https://doi.org/10.1016/j.trb.2019.07.017

Development of a Non-Parametric Classifier: Effective Identification, Algorithm, and Applications in Port State Control for Maritime Transportation

Shuaian Wang[1], Ran Yan[1], Xiaobo Qu[2*]

[1] *Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University, Hung Hom, Hong Kong*
[2] *Department of Architecture and Civil Engineering, Chalmers University of Technology, Gothenburg, Sweden*

Abstract

Maritime transportation plays a pivotal role in the economy and globalization, while it poses threats and risks to the maritime environment. In order to maintain maritime safety, one of the most important mitigation solutions is the Port State Control (PSC) inspection. In this paper, a data-driven Bayesian network classifier named Tree Augmented Naive Bayes (TAN) classifier is developed to identify high-risk foreign vessels coming to the PSC inspection authorities. By using data on 250 PSC inspection records from Hong Kong port in 2017, we construct the structure and quantitative parts of the TAN classifier. Then the proposed classifier is validated by another 50 PSC inspection records from the same port. The results show that, compared with the Ship Risk Profile selection scheme that is currently implemented in practice, the TAN classifier can discover 130% more deficiencies on average. The proposed classifier can help the PSC authorities to better identify substandard ships as well as to allocate inspection resources.

Keywords: Maritime transportation, Maritime safety, Port state control (PSC), Bayesian network (BN), TAN classifier

1. Introduction

Maritime transportation plays a pivotal role in the economic development and globalization (Teye et al., 2017; Tan et al., 2018; Zhang and Lam, 2018). According to UNCTAD (2017), over 80% of global trade by volume and more than 70% of its value are carried on board ships and handled by seaports worldwide. Maritime transport is relatively safe, but once a maritime accident occurs, the costs and loss can be huge to both the shipping industry and society (Qu and Meng, 2012; Chauvin et al., 2013; Zheng et al., 2017; Zheng et al., 2018; Sun et al., 2018). To reduce maritime risks, various international rules have been formulated under the auspices of the International Maritime Organization (IMO) and International Labour Organization (ILO), such as the International Convention for the Safety of Life at Sea (SOLAS), the International Convention for the Prevention of Pollution from Ships (MARPOL), the International Convention on Standards

* Corresponding author: Xiaobo Qu, drxiaoboqu@gmail.com; xiaobo@chalmers.se

of Training, the International Convention on Certification and Watchkeeping for Seafarers (STCW), the International Convention on Tonnage Measurement of Ships, and the International Convention on Load Lines (CLL) (IMO, 2018; Knapp and Franses, 2008).

Ships that cannot comply with these conventions are called substandard ships (Li and Zheng, 2008). In the maritime industry, flag states, which are deemed as the nationality of a vessel and under whose laws the vessel is registered, are seen as the first line of defence against substandard ships (Knapp and Velden, 2009; Cariou et al., 2007). However, it is widely believed that many flag states are unable to perform well their mandated duties of ensuring that ships flying their flags are fully compliant with the international rules, as these ships may visit their flag state ports only irregularly. The situation can be worse in the open registry countries, as these flag states often have insufficient or substandard regulations and those regulations are poorly enforced (Li and Wonham, 1999). As a result, port state control (PSC), which is an internationally agreed regime to inspect foreign ships coming to the port state, is proposed. It acts as the "second line of defence" and "last safety net" to eliminate substandard vessels, and is a complement instead of a substitute, to consolidate the safety net of the former maritime safety administration by the flag state (Cariou et al., 2008; Li and Zheng, 2008).

The Memorandum of Understanding (MoU) on PSC, which is an organization consisting of several PSC member authorities in a certain region, was first established in Europe in 1982 (often referred to as the "Paris MoU"), and by the end of 2018, nine MoUs on PSC have been signed around the world. The goal of the MoUs on PSC is the same: to verify that the incoming ships meet the requirements of the international agreements through a harmonized system of port state control which allows for information sharing (Kasoulides, 1993; Paris MoU, 2019). In each MoU, the member authorities are responsible for inspecting incoming foreign ships and should adopt the same set of inspection rules. In 2016, the number of inspections conducted by the nine PSC MoUs was 63,805 in total (Indian Ocean MoU, 2017; Caribbean MoU, 2017; Abuja MoU, 2017; Black Sea MoU, 2017; Viña del Mar Agreement, 2017; Tokyo MoU, 2017a; Mediterranean MoU, 2017; Riyadh MoU, 2017; Paris MoU, 2017), while the total number of merchant vessels in the whole world was 96,161 (UNCTAD, 2017). During a PSC inspection, the conditions on board that are not in compliance with the requirements are recorded as deficiencies and are required to be rectified. The PSC authorities also have the right to detain a ship until the deficiencies are rectified if those deficiencies might pose a danger to the crew and the marine environment (Tokyo MoU, 2017a). After the inspection, a report on the inspected ship, including ship information (e.g., ship name, ship flag, ship company, etc.) and inspection information (e.g., inspection date, inspection authority, the types and total number of deficiencies detected and ship detention information, etc.), is generated and kept in the database of the corresponding MoU.

One of the key issues faced by PSC authorities is how to select ships on which to conduct PSC inspections (IMO, 2018). On the one hand, the cost of PSC inspection to the port authorities is high. It is estimated by Knapp (2007) that the costs for a PSC inspection with and without deficiencies are 759 USD and 509 USD, respectively. Further, non-essential inspections may also delay the fast turnover of the maritime logistics system. On the other hand, not all ships are substandard. Tokyo MoU, which is the MoU on PSC in the Asia-Pacific Region and was signed in December 1993, reported that the total number of inspections conducted by its 20 member authorities in 2017 was 41,616, while only 18,113 inspections found deficiencies (Tokyo MoU, 2018). Due to the high costs and limited time and resources, it is impossible and unnecessary to inspect all coming ships. In order to identify as many substandard ships and ship deficiencies as possible after inspecting a certain number of ships, different PSC MoUs adopt different ship selection schemes. Taking Tokyo MoU as an example, it introduced a New Inspection Regime (NIR) from 2014 (Tokyo MoU, 2014) to calculate the ship risk profile (SRP) using criteria on an information sheet. The information sheet takes into consideration several parameters including ship type, age, ship company performance, previous detentions, etc. Each parameter is given a fixed weighting point and the SRP is determined by the total weighting points (Tokyo MoU, 2014). Based on the total points, all ships are divided into three types: low risk ship (LRS), standard risk ship (SRS) and high risk ship (HRS). The higher risk a ship has, the more frequently it will be inspected. As the SRP adopts a simple weighted sum model to classify the incoming ships, the weight of each parameter is determined simply by expert judgement. In addition, it does not take the dependencies between different parameters into account. Another issue is that even if each incoming ship is given a risk profile, there is no further information about the risk level of the ships in the same risk profiles. As a result, when ships of the same SRP come to the port state, the selection of ships to be inspected is dependent on the PSC officers' subjective judgements.

To address the abovementioned problems, this paper aims to propose a data-driven Bayesian network classifier called a Tree Augmented Naive (TAN) Bayes classifier as a new scheme to select ships for PSC inspection. The TAN classifier is constructed and validated from a case data set which is built based on the online database of Tokyo MoU. It takes into account factors related to a ship itself and its inspection history, and calculates their mutual dependencies and contributions to the total number of ship deficiencies. The TAN classifier provides PSC officers with an informed estimate of the number of deficiencies an incoming ship will have, which helps them to identify higher risk ships and better allocate resources to PSC inspections. The contribution of the paper is as follows. (i) The proposed TAN classifier is one of the first few models to take into consideration historical factors (including the number of previous detentions, last inspection time, number of deficiencies in the last inspection and number of flag changes) and the performance of the shipping company (which is responsible for verifying that the ship complies with the International Safety Management (ISM) code) when analyzing PSC inspection from a quantitative perspective. After inputting the above-mentioned information of a coming ship, the TAN classifier can generate the probabilities for the ship to have 0 to 2, 3 to 6, and more than 7 deficiencies immediately based on the trained CPTs, and the timely risk index of this ship can also be given for the PSCOs' reference. Thus, the proposed classifier can act as a real-time predictor of the number of deficiencies before conducting the PSC inspection. (ii) The newly proposed ship selection scheme for PSC inspection adopts a data-driven non-parametric model. This is the very first model that makes predictions on the possible number of deficiencies of incoming ships for PSC inspection. Compared with the currently used SRP ship selection scheme, it can identify an average of 130% more deficiencies in ships. (iii) Theoretically, our paper proposes a dynamic programming approach to optimally discretize input data into discrete states so that they can be analyzed by the TAN classifier. Moreover, by induction, it is rigorously proved that in the TAN classifier, random selection of root attribute variables will not influence the classification process.

2. Literature review

2.1 Studies on PSC inspection

PSC inspections have received increasing attention in the literature. One stream of related study concerns inspection target factors, which mainly include generic characteristics such as ship age, ship size and ship type. Several related studies reach a concordance that ship age, ship flag and ship type are the main determinants of ship deficiencies and detention (Cariou et al., 2007; Cariou et al., 2009; Cariou and Wolff, 2015; Huang et al., 2016; Kara., 2016; Tsou, 2018). More specifically, some studies have also identified the extent to which the target factors would contribute to the deficiencies and detention (e.g. Cariou et al., 2007). Based on the target factors, various and innovative ship selection schemes for PSC inspection are proposed. Zhou and Sun (2010) proposed an automatically optimized and self-evolutional ship target system based on the target factors using the Generalized Additive Modelling (GAM) approach. Xu et al. (2007a) introduced a risk assessment system based on a Support Vector Machine (SVM) to classify incoming ships as either high risk or low risk according to the target factors. If a ship is decided to be high risk and the inspection leads to a detention, this prediction is viewed as accurate. Numerical experience shows that the prediction accuracy of their proposed model was no more than 14%. Then, they combined the web-mining technology to improve the classification accuracy, which was improved to about 20% (Xu et al., 2007b). Based on these two studies conducted by Xu et al. (2007a, b), Gao et al. (2008) combined the K-nearest neighbour (KNN) and SVM to remove noisy training examples to improve the ship selection accuracy. After the improvement, the prediction accuracy could be more than 20%. Recently, Yang et al. (2018a) proposed a data-driven Bayesian network to analyze factors influencing PSC inspection and then used the model to predict the detention rate of bulk carriers. After that, Yang et al. (2018b) combined the Bayesian network model with the game model between PSC port authorities and ship owners to present an optimal PSC inspection scheme.

