Liang, X, Fan, S, Lucy, J and Yang, Z

Risk analysis of cargo theft from freight supply chains using a data-driven Bayesian network

Reliability Engineering and System Safety 226 (2022) 108702

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Risk analysis of cargo theft from freight supply chains using a data-driven Bayesian network

Xinrui Liang [a], Shiqi Fan [a], John Lucy [b], Zaili Yang [a,*]

[a] *Liverpool Logistics, Offshore and Marine Research Institute, Liverpool John Moores University, Liverpool, UK*
[b] *Liverpool Combined Authority, Liverpool, UK*

ARTICLE INFO

Keywords:
Cargo theft
Risk analysis
Bayesian network
Big data
Accident analysis
Supply chains

ABSTRACT

Cargo theft has been among the most concerning risks influencing global freight supply chains, which causes serious supply chain disruptions, injuries/deaths, economic loss, and environmental damage. However, there are very few studies on the risk analysis of cargo theft, particularly in a quantitative manner, and fewer on the relevant risk factors affecting theft-related accidents in the current literature. This paper aims to analyse the risk influential factors (RIFs) of cargo theft and predict the occurrence likelihood of different types of cargo theft accidents. The historical data of 9316 cargo theft accidents that happened in the UK from 2009 to 2021 were first collected from the TAPA IIS database, and then purified and trained to construct a Bayesian network (BN) based cargo theft risk analysis model. The data-driven BN interprets the interdependency of RIFs and their combined effects on the occurrence of different types of cargo theft accidents. Compared with the previous studies, this paper makes new contributions, including that (1) The cargo theft RIFs are identified from the literature and accident records. (2) A data-driven BN is proposed to construct the model with uncertainty to realise cargo theft risk prediction and diagnosis. (3) The critical RIFs contributing to cargo theft are evaluated and prioritised to predict the occurrence of possible cargo theft accidents. (4) The real accidents are investigated to verify the model and draw useful insights for cargo theft prevention. The findings show that the most influential RIFs for the occurrence of cargo theft accidents are product category, year, location type, modus operandi (MO), and region. The findings also reveal the combined risk contributions of the RIFs, hence providing useful insights for cost-effective theft risk control in practice.

1. Introduction

Among all emerging supply chain risks, the statistic shows that cargo theft has caused increasing concerns. In the last two years alone, TAPA EMEA has recorded over 15,000 cargo losses from supply chains, with a total loss value of over €310m ($373m), which is the equivalent of €424,000 of goods being stolen every single day of 2019 and 2020 (TAPA EMEA, 2021). In 2020, cargo theft throughout the United States hit the highest recorded volume in the last five years, according to an annual report by Sensitech (2021). Cargo theft is becoming a global problem that must be well addressed to avoid financial loss and disruptions in supply chain operations [40]. When a single cargo theft accident occurs, the involved supply chain costs six times the cargo value because the accident affects the costs of product replacement, accident handling, increased insurance premium, loss of sales, and negative impact on the business reputation [8]. Along with the financial cost, cargo theft may also lead to injuries/death and environmental damage when it involves dangerous cargo and violence in operations. Given such a high-risk stake, the relevant research in cargo theft risk analysis has not been undertaken sufficiently and in a good proportion to the risk level.

Within the context of cargo theft, the risks are diversified, involving classical and nonclassical events. For instance, the spread of COVID-19 in 2020 brought increased and more specific theft targets on cargo such as personal protection equipment (PPE) and medicines [44]. For other types of cargo, cargo theft trends stay stable in 2020 compared to

Abbreviations: RIF, risk influential factor; TAPA, Transported Asset Protection Association; BN, Bayesian network; MO, modus operandi; PPE, personal protection equipment; CLSC, Container Line Supply Chains; TAN, Tree Augmented Naive Bayes; CPT, Conditional Probability Table; TRI, True Risk Influence; HRI, High-Risk Inference; LRI, Low-Risk Inference.

* Corresponding author.
E-mail address: z.yang@ljmu.ac.uk (Z. Yang).

https://doi.org/10.1016/j.ress.2022.108702
Received 13 April 2022; Received in revised form 20 June 2022; Accepted 27 June 2022
Available online 1 July 2022

the volatile records in previous years, despite the implementation of many preventive measures in practice. Although real-time monitoring devices attached to the cargo are used in practice, it is revealed not effective enough to reduce the interests and attempts of thieves targeting the cargo. Drone monitoring has been seen as a new solution to cargo security, but its applicability is arguable for certain types of shipments due to the high cost [33]. The literature related to the risk analysis of cargo theft is little in general and less in supply chains. Most studies focused on the countermeasure against cargo theft, while others investigated the characteristics of cargo theft such as the main causes, hot spots, and seasonal patterns. However, without understanding the influential factors of cargo theft, preventative measures against theft and the related resources to support the preventative measures will not be allocated systematically and efficiently [45].

Typically, cargo thieves seek the opportunity of stealing depending on time, location, and objective (cargo type). Besides, they may choose different methods to commit a crime in different scenarios such as breaking and entering a vehicle/truck/warehouse, forcing a vehicle to stop. In addition, the occurrence of cargo theft from supply chains is complicated to understand because it involves various uncertainties such as transportation modes, product types, locations, and facility types. Hence, the occurrence of cargo theft accidents is dynamic depending on the situations in which the relevant risk factors are presented in an interactive way.

To fill the research gap, this study aims to develop a data-driven risk analysis model for the diagnosis of the effect of relevant RIFs on cargo theft and prediction of the occurrence likelihood of different types of theft accidents. To achieve this aim, this paper firstly describes the identification of the RIFs influencing cargo theft from both the relevant literature and historical database. Secondly, it uses a data-driven BN approach to evaluate the effects of the identified RIFs on the occurrence of different categories of cargo theft accidents. Furthermore, the model is verified using multiple methods including a test using real cases, sensitivity analysis, and scenario analysis. In this sense, the accident data is collected from the UK as it presents the riskiest area in terms of cargo theft, evident by the fact that more than half of the accidents (i.e., 50.9%) reported to Transported Asset Protection Association (TAPA) in the first half of 2020 occurred in the UK [48].

The rest of the paper is structured as follows. In Section 2, the literature on cargo theft-related risk studies and influential factors is reviewed to define the state-of-the-art in the field. Section 3 describes the development and application of a new methodology including RIF identification, BN structure learning, and sensitivity analysis to address the research aim. Section 4 presents the model validation methods. The analysis and results are presented and discussed for insightful implications in Section 5, while the conclusion is drawn in Section 6.

2. Literature review

2.1. Risk studies on cargo theft

Among the limited studies on cargo theft in the current literature, the majority focused on the countermeasures against different types of cargo theft accidents and the others were related to the exploration of the nature of cargo theft including the probabilities of the accidents and the relevant influential factors. More specifically, the issues such as a prevention approach [46], a communication system [20,23,26,37,60], and vehicular technology [10] were addressed. In 2009, Ekwall analysed and explained the reason why cargo theft continued to occur despite all the implemented countermeasures. Since then, the awareness of the significance of investigating the nature of cargo theft has been growing. Even though most subsequent studies still focused on the development of technologies, tools, and systems against cargo theft, there are still a few studies undertaken to capture the risk characteristics of cargo theft accidents. They are often treated as a part of the broad discussion of supply chain security, including the impact of low-wage labor on supply chain security [3], seasonality of cargo theft [14], the effects of modus operandi (MO) and location type on cargo theft, risk assessment of cargo theft [15], key factors behind cargo loss severity in logistics systems [51], geographical concentration of cargo theft [24], prediction of the cargo theft probability in rail transport [33,34].

Despite the evolution of themes in cargo theft research, the state-of-the-art methodologies in the field are mainly based on qualitative and/or basic statistical methods to investigate cargo theft factors. In other words, very few studies involve quantitative methods, and from the applied research perspective, the cases in such studies often represent a single component of a whole supply chain. As a result, the current cargo theft risk studies have revealed significant limitations from empirical and methodological perspectives. To be specific, Tang et al. [45] used a hierarchical structure of criteria to evaluate the security levels against theft in a port storage area in Container Line Supply Chains (CLSC). Based on the structure, a belief Rule-base Inference Methodology using the Evidential Reasoning (RIMER) algorithm was applied to handle the various kinds of uncertainties involved during the evaluation process and generate the evaluation result. Using a hierarchical structure to model the risk factors/variables of cargo theft can easily overlook the interdependency among the factors/variables and hence affect the model validity. Wu et al. [51] utilized data-driven business analytics involving descriptive, predictive, and prescriptive analysis to investigate cargo loss severity in logistics systems based on the data from an electronics company. Again, it overlooked the interdependency among the factors/variables, and thus the reflection of the result to the reality became questionable. Song et al. [43] used a data-driven approach to predict the theft risk of bulk cargo in ports based on the data from Guangzhou Port Group and Guangzhou Port Security Bureau in China. Various binary classifiers including OneR, Decision Tree (DT), Random Forest (RF), Naïve Bayesian (BN), and BN were compared, and the result showed that BN was a suitable predictive model. However, the BN structures derived from two different structure-learning algorithms were different, requiring subjective knowledge to configure the final structure. In addition, the results could not reflect the effects of multiple states of the identified risk factors. Lorenc and Kuznar [33] used Artificial Neural Networks (ANN) and Machine Learning (ML) methods to predict the probability of cargo theft in railway transport, respectively. Although showing some attractiveness, the methods failed to disclose the joint significance of multiple risk factors and their interdependency, leading to limited insights on prevention measures development.

Clearly, previous studies have revealed some theoretical implications on quantitative cargo theft risk analysis that have not been well addressed in the current literature and they could not be achieved without the analysis of the interdependency of the RIFs from a whole supply chain perspective. To fulfill this gap, this study aims to develop an advanced quantitative method to analyse the interdependence among the RIFs of cargo theft and pioneer a risk analysis model to realise the cargo theft risk prediction and diagnosis.

2.2. Risk factors influencing cargo theft

A cargo theft accident could occur in any part of a freight supply chain along with the cargo flows. However, the occurrence of cargo theft accidents in terms of time, place, MO, and some other factors follows certain rules to be explored. It is therefore crucial to identify and analyse the relevant RIFs. To do so, 92 relevant papers published from 1970 to 2021 were first found by searching the keyword “cargo theft” on the Web of Science. Secondly, book chapters were excluded. By the screen of titles, abstracts, and conclusions, we also excluded the papers (1) that addressed the development of security means and systems against cargo theft and (2) that focused on the evaluation of logistics performance without discussing the causes of cargo theft. As a result, 28 papers are finally selected, among which 22 risk factors appear frequently and are chosen for further analysis. The identified factors and their appearance frequencies are shown in Fig. 1. Moreover, such factors are analysed at

