5-25-2023

Evidence-Based Managerial Decision-Making With Machine Learning: The Case of Bayesian Inference in Aviation Incidents

Aaron Glassman
Embry-Riddle Aeronautical University, glassf10@erau.edu

Burak Cankaya
Embry-Riddle Aeronautical University, cankayam@erau.edu

Kazim Topuz
University of Tulsa

Dursun Delen
Spears School of Business, Istinye University

Omega 120 (2023) 102906

Contents lists available at ScienceDirect

Omega

journal homepage: www.elsevier.com/locate/omega

Evidence-based managerial decision-making with machine learning: The case of Bayesian inference in aviation incidents☆

Burak Cankaya[a], Kazim Topuz[b], Dursun Delen[c,d,*], Aaron Glassman[a]

[a] College of Business, Department of Management and Technology, Embry-Riddle Aeronautical University, Worldwide, Daytona Beach, FL, USA
[b] School of Finance and Operations Management, University of Tulsa, Tulsa, OK, USA
[c] Spears School of Business, Department of Management Science & Information Systems Regents Professor, Tulsa, OK, USA
[d] Faculty of Engineering and Natural Sciences, Department of Industrial Engineering, Istinye University, Istanbul, Turkey

A R T I C L E I N F O

Article history:
Received 20 April 2022
Accepted 25 May 2023
Available online 1 June 2023

Editor: Asil Oztekin

Keywords:
Decision support systems
Business analytics
Big data
Aviation incidents
Bayesian belief networks

A B S T R A C T

Understanding the factors behind aviation incidents is essential, not only because of the lethality of the accidents but also the incidents' direct and indirect economic impact. Even minor incidents trigger significant economic damage and create disruptions to aviation operations. It is crucial to investigate these incidents to understand the underlying reasons and hence, reduce the risk associated with physical and financial safety in a precarious industry like aviation. The findings may provide decision-makers with a causally accurate means of investigating the topic while untangling the difficulties concerning the statistical associations and causal effects. This research aims to identify the significant variables and their probabilistic dependencies/relationships determining the degree of aircraft damage. The value and the contribution of this study include (1) developing a fully automatic ML prediction-based DSS for aircraft damage severity, (2) conducting a deep network analysis of affinity between predicting variables using probabilistic graphical modeling (PGM), and (3) implementing a user-friendly dashboard to interpret the business insight coming from the design and development of the Bayesian Belief Network (BBN). By leveraging a large, real-world dataset, the proposed methodology captures the probability-based interrelations among air terminal, flight, flight crew, and air-vehicle-related characteristics as explanatory variables, thereby revealing the underlying, complex interactions in accident severity. This research contributes significantly to the current body of knowledge by defining and proving a methodology for automatically categorizing aircraft damage severity based on flight, aircraft, and PIC (pilot in command) information. Moreover, the study combines the findings of the Bayesian Belief Networks with decades of aviation expertise of the subject matter expert, drawing and explaining the association map to find the root causes of the problems and accident relayed variables.

1. Introduction

Flight safety is crucial in global aviation because lives are at stake, and the capital burden is enormous [1]. The economic impact of the incidents has a direct and indirect effect on the organizations. Immediate results include the cost of the parts and maintenance, the cost of the disruption to the utilization of the aircraft, and the cost of the trouble to airport gates, runways, and the whole airport system. Indirect impacts include brand damage, unplanned changes on the flight, maintenance, crew and schedules, depreciation costs, and business and personal life disruptions to the crew and passengers. The aviation business has a low-profit margin in the return-on-investment with significant investments up-front [2].

Considering the fatality and severity of aviation incidents, the Federal Aviation Administration (FAA) gathers, evaluates, and releases valuable information about reportable incidents. Even minor incidents have millions of dollars of impact [3]. For example, incidents involving engine failure or landing gear flaws are classified as minor, yet they have a multimillion-dollar monetary impact on corporations and the entire aviation industry. The aviation industry aims for zero mishaps and makes efforts to predict the incidents before they happen. Thus, understanding the root causes of the incidents is crucial [4].

The strategic decisions made in the aviation industry desperately need better decision support systems (DSSs) powered by expert judgment and machine-learning, especially explanatory methods such as Bayesian Belief Networks. Since aviation incidents rarely happen, most small aviation companies need more data on

☆ Area: Data-Driven Analytics. This manuscript was processed by Associate Editor Asil Oztekin.

* Corresponding author.
E-mail address: dursun.delen@okstate.edu (D. Delen).

https://doi.org/10.1016/j.omega.2023.102906

the incidents, especially those not reportable via government channels (e.g., FAA, NTSB, etc.). Many of them do not have the opportunity to hire data scientists specializing in aviation to evaluate their flight data and merge it with the findings from other incidents. They can make competent business decisions with a holistic view and datasets. A collective dataset fed DSS that considers relevant incidents in the US can guide stakeholders to make better strategic decisions. The holistic approach to collecting all of the relevant datasets enhances the power of Machine Learning algorithms. Machine Learning algorithms map the intricate relations between the variables and predict the target variable, the aircraft damage severity in this case. Machine Learning algorithms are typically known as "black-box" because they are good at predictions but hard to explain. Bayesian Belief Networks are contrary to this stereotype because they are good at making predictions and explaining the patterns resulting in incidents. Bayesian Belief Networks define the affinity between the most impactful categories of the most impactful variables. The size and complexity of the dataset require experts to utilize these Machine Learning methods because only experts can make these predictions and produce these insights with these technologies. Manual expert-based analysis can be shallow, frustrating, expensive, and miss critical relationships.

Even though Machine Learning is not new to predicting incidents, combining algorithmic root cause analysis and expert opinions to enlighten the business decisions inferred from aviation incidents is a comprehensive and novel approach to the aviation industry and the relevant body of knowledge. The current study's primary contributions are the design and development of a fully automatic ML prediction-based DSS for aircraft damage severity, conducting a deep network analysis of affinity between predicting variables using probabilistic graphical modeling (PGM) and developing a user-friendly dashboard to interpret the business insight coming from Bayesian Belief Network (BBN). Such summarizing approach to predict and explain the complex relationship between multiple variables with a simple explanatory tool brings evidence-based insights into the incident patterns and enables decision-makers to understand the incident patterns from numerous similar incidents.

This paper provides an analytical framework for identifying and understanding the severity of aircraft damage when a probability-based graphical model makes much sense for simpler, more informed decision-making and intuitive explanation of complicated relationships. Using publicly accessible data gathered and structured by the Federal Aviation Administration (FAA), this study analyzes the high-risk characteristics contributing to the severity of aviation accidents. This study proposes a multi-step skillful probabilistic Bayesian framework. It provides insight into probabilistic tracing accidents to multiple causes via the Bayesian Network (BN) framework. Additionally, this study introduces an omnidirectional expert system, a BN decision analysis tool, to assist aviation business managers in understanding interconnected relations.

The remainder of the manuscript is organized as follows. Section 2 summarizes the pertinent literature to identify the current study's novelty and contribution, Section 3 explains the methodology and data, Section 4 presents details behind the model design and development, and Section 5 describes the findings and implications. The last section, Section 6, is dedicated to providing concluding remarks, limitations of the study, and future research suggestions.

2. Background

Aviation incidents are generally considered low-probability/high-consequence (lp/hc) events. Even though the likelihood of the occurrence of events is low in terms of probability, the impact severity of the events is vital by risking human life and costly at the expense of the properties in the industry. Also, the direct and indirect disruptions to the aviation system have a high impact. Risk modeling of lp/hc events is a critical subject gaining professional and public attention, such as transportation and power plant accidents [5]. Aviation accidents are of particular interest in this study. The need to avert lp/hc events and the associated risks has grown in the past decades, leading to research efforts in risk modeling aircraft accidents. Several academic outlets have featured articles on investigating aircraft damage severity by analytical methods. These publications can be grouped into four categories: expert judgment and survey-based research, traditional statistical models, Machine Learning techniques, and PGMs.

2.1. Literature on expert judgment and survey studies

Expert judgment and survey studies employ traditional approaches with pre-sampled data, and they frequently aim at subject matter experts, such as accident analysts, pilots, designers, and policymakers. Many of these studies are aimed at individual reports, not summarizing the bulk of similar events and their occurrence patterns. There are institutional reports that experts write due to their incident analysis. These long institutional reports analyze the incident data, pilot communication, and environmental conditions and share their judgmental results and detail of the individual incidents with the public. These expert studies often contain complex information that only aviation specialists can understand [6].

Survey studies with pilots and aviation experts are also common in the literature. Expert-specific information such as; knowledge of the technical standards of the instruments, environmental conditions, the experience of the pilots and operators, specialized training of the pilots, emotional state of the pilots, and sleep quality of the pilots have been examined with expert surveys [7,8]. In addition, another study conducted a survey to compare pilot-related accidents for male and female pilots [9]. Meanwhile, some studies focused on probabilistic risk assessment to predict the responses in lp/hc environments such as nuclear power plant control rooms [10,11]. In addition, some studies elaborate on combining subject matter expertise (SME) with operations research (OR) and ML to predict lp/hc events [12–14] in other transportation mode contexts. Also, review studies have examined the human factors-related causations in airline incidents [15]. In another psychology and human factor analysis study, human factor causations are reported as not paying attention, distraction, complacency, fatigue, task over and undersaturation, misinterpretation of auditory, and inadequate mission briefing, but these human factors are not associated with incidents their specifications such as flight phase, model, flight hours by using Machine Learning [16].

2.2. Literature on traditional statistical models

Traditional statistical models such as simple linear regression, multiple linear regression, and multinomial logistics regression focus on finding primitive relationships between variables and finding how significant these relationships are. The relation between pilot errors and characteristics of the pilot in command (PIC), crash circumstances, and specifications of aircraft are evaluated with multivariate logistics regression modeling with NTSB data with a traditional descriptive statistical approach [17]. Li [18] conducted a study to summarize historic aviation accidents with traditional statistical methods and employed ARIMA (Autoregressive Integrated Moving Average model) to do a time series forecasting of possible aviation accidents. Koteeswaran et al. [19] describe an improved oscillated correlation feature selection (IOCFS) based on correlation-based feature selection and Oscillation Search Technique when data mining FAA accident and incident data. Their goal was to identify key attributes to help prevent accidents/incidents. This method showed some promise when compared against Naïve Bayes, artificial neural networks, support vector machines, decision trees, and like conventional methods.