The second stream of studies focuses on the effects of PSC inspections. There are three research sub-areas in this stream: the effect on maritime safety, on maritime pollution and on later PSC inspections. Regarding the first sub-area, several papers point out that the PSC inspection can help reduce the probability of maritime casualties and accidents. Knapp and Franses (2007, 2008) used a binary logistic regression model to measure the effect of PSC inspections on the casualty probability and found that the casualty probability was significantly reduced in some areas such as the Indian Ocean Region, while there was no evidence that the casualty probability was reduced in other areas such as Northern Europe. Hänninen and Kujala (2014) pointed out that knowledge of the ship type, the PSC inspection type and the number of structural conditions related deficiencies could provide the most information regarding future accident involvement and the true ship safety state. Heij and Knapp (2018) argued that the probability of future shipping accidents was related to the past PSC inspection deficiencies. In addition, the PSC inspection can help reduce maritime accident loss. Li and Zheng (2008) pointed out that the PSC inspection could reduce the total number of maritime accidents and the number of ships involved in consequential maritime accidents. Knapp et al. (2011) claimed that the PSC inspections could bring about monetary benefit by reducing maritime accident loss. As to the second sub-area, Titz (1989) and Heij et al. (2011) both pointed out that PSC inspections contribute to protecting the maritime environment. Concerning the third research sub-area, Cariou et al. (2007) suggested that the number of deficiencies in the next PSC inspection would be reduced by 63% compared with the former. Cariou and Wolff (2011) pointed out that a vessel that was subject to detention and/or a high number of deficiencies in the previous PSC inspection was more likely to change its ship flag and/or classification society before the next PSC inspection to avoid future inspection.

Despite a large number of studies on the PSC inspection, one main drawback is that most of them just serve as a summary of the factors influencing the results of PSC inspections instead of generating real-time deficiency information about new incoming vessels. On real-time prediction, there are some papers focusing on ship detention in PSC inspection. Since the detention rate is low (for example, in Hong Kong the detention rate was 4.07% in 2017, while 599 out of the total 908 inspections found deficiencies (Tokyo MoU, 2018)), it is more reasonable to make predictions on the number of deficiencies for better ship selection. Another shortcoming is that in most of the proposed ship selecting models, most of the target factors taken into account are ship generic factors, while the dynamic factors (number of ship flag changes) and ship inspection history (PSC detention history and last PSC inspection information) are rarely considered. To address these shortcomings, a TAN classifier is proposed, which takes the number of flag changes as the ship dynamic factor, as well as the previous detention times and the information from the last PSC inspection as the ship inspection history into consideration to improve the classification accuracy. The proposed TAN classifier can be used as a real-time predictor of the possible number of deficiencies of the coming ships.

Regarding the methodologies that are used in the relevant literature, some of the studies focus on PSC policy itself and adopt qualitative research methods or statistical methods, such as the combined t-test, decomposition analysis and econometric analysis model, to analyze the target factors used by PSC inspection authorities (Cariou et al., 2009), and to calculate the quantitative link between past PSC inspection outcomes and future shipping accidents (Heij and Knapp, 2018). Some studies use regression methods, including but not limited to quantile regression (Cariou and Wolff, 2015) and binary logistic regression (Knapp and Franses, 2007; Knapp et al., 2011; Knapp and Hänninen, 2014) to calculate the relationships between ship target factors and the deficiencies detected and between PSC ship inspection results and future accident involvement. Another methodology adopted by researchers is the classic machine learning models, such as the GAM approach (Zhou and Sun, 2010), SVM model (Xu et al., 2007a; Xu et al., 2007b), and the combination of KNN and SVM methods (Gao et al., 2008). Several game models are also used in PSC-related studies (Yang et al., 2018b; Gan et al., 2010). Some studies adopt expert-based Bayesian network models to predict future accident involvement based on past PSC inspection results (Hänninen and Kujala, 2014; Li et al., 2014; Hännine et al., 2014). In particular, Yang et al. (2018a) used a model similar to our TAN classifier but their model differs in the model construction process (which involves subjective variables and manually changing the structure of the network), the model training process (which is trained by the gradient descent approach) and the prediction target (i.e., ship detention).

2.2 Studies on the Bayesian network in maritime risk analysis

In recent years, we have witnessed a fast-growing number of maritime risk studies based on the Bayesian networks (BNs). Hänninen (2014) searched for and presented papers related to BNs applied to maritime safety. She concluded that BNs are rather well-suited tools for maritime safety management and development. There is a growing interest in and promising development of using BNs to conduct maritime risk analysis. In order to integrate different stakeholders' views and foundational perspectives on a risk ranking which could be used in complex systems such as the maritime transportation system, Goerlandt and Reniers (2017) proposed a BN model to combine the ranking methods based on the expected values, uncertainty, and moral perspective. Trucco et al. (2008) proposed a Bayesian belief network with conditional probabilities estimated using expert knowledge to model the Maritime Transportation System (MTS). Li et al. (2014) integrated logistic regression and BN to analyze maritime risks. The logistic regression model was able to provide parameters for the BN model to alleviate the bias brought by the expert estimation. Zhang et al. (2016) synthesized the statistics of historical accident data from 2008 to 2013 and expert judgement in the Bayesian belief network to express the dependencies between the indicator variables. Zhang et al. (2013) applied a formal safety assessment to evaluate the navigation risk of the Yangtze River and then constructed a data-based BN model to identify accident consequences. To reduce ship risk in ice-covered waters, Li et al. (2017) developed a BN model to link the ice conditions with the ship speed. The model could be used to generate the probability of a certain speed when the ice conditions were given and could be applied in risk assessment of route finding problems. Wróbel et al. (2016) analyzed the risk associated with unmanned ships by using a three-level BN model whose structure was determined based on the causes and effects of unfortunate events affecting ships' safety. Lu et al. (2019) proposed a BN model for assessing the effectiveness of oil spill recovery in icy conditions. A systematic approach was applied to establish the content and structure of the model, while various datasets were combined to estimate the probabilities of the model variables.

A serious drawback of the abovementioned BNs is that, due to the lack of historic data, most of the proposed BN models rely on expert knowledge in structure construction or model parameterization. The involvement of subjective judgements may bring about uncertainty and biases. Zhang and Thai (2016) thus pointed out that data-driven BNs are considered to be more objective since they are based on empirical data.

3. Methodology

3.1 Bayesian network (BN)

A Bayesian network (BN) is a directed acyclic graph containing a set of nodes and a set of directed arcs (Friedman et al., 1997). The nodes in the network represent the variables. The node at the tail of an arc is the parent node, which acts as the condition, while the node at the head is the child node of that parent node and is the consequence of that condition (Wang and Vassileva, 2003). The arcs from one node to its child nodes represent their dependencies. The BN is acyclic, which means that from any node, there must not be a way back to the same node. All the nodes in the network have a finite number of mutually exclusive states that represent the values of the corresponding variables. The values of a node can be either continuous or discrete, and our paper only focuses on the discrete values. A BN contains a network structure as the qualitative part and several probability parameters as the quantitative part. Compared to other prediction models, BNs have a solid mathematical background and present a graphical relationship that is easy to understand. In addition, the Bayesian approach performs well in coping with unknown probability parameters (Yu et al., 2012). It is therefore a commonly used method to analyze and predict maritime risks (Li et al., 2014; Zhang et al., 2016; Hänninen and Kujala, 2014).

3.2 The structure of the Tree Augmented Naive Bayes (TAN) classifier

Statistical classification identifies to which of a set of categories a new observation belongs based on the data training observations (Warfield et al., 2000). In the classification, the classifier is built from a set of training data and can be used to perform prediction on the testing data. One of the most widely used classifiers is the Naive Bayesian classifier, which is a simple probabilistic classifier based on Bayes' Theorem with strong (naive) independence assumptions between the features (Domingos and Pazzani, 1997; Hänninen, 2014; Zhang and Thai, 2016; Hazelton, 2010)). An example of a Naive Bayesian classier is illustrated in Figure 1. The Naive Bayesian classifier contains a class variable C , which is the classification target, e.g., the total number of deficiencies in a PSC inspection, and several attribute variables A_1 to A_4. Usually, the attribute variables are the properties and characteristics used to describe the cases, e.g., ship age, ship type, ship flag, and ship recognized organization. They are easy to access and thus act as the evidence for classifying. The classifier will be trained using a set of cases with known states of attribute variables and class variable (e.g., 250 past records of PSC inspection). Then, a new case with a set of attribute variables can be classified by the classifier to one state of the class variable (e.g., a ship visits a port and the PSC authority knows its age, type, flag and recognized organization, so the PSC authority can have an estimate of the number of deficiencies the ship has). In the Naive

Bayesian classifier, it is assumed that, given the class variable, every attribute variable is conditionally independent of the other attribute variables (Cheng and Greiner, 1999). However, there are actually more or fewer connections between the attribute variables (e.g., the flag states can authorize some certain recognized organizations to act on their behalf to carry out statutory survey and certification work of their ships). Hence, this assumption will influence the classification accuracy of the Naive Bayesian classifier (Dong et al., 2007).

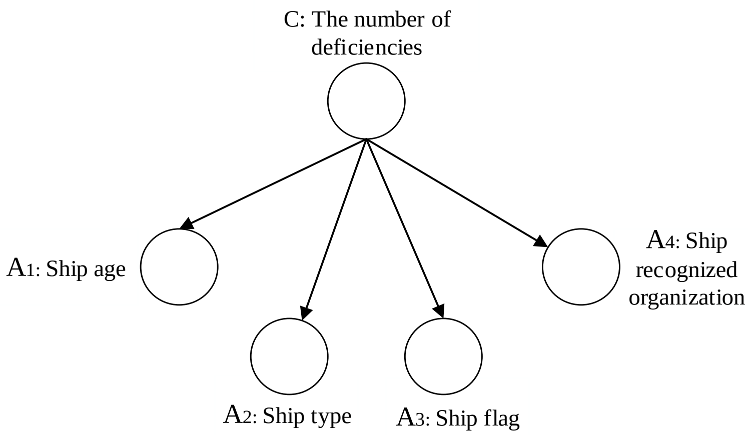


Figure 1. Example of a Naive Bayesian model.

To deal with the over-simplified assumption, the Tree Augmented Naive Bayes (TAN) classifier is proposed to identify the interactions between the attribute variables by using a tree structure (Friedman, 1997). An example of the TAN model is presented in Figure 2. As illustrated in the figure, a typical TAN classifier contains a class variable and several attribute variables. The class variable has no parent and is the parent of every attribute variable. Each attribute variable can have at most two parent variables including the class variable (Pernkopf, 2005). In this example, for instance, a flag state has expertise for registering certain types of ships, and it can authorize certain recognized organizations to act on its behalf to carry out statutory survey and certification work of their ships, so the node “Ship flag” depends on the node “Ship type”, and “Ship recognized organization” depends on “Ship flag”.

We now describe the TAN classifier mathematically. The class variable C has a total of N_C states; the set of these states is denoted by $S_C = \{c_1, ..., c_{N_C}\}$. The number of attribute variables is denoted by I and all the attribute variables are presented by a vector $A = (A_1, ..., A_I)$. The i th attribute, A_i, $i = 1, ..., I$, can take a total of N_i states, denoted by a state set $S_i = \{a_{i,1}, a_{i,2}, ..., a_{i,N_i}\}$.

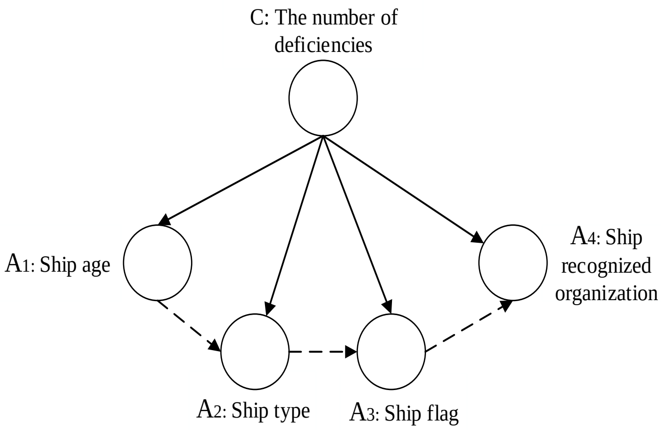


Figure 2. Example of TAN model.