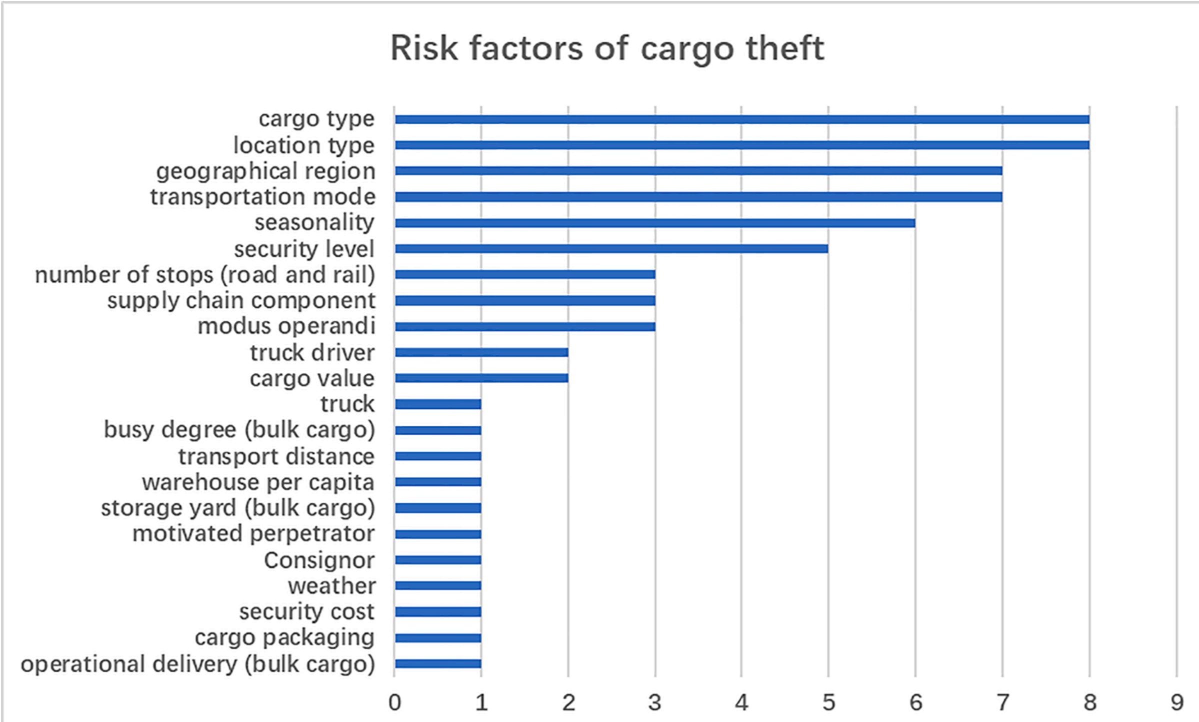

Fig. 1. Risk factors from the literature.

different levels in the selected literature. We use class I to represent the factors if their impacts are evidentially evaluated using mathematical methods and class II to represent the factors that appear in the selected literature just to support the research background or used in a specialized segment (i.e., bulk cargo). With reference to this classification, Table 1 shows how the 22 risk factors are analysed from the 28 references, and numbers 1 to 22 are used to represent the 22 risk factors in the front row. Obviously, some factors appear in both class I and class II because they receive different levels of analysis in different literature. There are 8 factors out of the 22 factors that are analysed at the level of class I and these 8 factors and their appearance frequencies as class I factors are shown in Fig. 2. The rest of this section summarizes how these important RIFs are described in the selected literature.

Cargo type. Cargo type is one of the most frequently observed influential factors influencing cargo theft in terms of the occurrence likelihood of accidents and consequences (e.g. stolen value). Ekwall [17] found that the thieves target on the type of goods more than anything else such as those relating to the theft opportunities exposed in a transport network. In other words, perpetrators tend to change different MOs to target the same product that they are interested in. According to 2021 data from CargoNet, the prime targets of thieves are electronics amid the chip shortage in the world and refrigerated food. BSI-recorded cargo thefts of medical devices and supplies, including PPE, jumped by over 5000% in 2020 compared to 2019 due to the Covid-19 pandemic.

Location type. Location type is one of the two most frequently used risk factors in cargo theft studies. Cargo can be stolen when it takes a stop in some places such as warehouses, terminals, equipment, and truck stops. Besides, trailers and containers have become virtual warehouses on wheels and easy targets for thieves with the Just-In-Time delivery replacing the on-hand inventory of most businesses [46]. 97% of all attacks during a stop occur at non-secure parking locations [12]. Cargo thefts at these locations are more of a volume crime than high-value thefts according to the TAPA EMEA data. The risk levels of different combinations of location types and accident categories in terms of both impact and probability were examined [15]. According to the BSI and NMU cargo theft report of Q1 2021, a wide variety of tactics were involved in cargo thefts throughout Europe. The United Kingdom, Germany, Russia, Italy and France generally record some of the greatest numbers of thefts in the region. As noted at the beginning of the outbreak of the COVID-19 pandemic, a higher-than-usual number of thefts continue to occur from warehouses and facilities. As a result of disruptions to movement caused by the pandemic, stockpiled goods and trucks parked outside of warehouses and facilities became more accessible targets for thieves.

Seasonality. The seasonal variation in theft accidents was observed during particular months of the year and days of the week for many location types along transport chains [14] and MOs [13]. The seasonal effect was also observed in cargo theft accidents that occurred in the São Paulo State of Brazil (Justus et al., 2018). Nevertheless, the patterns depend on different categories, e.g., the variation over a year is approximately the same for all location types, while the variation over a week is different [14]; the seasonal effect on violent cargo thefts is evident to be small [16].

Geographical region. There is ample evidence that the nature of cargo theft differs among geographical regions. Cargo theft involving violence is rare in the United States, however, violence (such as intrusion, pilferage, and hijackings) is more common in Europe. In Mexico, cargo theft is an extremely violent crime occurred by gangs (Burges, 2013). In Brazil, it mainly occurs in the most economically dynamic regions, such as the states of São Paulo, Minas Gerais, and Rio de Janeiro. Although São Paulo's capital shows the highest levels of cargo theft, it is in non-metropolitan areas that records of this offense are on the rise (Justus et al., 2018). Based on the data from a case company, the research by Wu et al. (2016) found that when products were shipped using sea transport to Australia or the Middle East, cargo loss of medium severity was likely to occur.

Transportation mode. Cargo theft occurs while it is in the logistical cycle of being transported by a mode of transportation [46]. Among the critical logistics factors (transit types, product categories, and shipping destinations) influencing the severity of cargo loss, transit type was determined to be the most influential factor in the severity model (Wu et al., 2016). The case company investigated by Wu et al. (2016) suffered

Table 1
References of the identified risk factors.

Refs.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
[3]											II											
[4]								II														
[8]		II																				
[7]																				II		
[14]		I			I		I															
[12]		I			I																	
[13]					I		I															
[15]		I																				
[16]	I	I		I			I															
[11]			I			I																
[17]	I					I																I
[22]																		II				
[24]	I		I		I																	
[25]				II																		
[27]				II																		
[28]		II																				
[34]	II								II	II									II			
[35]				II																		
[63]	II	II	II	II	II	II			II													
[33]	II	II	II	II	II	II			II													
[38]			II																			
[39]			II					II														
[42]										II												
[43]											II	II	II	II	II	II	II				II	
[45]						I																
[46]	II																					
[29]								II														
[51]	I		I	I																		

Note: I is class I, II is class II, * for factors once appear in class I, ** for factors only appear in class II.
1.
cargo type* 2.
location type* 3.
geographical region* 4.
transportation mode* 5.
Seasonality* 6.
security level* 7.
MO*
8.
supply chain component** 9.
number of stops (road and rail)** 10.
cargo value** 11.
truck driver** 12.
cargo packaging** 13.
Weather** 14.
Truck** 15.
operational delivery (bulk cargo)** 16.
Consignor** 17.
busy degree (bulk cargo)** 18.
warehouse per capita** 19.
transport distance** 20.
security cost** 21.
storage yard (bulk cargo)** 22.
motivated perpetrator

cargo loss (claim payments) during air, sea, and land transportation. The primary cargo loss in terms of loss value was correlated with sea transport, followed by air transport and truck transport. However, the occurrence of cargo loss accidents by air transport was much higher than that by sea transport.

Security level. Transport security means the measures to prevent both terrorist attacks and ordinary crime, especially theft (EUR, 2005). Tang et al. [45] studied on security evaluation of a port storage area against theft in CLSC, stating that security analysis is critical in CLSC operation as CLSC is a dominant way to transport cargo worldwide and at the same time it is also subject to many threats.

Other factors. Previous studies have also identified other factors influencing the occurrence of cargo theft accidents. Based on the theory of crime displacement, Ekwall [17] identified the three elements of cargo theft including the motivated perpetrators, transported goods (object), and preventive measures. Furthermore, MOs for cargo theft exhibit seasonal patterns by time of the year and day of the week [13]. Song et al. [43] identified the influential factors of bulk cargo theft such as truck drive, truck type, weather, cargo packaging, storage yard type, consignor, and operational setting.

3. Methodology

To identify the RIFs influencing cargo theft occurrences and assess the importance rankings of RIFs, this study uses a data-driven BN method to train and learn the big cargo theft data from TAPA. The flowchart of the methodology is presented in Fig. 3. Firstly, the data on cargo theft accidents that happened in the UK is collected from TAPA and a necessary process of data management and purification is conducted. Secondly, the identified RIFs of cargo theft from the cleaned data

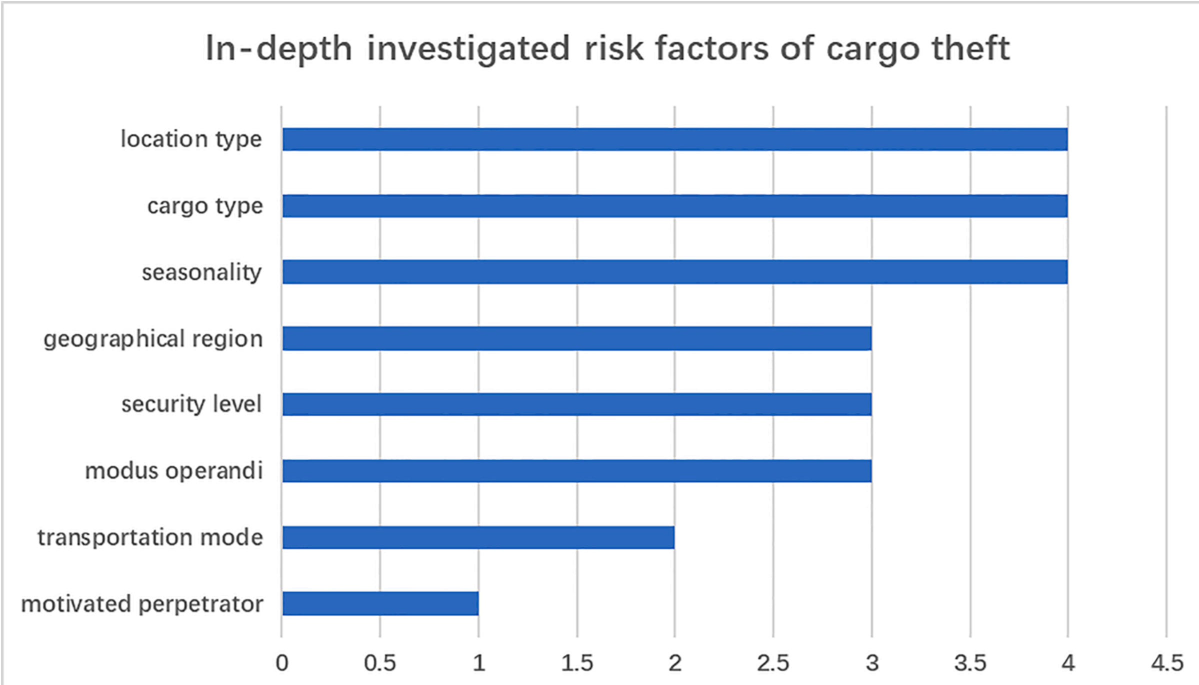

Fig. 2. Risk factors investigated in-depth from the literature.

according to the TAPA accident reports form of TAPA are verified by the knowledge gained from the literature. Thirdly, the identified RIFs and the datasheet are used as inputs to construct the model using the data-driven BN approach. Next, the model is validated in terms of its predicting ability and consistency. In this process, the real case test and sensitivity analysis are undertaken, and further, from the sensitivity analysis, the results of the importance rankings of RIFs, their interrelationships, and the effects of their multiple states are obtained for research implications. Finally, the results are presented and discussed thoroughly.