There have been location-specific studies such as Sun et al. [20] and Wang et al. [21] to utilize statistical models such as ordered weighted averaging (OWA) to predict civil aviation incident rates in China. Likewise, Gürbüz et al. [22] utilized regression to predict failures for component analysis report results of a local airline in Turkey. Nazeri et al. [23] described the impact of severe weather on the National Airspace System by utilizing classifications, regression, and clustering. Some studies focused on predicting the probabilities of a broader spectrum of accidents and incidents with statistical methods [24]. Bazargan and Guzhva [25] utilized NTSB data to investigate the statistical relations between pilot flight hours, experience, age, and gender in accidents. Similarly, Marais and Robichaud [26] evaluated NTSB accidents resulting in fatalities and their relationship with aircraft maintenance and found a correlation for their thesis.

2.3. Literature on conventional machine learning techniques

Data mining and its key enabler, machine learning, techniques, are superior tools to predict and explain aviation incidents because these algorithms effectively identify and capture complex patterns and relationships, as aviation incidents usually have highly intricate occurrence patterns. However, they can account for variable-related assumptions and other limitations, such as a lack of explainability and transparency. Conventional machine learning methods such as Decision Trees, Linear Regression, and Neural Networks are employed by Rehm [27] to predict the aircraft delay categories in Frankfurt Airport by considering the severity of the weather conditions. Lukáčová et al. [28] utilized a variety of decision tree algorithms (e.g., C5.0, CART, and CHAID) to predict fatalities in aviation accidents. Rao et al. [29] utilized several ML methods to predict and understand high-risk event chains for helicopter accidents in the US and found loss of control as the main reason for these accidents. Baugh [30] sampled 26,387 general aviation accidents using qualitative and quantitate data from the NTSB. This data was analyzed using a design tree, gradient boosting, logistic regression (text), neural network, and random forest (text and data) models. Using NTSB data, the gradient boosting and random forest models generate similar results; however, the logistic regression model had the lowest misclassification rate with the most predictive power. Data from the Philippine Aviation Incident Reporting System has also been analyzed using linear regression, gaussian process, multilayer perceptron, and SMO regression to describe present trends and predict causal factors of different accident profiles now and in the future [31]. Despite their high level of predictive accuracy, these conventional machine learning methods lack an in-depth explanation between multiple variables and complex fault patterns. What is needed is an intuitive, probability-based, graphical model that we can predict not only the outcomes but also simulate possible probabilities, sensitivities, and causal inferences.

Despite the recent overwhelming buzz, machine learning is not a new concept in the world of decision support systems; it has been used in a variety of domains, including healthcare and medicine [32,33], sports prediction [34], and drug court decisions [35]. However, it is the right time for its popularity and renewed value proposition due to the advancements in computation power, community-created unconventional new algorithms, and big data sources that feed into these algorithms [36]. One of the most challenging parts of big data analytics in solving complex business problems is to create actionable and understandable business insights because this phase in the analytics process is typically unstructured and often chaotic, as Delen and Ram [37] stated in their article where they explained the significant challenges of business analytics. Analyzing text and image data with ML methods and methods such as NLP to explain some of the variables can improve the accuracy of the predictions [38–40]. As ML methods become a "buzzword," many people use the methods to claim that they are making accurate predictions but have biases in many cases. Van Giffen et al. [41] summarized the common pitfalls and biases in ML research studies that benchmarked business research to claim accuracy in their predictions. Few studies analyze the business value creation out of ML models, which is supposed to be the most meaningful part for businesses. Reis et al. [42] surveyed 319 companies that use ML models and found that; ML use, analytics culture maturity, the expertise in the platform they are using, managerial approval, and the degree of sophistication of the analytics process are the drivers of business value creation in these organizations.

The latest trend in machine learning is deep learning, or more precisely, deep neural networks (DNN). Using a representation learning methodology (where the predictive features are determined by the learning algorithm itself, as opposed to explicitly provided by the data scientist), DNN progressively merges simple features via its multiple layers to create complex features and then uses these features to predict the target variable [43,44]. Although the predictive performance of DNN models has been impressive [45], due to their complex structure, DNN models suffer from explainability and interpretability [46]. Furthermore, designing and fine-tuning a proper DNN architecture (i.e., the number of layers and a number of processing elements in each layer) usually involves a large number of trial-and-error experiments with an in-depth understanding of the learning process. In addition to the number of layers and units, several other hyper-parameters (e.g., type of activation functions, initialization method, training batch size, number of training epochs, optimizer, learning rate, etc.) must be determined via fine-tuning to obtain the best possible model. Because of the vast number of possible combinations, obtaining a globally optimal set of all the hyper-parameters for the network architecture is computationally expensive and time-consuming.

2.4. Literature on probabilistic graphical models

The common anonymized phrase "correlation is not necessarily the causation" has been discussed among scientists in reference to prediction versus causal inferencing and understanding for decades. With the recent popularity of everyday use of machine learning (i.e., making accurate predictions by utilizing data as the pattern creator), understanding the causal inference and using this inference for explainability reignited the discussion as a crucial element of the analytics process, perhaps as important as (in some cases even more so) accurate prediction in many business settings.

In prediction problems needing causal inference, Bayesian Networks (BNs) are advantageous to other ML techniques by simply explaining complex inferences [47]. BN gives researchers a causally explanatory tool and separates statistical correlation and causal effects with input from subject matter experts [48]. BNs have become popular among PGMs because one can map previously disregarded data and complex relations in the network using graphical representation [49–51]. Weighting the variables is an essential component of Bayesian-based PGMs maned best-worst method (BWM) while making multicriteria decision-making by experts. Still, in our case, we did not need BWM because this study does not employ group decision-making [52]. These graphical representations are combined with risk categorization to make a comprehensive prediction and determine the patterns defining the risk categories [53].

Kraus et al. [54] conducted a study comparing the performances of deep learning algorithms with BN algorithms to predict and understand the aircraft engine sensor data collected from over 200 engine sensor data and successfully explained the prediction. Correspondingly, Chang et al. [55] examined and predicted wellhead fatigue failure for subsea oil rigs using Dynamic Bayesian Networks (DBN). Similarly, Arnaldo et al. [56] predicted the incident rate of events for different fleets of a generic airline and mid-air collisions using Bayesian Interference and Hierarchical Structured algorithms.

Some studies demonstrate probabilistic relationships between variables and injury severity. For example, Topuz and Delen [50] studied injury severity in car accidents using BNs. Their analysis revealed high-risk indicators of vehicle crashes using big data from a diverse range of car accidents and their severity categorization. Ancel et al. [57] focused on developing object-oriented BNs to incorporate the risks that result from US-based mid-air collisions (MACs) and flight loss-of-control events from 1987 to 2008. Arnaldo et al. [58] used BNs to predict the risk of loss of separation (LOS) near accidents, regarded as precursors of MACs. Bayesian Belief Networks effectively comprehend probabilistic interdependencies. There have been exciting findings with detailed technical reasoning-based investigation using consultation with experts [59] and NTSB incident reports [4].

This study differs from previous studies in the following ways. It solves inherent data challenges, blends subject matter expertise with aviation incident data, and leverages BN in an intuitive graphical network to explain the interdependencies between variables and their categories to understand causal inference. The study displays the Omnidirectional probabilistic dependencies among all variables, including direct and indirect relationships. This research also gives valuable information about the conditional likelihood of some of the most dangerous flight-type and flight-phase combinations. It leads experts through these scenarios to make policies and business decisions.

3. Method and data

This study provides a five-phase probability-based inference methodology for identifying the critical elements influencing the severity of aviation incidents by revealing the hidden relationships between all variables (input and output). The phases are shown in different colors in Fig. 1. Phase 1 starts with data comprehension, which entails merging, describing, and exploring data from diverse sources. The main database is the FAA incidents data, but other datasets, such as busy airports, major airlines, and aircraft manufacturers, are prepared and merged with the primary FAA data. After addressing data issues in the data preparation phase (2), the data is selected. The primary data issues are; inconsistencies like different names for the aircraft manufacturer names airport names, and changes in these names over time because of mergers and acquisitions. Some variables also have significant missing data that couldn't be used. Phase (3) compromises building a probabilistic graphical model, employing k-fold cross-validation on model assessment and selection, and estimating joint probabilities using data. Phase (4) evaluates model results. Phase (5) performs a sensitivity analysis to interpret the results and use the model for knowledge discovery using various what-if analyses to create recommendations for aviation decision-makers.

3.1. Data overview and preparation

The first step involves the consolidation of repositories to produce a complete data collection. The current study makes use of the FAA's Accident and Incident Data System (AIDS) database, which spans two decades, ending in August 2020. The data set includes all types of aviation incidents, from minor incidents such as bird-plane collisions, personal injuries, incidents that result in fatalities, and aircraft abandonment.

The FAA Aviation Accident and Incident Data System (AIDS) database includes accidents and incidents and may be sorted according to the categories listed above. The FAA database employs the exact definition of an aviation accident as 49 CFR 830.2, which is the definition used by the National Transportation Safety Board (NTSB) and requires the existence of severe injury and substantial damage as defined in those regulations. Additionally, the FAA database contains incidents that do not reach the aircraft damage or personal injury requirements specified in the NTSB definition of an accident [60].

It is essential to understand the concept of accident or incident as the definitions may only sometimes be intuitive. For example, a turbulence event in a “cruise” phase flight that led to a passenger head injury that did not require medical attention is likely neither an accident nor an incident if there were no other injuries or aircraft damage. In the same example, if a passenger received a broken nose that required medical attention, this would be classified as an incident since an accident does include broken bones but excludes simple fractures of the nose, toes, or fingers. Using a different example, a bird strike on takeoff that caused the aircraft to return to the departure airport is likely an incident. If the same bird hit penetrated the First Officer's windshield, it would be categorized as an accident since the aircraft's flying characteristics were harmed. A significant repair was necessary to return the aircraft to airworthiness.

The FAA defines "substantial" incidents as damage or failure that has a detrimental effect on the aircraft's structural strength, performance, or flying characteristics and would ordinarily need extensive repair or replacement of the affected component [60]. They characterized the "Destroyed" incidents as being unrepairable or scrap. "None" events are those that cause no damage to the aircraft, while "Minor" occurrences are those that cause damage between none and substantial. The "Substantial" and "Destroyed" categories were combined in this analysis due to their similar economic effect designations.