The TAN classifier will be trained by a full case data set, which is a case whose values of both the class variable and attribute variables are known. The full case data set is denoted by $\mathbb{K}=\{1,...,K\}$, and one certain case is denoted by $k\in\mathbb{K}$. The state of the class variable of case k is denoted by $c^k\in S_C$; in other words, case k is classified to c^k. The state of attribute variable A_i of case k is denoted by $a_i^k\in S_i$, and thus its state vector of the attribute variables is denoted by $ATT^k=(a_1^k,...,a_I^k)$.

Based on the full data set $\mathbb{K}$, we can evaluate the dependency between two attribute variables. The dependency between two attribute variables A_i and A_j given the class variable C, $i,j=1,...,I$, $i\neq j$, is described by the conditional mutual information $I(A_i;A_j\,|\,C)$, which is the expected value of the mutual information of two random variables given the value of the third (Wyner, 1978). For a data set $\mathbb{K}$, the conditional mutual information for two attribute variables A_i and A_j is defined as (Cover and Thomas, 2012)

$$I(A_i;A_j\,|\,C)=\sum_{s'=1}^{N_i}\sum_{s''=1}^{N_j}\sum_{s=1}^{N_C}P(a_{i,s'},a_{j,s''},c_s)\log\frac{P(a_{i,s'},a_{j,s''}\,|\,c_s)}{P(a_{i,s'}\,|\,c_s)P(a_{j,s''}\,|\,c_s)} \tag{1}$$

where the "log" means the logarithmic operation with base 2 in this study[1] and $P(a_{i,s'},a_{j,s''},c_s)$, $P(a_{i,s'},a_{j,s''}\,|\,c_s)$, and $P(a_{i,s'}\,|\,c_s)$ are abbreviated forms of $P(A_i=a_{i,s'},A_j=a_{j,s''},C=c_s)$,

[1] The base of the logarithmic operation can be any value greater than 1, as long as all pairs of attribute variables use the same base. This is because it is not the absolute values but the ratios of the conditional mutual information for each pair of attribute variables that will affect the result of the TAN classifier.

$P(A_i = a_{i,s'}, A_j = a_{j,s''} \mid C = c_s)$, and $P(A_i = a_{i,s'} \mid C = c_s)$, respectively. This also applies to the remainder of the paper. $P(a_{i,s'}, a_{j,s''}, c_s)$ is the joint probability and $P(a_{i,s'}, a_{j,s''} \mid c_s)$ and $P(a_{i,s'} \mid c_s)$ are conditional probabilities. Since we use the data set $\mathbb{K}$ to calibrate the TAN network, $P(a_{i,s'}, a_{j,s''}, c_s)$ should be understood as the *proportion* of cases in $\mathbb{K}$ whose states of attribute variable A_i, attribute variable A_j, and class variable C are $a_{i,s'}$, $a_{j,s''}$, and c_s, respectively. Similarly, $P(a_{i,s'}, a_{j,s''} \mid c_s)$ should be understood as: among cases in $\mathbb{K}$ whose class variable state is c_s, the *proportion* of cases whose states of attribute variable A_i and attribute variable A_j are $a_{i,s'}$ and $a_{j,s''}$, respectively.

A complete TAN classifier contains the structure part and the quantitative part (Hruschka Jr and Ebecken, 2007). To learn the structure of the TAN classifier containing $A_1, ..., A_I$ as the attribute variables and C as the class variable, let function $\pi : \{1, ..., I\} \mapsto \{0, ..., I\}$ identify the parent attribute variable index for each attribute variable, and

$$\pi(i) = \begin{cases} i', \text{ if } A_i \text{ has a parent attribute variable } A_{i'}, \ i = 1, ..., I, i' = 1, ..., I \text{ and } i' \neq i \\ 0, \text{ if } A_i \text{ has no parent attribute variable, } i = 1, ..., I. \end{cases} \quad (2)$$

The construction of the TAN classifier consists of an optimization problem to find a tree defining a function π over $A_1, ..., A_I$ such that the tree sum of mutual information is maximized (Chow and Liu, 1968). In this study, a procedure called Construct-TAN (Friedman, 1997) is adopted to identify the tree, which is the qualitative part of the TAN classifier. The conditional probability tables constitute the quantitative part of the TAN classifier, and the conditional probabilities are estimated based on the full case data set and the learned TAN structure. The detailed procedure of constructing the TAN classifier will be explained in Section 4.3.

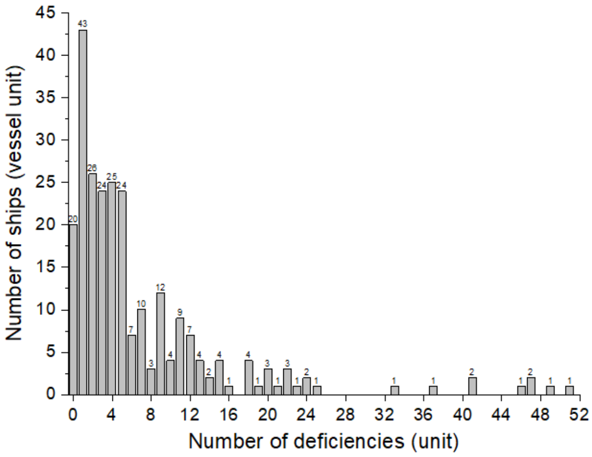
4. Model construction

4.1 Data

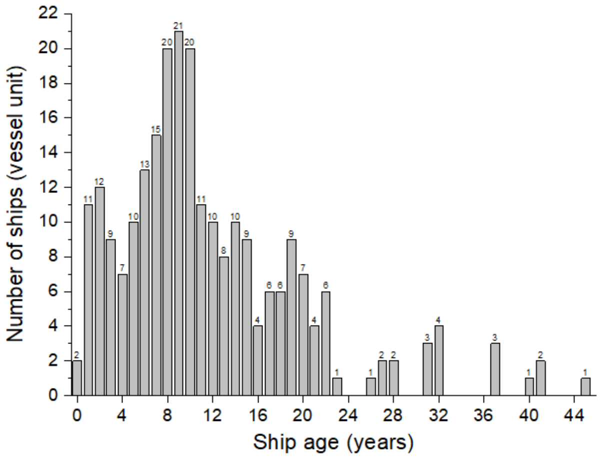
A case data set containing 250 PSC inspection records (full case data) from Hong Kong is denoted by $\mathbb{K}$ and established from the database of Tokyo MoU (http://www.tokyo-mou.org/inspections_detentions/psc_database.php). Inspected vessels with incomplete information are omitted. The inspection time range of these cases is from January 2017 to July 2017. Among the 250 records, 14 ships were inspected by PSC for the first time.

4.2 Identified variables

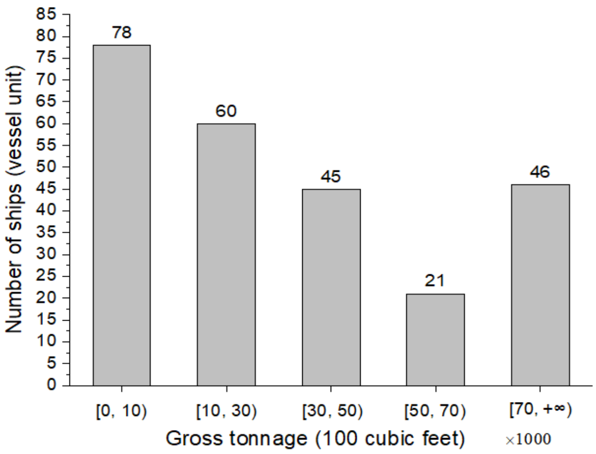
When a ship comes to the PSC inspection authority, it can be decided whether or not to inspect the ship if predictive information about the total number of deficiencies is available. To achieve this goal, we first choose the number of deficiencies as the class variable. According to the literature related to the factors influencing the inspection results (Yang et al., 2018a, b; Zhou and Sun, 2010; Xu et al., 2007; Gao et al., 2008), we select 10 attribute variables whose states are available once the ships come to the PSC authority and that may have an impact on the class variable (i.e. the number of deficiencies) to construct a TAN classifier. The 10 attribute variables are ship age, ship gross tonnage, number of previous detentions, last inspection time (months ago), number of deficiencies in last inspection, number of times of changing flag, ship type, ship flag, ship company, and ship recognized organization. The distribution of the 11 variables over the 250 cases is shown in Figure 3.



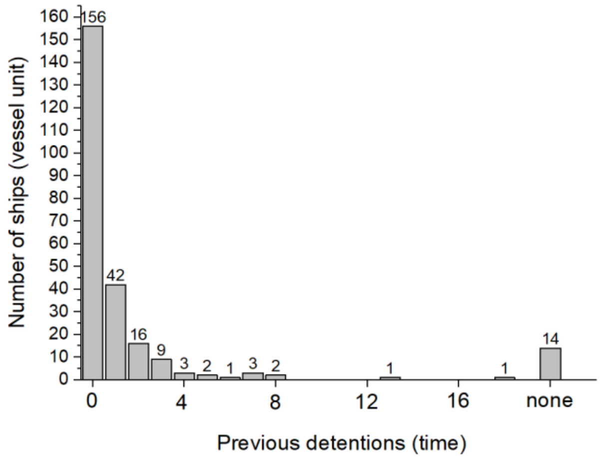
(a) Distribution of number of deficiencies



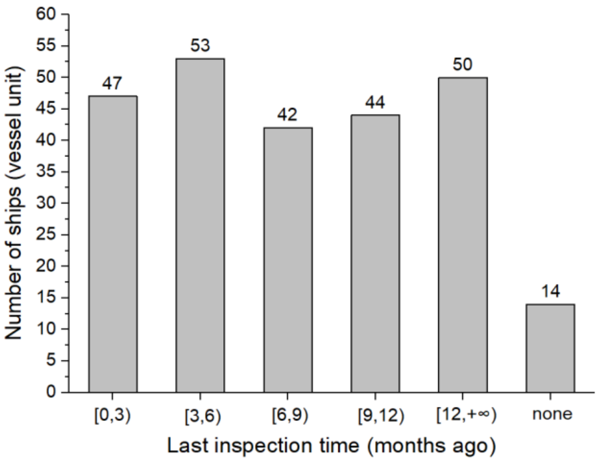
(b) Distribution of ship age



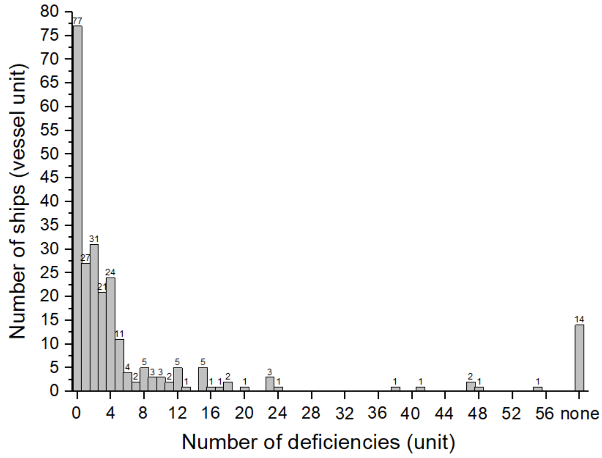
(c) Distribution of gross tonnage



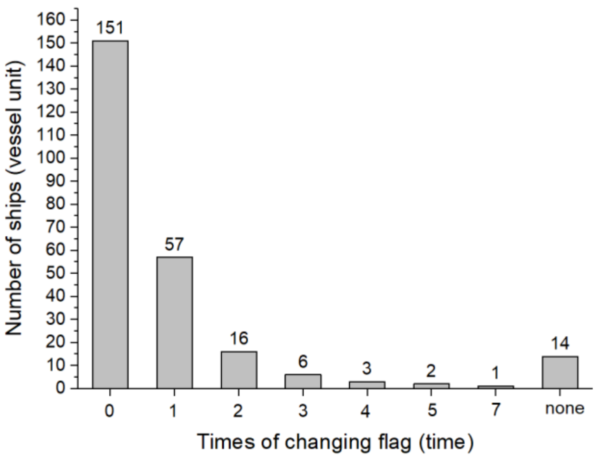
(d) Distribution of number of previous detentions