3.1. Data collection and cleaning

Compared to the one-year accident data used to support most of the previous studies in the field, however, 20,270 reported cargo theft accidents in the UK ranging from 2009 to 2021 have been collected from TAPA EMEA IIS to support the analysis in this paper. For each reported accident, entries can be made for the date of the accident, geographical location (including region, town, and district), location type (e.g., destination facility), type of accident (e.g., truck theft), type of MO, product category, loss value, major accident (by yes or no), attempt (by yes or no), last-mile delivery (by yes or no), and accident description. Given the fact that many accidents contain incomplete information such as unspecified products and unknown accident categories, a data cleaning process is conducted to ensure the data completeness and the accuracy of the developed model. Finally, 9316 accidents containing all complete data are used in this study. 8386 accidents (90%) are randomly chosen and used to build the model, while the other 930 accidents are reserved to test the model for its validation.

3.2. RIF identification

In the process of data purification, one variable 'last mile delivery' is removed because the character has been recognized and the relevant data has only been available since 2019. As a result, 9 RIFs influencing 'accident category' are identified, including major accident, attempt, MO, location type, product category, weekday, region, month, year, among which weekday, month, year are derived from the date column in the accident report.

Compared to the in-depth investigated RIFs in the previous literature (as presented in Fig. 2), there is a high harmony between the identified RIFs. To be specific, seasonality is related to month and weekday; the transportation mode is associated with the accident categories; security level is not incorporated in this study because of the lack of well-established definition and globally acceptable standards. Attempts and major accidents are additional RIFs identified from the accident reports. According to TAPA's explanation, an attempt means the act of trying to steal cargo/load/shipment unsuccessfully, while a major accident is defined as the one causing a loss value of over €100k.

Moreover, each RIF has various states. Table 2 shows the states of the 'accident category' and the 9 RIFs. This study uses the same definitions of states adopted by TAPA found online (https://tapaemea.org/iis-key-glossary). The states having very low percentages of the 9316 accidents are combined and categorized as new state 'other' because they are not of any critical mass statistically.

3.3. Model construction

Among various risk assessment methods, BN has attracted increasing interest owing to its advances in learning and inferencing. It combines visualization with mathematical knowledge and can help to analyse the importance of variables and the relationships among them, given the uncertainty in a system [53,62]. It has been widely used in the transportation area for risk factors analysis [18,31,47,52,55], and is increasingly popularised in recent years due to its aforementioned advantages (e.g. [9,32,36,49,50,57,59,61]). The BN structure can be constructed based on subjective and/or objective methods. This study relies on a data-driven method to build the BN structure using Tree Augmented Naive Bayes (TAN). Let $A_1, \ldots A_n$ be the risk variables, where n stands for the number of variables, TAN structure learning is the procedure of finding a tree defining function π over $A_1, \ldots A_n$ to maximize the log-likelihood. This procedure follows the general outline proposed by Chow and Liu [5]. Among various forms of Bayes network classifiers, Naive Bayes is the simplest and is competitive with other classifiers such as C4.5 [21]. However, its conditional independence assumption among features cannot well reflect the reality, which makes

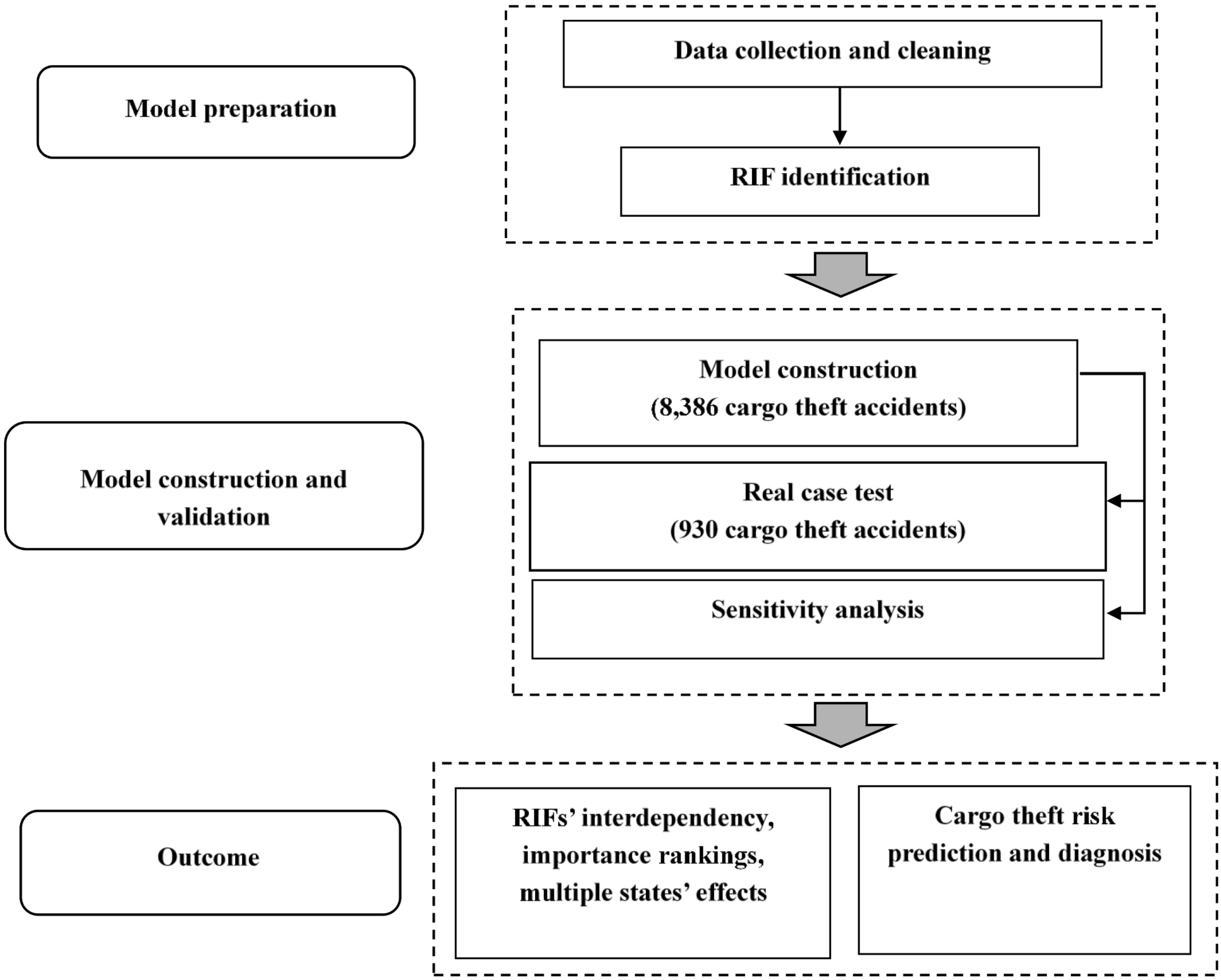

Fig. 3. Flow chart of data analysis.

a severe limitation on its application in empirical studies. TAN relaxes the independence assumption of naive Bayes, yet at the same time maintains the computational simplicity and robustness of naive Bayes [21]. One characteristic that differentiates the TAN model from the traditional BN lies in the class variables. Each class variable in the BN model must have at least one parent node. However, the links can go in either direction using Bayesian inference on the results to reflect reality [53]. Because of this superiority, TAN has been increasingly used to train big data to formulate BN risk models in transport (e.g., [18,53,54,58]).

Once the data is obtained and cleaned, the structure of BN can be generated through the process of TAN learning with the assistance of the Netica software. As a result, a new cargo theft risk BN model containing 10 nodes is formulated. The originally obtained structure is shown in Fig. 4, the links can go in either direction to fit the result in the reality.

Based on the TAN model, the Conditional Probability Tables (CPTs) of the involved nodes are then learned. Fig. 5 presents the results of TAN. It indicates that 'theft from vehicle' is the most frequent accident type, accounting for 64.2% of all accident categories, followed by truck theft and theft of vehicles accounting for 20.3% and 6.26%, respectively.

4. Model validation

The developed model is validated by three means, including (1) the comparative analysis of the historical statistics and the predicted results learned through 8,386 cargo theft accidents; (2) the real case tests using the reserved 930 cargo theft accidents; and (3) the logic inference validation by sensitivity analysis to see if the risk prediction results reflect the reality within the context of cargo theft.

4.1. Comparative analysis

The results of TAN have shown a very high reliability when compared to the historical statistics as shown in Table 3. To be specific, the predicted probability of 'truck theft' is the same with historical data (20.27%); the differences are 0.04% in 'theft from vehicle', and 0.01% in each other accident category. The very small variations are possibly caused by the introduction of the new state 'other'. It proves the prediction accuracy of the built model.

4.2. Real case tests

This study uses real cases to test the proposed model. A confusion matrix (see Appendix A) is generated to compare the prediction results with the true values of accident categories of real cases. Moreover, the kappa statistic is used to test the model consistency.

1) Prediction ability

Table 2
States of variables.

Variable	States	State - 'Other'
Accident category	Theft from Container/Trailer, Theft from Facility, Theft from Vehicle, Theft of Container/Trailer, Theft of Vehicle, Truck Theft, Other	Clandestine, Fraud, Hijacking, Robbery, Theft, Theft from Train
Year	13 years from 2009 to 2021	
Month	12 months	
Weekday	7 days	
Region	East Midlands, East of England, London, North East, North West, Northern Ireland, Scotland, South East, South West, Wales, West Midlands, Yorkshire and the Humber	
Product category	Clothing & Footwear, Food & Drink, Miscellaneous, No Load, Tobacco, Other	Agricultural Materials, Bicycles, Car parts, Cash, Computers/Laptops, Cosmetics & Hygiene, Furniture/Household Appliances, Jewellery/Precious Metals, Metal, Pharmaceuticals, Phones, Sports Equipment, Tools/Building Materials, Toys/Games, Tyres
Location type	Destination Facility, En Route, Origin Facility, Unclassified Parking, Other	Authorized 3rd Party Facility, Aviation Transportation Facility, Maritime Transportation Facility, Railway Operation Facility, Road Transportation Facility, Secured Parking, Services 3rd Party Facility,
Modus operandi (MO)	Intrusion, Theft from Moving Vehicles, Violent & Threat with Violence, Other	Internal, Forced Stop, Deceptive Stop, Deceptive Pick-up, Deception Other
Attempt	No, Yes	
Major accident	No, Yes	

930 accidents (10%) were reserved by random selection from the original database and used to test the prediction ability of the model, resulting in an overall accuracy rate of 89.14%. According to the confusion matrix in Appendix A, the prediction accuracy rates are 96.33% in 'theft from vehicles', and 97.91% in 'truck theft' by counting the number of correctly predicted accidents out of the actual accidents. Compared to the previous studies using BN in risk prediction [43,53], our result indicates that the model is robust for predicting the accident category of a cargo theft accident.

2) Kappa statistic for model consistency test

Kappa coefficient (k) was introduced by Cohen [6] as a statistic to measure the agreement between two raters. It has been applied in many fields and has been used in this study to measure the agreement between the predicted results and the real results. The definition of k is:

$$k = \frac{p_0 - p_e}{1 - p_e} \tag{1}$$

where p_o is the relative observed agreement between raters, and p_e is the hypothetical probability of chance agreement, using the observed data to calculate the probabilities of each observer randomly seeing each category. To calculate the k value for our confusion matrix, p_o is the sum of the correctly classified accidents divided by the total number of accidents. There are four steps to calculate p_e, including (1) multiplication of the marginal frequency for a certain accident type by the classifier (the sum of the predicted 'Other' accidents) and the marginal frequency for the same accident type by the true value (the sum of the actual 'Other' accidents), (2) division of the multiplied result from Step 1 by the total number of accidents, (3) repetition of the calculations in Steps 1 and 2 for each other accident type, and (4) division of the sum of values from the first three steps by the total number of accidents.