Data collectors review files from numerous agencies, evaluate them, and electronically store thousands of crash reports. The data includes when, where, who, and how the incidents happened. Even though the data has been taken care of by the organization standards, there are messy sides at the data collection phase that makes it hard to produce insights from the data; text columns that result in typos, missing data in many columns, aircraft, and operator companies change their names. Because this is a data-driven analytical study, the time spent on data preparation is critical to ensuring the quality and, more importantly, the credibility of the findings. The dataset has details of the incidents, such as the date, city, state, and airport where the incident was recorded. It has recorded aircraft type, make, operator, the number of engines, engine type, engine make, engine group code, flight type, flight conduct code, PIC flight experience in hours, damage severity, and fatality of the incident.

In phase 2, we address the data issues using the CRISP-DM procedure to prepare for Bayesian Analysis. The CRISP-DM processes are considered the cross-industry standard approach for data mining's "best practices." The data has been filtered for recent years and then examined for missing, erroneous, and imbalanced data. The data processing stage consumed a significant part of the project time, making it one of the most demanding components of the study. Machine Learning algorithms need clean data to give fast and accurate results, which means removing outliers, removing duplicates, removing unrelated variables and having considerable data size [61]. The data is cleaned, and some columns are merged with the help of the SME to create meaningful columns such as combining, changing company names, or different ver-

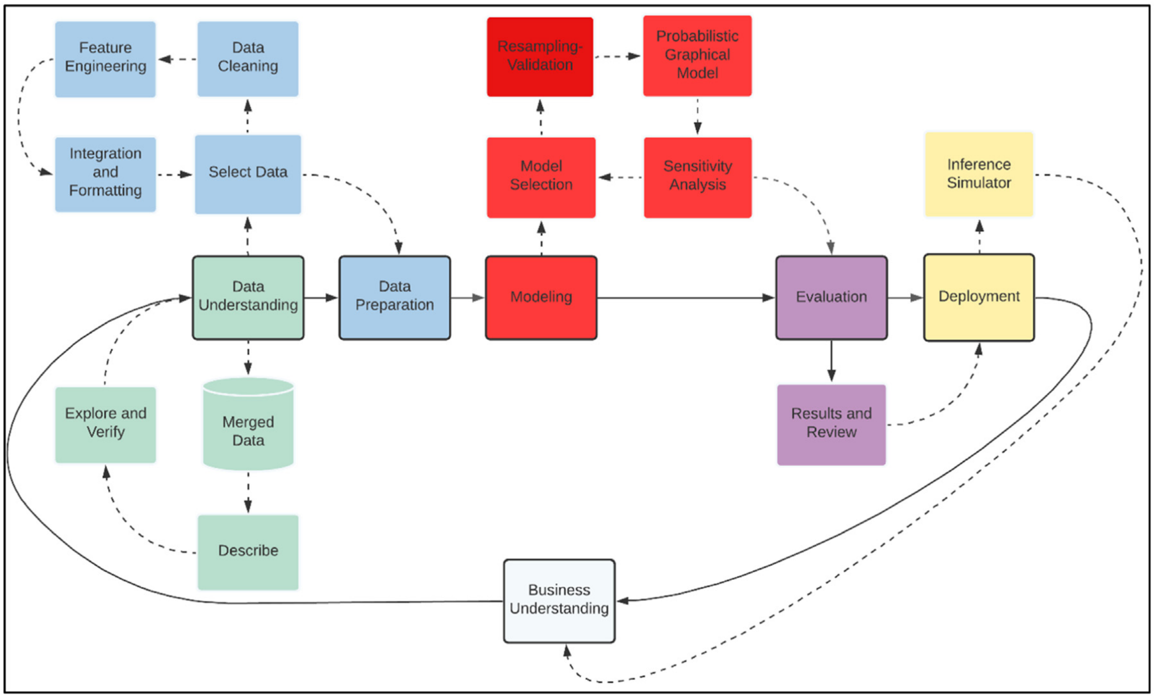

Fig. 1. Proposed methodology for the aviation DSS.

sions of brand names. The locally interpretable model explanations (LIME) method calculates the Euclidean distance between random values for that variable and the final prediction of all variables, then weights and selects the most impactful variables for the prediction [62]. Additional variables were created with manual and auto-variable creation methods to see influential variables combined with multiple variables. New columns have been added, such as whether the airport is congested, whether the aircraft is a helicopter, and whether the operator is a major airline. The aircraft damage column is the target variable; substantial and destroyed categories are merged, and final categories are reduced to minor, none, and destroyed/substantial. Minor incidents are four times more than other categories due to the nature of the dataset, definitions, and disparate probabilities of each category.

3.1.1. *Resolving the issue of missing data*

Preparing the data to handle the missing cases was a significant task. Missing values may be found in all data collection operations, especially when a human enters the data. Our study is no different; incorrect or misleading interpretations of missing observations, as well as an inaccurate perception of confidence in conclusions, can result from faulty treatment of missing observations [50]. Table 1 illustrates the missing percentages in our final cleaned data.

Rubin [63] distinguished three categories of missing data: To begin, “Missing Completely at Random (MCAR)” does not associate a missing data point with any value in the data collection. Second, “Missing at Random (MAR)” happens when missing values are impacted by known values and may be estimated by other data features. When data are “Missing not at Random (MNAR),” the fact that the data are missing is systematically connected to the unobserved data, i.e., the missingness is related to occurrences or circumstances that the researcher does not quantify. Using a systematic approach to imputation has led to significant advancements, according to MAR and MCAR [64]. More data should be addressed or presumed to represent MCAR in aviation literature, leading to ad hoc imputations [65].

Furthermore, we assume that our data is missing at random (MAR) because the missing variables are related to other variables in the dataset. For example, a large portion of the missing data is found in the variable “ModifiedAircraft Engine Model,” which is highly associated with another variable, “Aircraft Engine Make.” This suggests that the missingness is due to the observed values in the dataset rather than any unobservable factors that cannot be quantified. Assuming MAR is thus a reasonable approach for imputing missing data in this context.

This research addresses the problem using BNs rather than ad-hoc alternatives. As Heckerman [66] stated, BN offers a consistent framework for describing the overall joint distribution while concurrently capturing missing data links. Second, because BN is intrinsically probabilistic, the researchers manage missing data and imputations in a non-deterministic manner. It indicates that the necessary variation in the imputed data should be made naturally available rather than manufactured artificially. As with most other ad hoc imputation techniques (such as multiple imputations), BN’s imputation technique makes the MAR assumption. Although missing values in certain variables may be systematic, other variables may be conducive to imputing. The following procedure has been followed:

- *Initiation step:* We will use the Tree Augmented Naive Bayes (TAN) approach, discussed in the following section, to learn how to build an initial structure from a loose network and then populate its conditional probability tables. We will then use this structure learning algorithm to learn how to build a connected network and populate its conditional probability tables.
- *Expectation step:* Draw from probability distributions of the factors based on observed (non-missing or imputed) values using the newly learned network and parameters.
- *Maximization step:* Make use of this new dataset, which has no missing values, to learn about the structure of the data and estimate its parameters.
- *Convergence step:* Until convergence, the Expectation and Maximization processes alternate.

Table 1
Data description.

Variable	Explanation	Data Type	Descriptive Statistics*	Percent Missing
DOW	Day of Week	Nominal	Sunday(28.25), Saturday(22.46)	0
Weekday	Weekday of Weekend	Nominal	Weekday(68.25), Weekend(31.75)	0
Month	Month of Year	Nominal	July(10.41), August(9.53)	0
Day	Day of Month	Nominal	29th(3.47)- 17th(3.45)	0
Event_State	State the Incident Happened	Nominal	CA(11.14),FL(10.78)	3.29
Busy	If the Airport is one of the Busy Airports	Nominal	Yes (JFK, MCO)(6.0), No(94.0)	0
Aircraft Damage	Type of Damage	Nominal	Minor (82.1), None(13.3), Substantial_Destroyed(4.61)	0
Merged_Flight_Phase	Flight Phase	Nominal	Landing(26.86), Touchdown(17.59)	0
Helicopter	If the Aircraft is a Helicopter	Binary	Yes (2.81), No (97.19)	0.9
Primary_Flight_Type	Flight Type	Nominal	Personal(61.69), Instruction(14.34)	23.3
MajorCarrier	If the fight is managed by a major carrier	Binary	Yes (8.33), No (91.67)	4.14
Flight_Conduct_Code	Flight Conduct Code	Nominal	General(78.99), AirCarrier(8.36)	0.39
Aircraft_Engine_Make	Maker of Engine	Nominal	Lyoming(39.98), Continental(29.31)	40.98
ModifiedAircraft_Engine_Model	Engine Model	Nominal	IOSeries(65.59), PW6(5.43)	41.78
Nbr_of_Engines	Number of Engines	Nominal	1(70.14), 2(27.67)	26.33
PIC_Certificate_Type	PIC Certificate Type	Nominal	Private Pilot(38.81), Commercial(16.92)	14.77
Merged_PIC_Category	PIC Category Hours	Numeric	3955.3 (6134.7)	16.38
Merged_PIC_Model	PIC Model Hours	Numeric	738.0 (1657.2)	18.95
Merged_90	Flight hours in the last 90 days	Numeric	54.0 (72.5)	18.95

* Descriptive statistics: Binary-% of each category; Nominal-% of most common two categories; Numeric- mean (standard deviation);.

As an outcome of this process, the BN evolves from an initially disjointed network to its final form. A large number of iterations could be required based on the variables, method, and network complexity. Following that, the fully imputed dataset can be studied or extracted. Records with missing aircraft damage are removed from the study because it is the target variable. BBN method is more meaningful with categorical or ordinal variables, so continuous variables such as date are broken into month, day, weekday, and day of the week. Pilot flight hours are broken into expertise categories as Merged PIC Category as over and under 2000 h, PIC Model over and under 500 h, PIC in last 90 days is broken into over and under 20 h categories to reflect pilot expertise. The number of engines is converted to categorical variables. These data conversion judgments are needed for SME. The combined data set consists of 25,527 records; 21,065 of these records are categorized as “Minor,” 3264 records are “None,” and 1199 are “Substantial/Destroyed” in the target variable. There are twenty variables; five are eliminated in the prediction because of ID-ness(likelihood of the variable being an ID variable), the number of missing data, and low impact with a substantial number of categories whose effect on the prediction is negligible (Table 1).