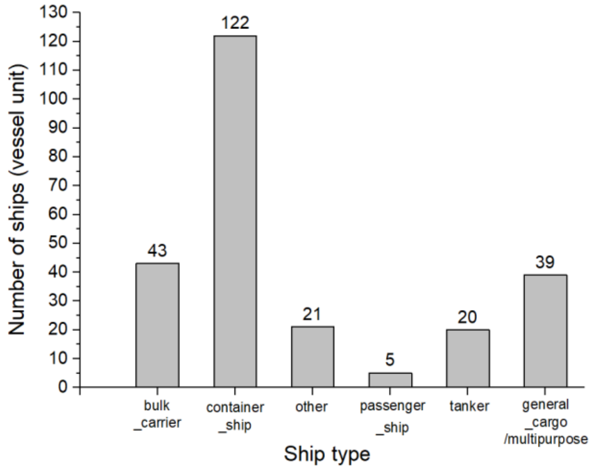
(e) Distribution of last inspection time



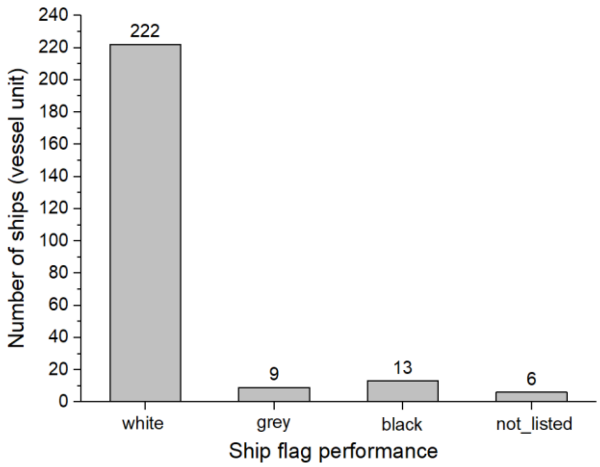
(f) Distribution of number of deficiencies in last inspection



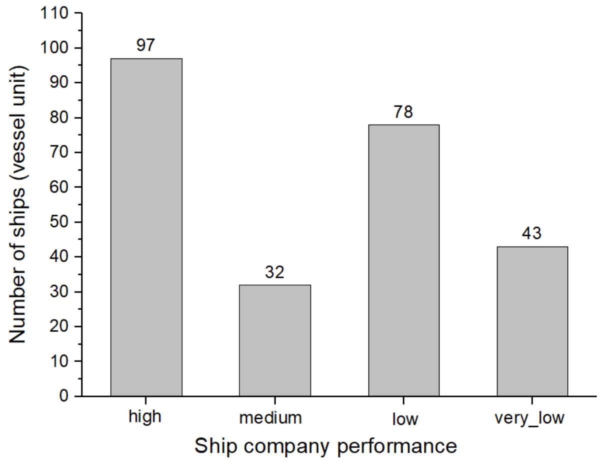
(g) Distribution of times of changing flag



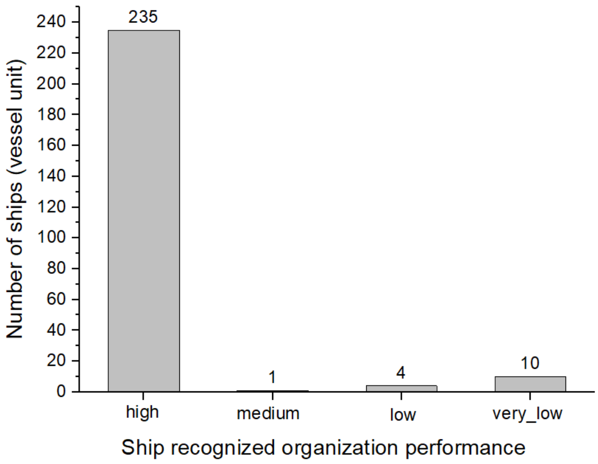
(h) Distribution of ship type



(i) Distribution of ship flag performance



(j) Distribution of ship company performance



(k) Distribution of ship RO performance

Figure 3. Distribution of the variables of all cases in the data set

(1) Number of deficiencies (class variable)

The number of deficiencies is the total number of deficiencies identified after the PSC inspection is conducted. It is the only variable that cannot be obtained when a ship comes to the PSC inspection authority. In the 250 inspection records, the number of deficiencies is between 0 and 51.

(2) Ship age

The age of a ship is the time difference (in years) between the keel laid date and the PSC inspection date. In the 250 inspection records, the ship age is between 0 and 45.

(3) Gross tonnage

The gross tonnage (GT) is a nonlinear measure of a ship's overall internal volume, with 100 cubic feet as the unit. In the 250 inspection records, the ship GT is between 299 and 194,308.

(4) Number of previous detentions

The number of previous detentions of a ship is the sum of the detentions from the first time the ship went through a PSC inspection. We use "none" as the state for this attribute variable for the 14 ships that were inspected for the first time. In the other 236 inspection records, the number of previous detentions is between 0 and 18.

(5) Last inspection time

The last inspection time of a ship is the time interval (in month) from the last PSC inspection to the time of the current PSC inspection. For the 14 ships that were inspected for the first time, we use "none" to represent the state of this attribute variable. In the other 236 inspection records, the last inspection time is between 0 and 180.7 months.

(6) Number of deficiencies in last inspection

The number of deficiencies in the last inspection is the number of deficiencies identified in the last PSC inspection. Similarly, we use “none” to denote the state for this attribute variable for the 14 ships that were inspected for the first time. In the other 236 inspection records, the deficiency number in the last PSC inspection is between 0 and 55.

(7) Number of times of changing flag

The number of times of changing flag is the sum of the times the ship’s flag has been changed since the first PSC inspection. Cariou and Wolff (2011) pointed out that vessels in relatively bad condition (resulting in detention or a high number of deficiencies) were more likely to be involved in flag changing activities to reduce the PSC inspection rate. In addition, Fan et al. (2014) concluded that a high PSC inspection rate would motivate ship flagging-out, i.e., changing the flag of the ship by registering the ship in a country other than the one in which it operates. Thus, we include this attribute variable in the TAN classifier. For the 14 ships that were inspected for the first time, we use “none” to represent the state of this attribute variable. In the other 236 inspection records, the flags of the ships were changed between 0 and 7 times.

(8) Ship type

According to the annual report on PSC from Tokyo MoU (Tokyo MoU, 2017a), the main types of ships that have been inspected in the Asia-Pacific region in 2017 are bulk carrier, container ship, general cargo/multipurpose, passenger ship, and tanker. Thus, the states of this variable are bulk carrier, container ship, general cargo/multipurpose, passenger ship, tanker and others.

(9) Ship flag

The performances of ship flags are reported in the annual report from Tokyo MoU (Tokyo MoU, 2017b). Assessment of the performance of each flag state takes into account the inspection and detention history over the preceding three calendar years and the flags are classified to be on the black list, grey list or white list. Only flags that have been involved in more than 30 PSC inspections during the previous three years are listed in the black-grey-white lists; otherwise the performance of the flag will not be listed (Tokyo MoU, 2017b). Thus, the states of this variable are white, grey, black and not listed.

(10) Ship company

The ship company refers to the ISM company for the ship (Tokyo MoU, 2017c), i.e., the ship operating company which is responsible for implementing the International Safety Management (ISM) code on ships. The performance of each company is judged by Tokyo MoU based on the company’s deficiency and detention performance and can be obtained by searching for the company IMO number in the Tokyo MoU database (Tokyo MoU, 2014). The states of ship company performance are high, medium, low and very low (Yang et al., 2018b).

(11) Ship recognized organization

Ship recognized organization (RO) is the classification society that carries out surveys and issues or endorses statutory certificates on behalf of a flag state. The performance of ROs is established annually and determined by the inspection and detention history over the last three calendar years (Paris MoU, 2013). The states of performance of the ship recognized organization are high, medium, low and not listed.

4.2.1 Discretizing the values of the variables into discrete states

As mentioned above, the TAN classifier works on discrete states of variables. The state of a variable can be represented by nominal data (nominal data has no order of rank), ordinal data (the order of rank is meaningful, e.g., strongly agree, agree, neutral, disagree, strongly disagree), and quantitative data. Quantitative data can be classified as discrete data and continuous data. The class variable and attribute variables in this study belong to the following categories: (i) "Ship type" is nominal data, and "ship flag", "ship company" and "ship recognized organization" are all ordinal data if we exclude the value "not listed". For nominal and ordinal states of variables, we consider each category of the values of a variable as a state of the variable. (ii) Gross tonnage and last inspection time are continuous quantitative data. Since the TAN classifier only deals with discrete states, we need to discretize the values of each continuous variable into a few states. Intuitively, we should discretize the possible values of a continuous variable into states of equal proportion. (iii) Ship age[2], number of previous detentions, number of times of changing flag, and number of deficiencies in last inspection are discrete quantitative data. Although they are discrete variables, their sets of possible values are too large and we need to propose a method to group the possible values into a smaller number of states.

For continuous variables (i.e., gross tonnage and last inspection time), the procedure of discretization into states of equal proportion is straightforward, because the values of the variable for all cases in $\mathbb{K}$ are different[3]. For example, suppose we want to discretize the possible values of a variable into states N of equal proportion, and the values of the variable in the K cases in $\mathbb{K}$ are listed in ascending order $v_1,...,v_K$, $K \geq N$. Then, defining $\lceil x \rceil$ as the smallest integer greater than or equal to x, values in the interval $[v_1, v_{\lceil K/N \rceil}]$ should be in the first state, values in the

[2] Ship age should normally be continuous data. But in our study the ship age is recorded as an integer number of years and hence it is considered to be discrete data.