Therefore, the k value for the confusion matrix in Appendix A is calculated as follows:

$$p_e = (19 \times 25 + 16 \times 37 + \ldots + 190 \times 191)/930 \times 930 = 0.4839, \; p_0 = 0.8914$$

$$k = (0.8914 - 0.4839)/(1 - 0.4839) = 0.7896$$

Although there is not a standardized interpretation of the kappa

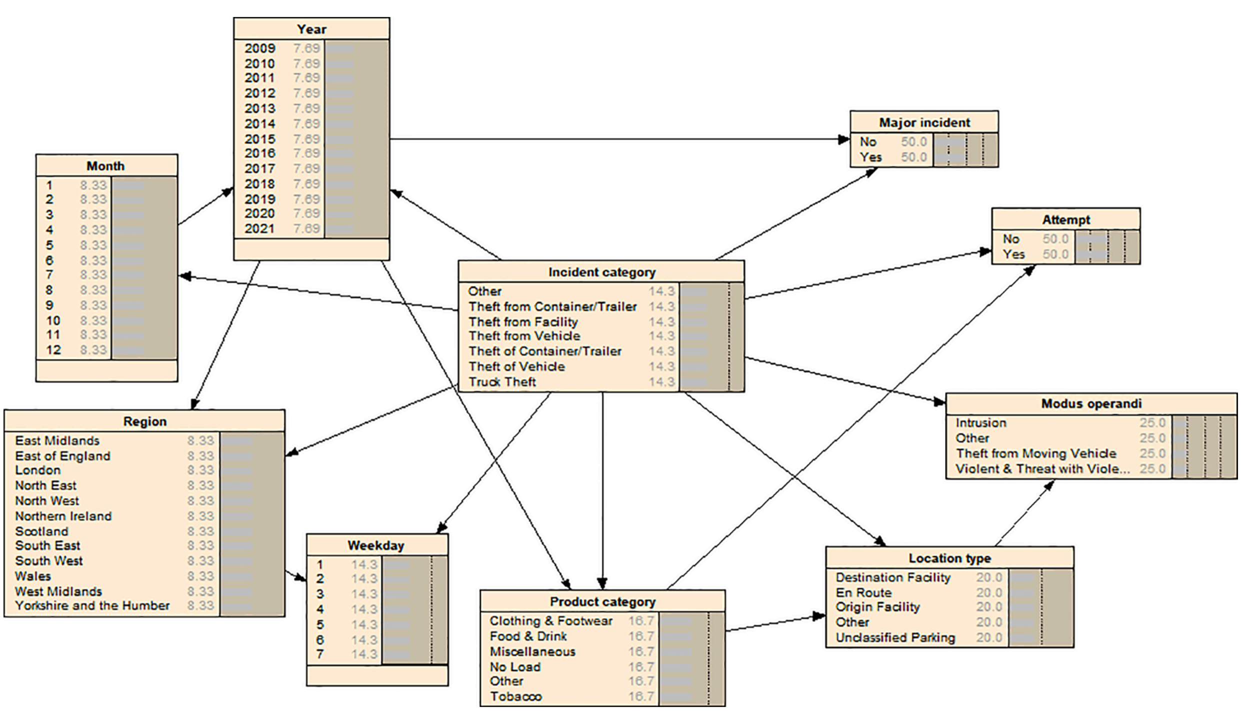

Fig. 4. TAN structure for theft accident category.

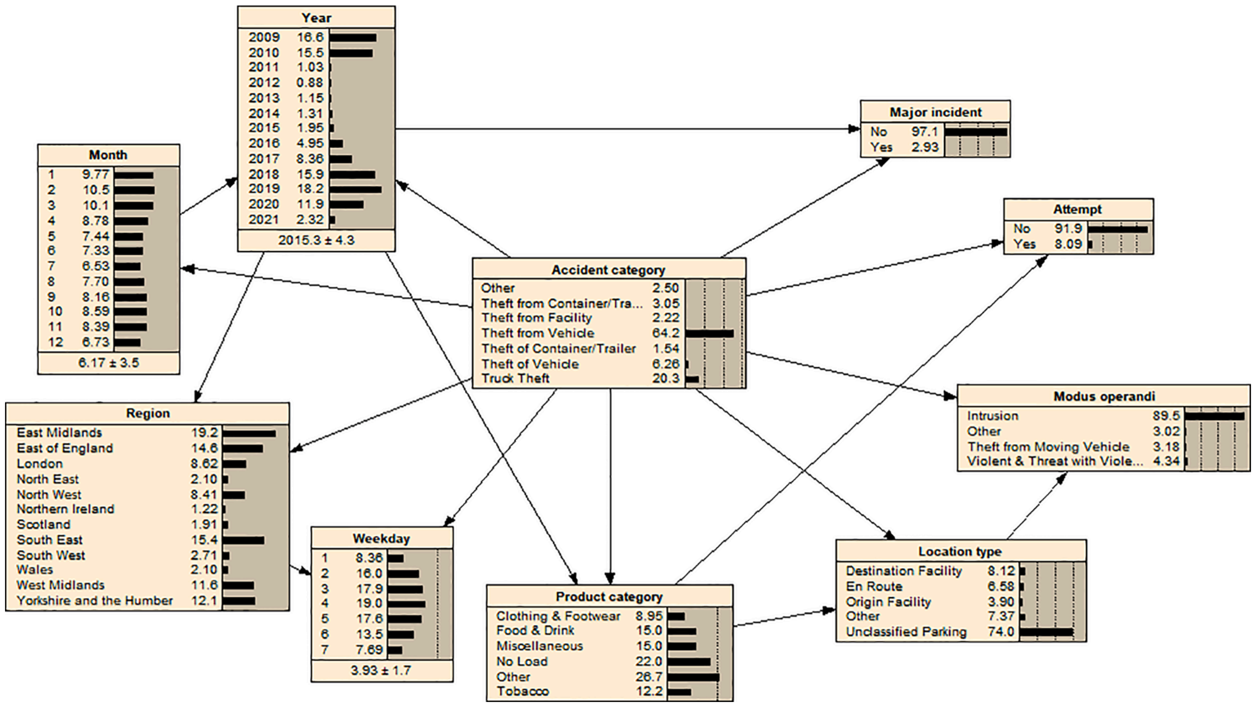

Fig. 5. Results of TAN.

Table 3
Comparative results of the historical data and TAN.

Accident category	Historical data (%)	Results of TAN (%)
Other	2.49	2.50
Theft from Container/Trailer	3.04	3.05
Theft from Facility	2.21	2.22
Theft from Vehicle	64.21	64.17
Theft of Container/Trailer	1.53	1.54
Theft of Vehicle	6.25	6.26
Truck Theft	20.27	20.27
Grand Total	100	100

statistic, a kappa (k) of 0.7896 indicates a strong strength of agreement according to Altman [1]. Landis and Koch [30] consider 0-0.20 as slight, 0.21-0.40 as fair, 0.41-0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1 as almost perfect. Further, according to Fleiss [19], 0.7896 (>0.75) is excellent.

4.3. Sensitivity analysis

To measure the dependence between accident category and RIFs and validate the model, the sensitivity analysis in this study is conducted based on mutual information, True Risk Influence (TRI) [2], and a joint probability. Besides, a sensitivity analysis can also help validate the model [56,62].

Mutual information. The concept of mutual information is intimately linked to that of entropy. The entropy of a random variable represents the average level of "information", "surprise", or "uncertainty" of its possible outcomes. The concept of information entropy was introduced by Claude [41]. Mutual information is the reduction of uncertainty about a variable, quantifying the amount of information obtained about one random variable based on the other ones. Therefore, mutual information is used in this study to measure the mutual dependence between the 'accident category' and RIFs, it can be defined as:

$$I(S,\beta) = -\sum_{s,i} P(s,\beta_i)\log_b \frac{P(s,\beta_i)}{P(s)P(\beta_i)} \tag{2}$$

where S represents 'accident category' of cargo theft, β represents a random RIF (e.g. location type), β_i represents the ith state of β, $I(S,\beta)$ represents the mutual information between accident category and RIFs. The RIFs having higher values of mutual information with the accident category are considered as more essential RIFs influencing the accident category of cargo theft. Thus, the overall importance ranking of RIFs can be obtained (see Table 4). When 'accident category' is the target node, the 'percentage' column in the table indicates the extent to which each RIF influences the 'accident type'. For instance, the influence level of the 'accident category' on itself is 100%. It can be seen from the 'mutual info' column, that the most essential factor among all RIFs influencing the 'accident category' is the 'product category', with a mutual information value of 0.55028.

True Risk Influence (TRI). TRI as a new method of sensitivity analysis was proposed by Alyami et al. [2]. In nature, the index is generated by the average of the highest and lowest possible influence of a variable on the target node in the investigated risk-oriented BN. It is used in this study because of its ability to evaluate the risk impacts of RIFs in multiple states. Specifically speaking, there are four steps to calculate the value of TRI of a random RIF (e.g., product category) with respect to an accident category (e.g., truck theft). Firstly, it is to increase the probability of each state of a selected RIF (e.g. each product category) to 100%, respectively. Secondly, it is to identify the two states (product types) generating the highest and the lowest probabilities of truck theft,

Table 4
Mutual information of 'accident category'.

Node	Mutual Info	Percentage (%)	Variance of Belief
Accident category	1.6286	100	0.3547
Product category	0.5504	33.80	0.1096
Year	0.3810	23.40	0.0632
Location type	0.1844	11.30	0.0090
MO	0.1289	7.91	0.0069
Region	0.0933	5.73	0.0148
Month	0.0429	2.63	0.0043
Major accident	0.0303	1.86	0.0014
Weekday	0.0260	1.60	0.0043
Attempt	0.0072	0.44	0.0001

respectively. Thirdly, it is to calculate the absolute difference value between the highest probability (86.60) generated from the second step and the original probability (20.3) of truck theft to obtain the High-Risk Inference (HRI) value; and calculate the absolute difference value between the lowest probability (0.92) generated from the second step and the original probability (20.3) of truck theft to obtain the Low-Risk Inference (LRI) value. Lastly, it calculates the TRI (42.84) of the product category for truck theft by taking the average value of HRI (66.30) and LRI (19.38). The RIFs with higher TRI values have stronger impact on the investigated accident category. Therefore, the importance rankings of RIFs for different accident categories can be generated. According to the above procedure, the TRI values of 'product category' for 7 different accident categories are calculated by adjusting the probability of each product type to 100%, respectively, as displayed in Table 5. Scenarios 1-6 represent the results of adjusting the probabilities of the six product types to 100%, respectively. A similar procedure is then applied to other RIFs. Eventually, TRI values of all RIFs for the 7 accident categories are obtained and displayed in Table 6. Accordingly, Table 7 shows the importance rankings of RIFs in each investigated accident category, it is obvious that the influence of a RIF on cargo theft varies with the accident type. For instance, 'product category' is the most important RIF for theft from vehicles and truck theft, it is less important for other accident types though.