Since the number of records for target variable categories is imbalanced and to aim for low bias and low deviation, 10-fold cross-validation has been applied in phase 3, and training to testing data has been partitioned to nine to one. In 10-fold cross-validation, in every iteration, one of these pieces of data is held as test data, and the other nine are held as training. The process is reiterated for every piece combination, and their performance is compared with each other and the whole original data. This study utilized $k = 10$ folds to balance effort for computation and objectivity, a common practice in the data mining community [39]. The cross-validation performance metrics may be expressed as follows:

$$Performance = \frac{1}{\mathrm{k}} \sum P_i \tag{1}$$

where k is the number of folds and $\mathrm{P_i}$ is the validation performance [67]

Phase 4 involves the creation of a PGM to display the BN inference diagram. The tree augmentation describes the relationship of multiple factors to their parents.

In phase 5, the results of the BBN analysis are tested to see the sensitivity of the influence utilized for the incident risk factors. The model results are compared to validate the results of the data sampling, balancing data, and variable selection. All predictor variables are included in these comparisons, and there is no data balancing. The synthetic minority oversampling technique (SMOTE) is a data balancing method where the target variable is not balanced. SMOTE is used to oversample minority classes and balance their weight in the data. This study presents selected features without data balancing since SMOTE did not significantly increase the accuracy of "Minor" incidences under this scenario [68].

3.1.2. Entropy and mutual information

When conducting analysis, we often investigate the correlation and covariance among factors to identify their relative importance and link to the targeted variable. In this study, aircraft damage is the categorical target variable, and we examine correlations and covariance based on the information theory and evaluate the results with SME. We predict the uncertainty of the states of the variables and consider the technical meanings of these states. Entropy is a quantifiable measure of uncertainty in information theory [69]. This study uses entropy to measure uncertainty in the form of probability distributions. Our case explains the uncertainty of aircraft damage in aviation incidents. Incident investigation is a deep field in terms of SMEs. To present a case, we need detailed information about the case. When the data is missing, we can comprehend the information we already have with the help of the SME and clarify the uncertainty. For example, an incident at a scheduled carrier will have a significantly different likelihood of the flight phase from a personal recreational flight because of many reasons such as the number of people flying, the congestion difference of the airports, and experience. The information about these variables significantly increases or reduces the probability.

We can calculate the entropy using the probability distribution. Where H represents the marginal entropy of a random variable X, discrete distribution entropy is defined as follows:

$$H(X) = -\sum_{x} P(x) log P(x) \tag{2}$$

In terms of probability distributions, the conditional entropy of X given Y is as follows:

$$H(X|Y) = -\sum_{x}\sum_{y} P(x, y) log P(x, y) \tag{3}$$

Alternatively, the conditional entropy of X given Y can also be expressed as:

$$H(X|Y) = H(X,Y) - H(Y) \tag{4}$$

where H(X,Y) is the joint entropy of X and Y, and H(Y) is the marginal entropy of Y. And the information gain for a random variable X with respect to Y is given by:

$$I(X|Y) = H(Y) - \ H(X|Y) \tag{5}$$

For example, the information gain or entropy decline associated with knowledge regarding "FlightType" is self-evident. Observing another event with a more common flight type, such as personal recreation, typically provides less knowledge and so has less predictive ability for that event. However, we want to know how much information we would get on average if we considered all potential values of "FlightType" and their associated probabilities—in other words, if we used it as a predictive variable for "Incident Severity." The predictive value of observing the variable "FlightType" would be shown by this "average information gain" statistics. Assume we have a random variable X that represents the severity of an incident, with three levels of severity: none, minor, and major. We also have a random variable Y that represents the flight type, which has two options: personal or instructional. Conditional entropy of Severity given FlightType:

$$\begin{aligned} &H(Severity \mid FlightType) = \\ &-\ [P(X = none, Y = personal)\ \log P(X = none|Y = personal) \\ &+P(X = minor, Y = personal) \log P(X = minor|Y = personal) \\ &+P(X = major, Y = personal) \log P(X = major|Y = personal) \\ &+P(X = none, Y = instructional) \log P(X = none|Y = instructional) \\ &+\ P(X = major, Y = instructional)\ log\ P(X{=}major|Y{=}instructional)] \end{aligned} \tag{6}$$

Mutual Information, denoted by I, is the gap between the target variable's marginal entropy and the target's conditional entropy given the predictive variable. Mutual Information between "Incident Severity" and "FlightType" is defined as "Incident Severity" marginal entropy minus "Incident Severity" conditional entropy given "Flight Type" in our example:

$$I(Severity, FlightType) = H(Severity) - H(Severity \mid FlightType) \tag{7}$$

With this method, we can calculate the information gain for each independent variable and its predictive importance.

3.2. Bayesian network probabilistic inference model

DAG (Directed Acyclic Graphs) were the ancestors of the PGMs. Early examples from the 20th century, such as Seawall Wright's works, have been utilized in different application areas [70]. Over time, the PGMs are developed and updated as BBNs, referred to as the directed graphical models that utilize statistics and Machine Learning to build belief networks between variables.

By the late seventies and early eighties, decision sciences started to make bottom-up and top-down reasoning to find answers to research questions [48]. The reasoning process started with more qualitative methods such as Root Cause Analysis, Five Why's, DELPHI, etc. Using expert judgment, the reasoning process evolved to use survey results and other data collection to understand the reasoning using statistical methods [71]. Later, BN replaced the reasoning methods mentioned above. The qualitative and primitive qualitative reasoning methods and rule-based methods are improved by utilizing artificial intelligence's power and enabling experts to make deeper reasoning by utilizing BN-based Machine Learning algorithms and end up with indeterminate inferences [72,73].

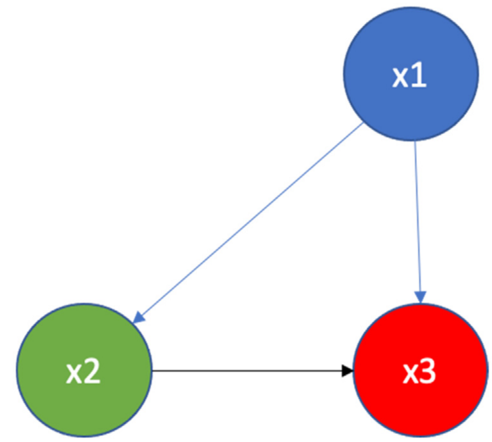

Fig. 2. An illustration of simple directed acyclic graphs (DAG).

The BN model is a directed acyclic graph (DAG) having nodes matching certain variables. For instance, the purpose or nature of a flight operation is referred to as its flight type. Flight types vary depending on the purpose of the flight, such as personal, instructional, industrial, and air taxi flights. The flight Phase describes the various stages of a flight operation. Takeoff, cruise, touch, and roll-out are all flight phases based on the position and movement of the aircraft during a flight. Assuming no additional variables are present, Fig. 2 depicts a basic DAG representing variable conditional dependencies using arcs, with the direction of the arcs indicating specific parent-child relationships. Suppose a connection between x_1: Aircraft_Damage, x_2: Primary_Flight_Type, and x_3: Flight Phase. Three types of connections between variables are commonly observed in causal inference: common effect, common cause, and indirect effect. The variables x1: Aircraft Damage, x2: Primary Flight Type, and x3: Flight Phase may have different types of connections in the context of flight operations. When two or more variables have a common effect on a third variable, this is referred to as a common effect connection. For example, x2: Primary Flight Type and x3: Flight Phase may have a common effect on x1: Aircraft Damage, such as when a personal flight during the takeoff phase is more likely to cause aircraft damage. A common cause connection occurs when a common cause affects two or more variables. For example, x2: Primary Flight Type and x3: Flight Phase may share a common cause, weather conditions, which can affect the likelihood of aircraft damage during a flight. When one or more intermediate variables mediate the effect of one variable on another, this is referred to as an indirect effect connection. For example, x2: Primary Flight Type may have an indirect effect on x1: Aircraft Damage via x3: Flight Phase. A flight operation with an industrial primary flight type may be more likely to be in the cruise phase, which may reduce the risk of aircraft damage. The type of connection between variables in a DAG can have significant consequences for causal inference and decision-making in flight operations. Flight operators can make more informed decisions about safety protocols and risk management if they understand the direction and nature of these connections.

There are two types of probability distributions: marginal and conditional. It is marginal for parentless nodes and conditional for those with parents. In conditional, the dependencies will be determined using conditional probability tables (CPT) for each node in the graph with a parent. Following the specification, the BN efficiently depicts the JPD (joint probability distribution) and can therefore be used to calculate the posterior probabilities of any subset of variables. BN chain rule is often employed to represent

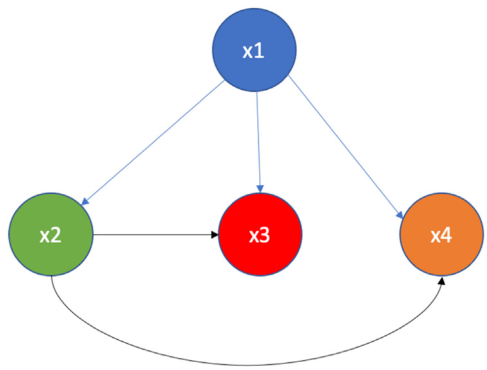

Fig. 3. Depiction of a simple tan network structure.

complicated probability distributions [63]:

$$P(x_1 \ldots, x_n) = \prod_{i=1}^{n} P(x_i|A_{x_i}) \tag{8}$$

where x_i is a variable, and A_{x_i} is the parent, in Fig. 3, P (x_3 |x_2, x_1) is the probability of flight phase given the values of flight type and damage.

Predicting the exact inference using BN is an NP-hard problem to solve, as Pearl 47] eloquently explains. It is more beneficial to use BN since it is simple to transfer earlier knowledge into a network structure by prohibiting relations, leveraging prior distribution across network constraints, or modifying structural components.