[3] Due to the limited precision of measurement and recording, it is possible that two values are equal. The chance that two values are equal is small and has little effect on our model.

interval $[v_{\lceil K/N \rceil+1}, v_{\lceil 2K/N \rceil}]$ should be in the second state, and values in $[v_{\lceil (N-1)K/N \rceil+1}, v_K]$ should be in the N th state. To ensure that the states cover all possible values of the variable, including values that are not included in the full data set but may appear in future cases, we can define the first state as $(-\infty, \ (v_{\lceil K/N \rceil} + v_{\lceil K/N \rceil+1})/2]$, the second state as $((v_{\lceil K/N \rceil} + v_{\lceil K/N \rceil+1})/2, \ (v_{\lceil 2K/N \rceil} + v_{\lceil 2K/N \rceil+1})/2]$, and the N th state as $((v_{\lceil (N-1)K/N \rceil} + v_{\lceil (N-1)K/N \rceil+1})/2, \ +\infty)$.

For discrete variables (i.e., ship age, number of previous detentions, number of times of changing flag, and number of deficiencies in last inspection), a natural way is to consider each possible value (e.g., 1, 2, … for ship age) as a state. However, this will lead to a large number of combinations of states considering that the TAN classifier accounts for the dependencies between variables. A large number of combinations of states require an extremely large full data set (e.g., billions of records), otherwise the number of cases in some states will be extremely small. Since we have only 250 records, we combine several possible values of a variable into one state; for example, ages between 0 and 5 can be considered as one state, ages between 6 and 10 can be considered as another state. Aggregating values of a variable into states should not be conducted in an arbitrary way. Instead, the possible values of a variable should be discretized into states of equal or approximately equal proportion. The process of discretizing the values of a discrete variable into a few states of equal proportion is not as straightforward as that of discretizing the values of a continuous variable. For a discrete variable, it is highly probable that some cases have exactly the same value and these cases should be in the same state. It should be noted that although the idea of the equal-frequency discretization method has been used in the BN-related literature (Dougherty and Sahami, 1995; Flores et al., 2011), no rigorous discretization method is proposed and there are ambiguities in implementation. We formally state the problem of discretizing the values of a discrete variable into states of as equal proportion as possible:

Data discretization problem: A data set of K cases has a discrete variable. There are V categories of values in ascending order for the discrete variable in the K cases and the number of cases in category $v = 1,...,V$ is θ_v. $K = \sum_{v=1}^{V} \theta_v$. The data discretization problem aims to discretize the V categories into N states of consecutive categories, $N \le V$, such that each state has at least one category and the proportion of cases that fall into each state is as close to $1/N$ as possible. Letting Z^+ be the set of non-negative integers, the problem is to find integer values $s_0, s_1, s_2, ..., s_N$ that solve the following optimization problem:

$$\text{Min} \sum_{n=1}^{N} \left(\frac{\sum_{v=s_{n-1}+1}^{s_n} \theta_v}{K} - \frac{1}{N} \right)^2 \tag{3}$$

subject to

$$s_n \geq s_{n-1} + 1, n = 1, ..., N \tag{4}$$

$$s_n \in Z^+, n = 1, ..., N \tag{5}$$

$$s_0 = 0 \tag{6}$$

$$s_N = N \ . \tag{7}$$

The objective function (3) minimizes the sum of squared deviations of the proportion of each state from the average proportion $1/N$. The first state will be $(-\infty, \quad (v_{s_1} + v_{s_1+1})/2]$, the second state will be $((v_{s_1} + v_{s_1+1})/2, \quad (v_{s_2} + v_{s_2+1})/2]$, and the N th state will be $((v_{s_{N-1}} + v_{s_{N-1}+1})/2, \quad +\infty)$[4].

Theorem 1: The data discretization problem can be solved in time bounded by $O(NV^2)$.■

The idea of proving Theorem 1 is to use dynamic programming to solve model (2). The detailed proof is in Appendix A.

4.2.2 States of the variables

The total data set contains $K = 250$ inspected ships, where there are 14 ships without previous PSC inspections. For the variables "the number of deficiencies", "ship age" and "ship gross tonnage", which are irrelevant to the previous inspections, we discretize their states into $N = 3$ states. For the variables that are related to previous PSC inspections, including "the number of previous detentions", "last inspection time", "the number of deficiencies in last inspection" and "the number of times of changing flag", we discretize them into $N' = 4$ states, with the state "none" for the 14 ships without former inspection, while the remaining three states contain $K' = 250 - 14 = 236$ ships. The states of the variables are in Table 1.

4.3 Construct the qualitative part of the TAN classifier

There are six steps to construct the qualitative part of a TAN classifier in PSC inspection according to the Construct-TAN procedure (Friedman et al., 1997).

[4] If the values of the variable can only be integers, then the intervals for the states can be truncated so that the end points of each interval are both integers.

Table 1 Variables in TAN classifier

Variable	Unit	Type	Node name	States
Number of deficiencies		discrete	deficiency_no	S1:0to2, S2:3to6, S3:7+
Ship age	year	discrete	age	S1:0to7, S2:8to12, S3:13+
Gross tonnage	100 cubic feet	continuous	GT	S1:0to11228, S2:11229to40053, S3:40054+
Number of previous detentions		discrete	pre_detention	S1:zero, S2:one, S3:2+, S4:none
Last inspection time	month	continuous	last_inspection	S1:0to5.5, S2:5.6to9.6, S3:9.7+, S4:none
Number of deficiencies in last inspection		discrete	last_deficiency_no	S1:zero, S2:1to3, S3:4+, S4:none
Times of changing flag		discrete	change_flag	S1:zero, S2:one, S3:2+, S4:none
Ship type		nominal data	type	S1:bulk_carrier, S2: container_ship, S3:general_cargo/multipurpose, S4:passenger_ship, S5:tanker, S6:other
Ship flag		ordinal data	flag	S1:white, S2:grey, S3:black, S4:not_listed
Ship company		ordinal data	company	S1:high, S2:medium, S3:low, S4:very_low
Ship recognized organization		ordinal data	RO	S1:high, S2:medium, S3:low, S4:not_listed

4.4 Constructing the quantitative part of the TAN classifier

There are two components in the quantitative part of the TAN classifier: the marginal probability distribution of each variable and the conditional probability table (CPT) for each variable. Marginal probability, denoted by $P(X = x)$, is an unconditional probability of the occurrence of state x of event X. The probabilities of states corresponding to each variable are the marginal probabilities in percentage form, as shown in Figure 5.

Conditional probability $P(A|B)$ is the probability of A under condition B. In the BN models, the conditional probabilities of each attribute variable are presented in conditional probability tables (CPTs). The method used to calculate the CPTs is presented in Appendix C. The root variable (i.e., the class variable deficiency_no) has no parent and therefore its conditional probabilities are reduced to prior probabilities. Now, the construction process of the quantitative part of the TAN classifier is done, which involves generating the marginal probability distribution of each variable as presented in Figure 5 and the CPT for each variable as presented in Appendix C.

Procedure 1. *Construct-TAN procedure.*

Step 1: Select deficiency_no as the class variable, and age, GT, type, flag, company, RO, pre_detention, last_inspection, last_deficiency_no and change_flag as attribute variables.

Step 2: Compute the conditional mutual information between all pairs of attribute variables given the class variable $I(A_i; A_j \mid C)$ to identify their dependency, $A_i \neq A_j$, $i = 1, \ldots, 10,\ j = 1, \ldots, 10, i \neq j$.

Step 3: Build a complete undirected graph with attribute variables as the nodes and the conditional mutual information $I(A_i; A_j \mid C)$ as the weight of the edge of A_i and A_j. The results are shown in Table 2.

Step 4: Build the maximum weighted spanning tree by sorting the weights of the edges from large to small, and then choose the edges from the largest weight to the smallest weight without forming a circle. For each chosen edge, if adding this edge forms a circle, it will not be chosen anymore; instead, edges with weights smaller than this edge will be chosen from larger weight to smaller weight. Keep the chosen edges and delete the others. The selected edge weights are in bold in Table 2.

Step 5: Transform the undirected spanning tree into a directed tree by choosing age as the root variable and setting the directions of all arcs to other attribute variables to be outward from it.

Step 6: Add the class variable C (i.e. deficiency_no) to the tree and arcs from the class variable to every attribute variable. The structure of the TAN classifier is presented in Figure 4.

Table 2 Conditional mutual information of attribute variables

	age	GT	type	flag	company	RO	pre_ detention	last_ inspection	last_ deficiency _no	change_ flag
age		0.051	0.081	0.030	0.031	0.032	0.073	0.063	0.069	**0.108**
GT			**0.198**	0.063	0.075	0.013	0.069	0.092	0.043	0.069
type				**0.099**	**0.141**	0.054	0.110	0.083	0.103	0.074
flag					0.080	**0.067**	0.068	0.067	0.065	0.089
company						0.033	0.108	**0.122**	0.060	0.118
RO							0.044	0.045	0.036	0.045
pre_detention								**0.276**	**0.250**	**0.268**
last_inspection									0.246	0.247
last_deficiency_no										0.243
change_flag										

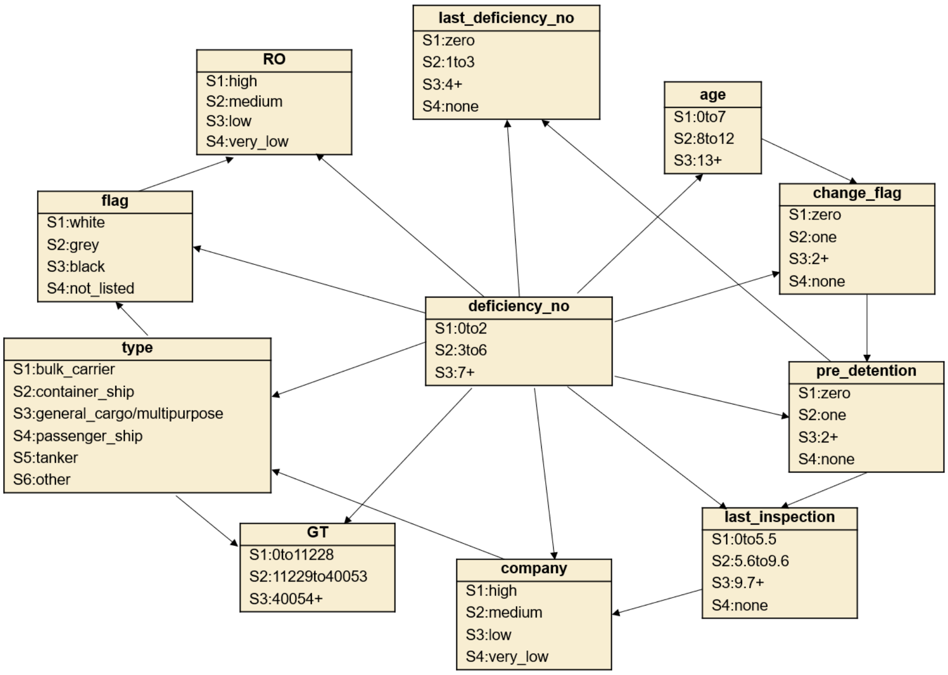


Figure 4. Structure of the TAN classifier for PSC inspection.