Furthermore, the network joint probability is generated (as presented in Table 8) to reflect the states' effects of RIFs and enable the analysis of the joint effect of multiple RIFs on accident categories. Let X, Y represent a random RIF and accident category, respectively, X_i represents the ith state of X, Y_j represents the jth state of Y. The joint probability that events X_i and Y_j both occur is calculated by:

$$P(X = X_i,\ Y = Y_j) = P(X = X_i)P(Y = Y_j | X = X_i) \tag{3}$$

where $P(X{=}X_i)$ is the prior probability of the ith state of a random RIF X, and $P(Y{=}Y_j|X{=}X_i)$ is the conditional probability that the jth state of accident category Y occurs given that the ith state of X has already occurred. The highest joint probability value in each column indicates the most influential state for a particular accident category. For instance, for theft from vehicles, 'tobacco' is the most targeted product (88.2%) and 'en route' is the most influential location type. In each column, both the highest and the lowest values are highlighted as bold and italic values. Thus, it helps understand the influence level of each state on various accident types compared to other states. More analytical results are to be found in the next section.

5. Result discussion and implications

5.1. Analytical results

The overall ranking of risk impacts of RIFs on cargo theft accident category shows that 'product category' is the most important RIF out of the 9 RIFs, followed by 'year', 'location type', 'MO', 'region', and the other four RIFs ('month', 'major accident', 'weekday', 'attempt'). Furthermore, the essential RIFs and their significant states with respect to each accident type are evidentially evaluated. The product category is the most important RIF for the overall accident category mainly because of its significant effects on theft from vehicles and truck theft. Furthermore, 'tobacco' is the most targeted product in theft from vehicles with a joint probability of 88.2% being the highest among all product categories. Besides, a 'no-load' truck is more attractive (86.6%) than a 'no-load' vehicle/container/trailer/facility.

Location type has significant impacts on many accident types including theft from facility, theft of container/trailer and vehicle, theft from container/trailer and vehicle. Regarding the significant states of location type, the most contributed location type of theft from facilities for example is 'origin facility'. The most influential location type of theft of vehicles and theft from vehicles are 'destination facility' and 'en route', respectively, which indicates that it is most likely for vehicles to be stolen at destination facilities, and for cargoes to be stolen in motion. However, direct statistics show that 'unclassified parking' is the riskiest location type accounting for around 75% of all the investigated cargo theft accidents. A possible explanation is that the correlations between location type and other RIFs (i.e., product category, MO) have more contributions compared to its direct contribution to each accident category. A similarity applies to the states' effects of MO, although 90% of the accidents use 'intrusion' according to direct statistics, 'intrusion' is not the most effective MO for the investigated accident types except truck theft.

The region is the fifth most important RIF influencing the occurrence likelihood of cargo theft accidents, it still reveals some useful information. For example, in 'East Midlands', the probabilities of theft from facility (1.29%), theft of vehicle (2.47%), and truck theft (8.8%) are the lowest among all regions in the UK, while the risk of theft from vehicle (82.3%) is the highest. On the contrary, in 'Northern Ireland', the risk of theft from vehicle is the lowest in the UK, while the risks of other accident types are higher than that in most of the other regions. Previous studies have also discussed the dynamic character of cargo theft occurrences in geographical regions [8,24]. While this study further investigates this character by differentiating the accident categories of cargo theft.

Month, major accident, weekday, and attempt are less significant RIFs than the other five RIFs. Overall, the seasonal pattern in terms of the month of a year and the day of a week is insignificant. Fig. 6 displays the trends of probabilities of different cargo theft accident categories in months and weekdays. Looking at the most frequent accident category i.e., theft from vehicle, the riskiest months are 'October' and 'November', and the peak days during the week are 'Tuesday to Friday'. For other accident categories e.g., theft from/of container/trailer, theft from

Table 5
TRI of product category for all accident categories.

	Scenario									
Product category	Original	1	2	3	4	5	6			
Clothing & Footwear	8.95	100.00	0.00	0.00	0.00	0.00	0.00			
Food & Drink	15.00	0.00	100.00	0.00	0.00	0.00	0.00			
Miscellaneous	15.00	0.00	0.00	100.00	0.00	0.00	0.00			
No Load	22.00	0.00	0.00	0.00	100.00	0.00	0.00			
Other	26.70	0.00	0.00	0.00	0.00	100.00	0.00			
Tobacco	12.20	0.00	0.00	0.00	0.00	0.00	100.00			
Accident category								HRI	LRI	TRI
Other	2.50	2.86	3.15	3.15	0.89	2.72	3.08	0.65	1.61	1.13
Theft from Container/Trailer	3.05	3.74	2.79	6.07	1.04	3.28	2.27	3.02	2.01	2.52
Theft from Facility	2.22	3.19	2.40	2.48	0.52	3.53	1.14	1.31	1.70	1.51
Theft from Vehicle	64.20	82.60	82.60	67.40	7.20	81.80	88.20	24.00	57.00	40.50
Theft of Container/Trailer	1.54	2.09	2.30	1.37	1.26	1.47	1.04	0.76	0.50	0.63
Theft of Vehicle	6.26	2.86	5.21	18.00	2.54	6.26	2.35	11.74	3.91	7.83
Truck Theft	20.30	2.64	1.57	1.58	86.60	0.92	1.94	66.30	19.38	42.84

Table 6
TRI of all RIFs for all accident categories.

	Product category	Year	Location type	MO	Region	Month	Major accident	Weekday	Attempt
Other	1.13	5.13	3.13	15.06	3.48	1.00	7.42	0.46	0.14
Theft from Container/Trailer	2.52	10.09	4.19	0.69	3.12	2.18	4.20	0.66	3.46
Theft from Facility	1.51	7.02	10.05	1.93	3.96	0.98	4.53	1.48	0.25
Theft from Vehicle	40.50	31.90	32.40	33.80	31.16	12.45	23.90	12.65	0.00
Theft of Container/Trailer	0.63	3.44	4.91	1.39	2.71	0.97	3.74	1.47	0.11
Theft of Vehicle	7.83	10.77	12.86	13.53	6.47	7.02	3.99	3.50	1.00
Truck Theft	42.84	32.41	15.62	9.27	16.40	10.22	0.05	7.70	3.00

Table 7
The importance rankings of RIFs for the accident categories.

	Product category	Year	Location type	MO	Region	Month	Major accident	Weekday	Attempt
Other	6	3	5	1	4	7	2	8	9
Theft from Container/Trailer	6	1	3	8	5	7	2	9	4
Theft from Facility	6	2	1	5	4	8	3	7	9
Theft from Vehicle	1	4	3	2	5	8	6	7	9
Theft of Container/Trailer	8	3	1	6	4	7	2	5	9
Theft of Vehicle	4	3	2	1	6	5	7	8	9
Truck Theft	1	2	4	6	3	5	9	7	8

facility, and truck theft, 'Sunday' tends to be the peak day during the week. Another finding is that even though direct statistics show that the overall probability of truck theft accidents (20.3%) is much lower than that of theft from vehicles (64.2%), truck theft is the most likely accident type to cause a major accident with a loss value over €100k.

In addition to the correlations between accident category and each RIF, BN can reflect the combined effects of multiple RIFs in each accident category to simulate reality. For instance, the riskiest scenario in 'theft from vehicles' is demonstrated when each RIF is assigned with the state generating the highest joint probability with 'theft from vehicle', as seen in Fig. 7. In that scenario, the probability of 'theft from vehicles' significantly increases from the initial 64.2% to 99.9%. If knowing the probability of a cargo theft accident is 99.9% in advance, freight owners would not deliver their cargoes in that scenario without taking special protection measures. Such high-level risks could be avoided in the future with the availability of our proposed risk prediction and diagnosis model in this paper.

Whereas the states of some RIFs are unknown in practice such as 'MO', 'attempt', and 'major accident'; besides, 'year' presents a historical character in this study. In this circumstance, the abovementioned scenario analysis in BN can be adopted to predict the accident category based on the known information to better simulate the reality. As presented in Figs. 8 and 9, given the above four RIFs unknown and the month and weekday both assigned with the states generating the highest impacts in Fig. 8 and the lowest impacts in Fig. 9, it is observed that the predicted probability that tobacco to be stolen from vehicles while in transportation in East Midlands is between 89.9% and 98.6%, with 'intrusion' and 'theft from moving vehicles' being the most likely used MOs. In other accident categories, the combined effects of the 'hot spots' and 'popular products' can be examined as well.

5.2. Implications

From the above analytical results, the most important factors and their significant states influencing the occurrence likelihood of cargo theft accidents from freight supply chains have been identified. Accordingly, decision-makers in supply chains can gain useful insights on how to prioritize the resource allocations for various products, location types, and regions where the cargo security level is relatively low. For instance, the highest security setting should be allocated to moving cargoes (e.g., tobacco, clothing & footwear, food & drink) from vehicles to use high-tech real-time monitoring equipment such as drones, considering the significantly high probability of tobacco theft accidents from vehicles. Besides, the highest types of cargo theft accidents in different regions vary, indicating cargo protection associations should enhance cooperation with local transport authorities to develop different safety policies for cargo transportation with respect to the major accident types in different regions.

Based on the known information on cargo type, conveyance mode, location type, and destination region, multiple supply chain stakeholders can use the developed model of this study to make optimized decisions against cargo theft. Logistics companies, for the first time, can evaluate their logistics solutions made for the shippers and/or consignees from a safety perspective beyond the traditional cost and transit time aspects. Insurance companies can make diversified pricing strategies considering not only the cargo value factor but also the risk level of theft crimes derived from the model, simultaneously, freight owners and carriers can select the best-fitted insurance product for their shipped cargo.

This is a pioneering study advising supply chain stakeholders to not only pay attention to the high-valued product, location type, MO, and region contributing to the occurrence of cargo theft accidents but also give special consideration to the direct causal relationships between the states of the essential RIFs and each accident type.

6. Conclusion

This paper describes a new cargo theft risk analysis from both empirical and methodological perspectives. It develops an advanced quantitative risk analysis method to analyse the interdependency of the RIFs influencing cargo theft from a whole supply chain perspective. First, the cargo theft RIFs are identified from the literature and accident records. Second, a data-driven BN is proposed to construct the model with uncertainty to realise cargo theft risk prediction and diagnosis. Despite BN's popularity in such sectors as transportation and energy for accident investigation, its application in freight supply chains is new. Third, the critical RIFs contributing to cargo theft are evaluated to predict the occurrence of possible cargo theft accidents. Lastly, the real accidents are investigated to test and verify the model with an accuracy rate of 89.14%. Furthermore, the model is validated using sensitivity

Table 8
The joint probability.