This process begins with Naive Bayes, which provides conclusions for all examined variables to be anticipated. The Naive Bayes classifier keeps to the Bayes principle by constraining the network under the strict assumption that all factors except the target variable are independent. In the development of naive Bayes classifiers, the TAN approach provides a tree-like graphical model for predicting the interactions among several predictors [74].

The TAN structure outperforms the Naive Bayes approach and requires no search in the calculation. TAN employs a parentless class variable; nonetheless, the conditional probability for the class variable and another feature is determined for each attribute. TAN fits this research problem better because we need tree augmentation to understand the relationship between incident-creating variables [76]. The TAN structure is shown in Fig. 3, where the tree is a function over $x(i) > 0$, x_1 is the target variable with no parents, i.e., $A_{x_1} = \emptyset$. Chow and Liu's [77]. Tree Bayesian concept is used to assemble the TAN structure. The elaborated conceptual proof and formulation of the tree Bayesian concept can be found in detail in Chow and Liu's [77] and TAN structure in the work of Friedman et al. [75].

Safety is the first factor considered while designing any system related to aviation, such as an airport or aircraft. Aviation systems are designed and regulated to have multiple safety futures with multiple assurance systems. The aviation systems are carefully designed to back up one another in an emergency, so understanding the series of events and pattern of failed systems in a tree-based augmentation network is critical to remedying reverse-engineered policymaking for the aviation industry. Unlike alternatives, we aimed to determine conditional dependencies of incident-creating variables due to their enhanced performance in understanding the aviation incidents. We utilized the BN's TAN model for its outstanding performance compared to other structural learning algorithms.

Table 2
Notation used in this study.

Symbol	Definition
X	Random variable
Y	Random variable
P(x)	Probability mass function of X
P(y)	Probability mass function of Y
P(x,y)	Joint probability mass function of X and Y
P(x|y)	Conditional probability mass function of X given Y
H(X)	Marginal entropy of X
H(Y)	Marginal entropy of Y
H(X|Y)	Conditional entropy of X given Y
H(X,Y)	Joint entropy of X and Y
I(X|Y)	Information gain of X with respect to Y

4. Model design and results

4.1. Model design

Making solid expert judgments about aviation incidents is challenging. We need to make inferences with components and findings contrary to each other and produce evidential results. BNs can do these computations under uncertainty and create reliable inferences. Even though we have created a predictive model in this study, the focus is on creating an insightful explanatory model to enable decision-makers to test different aviation incident scenario combinations and take preventive actions accordingly.

This study gives the metrics about the predictive performance, but the main spotlight will be on interpreting the aviation incidents with a multiclass probabilistic Bayesian Belief Model. The data used in the study is gathered from FAA public records, and it is assumed that the data is MAR (Missing at Random). We applied the routine missing data resolving methods for stochastic conditional BN methods depending on the properties of the variables that can be better understood by looking at the data properties in Table 2. Due to the nature of the aviation incidents, the Substantial and Destroyed (% 4.72), Minor (%82.25), and None (%12.71) category incidents are imbalanced in the data. We applied the cross-entropy loss function approach to offset the data. Minor incidents are the main focus of the research question accordingly. Under Sampling for the "minor incident" category has been implemented to align the cross-entropy loss function by balancing the data.

4.2. Sensitivity analysis

When we look at the model's predictive performance, comparing the balanced and imbalanced datasets is essential because there is a significant imbalance in the target class categories. The receiver operating characteristic (ROC) curve is a summarizing model performance metric that visualizes the True Positive Rate over False Positive Rate. As a result of the comparison, imbalanced datasets performed better than balanced datasets.

(ROC 73.74% vs 73.98%). In addition, the business scenarios we mainly evaluated depend on the "Minor" severity class because of the economic value of the events that happen in the class, and the imbalanced dataset gave better in-class precision for the "Minor" class values (93.35% vs. 67.55%).

A ten-fold cross-validation procedure was used in this study. Cross-validation is necessary when it is a classification problem and an imbalance in the data. Table 3 reflects the cross-validation test results for the target variable. The aircraft damage variable has three categories for damage severity. The categories are unbalanced. Experiments are repeated with and without SMOTE data balanced datasets to overcome the balance issue. The study found that SMOTE data balancing did not significantly improve the overall metric compared to no balancing. Since our primary focus is on

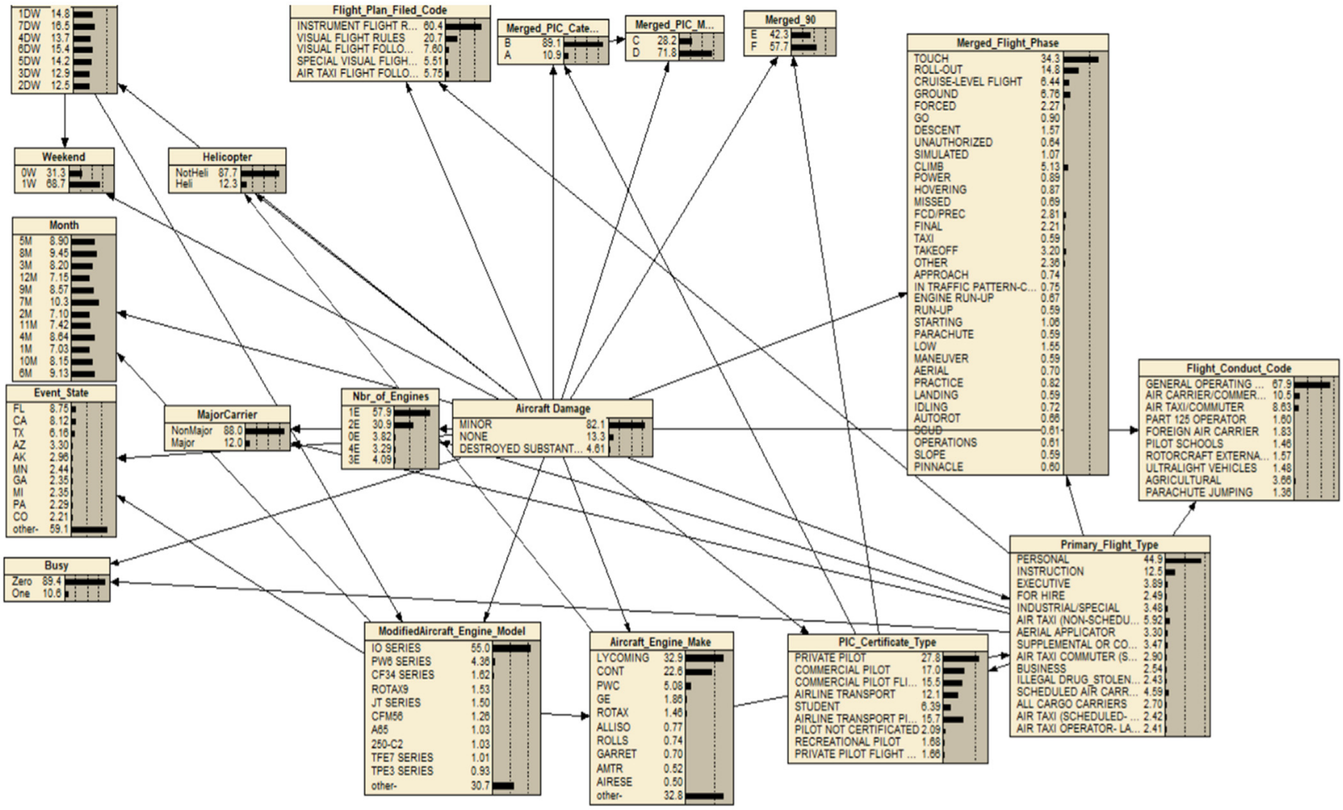

Fig. 4. Bayesian network for prediction of injury severity.

Table 3
Ten-fold cross-validation prediction performance measures for the models with and without SMOTE data balancing.

	No Balancing	SMOTE Data Balancing
Mean ROC	73.74%	73.98%
Overall Precision	83.04%	81.28%
Overall Reliability	79.10%	78.26%
“Minor” Damage In-Class Precision	93.35%	67.55%

the” Minor” category incidents and their business inferences, we proceeded with imbalanced data.

5. Discussion of the findings

The result of the BN structure enables us to simulate testing different event scenarios. Depending on the business type, organizations can test the interdependency of variables and comprehend beliefs to develop findings to create a competitive advantage by having fewer incidents. In addition, the reduced risks using the decision support system contribute to the organization by reducing the insurance costs and increasing public brand value, operational efficiency, and profitability in a business where every little incident has significant direct and indirect costs. We utilized some of the most reliable and interpretable data preparation, prediction, and explanation methods to reduce the prediction bias. Fig. 4 explains the Baseline model for the Bayesian belief network (BBN). It visualizes the prior probabilities and interdependencies of the BBN. It shows the interdependencies of all variables and how changing one category impacts other flight-related characteristics.

The arrows indicate the relations between variables; the TAN network explains the interdependency between variables. Uncertainty is an essential factor in predictive analytics problems. Fig. 4 is a summary visualization that includes the descriptive nature of the situation by the distribution of the data and the predictive solution by the parental relations between the variables visualized by the arrows between variables. For example, Flight type is parent to flight phase, conduct code, number of engines, and Major Carrier and PIC certificate type. On the other hand, flight type is also a child of Aircraft Engine Make and Aircraft Damage.