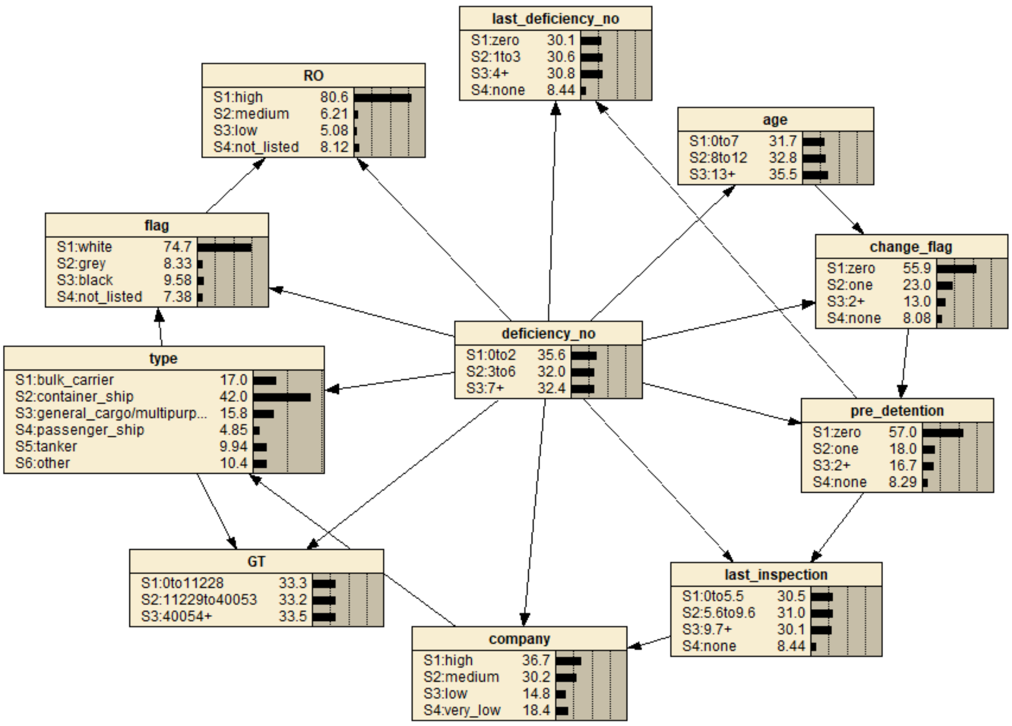


Figure 5. TAN model with marginal probabilities for PSC inspection.

4.5 Classification process for coming vessels

The TAN classifier obtained in the previous subsections has $I = 10$ attribute variables, and its class variable has $N_C = 3$ states: "0to2" is the first state, "3to6" is the second state, and "7+" is the third state. We define A_{i_0} as the root attribute variable (A_{i_0} is "age" for this classifier). Recall that $A_{\pi(i)}$ is the parent attribute variable of attribute variable $A_i, i=1,...,I, i \neq i_0$. For a specific incoming vessel k with attribute variable set $ATT^k = (a_1^k,...,a_I^k)$, the TAN classifier can calculate the probability for it to belong to each state $c_s \in S_C$ of the class variable. For ease of exposition, we define

$$\begin{aligned} &\tilde{P}^I(a_{1,k^{(1)}},...,a_{I,k^{(I)}},c_{\bar{s}}) \\ &= P^I(c_{\bar{s}}) \times P^I(a_{i_0,k^{(i_0)}} \mid c_{\bar{s}}) \times \prod_{i=1,i\neq i_0}^{I} P^I(a_{i,k^{(i)}} \mid a_{\pi(i),k^{(\pi(i))}}, c_{\bar{s}}) \\ &= P^I(c_{\bar{s}}) \times P^I(a_{i_0,k^{(i_0)}} \mid c_{\bar{s}}) \times \prod_{i=1,i\neq i_0}^{I} \frac{P^I(a_{i,k^{(i)}}, a_{\pi(i),k^{(\pi(i))}} \mid c_{\bar{s}})}{P^I(a_{\pi(i),k^{(\pi(i))}} \mid c_{\bar{s}})}, \bar{s}=1,...,N_C \end{aligned} \tag{8}$$

where the superscript " I " means the TAN has I attribute variables and $\bar{s}=1,...,N_C$ refers to the three states of the class variable. Then, the probability that vessel k belongs to $c_s \in S_C$ is calculated by the following posterior probabilities formula:

$$P^I(c_s \mid a_{1,k^{(1)}},...,a_{I,k^{(I)}}) = \frac{\tilde{P}^I(a_{1,k^{(1)}},...,a_{I,k^{(I)}},c_s)}{\sum_{\bar{s}=1}^{N_C}\tilde{P}^I(a_{1,k^{(1)}},...,a_{I,k^{(I)}},c_{\bar{s}})}, s=1,...,N_C. \tag{9}$$

Two ships chosen from the testing data set are used to show the deficiency number classification process. The detailed information of the attribute variables of the two incoming ships is shown in Table 3. The results of the classification process are shown in Table 4.

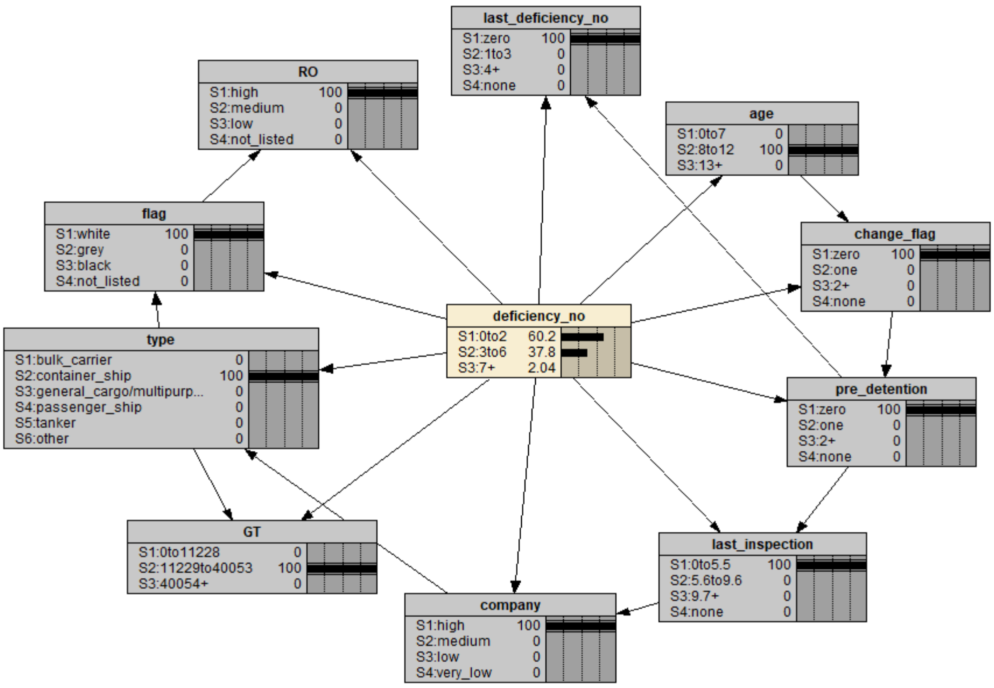
Table 3 Information on the incoming vessels

Ship 1:		**Ship 2**:	
Attribute variables	State	Attribute variables	State
age	S2:8to12	age	S3:13+
type	S2:container_ship	type	S3:general_cargo /multipurpose
GT	S2:11229 to40053	GT	S3:0to11228
RO	S1:high	RO	S2:medium
flag	S1:white	flag	S4:not_listed
company	S1:high	company	S4:very_low
change_flag	S1:zero	change_flag	S1:zero
pre_detention	S1:zero	pre_detention	S2:one
last_inspection	S1:0to5.5	last_inspection	S1:0to5.5
last_deficiency_no	S1:zero	last_deficiency_no	S3:4+

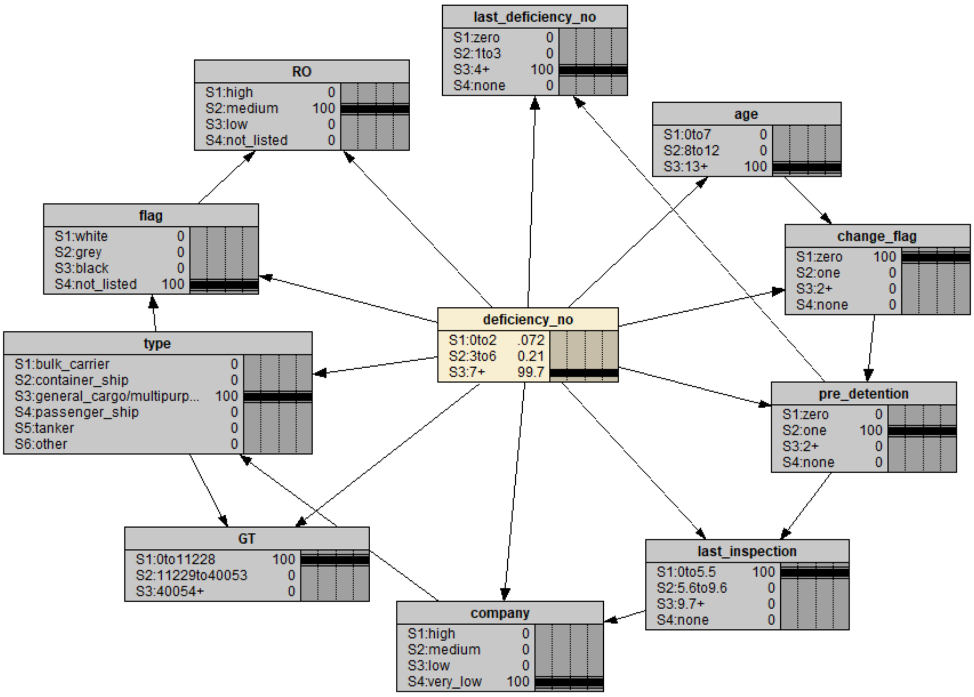
Table 4 Classification results of the incoming vessels

Ship 1		**Ship 2**	
$\tilde{P}(S1:0to2)$ in Eq. (8)	3.46×10^{-4}	$\tilde{P}(S1:0to2)$ in Eq. (8)	3.44×10^{-8}
$\tilde{P}(S2:3to6)$ in Eq. (8)	2.17×10^{-4}	$\tilde{P}(S2:3to6)$ in Eq. (8)	1.00×10^{-7}
$\tilde{P}(S3:7+)$ in Eq. (8)	1.17×10^{-5}	$\tilde{P}(S3:7+)$ in Eq. (8)	4.77×10^{-5}
$P(S1:0to2)$ in Eq. (9)	60.17%	$P(S1:0to2)$ in Eq. (9)	0.07%
$P(S2:3to6)$ in Eq. (9)	37.79%	$P(S2:3to6)$ in Eq. (9)	0.21%
$P(S3:7+)$ in Eq. (9)	2.04%	$P(S3:7+)$ in Eq. (9)	99.70%

This classification process can also be shown visually by selecting the corresponding states of each variable in Figure 6. The posterior probability distribution of the deficiency_no is shown in the corresponding node.