Product category	Other	Theft from Container/ Trailer	Theft from Facility	Theft from Vehicle	Theft of Container/ Trailer	Theft of Vehicle	Truck Theft
Clothing & Footwear	2.86	3.74	3.19	82.60	***2.09***	2.86	2.64
Food & Drink	***3.149***	2.79	2.40	82.60	2.30	5.21	1.57
Miscellaneous	3.15	***6.07***	2.48	67.40	1.37	***18***	1.58
No Load	***0.89***	***1.04***	***0.52***	***7.2***	1.26	2.54	***86.6***
Other	2.72	3.28	***3.53***	81.80	1.47	6.26	***0.92***
Tobacco	3.08	2.27	1.14	***88.2***	***1.04***	***2.35***	1.94
Year							
2009	2.05	***0.52***	1.4	39.7	***0.38***	6.13	49.8
2010	1.67	0.56	***0.77***	***27.8***	0.41	3.12	***65.7***
2011	11.5	8.44	12	31.6	6.21	12.3	18
2012	9.17	9.85	11.5	31.6	***7.25***	15.8	14.8
2013	8.35	7.56	***14.8***	40.5	5.56	11.9	11.3
2014	***9.51***	9.08	12.7	39	4.88	14.9	9.94
2015	5.94	***20.7***	8.89	44	4.38	9.46	6.67
2016	3.35	12.3	4.95	69.8	2.28	4.62	2.63
2017	2.27	2.63	1.14	87	2.03	3.26	1.69
2018	***1.24***	1.29	0.94	***91.6***	1.66	***2.37***	***0.88***
2019	1.63	3.54	1.59	81.4	1.21	8.93	1.71
2020	2.58	1.5	1.82	82	1.8	7.7	2.56
2021	9.15	10.2	5.89	41.5	3.8	***23.9***	5.59
Location type							
Destination Facility	***7.53***	***1.01***	2.52	56.1	0.83	***27.9***	4.07
En Route	3.61	2.72	1.22	***80.3***	1.02	7.49	***3.67***
Origin Facility	5.19	2.24	***20.4***	***15.5***	3.34	18.4	***34.9***
Other	6.79	***9.39***	12.4	35.7	***10.5***	15.7	9.5
Unclassified Parking	***1.28***	2.71	***0.3***	69	***0.68***	***2.19***	23.8
MO							
Intrusion	***0.28***	3.13	***2.05***	66.5	***1.41***	4.64	***22***
Other	***30.4***	***3.3***	***5.91***	***18.1***	***4.18***	***28.9***	9.15
Theft from Moving Vehicles	2.1	2.19	2.07	***85.7***	2.05	***1.84***	4.02
Violent & Threat with Violence	29.2	***1.92***	3.17	33.1	1.96	27.1	***3.47***
Region							
East Midlands	***1.46***	2.54	***1.29***	***82.3***	1.1	***2.47***	***8.8***
East of England	2.02	4.54	1.34	73.9	1.04	3.65	13.6
London	3.75	1.71	1.73	63.1	0.93	9.09	19.6
North East	5.2	4.89	4.62	45.7	3.85	11.4	24.4
North West	3.83	2.38	3.57	39.8	2.27	9.95	***38.2***
Northern Ireland	***8.42***	***7.83***	9.04	***19.98***	***6.16***	***15.4***	33.1
Scotland	5.39	6.16	***9.21***	34	5.09	15	25.1
South East	1.56	4.09	1.65	72.3	***0.74***	3.54	16.1
South West	3.78	3.8	5.98	41.5	4.18	11.8	29
Wales	5.57	4.9	5.23	26.5	3.59	12.6	41.6
West Midlands	2.6	***1.6***	1.96	54.6	1.73	10.4	27.2
Yorkshire and the Humber	1.64	1.77	1.55	64.2	1.2	4.79	24.8
Month							
1	2.67	3.39	2.19	68.60	1.57	10.75	***10.76***
2	2.70	3.48	2.25	61.90	***0.84***	8.31	20.50
3	2.36	2.72	2.12	61.10	1.41	9.48	20.80
4	2.97	2.73	2.44	61.40	1.00	4.52	25.00
5	***3.65***	***5.99***	1.51	***51.6***	1.18	6.28	29.80
6	3.09	3.74	2.61	63.80	1.95	***2.07***	22.80
7	1.73	2.45	2.76	61.90	1.51	4.11	25.60
8	2.06	2.37	2.05	56.40	2.28	3.64	***31.2***
9	***1.66***	2.66	***1.38***	68.00	1.21	2.86	22.20
10	1.71	3.19	1.70	76.40	1.28	2.72	13.00
11	2.83	***1.63***	2.55	***76.5***	1.96	3.06	11.40
12	2.52	2.21	***3.34***	59.20	***2.77***	***16.1***	13.90
Major accident							
No	***2.07***	***2.8***	***1.95***	***65.6***	***1.32***	***6.02***	***20.3***
Yes	***16.9***	***11.2***	***11***	***17.8***	***8.8***	***14***	***20.4***
Weekday							
1	2.41	2.63	3.08	48.1	3.56	***11.4***	28.9
2	2.21	***2.35***	2.41	66.1	1.18	6.05	19.7
3	2.49	3.17	***1.51***	67.6	1.13	5.25	18.8
4	***2.02***	2.83	1.58	70.4	***0.86***	4.87	17.5
5	2.9375	3.63	1.6	***71.3***	1.23	***4.41***	***14.9***
6	***2.9434***	3.16	2.81	59.6	1.31	7.39	22.8
7	2.66	***3.66***	***4.46***	***46***	***3.8***	9.16	***30.3***
Attempt							
No	***2.48***	***2.49***	***2.18***	***64.17***	***1.52***	***6.42***	***20.8***
Yes	***2.76***	***9.4***	***2.67***	***64.2***	***1.74***	***4.42***	***14.8***

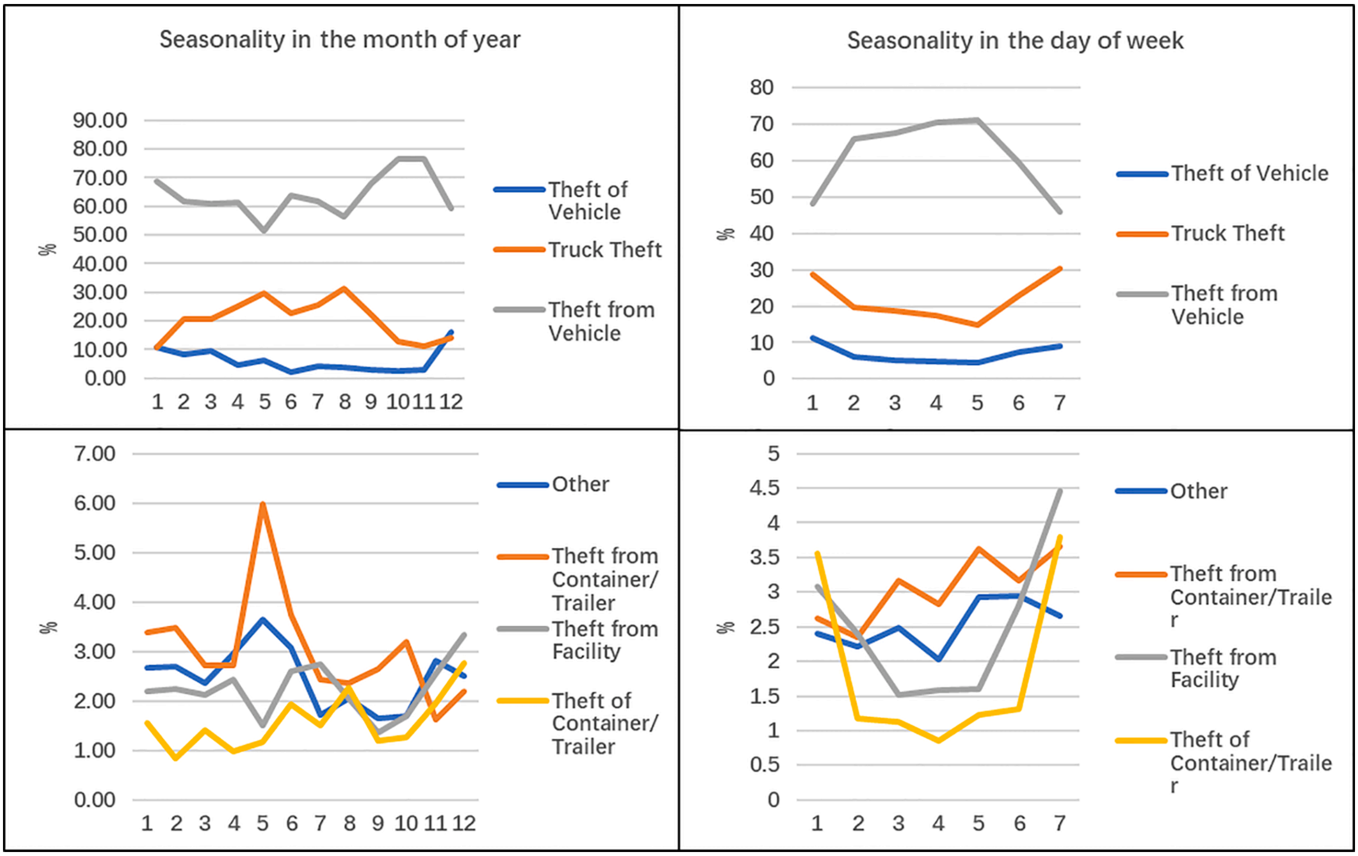

Fig. 6. Seasonality in accident categories.

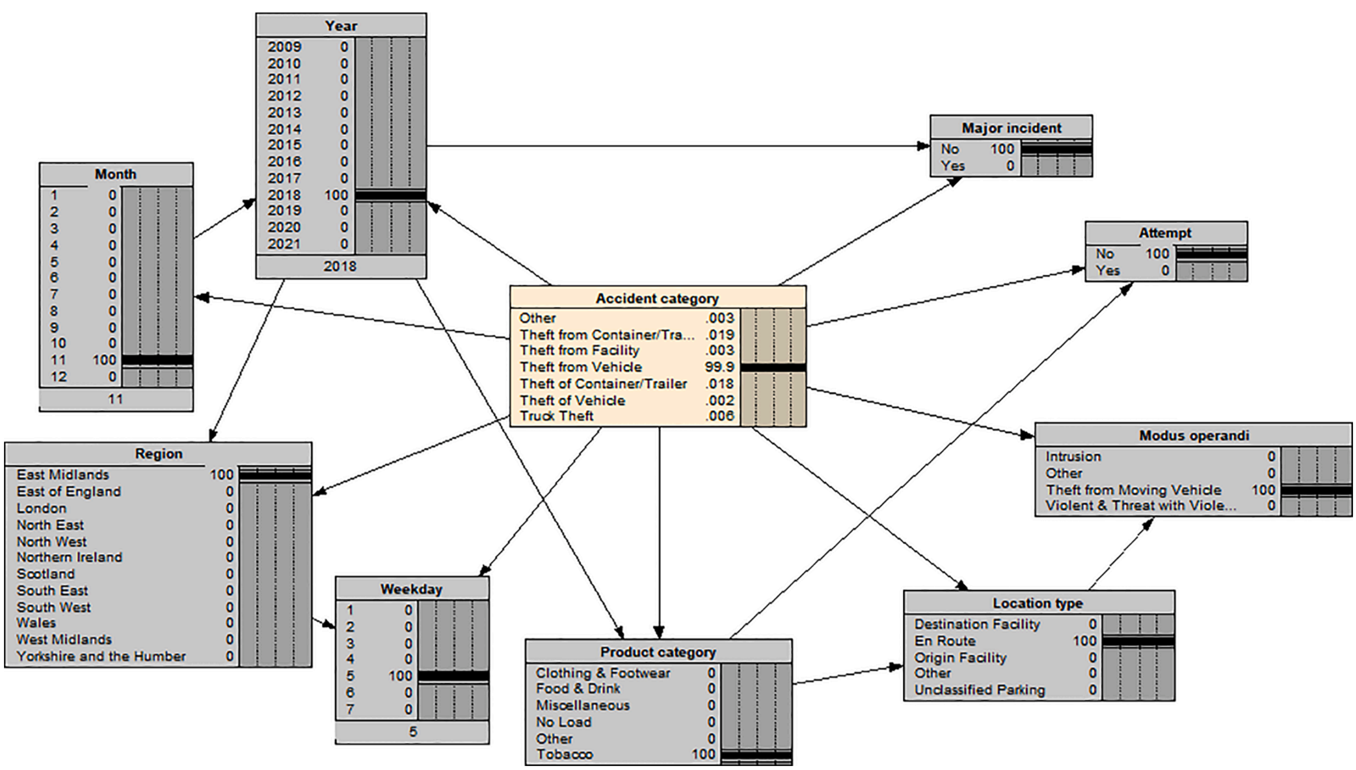

Fig. 7. Scenario A.