The sensitivity analysis of the target variable (Aircraft Damage) in Fig. 5 is a critical metric to visualize the comparative importance of the variables to the target variables. The sensitivity analysis means the order of the sensitivity means relative significance for the prediction. Fig. 5 shows that Flight Phase (16.6%) is the most critical variable. We know from the aviation community that landing and departure are the most problematic flight phases in that incidents happen [78,79]. The aviation community focuses on improving flight safety in these phases. After the Flight Phase, the Flight Conduct Code (10%) defines the flight type, pilot certification, and schedule flexibility of the flights. Later, the Engine Model (8.92%) and the Engine Brand (8.44%) can define the aircraft brand, aircraft type, maintenance routines, brand specifications, and reliability. Flight Type (7.83%) is similar to the conduct code but more specific to explain personal flights, instructional flights, executive flights, for-hire flights, and industrial flights, which gives the context of the flight. The aircraft type in terms of the number of engines (4.96%) is related to the flight type, pilot expertise, and airport specifications. These variables are the most important variables to predict the damage. Whereas PIC experience is one of the first factors that come to mind, shockingly, they could be more impactful. The sensitivity of PIC-related variables is PIC Certificate Type (4.9%), PIC Total Flight Hour (0.83%), and PIC hour in the last 90 days (0.26%). PIC hour in the model (0.085%)

Fig. 6 adds another level and investigates Aircraft Damage’s sensitivity to the Flight Phase. When we look at the business infer-

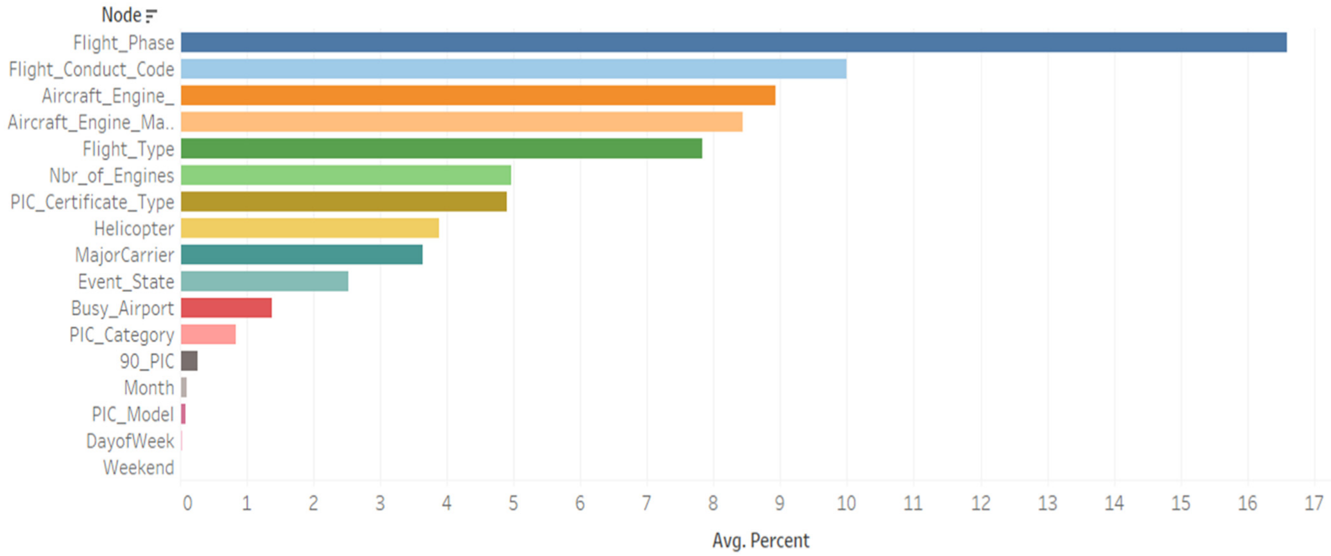

Fig. 5. Sensitivity analysis of aircraft damage-related factors.

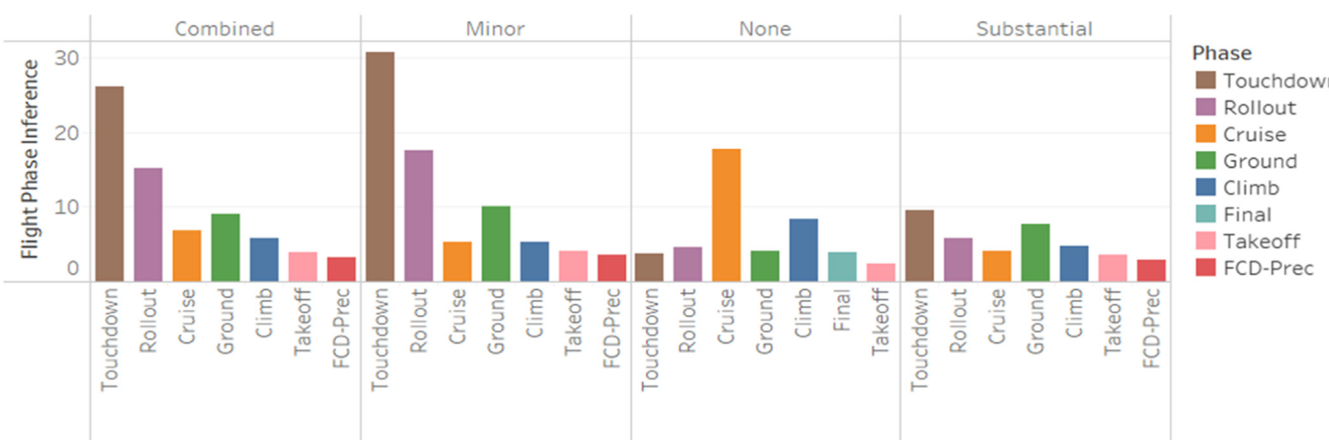

Fig. 6. Sensitivity of aircraft damage to flight phases.

ence by testing the sensitivity of Aircraft Damage to Flight Phases for all aviation incidents, we can see in Fig. 6 that there is a significant difference between damage categories and flight phases in the incidents that happen. Touchdown is a landing base phase, and touchdown phase incidents are more likely to end up in “Minor” (30.7%), “None” (3.77%), and “Substantial_Destroyed” (3.6%), and combined (26.1%) category incidents. Similarly, in another landing-related phase, the Rollout phase incidents are more likely at “Minor” (17.5%) incidents than “None” (4.57%), “Substantial_Destroyed” (5.77%), and combined (15.2%) category events. Historically, landing is the most challenging part of flying an airplane [80]. Both touchdown and rollout are landing phase events where the rollout is just after the touchdown. In this comparison of Touchdown and Rollout phase incidents, it is observed that Rollout incidents are more likely to cause “None” category incidents than touchdown incidents because they primarily impact people, not aircraft.

Although, the “Cruise” phase has more “None” (17.7%) Aircraft Damage category events than “Minor” (5.28%), “Substantial_Destroyed” (4.08%), and “Combined” (6.87%). Similarly, there is a significant difference for “Climb” phase incidents at “None” (8.3%) Aircraft Damage in comparison with Minor (5.77%), “Substantial_Destroyed” (4.8%), and “Combined” (5.73%)

To illustrate some of the incident cases and model findings of cases, we came up with an omnidirectional inference diagram. In Fig. 7, we visualize the omnidirectional probabilistic relationship between Flight Type, Flight Phase, and Aircraft Damage to compare personal flights and scheduled air carrier flights on cruise-level incidents. We run the model for the probabilistic value of chosen category combinations called omnidirectional inference. If a variable category is chosen, that is shown to have 100% probability. Case 1 shows the overall model with flight and flight phases. Case 2 offers personal flights. Case 3 shows scheduled air carriers. Case 4 shows scheduled air carriers on crise level incidents. And Case 5 shows personal typed flights on cruise level. We observed the “Flight Type”-Scheduled Carrier vs. Personal Flights-the probability of Minor Incidents in Personal Flights (Case 2–92.7%) is higher than in Scheduled Carriers (Case 3–74.1%). On the other hand, the likelihood of Substantial_Destroyed Incidents in Personal Flights (1.83%) is lower than in Scheduled Carriers (6.36%). In all flights, the probable flight phase is sorted in order as; Level-of-Touchdown (26.1%), Roll-Out (15.2%), Ground (9.11%), Cruise (6.78%), Climb (5.73%). In Personal Flights, the probable flight phase is sorted in order as; Level-of-Touchdown (40.1%), Roll-Out (21.5%), Ground (9.41%), Cruise (6.19%), Climb (4.50%). The ranking of probabilities for Flight Phases changes for Scheduled Carrier vs. Personal Flights. Whereas in Scheduled Carriers, the flight phase probabilities for incidents are sorted as follows; Ground (%17.3), Climb (14.5%), Cruise (11.7%), Roll-Out (7.44%), Level-of-Touchdown (6.79%), Takeoff (5.76%).

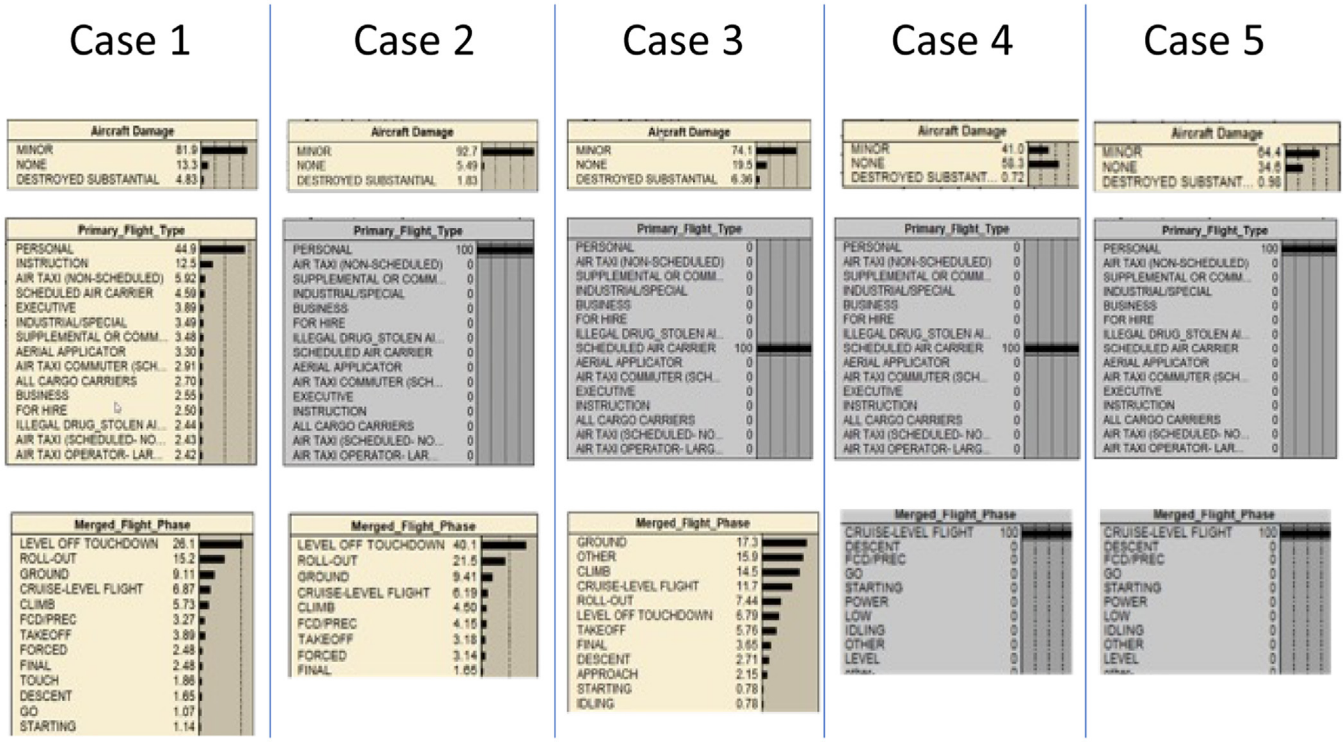

Fig. 7. Omnidirectional inference update for flight type and flight phase.