(a). Classification process of ship 1



(b). Classification process of ship 2

Figure 6. Illustration of the classification process of the new incoming ships

Now we are ready to present the results: the probabilities for ship 1 to have 0 to 2 deficiencies, 3 to 6 deficiencies and more than 7 deficiencies are 60.17%, 37.79%, and 2.04% respectively. As the state with the highest probability is the predicted range of the number of deficiencies, we can conclude that the incoming vessel is most likely to have 0 to 2 deficiencies. Meanwhile, the probabilities for ship 2 to have 0 to 2 deficiencies, 3 to 6 deficiencies and more than 7 deficiencies are 0.72%, 0.21%, and 99.70% respectively, and thus the estimated deficiency number of this vessel is more than 7.

4.6 Effect of the choice of root attribute variable

Based on the construction of the TAN classifier and the posterior probabilities formulae (7) and (8) for classifying a case k, we have the following theorem:

Theorem 2: To construct a TAN classifier with I attribute variables, $I \geq 2$, different choices of root attribute variable node in Step 5 of the Construct-TAN procedure all have the same posterior probability of classifying a case k into a state to $c_s \in S_C$ in Eq. (8).■

We use mathematical induction to prove Theorem 2. The detailed proof is in Appendix B.

5. Model validation and results

As a classifier, a typical way to validate the model is to evaluate how well it performs on unseen data, i.e., to check the classification accuracy using a testing data set (Hänninen, 2014; Hänninen and Kujala, 2014). We construct the TAN model by inputting the ships' attribute variable states (i.e., states of age, flag, GT, etc.) and the class variable state (i.e., state of deficiency_no) in the training case set to learn the structure and parameters of the TAN classifier. To validate the model, in addition to the 250 cases in set Ψ, we collected a set of another 50 cases, denoted by Ψ', which is mainly used as the testing data set.

5.1 Classification accuracy

To analyze the classification accuracy of the TAN model, we first construct a test case set containing the first $m \in \{50, 100, 150, 200, 250\}$ inspections in Ψ. Then, we put in the attribute variable states of each ship in Ψ' and use the TAN classifier to calculate the state of deficiency_no. If the ship is indeed in the deficiency_no state, then the classification is accurate; otherwise it is inaccurate. The classification accuracy results for $m \in \{50, 100, 150, 200, 250\}$ training cases are listed in Table 5. It can be seen from the table that as the scale of the training set increases, the classification accuracy shows an upward trend. When the training set contains more than 200 cases, the prediction accuracy is beyond 60%. This is almost twice as accurate as a random guess.

Table 5 TAN classifier accuracy

Number of training cases	Number of testing cases	Error rate	Accuracy rate
50	50	50%	50%
100	50	48%	52%
150	50	42%	58%
200	50	40%	60%
250	50	38%	62%

5.2 Comparison between TAN classifier and Ship Risk Profile (SRP)

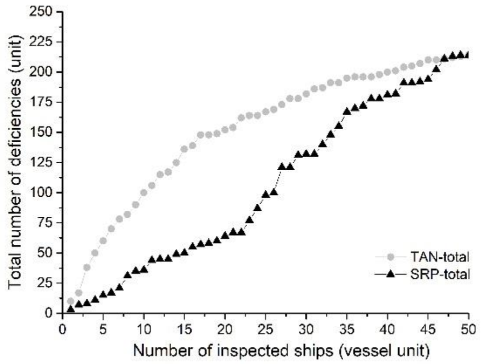
5.2.1 Introduction to SRP and comparison method

The Ship Risk Profile (SRP) is the method currently used by Tokyo MoU for selecting ships to conduct PSC inspections, which is calculated daily in the corresponding PSC MoU's database (Tokyo MoU, 2014). Different weighting points are given to different states of ship type, ship age, ship flag performance, ship RO performance, ship company performance, previous number of deficiencies and detentions. Based on the total weighting points, the ships are classified into three risk profiles: high risk ship (HRS), standard risk ship (SRS) and low risk ship (LRS). At the same time, time windows of 2 to 4 months, 5 to 8 months, and 9 to 18 months, which refer to the time since the previous inspection, are attached to HRS, SRS, and LSR, respectively. The current inspection selection scheme is based on the ship inspection priority: ships without prior inspection are Priority I; incoming ships whose time window has been closed (i.e., HRS, SRS and LRS with last inspection time of more than 4 months, 8 months, and 18 months respectively) are Priority II. Ships within the time window (i.e., HRS, SRS and LRS with the last inspection time between 2 to 4 months, 5 to 8 months and 9 to 18 months respectively) are Priority III. Ships that do not enter the time window are of Priority IV.

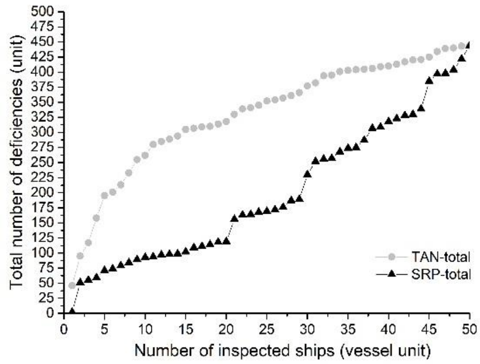
We compare the "effectiveness" of the currently used SRP inspection scheme and the newly constructed TAN classifier. The port authority wishes to identify as many deficiencies as possible after inspecting a certain number of ships for the following two reasons: first, the inspection results only contain ship deficiencies and ship detention, but the ship detention rate is low. A more direct approach to improve the inspection efficiency is to inspect ships with a larger number of deficiencies. Second, larger numbers of deficiencies are also supposed to have strong relationship with ship detention (Yang et al., 2018a; Cariou and Wolff, 2015). Thus, the "effectiveness" here refers to the "quickness" of identifying the ships with larger numbers of deficiencies. This can be reflected by the inspection sequence of the incoming ships generated by using the two selection methods. The TAN classifier used for comparison is the one proposed in Section 4, which is trained by data set Ψ (training set 1). Both SRP and the TAN classifier use the same testing data set Ψ' (testing set 1). Suppose that the ships in Ψ' arrive at the PSC authority at the same time, and the PSC authority has the resources to inspect $n = 1, 2, ..., 50$ ships. If the SRP selection scheme is used, a list of n ships will be chosen for inspection based on Procedure 2 in Appendix D; if the TAN classifier is used, another list of n ships will be chosen for inspection based on Procedure 3 in Appendix E. We can then calculate the total numbers of deficiencies they can detect after inspecting the same number of ships n to compare their efficiency.

5.2.2 Comparison results

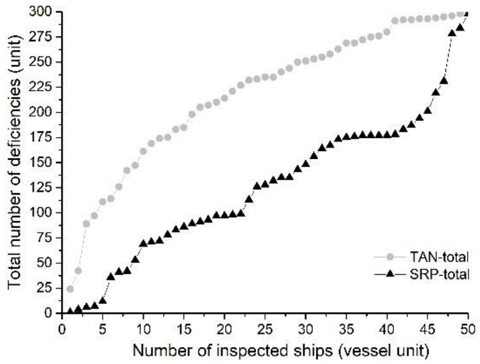
We enumerate the possible values of $n = 1, 2, ..., 50$ and draw the two total detected deficiency number curves in Figure 7.



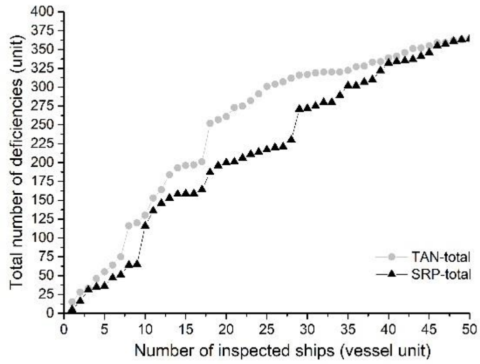
(a) Comparison results of Testing set 1



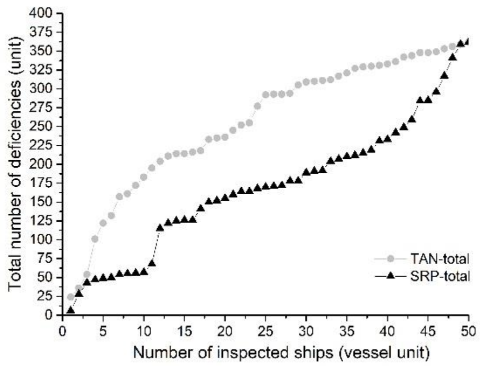
(b) Comparison results of Testing set 2



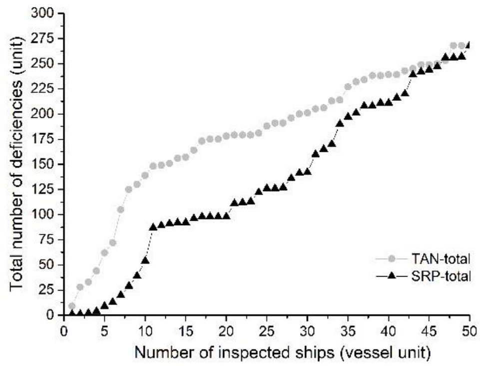
(c) Comparison results of Testing set 3



(d) Comparison results of Testing set 4



(e) Comparison results of Testing set 5



(f) Comparison results of Testing set 6

Figure 7. Comparisons of ship selection efficiency between SRP and TAN classifier

Figure 7(a) illustrates that the selection performance of the TAN classifier significantly outperforms the currently used SRP selection scheme. We define the improvement of the TAN classifier over the SRP selection scheme at the m th inspection (denoted by $I(m)$) and the average improvement (denoted by AI) after the total M inspections as follows:

$$I(m)=\frac{total\_de(TAN(m))-total\_de(SRP(m))}{total\_de(SRP(m))}\times 100\%$$

$$AI=\frac{\sum_{m=1}^{M} I(m)}{M}$$

where $total\_de(TAN(m))$ and $total\_de(SRP(m))$ are the total numbers of deficiencies detected after the m th inspection by the TAN classifier and SRP selection scheme respectively, and $M=50$. In Figure 7(a), the average improvement over the 50 ships in testing set 1 is 101.00%. We further assume that the port authority has the ability to inspect 10%, 20%, …, 60% of all the 50 incoming ships, and the improvements of the TAN classifier over the SRP ship selection scheme after inspecting 5, 10, 15, 20, 25, and 30 ships are 300%, 177.78%, 172%, 137.5%, 70.41%, and 37.88% respectively. These statistics tell us that when the PSC authority only has limited resources to inspect the incoming ships, the TAN classifier can help to identify ships with higher risks better.

It is worth mentioning that, in Figure 7(a), the ship with the largest number of deficiencies among the total 50 ships (i.e. 21 deficiencies) is ranked 3rd in the inspection list generated by the TAN classifier, while it is 27th on the inspection list in the SRP selection scheme. Although this ship is in the HRS category, it was inspected in Shandong, China, 2.3 months ago and is thus within the inspection time window. As the SRP only takes the inspection time window into consideration among all the high-risk ships, ships that have been inspected a short time ago would have lower risk indices than many other ships and are thus at the end of the SRP inspection list. In addition, the weighting points given to the risk parameters in SRP are rough; for example, all types of ships with age more than 12 will be given 1 weighting point, those with low or very low RO performance will be given 1 weighting point and those with low or very low company performance will be given 2 weighting points. Moreover, if the total weighting point is larger than or equal to 4, it is classified as an HRS, with no more information attached except for an inspection time window. On the contrary, the TAN classifier is more sensitive to the states of the attribute variables, as it treats them in a detailed manner (e.g., all the states of the attribute variables are taken into account instead of some extreme states) while also taking the dependencies between the variables into consideration. What is more, the TAN classifier can generate an expected number of deficiencies (i.e. $E(deficiency\_no)$) for each individual ship, which can better distinguish the ships instead of roughly classifying them into three risk profiles. For this ship, the age of 8 to 12, flag on the grey list, company of very low performance, RO of medium performance, more than two times of changing flag and previous detentions all give it a higher probability of having a larger number of deficiencies in the TAN classifier. As a consequence, the TAN classifier assigns a higher priority to this ship than the SRP selection scheme does.