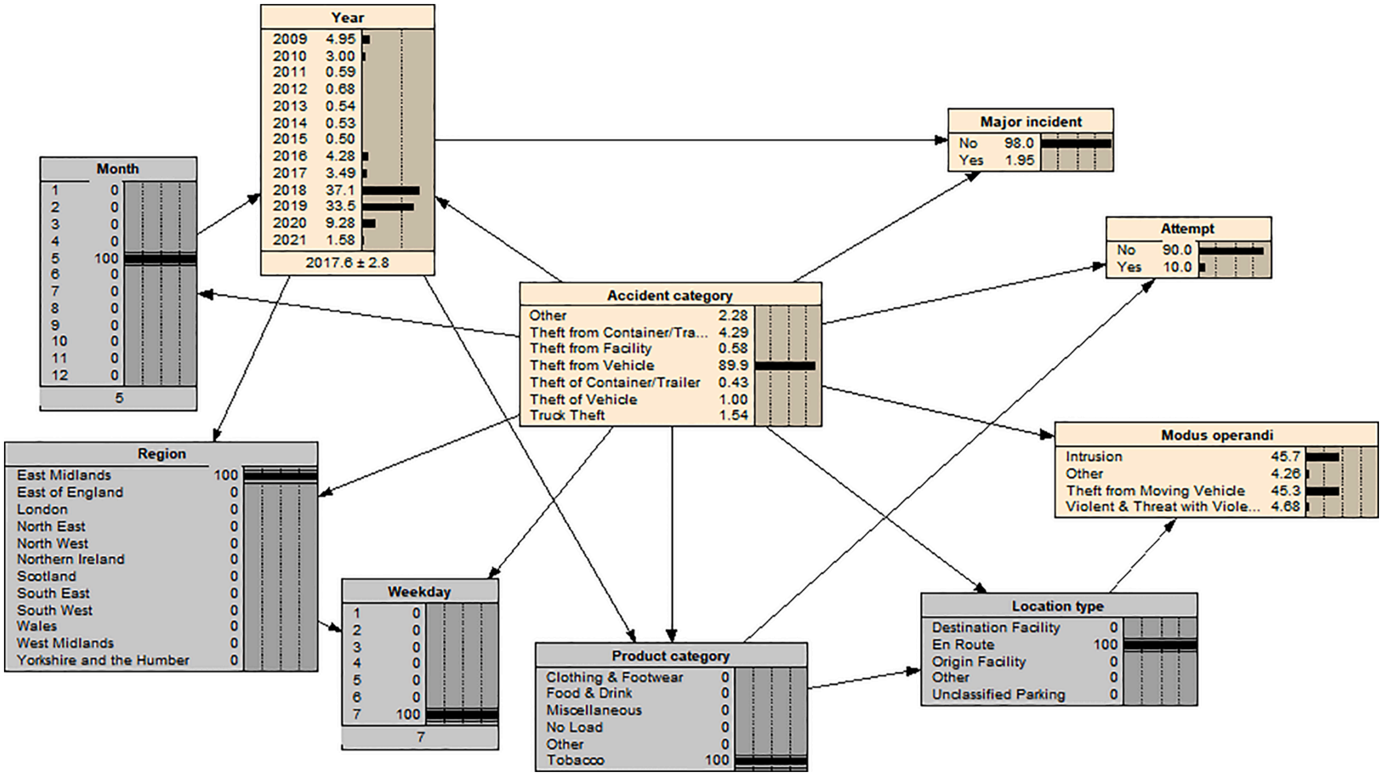

Fig. 8. Scenario B (1).

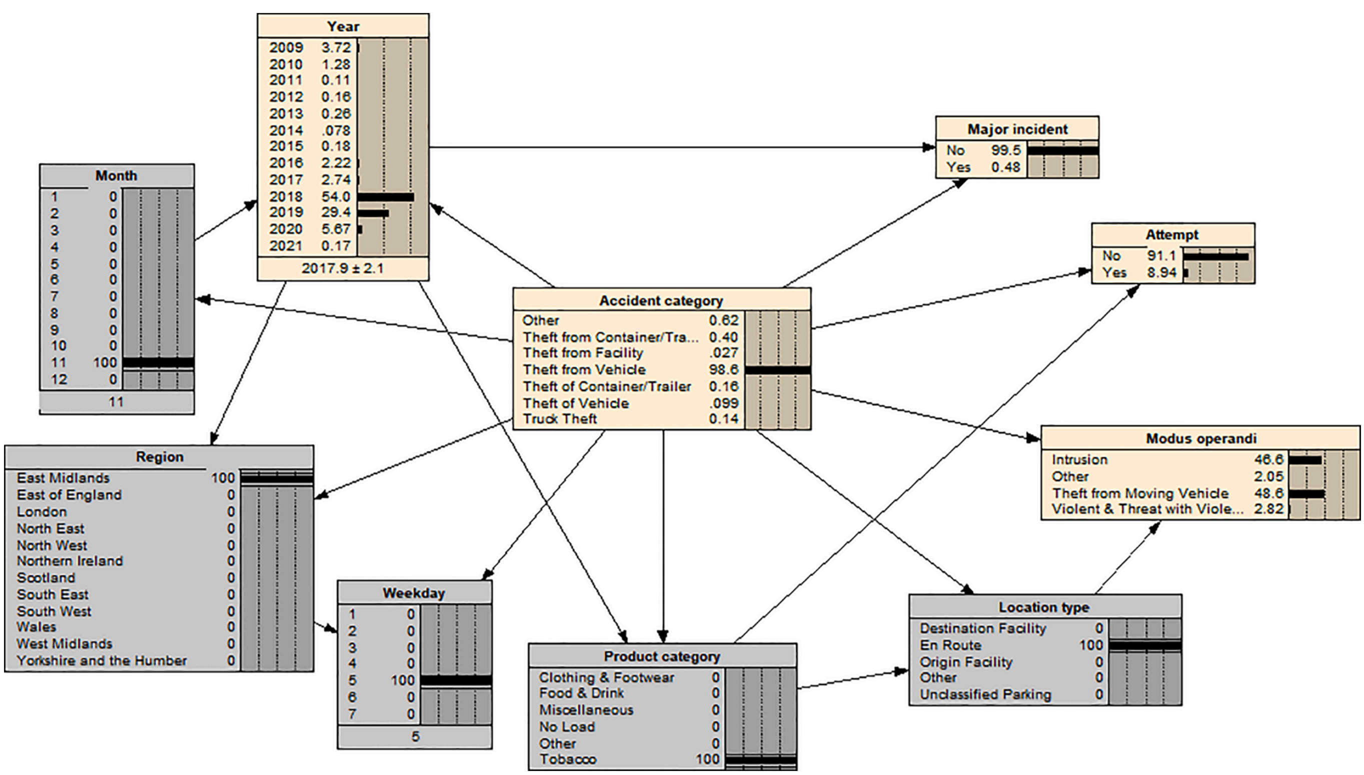

Fig. 9. Scenario B (2).

analysis and scenario analysis.

The findings of this study provide the most significant implications for the prevention of cargo theft in freight supply chains.

1) This study pioneers the development of an advanced model to predict the risk level of cargo theft accidents. The most influential RIFs of cargo theft accidents are identified as product category, year, location type, MO, and region from a UK case study.
2) This study reveals the combined effects of multiple RIFs and differentiates the states' effects of each RIF, which enables the scenario simulation of reality.
3) It is evident that the attention should not be only paid to the high-valued product, location type, MO, and region of occurrence of cargo theft accidents, but also to the root causes of the respective accident types derived from the analysis of causal relationships.
4) The developed model can benefit multiple supply chain stakeholders in prioritizing resource allocation and optimizing the decisions for cost-effective theft risk control in practice.

This study exposes some limitations to be addressed in future research, for instance, some states with very low probabilities in TAPA's database are combined as one state 'other'. However, some of these states (e.g., electronics in the product category and hijacking in the accident category) have achieved industrial attention, thus their effects combined with other RIFs need to be further investigated in the future. Furthermore, the designed flow of data analysis in this study focuses on the UK area and the model can be applied in other areas for cargo theft analysis to develop the best practice of protection through the benchmarking of the performance of different areas in future.

CRediT authorship contribution statement

Xinrui Liang: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Shiqi Fan:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration. **John Lucy:** Validation, Writing – review & editing, Supervision. **Zaili Yang:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

We confirm that this paper is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

Data availability

The authors do not have permission to share data.

Acknowledgment

This work is financially supported by a European Research Council project (TRUST CoG 2019 864724). We also thank TAPA for providing us with cargo theft microdata.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ress.2022.108702.

Appendix A. Confusion matrix

Predicted	Other	Theft from Container/Trailer	Theft from Facility	Theft from Vehicles	Theft of Container/ Trailer	Theft of Vehicles	Truck Theft	Actual total	Accuracy rate (%)
Actual									
Other	**14**	1	1	5	0	4	0	25	56.00
Theft from Container/Trailer	0	**13**	1	21	0	2	0	37	35.14
Theft from Facility	0	0	**4**	6	1	4	1	16	25.00
Theft from Vehicles	4	2	4	**577**	1	11	0	599	96.33
Theft of Container/ Trailer	0	0	0	6	**2**	0	1	9	22.22
Theft of Vehicles	1	0	2	15	2	**32**	1	53	60.38
Truck Theft	0	0	0	1	0	3	**187**	191	97.91
Predicted total	19	16	12	631	6	56	190	930	89.14

References

[1] Altman DG. Practical statistics for medical research. CRC Press; 1990.
[2] Alyami H, Yang Z, Riahi R, Bonsall S, Wang J. Advanced uncertainty modelling for container port risk analysis. Acc Anal Prevent 2019;123:411–21. https://doi.org/10.1016/j.aap.2016.08.007.
[3] Belzer MH, Swan PF. Supply chain security: agency theory and port drayage drivers. Econ Labour Relat Rev 2011;22:41–63. https://doi.org/10.1177/103530461102200103.
[4] Boone CA, Skipper JB, Murfield A, Murfield ML. Cargo theft in the motor carrier industry: an exploratory study. J Transp Secur 2016;9:57–70. https://doi.org/10.1007/s12198-016-0166-1.
[5] Chow CK, Liu CN. Approximating discrete probability distributions with dependence trees. IEEE Trans Inf Theory 1968;14:462–7. https://doi.org/10.1109/TIT.1968.1054142.
[6] Cohen J. A Coefficient of agreement for nominal scales. Educ Psychol Measur 1960;20:37–46. https://doi.org/10.1177/001316446002000104.
[7] da Silva ARA, da Cruz CFS, Fernandes JEL, da Costa JB, Chen X, Fabricio R, et al. RFID-based scheme for TV receiver control in case of theft. 2018 IEEE International Conference on Consumer Electronics (ICCE), 2018, p. 1–6. 10.1109/ICCE.2018.8326100.
[8] Dan Burges. Cargo Theft, Loss Prevention, and Supply Chain Security | ScienceDirect (2022). Available at: https://www.sciencedirect.com/book/9780124160071/cargo-theft-loss-prevention-and-supply-chain-security.