Flight Type "Scheduled Flights" have many "None" Aircraft Damage category incidents because most incidents happen at the terminal while the aircraft is grounded. In some cases, nothing is happening about the aircraft, but the crew or passengers fall off the stairs, etc. "Personal Flights" are mostly non-scheduled and for leisure due to their nature. Their schedule can easily be rescheduled or rerouted, but an airliner cannot easily be rescheduled in case of a low probability weather risk event. So, "Climb" phased incidents are inevitable and more likely to happen on Scheduled Flights (14.5%) than on Personal Flights (4.5%) because Bird Strikes, Engine Shutdowns, and "Cruise" phase incidents with turbulence cannot be easily avoided without changing the schedule or route. The main reason for the significant difference between scheduled and personal flights in the climb phase is the prior being fixed scheduled months ahead and later mostly being leisure and rescheduled in any risky weather condition.

6. Conclusion, limitation, and future research

Aviation authorities, airline operators, airport managers, flight schools, aircraft manufacturers, aviation researchers, and insurance companies benefit from this study to determine the landscape of related incidents and speculate on the physical safety of their assets and financial safety of their assets. Aviation authorities can use the tool to discover risky patterns and change aviation policies for a safer aviation system. The most repeated incident patterns, such as the landing phase (touchdown and rollout phases), are associated with personal flights. These personal flights are mostly recreational flights, and the flight schedule and location are changeable. These flights where the pilot has less experience than professional flight categories can be operated at better weather conditions and from less-risk airports. In addition, flight training schools should spend more training and simulation time on these types of landing pieces of their training. Government authorities should increase the responsibility and requirements of flight training schools to spend more time on the landing phase. The airline operators further investigate the incident patterns for their companies to define their crew planning, fleet planning, and flight routes built on stochastic demand [81]. "None" category aircraft damage incidents for scheduled carriers mostly have ground, climb, and cruise phase incidents. The ambiguity can be mitigated by the incident patterns the company is likely to have [82]. In addition, they can find out solutions for "Ground," "Climb," and "Cruise" phase incidents with schedule and route changes which holds the majority of their incidents. The taxi phase is part of the ground phase, and airlines need to train pilots for the airport-specific blind spots, hot spots, and airport layout. The pilots need to be more airports specific, trained, and assigned to reduce these incidents. Airport managers can evaluate the patterns of the incidents happening in their airports, find seasonality, and compare the results with runway blind-spot analysis reports done by FAA [83]. Airport managers can also plan for high-risk seasons, which can be winter scheduling operations, by predicting the stochastic variables such as when and where it is likely to have incidents [84].

Cruise phase incidents for scheduled carriers have mostly turbulence as the cause of the incident. In these events, there is no damage to the aircraft but to the people. Airlines should choose less-turbulent routes and reduce these cruise category events. Passengers record turbulent incidents, and their appearance on social media causes significant damage to the brand image. Flight Schools can use the tool to filter the airports and aircrafts similar to their specifications, comprehend training flight findings, and advance their operations. The air traffic control (ATC) schedule is very fragile and a bottleneck of all aviation operations. Utilizing this study to assess incident risks specified for that airport and make the ATC schedule accordingly so that the schedules become less vulnerable [85].

In addition, flight schools can improve the flight phases with their experienced instructors or search for what mistakes new PICs make. Aircraft Manufacturers use the decision support tool to test the specifications they plan to add to the aircraft by comparing the incidents with similar aircraft models that already have the specifications. Aircraft manufacturers also see the landscape of incidents as a summary in a BBN structure. Aviation Researchers build omnidirectional specific scenarios for various pilot expertise, aircraft model, and flight phase combinations. Similarly, they can add the TAN structure to visualize the research area they are interested. Insurance Companies use the DSS to test various combination events and the potential damage and injuries to find high-risk and low-risk flights to calibrate their pricing policies.

The limitation of the study is not combining the incident reports with other third-party incident investigation reports because the study targets incidents in an extensive timeframe and aims to find the patterns that build the big picture by combining BBN results with SMEs. We also did not consider NLP analysis in the expert comment column.

The study can be advanced by expanding the data set with other incident reports, such as NTSB data or third-party incident reports. Conducting a natural language processing (NLP) analysis and getting more information from the investigator report. In addition, the study can be combined with other expert guidance information about incidents, such as integrating with the airport runway blind-spot notes released by the FAA to find the root cause of related incident cases and understand the causality of such blind spot-related incident cases. Finally, many omnidirectional inferences update possibilities mean different possible variable combinations important for aviation stakeholders. These varying combinations mean other incident cases with free text format notes (categorical variables) are also possible combination changes in the flight equipment, pilot, or physical conditions of the flight. We only discussed a few of these crucial combinations in this study.

Declaration of Competing Interest

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

CRediT authorship contribution statement

Burak Cankaya: Conceptualization, Formal analysis, Investigation, Writing – original draft. **Kazim Topuz:** Methodology, Writing – original draft. **Dursun Delen:** Conceptualization, Methodology, Validation, Writing – review & editing. **Aaron Glassman:** Methodology, Validation, Writing – review & editing.

Data availability

Data will be made available on request.

References

[1] Akyildirim E, Corbet S, Efthymiou M, Guiomard C, O'Connell JF, Sensoy A. The financial market effects of international aviation disasters. Int Rev Finan Anal 2020;69:101468.
[2] Zuidberg J. Exploring the determinants for airport profitability: traffic characteristics, low-cost carriers, seasonality and cost efficiency. Transp Res Part A Pol Pract 2017;101:61–72.
[3] Nikolaykin NI, Nikolaykina NE, Sekerin VD, Gorokhva AE. Environmental and economic model of an aircraft accident evaluation. J Environ Manag Tourism 2017;8(5):21.
[4] Zhang X, Mahadevan S. Bayesian network modeling of accident investigation reports for aviation safety assessment. Reliab Eng Syst Saf 2021:209.
[5] Waller RA, Covello VT. Low-probability high-consequence. Risk analysis: issues, *methods, and case studies*. New York: Plenum Press; 1984.
[6] Thomson ME, Önkal D, Avcioğlu A, Goodwin P. Aviation risk perception: a comparison between experts and novices. Risk Anal Int J 2004;24(6):1585–95.
[7] Sieberichs S, Kluge A. Good sleep quality and ways to control fatigue risks in aviation—an empirical study with commercial airline pilots. In: Advances in physical ergonomics and human factors. Cham: Springer; 2016. p. 191–201.
[8] van Drongelen A, Boot CR, Hlobil H, Twisk JW, Smid T, van der Beek AJ. Evaluation of an mHealth intervention aiming to improve health-related behavior and sleep and reduce fatigue among airline pilots. Scand J Work Environ Health 2014:557–68.
[9] McFadden KL. Comparing pilot-error accident rates of male and female airline pilots. Omega (Oxford) 1996;24(4):443–50.
[10] Chang YHJ, Mosleh A. Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents: part 1: overview of the IDAC Model. Reliab Eng Syst Saf 2007;92(8):997–1013.
[11] Geldermann J, Bertsch V, Treitz M, French S, Papamichail KN, Hämäläinen RP. Multi-criteria decision support and evaluation of strategies for nuclear remediation management. Omega (Oxford) 2009;37(1):238–51.
[12] Cankaya B, Wari E, Eren Tokgoz B. Practical approaches to chemical tanker scheduling in ports: a case study on the port of Houston. Marit Econ Log 2019;21(4):559–75.
[13] Cankaya B, Wari E, Tokgoz BE. A chemical tanker scheduling problem: port of Houston case study. Int J Plann Sched 2019;3(1):47–67.
[14] Cankaya B, Tokgoz BE, Dag A, Santosh KC. Development of a machine-learning-based decision support mechanism for predicting chemical tanker cleaning activity. J Model Manag 2021.
[15] Kharoufah H, Murray J, Baxter G, Wild G. A review of human factors causations in commercial air transport accidents and incidents: from to 2000–2016. Prog Aerosp Sci 2018;99:1–13.
[16] Sieberichs S, Kluge A. Why learning opportunities from aviation incidents are lacking. Aviation Psychol Appl Human Factors 2021.
[17] Heinrich DJ. Safer* approaches and landings: a multivariate analysis of critical factors. Capella University; 2004.
[18] Li Y. Analysis and forecast of global civil aviation accidents for the period 1942–2016. Math Probl Eng 2019;2019.
[19] Koteeswaran S, Malarvizhi N, Kannan E, Sasikala S, Geetha S. Data mining application on aviation accident data for predicting topmost causes for accidents. Cluster Comput 2019;22(Suppl 5):11379–99. doi:10.1007/s10586-017-1394-2.
[20] Sun Y, Geng N. Forecasting civil aviation incident rate in China using a combined prediction model. J Adv Transp 2021:2021.
[21] Wang L, Yang Z. Bayesian network modelling and analysis of accident severity in waterborne transportation: a case study in China. Reliab Eng Syst Saf 2018;180:277–89.
[22] Gürbüz F, Özbakir L, Yapici H. Data mining and preprocessing application on component reports of an airline company in Turkey. Expert Syst Appl 2011;38(6):6618–26.
[23] Nazeri Z, Zhang J. Mining aviation data to understand impacts of severe weather on airspace system performance, IEEE. In: Proceedings. International Conference on Information Technology: Coding and Computing; 2002. p. 518–23.
[24] Li Y. Analysis and forecast of global civil aviation accidents for the period 1942-2016. Math Probl Eng 2019;2019.
[25] Bazargan M, Guzhva VS. Impact of gender, age and experience of pilots on general aviation accidents. Accident Anal Prevent 2011;43(3):962–70.
[26] Marais KB, Robichaud MR. Analysis of trends in aviation maintenance risk: an empirical approach. Reliab Eng Syst Saf 2012;106:104–18.
[27] Rehm F. Prediction of aircraft delay at frankfurt airport as a function of weather. Presentation from German Aerospace Center; 2004.
[28] Lukáčová A, Babič F, Paralič J. Building the prediction model from the aviation incident data. In: Proceedings of the IEEE 12th international symposium on applied machine intelligence and informatics (SAMI). IEEE; 2014. p. 365–9.
[29] Rao AH, Marais K. High-risk occurrence chains in helicopter accidents. Reliab Eng Syst Saf 2018;170:83–98.
[30] Baugh, B.S. (2020). Predicting General Aviation Accidents Using Machine Learning Algorithms.
[31] Abesamis PPR, Bulos RD, Ching M. Improving aviation incidents using association rule mining algorithm and time series analysis. IOP Conf Ser Mater Sci Eng 2020;946(1):12005. doi:10.1088/1757-899X/946/1/012005.
[32] Wang R, Miao Z, Liu T, Liu M, Grdinovac K, Song X, Paiva W. Derivation and validation of essential predictors and risk index for early detection of diabetic retinopathy using electronic health records. J Clinical Med 2021;10(7):1473.
[33] Davazdahemami B, Peng P, Delen D. A deep learning approach for predicting early bounce-backs to the emergency departments. Health Anal 2022;2:100018.
[34] Delen D, Zolbanin HM, Crosby D, Wright D. To imprison or not to imprison: an analytics model for drug courts. Ann Operat Res 2021;303:101–24.
[35] Delen D, Cogdell D, Kasap N. A comparative analysis of data mining methods in predicting NCAA bowl outcomes. Int J Forecast 2012;28(2):543–52.
[36] Kattan MW, Cooper RB. A simulation of factors affecting Machine Learning techniques: an examination of partitioning and class proportions. Omega (Oxford) 2000;28(5):501–12.
[37] Delen D, Ram S. Research challenges and opportunities in business analytics. J Bus Anal 2018;1(1):2–12.
[38] Potrawa T, Tetereva A. How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. J Bus Res 2022;144:50–65.
[39] Kuhn KD. Using structural topic modeling to identify latent topics and trends in aviation incident reports. Transp Res Part C Emerg Technol 2018;87:105–22.
[40] Zhang X, Srinivasan P, Mahadevan S. Sequential deep learning from NTSB reports for aviation safety prognosis. Saf Sci 2021;142:105390.
[41] van Giffen B, Herhausen D, Fahse T. Overcoming the pitfalls and perils of algorithms: a classification of Machine Learning biases and mitigation methods. J Bus Res 2022;144:93–106.
[42] Reis C, Ruivo P, Oliveira T, Faroleiro P. Assessing the drivers of machine learning business value. J Bus Res 2020;117:232–43.
[43] Goodfellow I, Bengio Y, Courville A. Deep learning. MIT Press; 2016.

