp 500ft	2.91
Eng (1) T20 Selected 2000.0ft	2.73	Pitch 500ft	2.81
Eng (1) T20 Selected 3000.0ft	2.01	Eng (2) T20 1000.0ft	2.69
Stabilizer 1000.0ft	1.99	Eng (1) Oil Temp 3500.0ft	2.57
Eng (2) T20 Selected 2500.0ft	1.99	Subframe Offset In Data 500ft	2.55
European airport 2 Features	**Importance**	**European airport 3 Features**	**Importance**
Flap Angle 1000.0ft	11.87	Eng (2) N2 1000.0ft	17.72
Longitude 500ft	4.07	Eng (1) N2 500ft	4.72
TCAS Sensitivity Level 2000.0ft	3.86	Altitude STD 500ft	4.57
Wind Direction 1000.0ft	2.47	Eng (1) N2 1000.0ft	3.3
Eng (1) N2 3500.0ft	2.14	Flap Angle 1000.0ft	2.96
Flap Angle 500ft	2.12	Airspeed Selected 1000.0ft	2.94
Acceleration Lateral 3500.0ft	2.08	Altitude STD 1000.0ft	2.89
Longitude 4000.0ft	1.93	Eng (2) N2 500ft	2.82
Eng (2) Fuel Flow 500ft	1.82	Longitude 500ft	2.11
Acceleration Normal 2000.0ft	1.8	V1 500ft	2.04

Table A.5
Ten most important features found for each airport (part 2).

North American airport 2 Features	Importance	Central American airport Features	Importance
Flap Angle 1000.0ft	22.12	Flap Lever 500ft	30.36
Mach 1000.0ft	2.45	Eng (1) Oil Qty 500ft	22.6
Drift 4000.0ft	2.35	AOA (R) 500ft	3.61
Airspeed 500ft	1.9	Eng (1) Oil Temp 3000.0ft	3.22
Hour 1000.0ft	1.65	AOA (R) 3500.0ft	3.08
Altitude Selected 2000.0ft	1.61	Eng (1) Oil Temp 2000.0ft	2.57
Gross Weight 500ft	1.57	AT Mode 1000.0ft	2.47
FD Roll 3500.0ft	1.46	Eng (1) Gas Temp 2500.0ft	1.93
FD (B) Engaged 1000.0ft	1.37	Eng (1) Oil Qty 3500.0ft	1.41
DME (2) 1000.0ft	1.34	Acceleration Normal 2000.0ft	1.31
South American airport 1		**South American airport 2**	
Features	**Importance**	**Features**	**Importance**
Flap Angle 1000.0ft	31.25	Fuel Qty 2500.0ft	6.65
Fuel Qty 2000.0ft	3.25	Wind Direction 1000.0ft	6.58
Airspeed 1000.0ft	2.66	Heading Selected 2000.0ft	4.53
Slat Angle 3000.0ft	1.6	Wind Speed 1000.0ft	4.12
Eng (1) T20 Selected 4000.0ft	1.51	Wind Direction 3000.0ft	3.42
Eng (1) T20 Selected 1000.0ft	1.38	Fuel Qty 1000.0ft	3.18
Fuel Qty 1500.0ft	1.34	Fuel Qty 4000.0ft	3.15
Eng (1) T20 Selected 500ft	1.34	Fuel Qty 3500.0ft	2.91
Drift 2000.0ft	1.31	Wind Direction 2000.0ft	2.68
Eng (2) T20 Selected 1000.0ft	1.31	Fuel Qty 2000.0ft	2.62
Asian airport			
Features	**Importance**		
Flap Lever 500ft	10.24		
Roll Mode 3000.0ft	5.5		
Stabilizer 500ft	3.81		
Eng (1) Gas Temp 1000.0ft	3.21		
Airspeed Selected 500ft	3.14		
Flap Angle 1500.0ft	2.79		
Roll Mode 3500.0ft	2.49		
Altitude Radio (C) 1500.0ft	2.18		
Airspeed 1000.0ft	1.93		
Gross Weight 500ft	1.9		

References

[1] G. Walker, Redefining the incidents to learn from: safety science insights acquired on the journey from black boxes to flight data monitoring, Saf Sci 99 (2017), https://doi.org/10.1016/j.ssci.2017.05.010.

[2] B. C. Airplanes, Statistical summary of commercial jet airplane accidents, 1959 - 2008.

[3] Bureau of aircraft accident archives, https://www.baaa-acro.com/, accessed on 2020-09-10.

[4] H. Taghipour, A.B. Parsa, A.K. Mohammadian, A dynamic approach to predict travel time in real time using data driven techniques and comprehensive data sources, Transportation Engineering 2 (2020) 100025, https://doi.org/10.1016/j.treng.2020.100025.