[9] Dindar S, Kaewunruen S, An M. A hierarchical Bayesian-based model for hazard analysis of climate effect on failures of railway turnout components. Reliab Eng Syst Saf 2022;218:108130. https://doi.org/10.1016/j.ress.2021.108130.
[10] Dong-li H, Zheng-wu Y, You BS, Cho SK, Bae HY. A novel architecture for cargos location and safety in networked RFID. In: Proceedings of the 5th Asian symposium on geographic information systems. Chongqing: Chongqing University Posts & Telecommunications; 2007. p. 386–91.
[11] Ekwall D, Brüls H, Wyer D. Theft of pharmaceuticals during transport in Europe. J Transp Secur 2016;9:1–16. https://doi.org/10.1007/s12198-015-0162-x.
[12] Ekwall D, Lantz B. Cargo theft at non-secure parking locations. Int J Retail Distribut Manag 2015;43:204–20. https://doi.org/10.1108/IJRDM-06-2013-0131.
[13] Ekwall D, Lantz B. Modi operandi for cargo theft in EMEA–a seasonality analysis. J Transp Secur 2015;8:99–113. https://doi.org/10.1007/s12198-015-0160-z.
[14] Ekwall D, Lantz B. Seasonality of cargo theft at transport chain locations. Int J Phys Distribut Logist Manag 2013;43:728–46. https://doi.org/10.1108/IJPDLM-06-2012-0175.
[15] Ekwall D, Lantz B. Supply chain risk analysis and assessment: cargo theft. Transp J 2016;55:400–19. https://doi.org/10.5325/transportationj.55.4.0400.
[16] Ekwall D, Lantz B. The use of violence in cargo theft – a supply chain disruption case. J Transp Secur 2018;11:3–21. https://doi.org/10.1007/s12198-018-0186-0.
[17] Ekwall D. The displacement effect in cargo theft. Int J Phys Distribut Logist Manag 2009;39:47–62. https://doi.org/10.1108/09600030910929183.
[18] Fan S, Blanco-Davis E, Yang Z, Zhang J, Yan X. Incorporation of human factors into maritime accident analysis using a data-driven Bayesian network. Reliab Eng Syst Saf 2020;203:107070. https://doi.org/10.1016/j.ress.2020.107070.
[19] Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76(5):378–82. https://doi.org/10.1037/h0031619.
[20] Fokum DT, Frost VS, DePardo D, Kuehnhausen M, Oguna AN, Searl LS, et al. Experiences from a transportation security sensor network field trial. IEEE Globecom Workshops. New York: IEEE; 2009. p. 302–7.
[21] Friedman N, Geiger D. Classifiers Goldszmidt M. Bayesian network. Mach Learn 1997;29:131–63. https://doi.org/10.1023/A:1007465528199.
[22] Guerin L, Vieira JGV, de Oliveira RLM, de Oliveira LK, de Miranda Vieira HE, Dablanc L. The geography of warehouses in the São Paulo Metropolitan Region and contributing factors to this spatial distribution. J Transp Geogr 2021;91:102976. https://doi.org/10.1016/j.jtrangeo.2021.102976.
[23] Harvey JM. The secure networked truck: protecting America's transportation infrastructure. In: Proceedings of the IEEE 60th vehicular technology conference. 7. IEEE; 2004. p. 5281–4. VTC2004-Fall20047.
[24] Justus M, Ceccato V, Kahn T, Moreira G. Crime against trading: the case of cargo theft in São. Paulo 2017. https://doi.org/10.13140/RG.2.2.21943.21923.
[25] Kit I, Fomenko A, Vyshnia V, Novosad I. Computer tools for cargo thefts fighting on railway transport. In: Proceedings of the 10th IEEE international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS). 1. IEEE; 2019. p. 172–5.
[26] Klodzinski J, Kerr P. Electronic freight theft management system for florida. Transp Res Rec 2007;2008:1–9. https://doi.org/10.3141/2008-01.
[27] Kubáňová J, Kubasakova I. Security risks in the trucking sector. Transp Res Procedia 2020;44:234–9. https://doi.org/10.1016/j.trpro.2020.02.048.
[28] Kubanova. Truck driver scheduling of the rest period as an essential element of safe transport. N.p 2016.
[29] Kuo-Yu T, Yang MH, Luo JN, Liew WT. Novel designated ownership transfer with grouping proof. Appl Sci 2019;9. https://doi.org/10.3390/app9040724.
[30] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159–74. https://doi.org/10.2307/2529310.
[31] Li KX, Yin J, Bang HS, Yang Z, Wang J. Bayesian network with quantitative input for maritime risk analysis. Transportmetrica A 2014;10:89–118. https://doi.org/10.1080/18128602.2012.675527.
[32] Liu K, Yu Q, Yang Z, Wan C, Yang Z. BN-based port state control inspection for Paris MoU: new risk factors and probability training using big data. Reliab Eng Syst Saf 2022;224:108530. https://doi.org/10.1016/j.ress.2022.108530.
[33] Lorenc A, Kuźnar M, Lerher T, Szkoda M. Predicting the probability of cargo theft for individual cases in railway transport. Tehnički Vjesnik 2020;27:773–80. https://doi.org/10.17559/TV-20190320194915.
[34] Lorenc A, Kuźnar M. An intelligent system to predict risk and costs of cargo thefts in road transport. Int J Eng Technol Innovat 2018;8:284–93.
[35] Lorenc A, Kuźnar M. The most common type of disruption in the supply chain: evaluation based on the method using artificial neural networks. Int J Shipp Transp Logist : IJSTL 2021;13.
[36] Nguyen S, Chen PSL, Du Y, Thai VV. An operational risk analysis model for container shipping systems considering uncertainty quantification. Reliab Eng Syst Saf 2021;209:107362. https://doi.org/10.1016/j.ress.2020.107362.
[37] Oranye N, Masupe S, Chuma JM, Sebusang SEM. Design of locomotive location indicator communication system. In: 2004 IEEE Africon 7th Africon Conference in Africa (IEEE Cat. No.04CH37590). 2; 2004. p. 1269–73. https://doi.org/10.1109/AFRICON.2004.1406896. Vol.2.
[38] Orfanus, Dominik. Cargo thefts as part of the black economy and their economic implications. Aktualne Pohl 'Ady Na Konkurencieschopnost' A Podnikanie – Nove Vyzvy.N.p., 2011. 376–380.
[39] Repolho HM, Marchesi JF, Júnior OSS, Bezerra RRR. Cargo theft weighted vehicle routing problem: modeling and application to the pharmaceutical distribution sector. Soft Comput 2019;23:5865–82. https://doi.org/10.1007/s00500-018-3250-6.
[40] Risk Management Magazine - Europe's Cargo Theft Troubles. Magazine n.d. https://www.rmmagazine.com/articles/article/2011/11/01/-Europe-s-Cargo-Theft-Troubles.
[41] Shannon CE. A Mathematical Theory of Communication. Bell Syst Tech J 1948;27:379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.
[42] Shoukry G. Criminals' response to changing crime lucre. Rochester, NY: Social Science Research Network; 2015. https://doi.org/10.2139/ssrn.2627288.
[43] Song R, Huang L, Cui W, Óskarsdóttir M, Vanthienen J. Fraud detection of bulk cargo theft in port using bayesian network models. Appl Sci 2020;10:1056. https://doi.org/10.3390/app10031056.
[44] Supply chain risk insights annual intelligence report (2021). Available at: https://www.bsigroup.com/en-GB/our-services/consulting/supply-chain-risk/supply-chain-reports/.
[45] Tang D, Xu L, Yang J, Chen Y. A model for security evaluation of a port storage area against theft in a container line supply chain. Research.Manchester.Ac.Uk 2010. https://www.research.manchester.ac.uk/portal/en/publications/a-model-for-security-evaluation-of-a-port-storage-area-against-theft-in-a-container-line-supply-chain(e2d2dd0c-16db-434d-a577-0f92473a198a).html.
[46] Toth GE. CargoTIPS: an innovative approach to combating cargo theft. Enforcement Secur Technol 1998;3575:315–9. https://doi.org/10.1117/12.335002. SPIE.
[47] Trucco P, Cagno E, Ruggeri F, Grande O. A Bayesian Belief Network modelling of organizational factors in risk analysis: a case study in maritime transportation. Reliab Eng Syst Saf 2008;93:845–56. https://doi.org/10.1016/j.ress.2007.03.035.
[48] UK tops supply chain thefts in 2020. Logistics Manager 2020. https://www.logisticsmanager.com/uk-tops-supply-chain-thefts-in-2020/.
[49] Ung ST. Navigation Risk estimation using a modified Bayesian Network modeling-a case study in Taiwan. Reliab Eng Syst Saf 2021;213:107777. https://doi.org/10.1016/j.ress.2021.107777.
[50] Wu B, Tang Y, Yan X, Guedes Soares C. Bayesian Network modelling for safety management of electric vehicles transported in RoPax ships. Reliab Eng Syst Saf 2021;209:107466. https://doi.org/10.1016/j.ress.2021.107466.
[51] Wu PJ, Chen MC, Tsau CK. The data-driven analytics for investigating cargo loss in logistics systems. Int J Phys Distribut Logistic Manag 2017;47:68–83. https://doi.org/10.1108/IJPDLM-02-2016-0061.
[52] Xie Y, Lord D, Zhang Y. Predicting motor vehicle collisions using Bayesian neural network models: an empirical analysis. Acc Anal Prevent 2007;39:922–33. https://doi.org/10.1016/j.aap.2006.12.014.
[53] Yang Z, Wan C, Yang Z, Yu Q. Using Bayesian network-based TOPSIS to aid dynamic port state control detention risk control decision. Reliab Eng Syst Saf 2021;213:107784. https://doi.org/10.1016/j.ress.2021.107784.
[54] Yang Z, Yang Z, Smith J, Robert BAP. Risk analysis of bicycle accidents: a Bayesian approach. Reliab Eng Syst Saf 2021;209:107460. https://doi.org/10.1016/j.ress.2021.107460.
[55] Yang Z, Yang Z, Yin J. Realising advanced risk-based port state control inspection using data-driven Bayesian networks. Transp Res Part A 2018;110:38–56. https://doi.org/10.1016/j.tra.2018.01.033.
[56] Yang ZL, Wang J, Bonsall S, Fang QG. Use of fuzzy evidential reasoning in maritime security assessment. Risk Anal 2009;29:95–120. https://doi.org/10.1111/j.1539-6924.2008.01158.x.
[57] Yin J, Ren X, Liu R, Tang T, Su S. Quantitative analysis for resilience-based urban rail systems: a hybrid knowledge-based and data-driven approach. Reliab Eng Syst Saf 2022;219:108183. https://doi.org/10.1016/j.ress.2021.108183.
[58] Yu Q, Liu K, Yang Z, Wang H, Yang Z. Geometrical risk evaluation of the collisions between ships and offshore installations using rule-based Bayesian reasoning. Reliab Eng Syst Saf 2021;210:107474. https://doi.org/10.1016/j.ress.2021.107474.
[59] Yu YC, Gardoni P. Predicting road blockage due to building damage following earthquakes. Reliab Eng Syst Saf 2022;219:108220. https://doi.org/10.1016/j.ress.2021.108220.
[60] Yuan Z, Huang D. A novel RFID-based shipping containers location and identification solution in multimodal transport. proceedings of the Canadian conference on electrical and computer engineering, 2008, p. 000267–72. 10.1109/CCECE.2008.4564537.
[61] Zhang X, Mahadevan S. Bayesian network modeling of accident investigation reports for aviation safety assessment. Reliab Eng Syst Saf 2021;209:107371. https://doi.org/10.1016/j.ress.2020.107371.
[62] Jones B, Jenkinson I, Yang Z, Wang J. The use of Bayesian network modelling for maintenance planning in a manufacturing industry. Reliab Eng Syst Saf 2010;95(3):267–77. https://doi.org/10.1016/j.ress.2009.10.007.
[63] Augustyn L, Małgorzata K. Modeling cargo theft probability in rail transport using arifiCial neural network. Proc CLC Carpathian Logist Congr 2019:306–11. https://www.confer.cz/clc/2018/2581-modeling-cargo-theft-probability-in-rail-transport-using-arifitial-neural-network.