[44] Zolbanin HM, Davazdahemami B, Delen D, Zadeh AH. Data analytics for the sustainable use of resources in hospitals: predicting the length of stay for patients with chronic diseases. Inf Manag 2022;59(5):103282.
[45] Yazici I, Beyca OF, Delen D. Deep-learning-based short-term electricity load forecasting: a real case application. Eng Appl Artif Intell 2022;109:104645.
[46] Davazdahemami B, Zolbanin HM, Delen D. An explanatory machine learning framework for studying pandemics: the case of COVID-19 emergency department readmissions. Decis Supp Syst 2022;161:113730.
[47] Pearl J. Causal inference in statistics: an overview. Stat Surv 2009;3:96–146.
[48] Pearl J. Interpretation and identification of causal mediation. Psychol Methods 2014;19(4):459.
[49] Topuz K, Uner H, Oztekin A, Yildirim MB. Predicting pediatric clinic no-shows: a decision analytic framework using elastic net and Bayesian belief network. Ann Operat Res 2018;263(1):479–99.
[50] Topuz K, Delen D. A probabilistic Bayesian inference model to investigate injury severity in automobile crashes. Decis Support Syst 2021;150:113557.
[51] Topuz K, Jones BD, Sahbaz S, Moqbel M. Methodology to combine theoretical knowledge with a data-driven probabilistic graphical model. J Bus Anal 2021;4(2):125–39.
[52] Mohammadi M, Rezaei J. Bayesian best-worst method: a probabilistic group decision making model. Omega (Oxford) 2020;96:102075.
[53] Simsek S, Dag A, Tiahrt T, Oztekin A. A bayesian belief network-based probabilistic mechanism to determine patient no-show risk categories. Omega (Oxford) 2021;100:102296.
[54] Kraus M, Feuerriegel S. Forecasting remaining useful life: interpretable deep learning approach via variational Bayesian inferences. Dec Support Syst 2019;125:113100.
[55] Chang Y, Wu X, Zhang C, Chen G, Liu X, Li J, Xu L. Dynamic Bayesian networks based approach for risk analysis of subsea wellhead fatigue failure during service life. Reliab Eng Syst Saf 2019;188:454–62.
[56] Arnaldo Valdés RM, Comendador VFG, Sanz LP, Sanz AR. Prediction of aircraft safety incidents using Bayesian inference and hierarchical structures. Saf Sci 2018;104:216–30.
[57] Ancel E, Shih AT, Jones SM, Reveley MS, Luxhøj JT, Evans JK. Predictive safety analytics: inferring aviation accident shaping factors and causation. J Risk Res 2015;18(4):428–51.
[58] Arnaldo Valdés RM, Liang Cheng SZ, Gómez Comendador VF, Sáez Nieto FJ. Application of Bayesian networks and information theory to estimate the occurrence of mid-air collisions based on accident precursors. Entropy 2018;20(12):969.
[59] Bandeira MCGSP, Correia AR, Martins MR. General model analysis of aeronautical accidents involving human and organizational factors. J Air Transp Manag 2018;69:137–46.
[60] AIDS system information. AIDS System Info Page. (n.d.). Retrieved February 3, 2022, from https://www.asias.faa.gov/apex/f?p=100%3A15%3A%3A%3ANO%3A%3AP15_REGION_VAR%3A1.
[61] Misiunas N, Oztekin A, Chen Y, Chandra K. DEANN: a healthcare analytic methodology of data envelopment analysis and artificial neural networks for the prediction of organ recipient functional status. Omega (Oxford) 2016;58:46–54.
[62] Plumb G, Molitor D, Talwalkar AS. Model agnostic supervised local explanations. Adv *Neural Inf Process Syst* 2018:2515–24.
[63] Rubin DB. Inference and missing data. Biometrika 1976;63(3):581–92.
[64] Afghari AP, Haque MM, Washington S. Applying a joint model of crash count and crash severity to identify road segments with high risk of fatal and serious injury crashes. Acc Anal Prevent 2020;144:105615.
[65] Gui G, Zhou Z, Wang J, Liu F, Sun J. Machine Learning aided air traffic flow analysis based on aviation big data. IEEE Trans Veh Technol 2020;69(5):4817–26.
[66] Heckerman D. A tutorial on learning with Bayesian networks. Innovat Bayesian Netw 2008:33–82.
[67] Olson DL, Delen D. Advanced data mining techniques. Springer Science & Business Media; 2008.
[68] Ordenes FV, Silipo R. Machine Learning for marketing on the KNIME Hub: the development of a live repository for marketing applications. J Bus Res 2021;137:393–410.
[69] van Valkenhoef G, Tervonen T. Entropy-optimal weight constraint elicitation with additive multi-attribute utility models. Omega (Oxford) 2016;64:1–12.
[70] Wright S. The analysis of variance and the correlations between relatives with respect to deviations from an optimum. J Genet 1935;30(2):243–56.
[71] Marks T. Useful risk management tools. In: The practitioner handbook of project controls. Routledge; 2020. p. 231–46.
[72] Coussement K, Phan M, De Caigny A, Benoit DF, Raes A. Predicting student dropout in subscription-based online learning environments: the beneficial impact of the logit leaf model. Decision Support Syst 2020;135:113325.
[73] Nadkarni S, Shenoy PP. A causal mapping approach to constructing Bayesian networks. Dec Support Syst 2004;38(2):259–81.
[74] Koller D, Friedman N. Probabilistic graphical models: principles and techniques. MIT press; 2009.
[75] Friedman N, Geiger D, Goldszmidt M. Bayesian network classifiers. Mach Learn 1997;29(2):131–63.
[76] Korb KB, Nicholson AE. Bayesian artificial intelligence. CRC press; 2010.
[77] Chow CKCN, Liu C. Approximating discrete probability distributions with dependence trees. IEEE Trans Inf Theory 1968;14(3):462–7.
[78] Benbassat D, Abramson CI. Landing flare accident reports and pilot perception analysis. Int J Aviat Psychol 2002;12(2):137–52.
[79] Gribkovskaia I, Halskau O, Kovalyov MY. Minimizing takeoff and landing risk in helicopter pickup and delivery operations. Omega (Oxford) 2015;55:73–80.
[80] Wang L, Ren Y, Wu C. Effects of flare operation on landing safety: a study based on ANOVA of real flight data. Saf Sci 2018;102:14–25.
[81] Sa CAA, Santos BF, Clarke JB. Portfolio-based airline fleet planning under stochastic demand. *Omega (Oxford)* 2020;97:102101.
[82] Rao AH, Marais K. High risk occurrence chains in helicopter accidents. Reliab Eng Syst Saf 2018;170:83–98.
[83] Jiang L, Yang P, Ma X, Yang H, Li T, Yang J. Comparison of detection technology for runway incursion prevention in airport hot spot. Journal of physics: conference series, 1570. IOP Publishing; 2020.
[84] Pohl M, Kolisch R, Schiffer M. Runway scheduling during winter operations. Omega (Oxford) 2021;102:102325.
[85] Samà M, D'Ariano A, D'Ariano P, Pacciarelli D. Scheduling models for optimal aircraft traffic control at busy airports: tardiness, priorities, equity and violations considerations. Omega (Oxford) 2017;67:81–98.