[5] K. Kourou, T.P. Exarchos, K.P. Exarchos, M.V. Karamouzis, D.I. Fotiadis, Machine learning applications in cancer prognosis and prediction, Comput Struct Biotechnol J 13 (2015) 8–17, https://doi.org/10.1016/j.csbj.2014.11.005.

[6] M.S. P, G. V, P. P, G. A, D. G, Prediction efficiency of artificial neural network for crdi engine output parameters, Transportation Engineering 3 (2021) 100041, https://doi.org/10.1016/j.treng.2020.100041.

[7] D. Kimera, F.N. Nangolo, Predictive maintenance for ballast pumps on ship repair yards via machine learning, Transportation Engineering 2 (2020) 100020, https://doi.org/10.1016/j.treng.2020.100020.

[8] S. Gupta, S.D. Langhans, S. Domisch, F. Fuso-Nerini, A. Felländer, M. Battaglini, M. Tegmark, R. Vinuesa, Assessing whether artificial intelligence is an enabler or an inhibitor of sustainability at indicator level, Transportation Engineering 4 (2021) 100064.

[9] Q. Fournier, G.M. Caron, D. Aloise, A practical survey on faster and lighter transformers, arXiv preprint arXiv:2103.14636 (2021).

[10] R. Vinuesa, H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, A. Felländer, S.D. Langhans, M. Tegmark, F.F. Nerini, The role of artificial intelligence in achieving the sustainable development goals, Nat Commun 11 (1) (2020) 1–10.

[11] J.S. Duanmu, X.S. Gan, J.G. Gao, Hybrid prediction method of flight accident based on arima and svm. Applied Mechanics and Materials volume 339, Trans Tech Publ, 2013, pp. 756–761.

[12] R.M.A. Valdés, V.F.G. Comendador, L.P. Sanz, A.R. Sanz, Prediction of aircraft safety incidents using bayesian inference and hierarchical structures, Saf Sci 104 (2018) 216–230.

[13] G. Xusheng, D. Jingshun, C. Wei, Flight accident modeling and predicting based on least squares support vector machine. 2010 International Conference on Educational and Information Technology volume 3, IEEE, 2010, pp. V3–256.

[14] X. Zhang, S. Mahadevan, Ensemble machine learning models for aviation incident risk prediction, Decis Support Syst 116 (2019) 48–63, https://doi.org/10.1016/j.dss.2018.10.009.

[15] L. Drees, F. Holzapfel, Predicting the occurrence of incidents based on flight operation data. AIAA Modeling and Simulation Technologies Conference, 2011, p. 6700.

[16] A. Nanduri, L. Sherry, Anomaly detection in aircraft data using recurrent neural networks (rnn), IEEE (2016), 5C2–1–5C2–8, https://doi.org/10.1109/ICNSURV.2016.7486356.

[17] M. Memarzadeh, B. Matthews, I. Avrekh, Unsupervised anomaly detection in flight data using convolutional variational auto-encoder, Aerospace 7 (8) (2020), https://doi.org/10.3390/aerospace7080115.

[18] L. Li, S. Das, R. John Hansman, R. Palacios, A.N. Srivastava, Analysis of flight data using clustering techniques for detecting abnormal operations, Journal of Aerospace Information Systems 12 (9) (2015) 587–598, https://doi.org/10.2514/1.I010329.

[19] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics and intelligent laboratory systems 2 (1–3) (1987) 37–52.

[20] T.G. Puranik, D.N. Mavris, Anomaly detection in general-aviation operations using energy metrics and flight-data records, Journal of Aerospace Information Systems 15 (1) (2018) 22–36, https://doi.org/10.2514/1.I010582.

[21] H. Midi, S. Sarkar, S. Rana, Collinearity diagnostics of binary logistic regression model, Journal of Interdisciplinary Mathematics 13 (3) (2010) 253–267, https://doi.org/10.1080/09720502.2010.10700699.

[22] V. Fonti, E. Belitser, Feature selection using lasso, VU Amsterdam Research Paper in Business Analytics 30 (2017) 1–25.

[23] J. Han, M. Kamber, J. Pei, Data mining concepts and techniques, The Morgan Kaufmann Series in Data Management Systems 5 (4) (2011) 83–124.

[24] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, in: KDD '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[25] S. Amari, et al., The handbook of brain theory and neural networks, MIT press, 2003.

[26] T.K. Ho, Random decision forests. Proceedings of 3rd international conference on document analysis and recognition volume 1, IEEE, 1995, pp. 278–282.

[27] L. Breiman, Random forests, Mach Learn 45 (1) (2001) 5–32.

[28] M.W. Browne, Cross-validation methods, J Math Psychol 44 (1) (2000) 108–132.

[29] C.J. van RIJSBERGEN, Information retrieval, London: Butterworths, 1975.