To further test the robustness of the performance of the TAN classifier, we randomly divide the 250 training data cases in Ψ into five mutually exclusive data sets, denoted by Ψ_1, Ψ_2, Ψ_3, Ψ_4, and Ψ_5, each containing 50 cases. Then, we obtain five new training sets and the corresponding testing sets: $\Psi' \cup \Psi_2 \cup \Psi_3 \cup \Psi_4 \cup \Psi_5$ (training set 2) and Ψ_1 (testing set 2), $\Psi_1 \cup \Psi' \cup \Psi_3 \cup \Psi_4 \cup \Psi_5$ (training set 3) and Ψ_2 (testing set 3), $\Psi_1 \cup \Psi_2 \cup \Psi' \cup \Psi_4 \cup \Psi_5$ (training set 4) and Ψ_3 (testing set 4), $\Psi_1 \cup \Psi_2 \cup \Psi_3 \cup \Psi' \cup \Psi_5$ (training set 5) and Ψ_4 (testing set 5), and $\Psi_1 \cup \Psi_2 \cup \Psi_3 \cup \Psi_4 \cup \Psi'$ (training set 6) and Ψ_5 (testing set 6). After comparing the TAN and SRP selection scheme by using the five training sets and the corresponding testing sets, we find that the TAN classifier can detect 141.29%, 215.54%, 25.83%, 75.31% and 193.76% more deficiencies on average each time, with 130.35% more deficiencies than the SRP selection scheme on average in total. After further assuming that the port state has the resources to inspect 10%, 20%, 30%, 40%, 50%, and 60% of the 50 total incoming ships, we can calculate that the average improvement of the TAN classifier is 348.38%, 147.23%, 108.32%, 98.29%, 70.33%, and 48.83% after inspecting 5, 10, 15, 20, 25, and 30 ships, respectively. The comparisons are illustrated in Figure 7. The reasons for the superior performance of the TAN classifier are as follows. First, the SRP selection scheme attaches a fixed time window for the ships and this will unconditionally give a high priority for ships out of the time window to be inspected first, even if some of them have fewer deficiencies. Meanwhile, in the TAN classifier, the last inspection time is just viewed as one attribute variable. Second, the weighting point given to each parameter is based on expert knowledge and is fixed in the SRP selection scheme. In contrast, the TAN classifier is based on a mathematical model, as the probabilities are all based on the statistical data and the classification process is based on Bayes' Theorem. Third, the ships are divided into three categories (excluding the small number of ships that have not been inspected before) in the SRP selection scheme, which means that there are 1/3 ships in each category on average and these ships will have the same inspection time window (i.e., the same inspection priority). On the contrary, the TAN classifier can generate a different risk index for each incoming ship to better distinguish them in order to identify the ships of higher risk.

5.3 Comparison between TAN classifier and ordered logistic regression

5.3.1 Introduction to ordered logistic regression

Among the most widely adopted methods in the research on PSC inspection are logistic regression models (Knapp and Franses, 2007; Knapp et al., 2011; Knapp and Hänninen, 2014; Li et al., 2014). Thus, we compare the performance of the TAN classifier and the logistic regression model in identifying ships with larger numbers of deficiencies. It should be noted that the logistic regression models proposed in the abovementioned studies in the brackets are all binary logistic regression models, in which the regression target has only two states. In our study, there are three states of "deficiency_no", and its states are ordinal (i.e., the conditions of ships with 0 to 2 deficiencies are better than ships with 3 to 6 deficiencies and are much better than ships with more than 7 deficiencies). We thus extend the binary logistic regression model to a multilevel ordered logistic regression model, which is a regression model used for ordinal dependent variables with multiple states (McCullagh, 1980). For more detail on the multilevel ordered logistic regression models, please refer to Menard (2002). We use the input data set that is used to construct the TAN classifier in Section 4, and the assumptions of the multilevel ordered logistic regression on the input data are guaranteed: (a) the input data are categorical; (b) there is no multicollinearity in the input data; (c) the input data are proportional odds, i.e., each independent input variable has an identical effect at each cumulative split of the ordinal dependent variable (Menard, 2002). It should be noted that the "independence" of the input data does not mean that the input variables are statistically independent of each other; instead, only the non-multicollinearity of the input data needs to be guaranteed. The verification of input data and the construction of the multilevel ordered logistic regression model are conducted in SPSS software.

5.3.2 Comparison results

After estimating the parameters in the multilevel ordered logistic regression model, we use testing set 1, which is used to test the TAN classifier, to test the performance of the logistic regression model. The testing method is almost the same as Procedure 3 proposed in Appendix E, which is used to test the performance of the TAN classifier, i.e., calculating the estimated deficiency number based on the probabilities and average deficiency numbers of different states of "deficiency_no". The comparison results are shown in Figure 8.

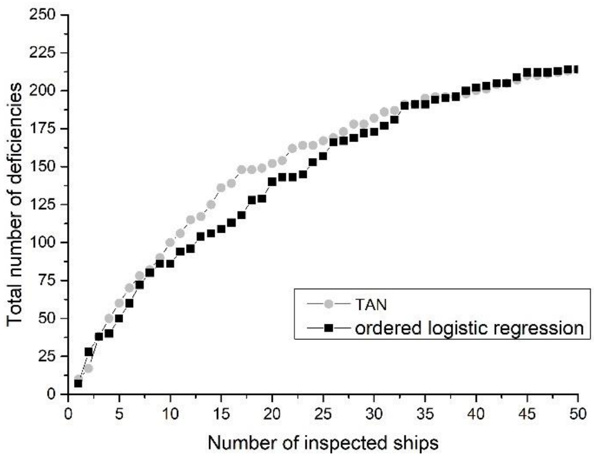


Figure 8. Comparison between TAN classifier and ordered logistic regression model

We can see from Figure 8 that the TAN classifier outperforms the multilevel ordered logistic regression model. It can detect 6.70% more deficiencies on average than the ordered logistic regression model. We further assume that the port state has the resources to inspect 10%, 20%, 30%, 40%, 50%, and 60% of all the incoming ships, and that the TAN classifier can detect 20%, 16.28%, 24.77%, 8.57%, 6.37%, and 5.20% more deficiencies than the ordered logistic regression model.

6. Variable analysis

6.1 Dependency of class variable on attribute variables

Mutual information on two random variables is a measure of the mutual dependence between two variables (Fraser and Swinney, 1986). In the proposed TAN classifier trained by 250 cases, we use the mutual information $I(A_i;C)$ between each attribute variable and the class variable to present the extent to which the attribute variables have an influence on the number of deficiencies. $I(A_i;C)$ can be calculated by the following formula:

$$I(A_i;C)=\sum_{s'=1}^{N_i}\sum_{s=1}^{N_c}P(a_{i,s'},c_s)\log\frac{P(a_{i,s'},c_s)}{P(a_{i,s'})P(c_s)} \tag{10}$$

where "log" means the logarithmic operation with base 2 in this study. $P(a_{i,s'},c_s)$ is the non-negative joint probability distribution of A_i and C. If A_i has a state $a_{i,\bar{s}'}$ with $P(a_{i,\bar{s}'},c_s)=0$, then $P(a_{i,s'},c_s)\log\frac{P(a_{i,s'},c_s)}{P(a_{i,s'})P(c_s)}=0$. $P(a_{i,s'})$ and $P(c_s)$ are marginal probability distributions of A_i

and C. $I(A_i;C)$ is non-negative if and only if A_i and C are independent, then $I(A_i;C)=0$. Larger $I(A_i;C)$ means that A_i and C are more dependent on each other. Table 6 presents the mutual information between each attribute variable and the class variable.

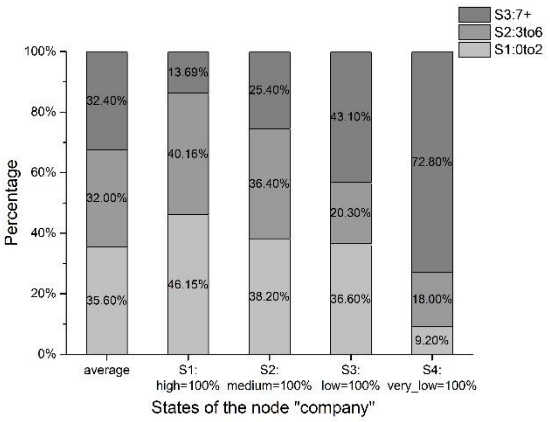
Table 6 Mutual information between attribute variables and class variable

Mutual information	Value
$I(\text{company}; \text{deficiency\_no})$	0.15904
$I(\text{last\_deficiency\_no}; \text{deficiency\_no})$	0.14938
$I(\text{age}; \text{deficiency\_no})$	0.11398
$I(\text{pre\_detention}; \text{deficiency\_no})$	0.10494
$I(GT; \text{deficiency\_no})$	0.09490
$I(\text{type}; \text{deficiency\_no})$	0.08780
$I(\text{flag}; \text{deficiency\_no})$	0.04956
$I(\text{change\_flag}; \text{deficiency\_no})$	0.04946
$I(\text{last\_inspection}; \text{deficiency\_no})$	0.04075
$I(RO; \text{deficiency\_no})$	0.00625

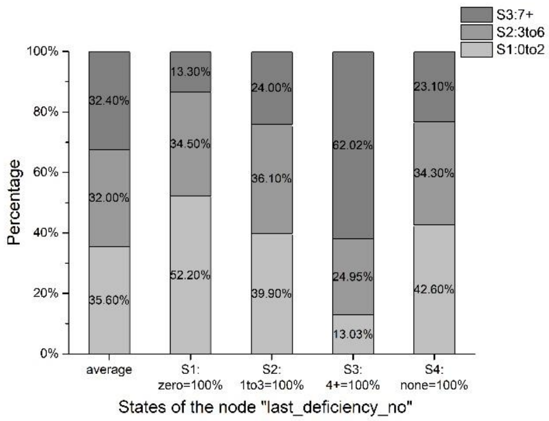
It can be seen from Table 6 that the ship company has the most significant influence on the number of deficiencies detected in the PSC inspection. This may be because, after the NIR was introduced in 2014, the performance of the companies was divided into four grades according to the inspection results of their ships in the PSC inspection. In addition, company performance is also a determinant of the ship risk profile. As a result, low performance will give the company a bad reputation, and may thus decrease its revenue. Also, the number of deficiencies in the last PSC inspection is one of the dominant predictors of the number of deficiencies in the next inspection. Ship age and previous detention times can also have a big impact on the ship deficiency number. Meanwhile, the last inspection time and the performance of ship RO have the least influence on the number of deficiencies of a ship.

6.2 Effects of attribute variables on class variable

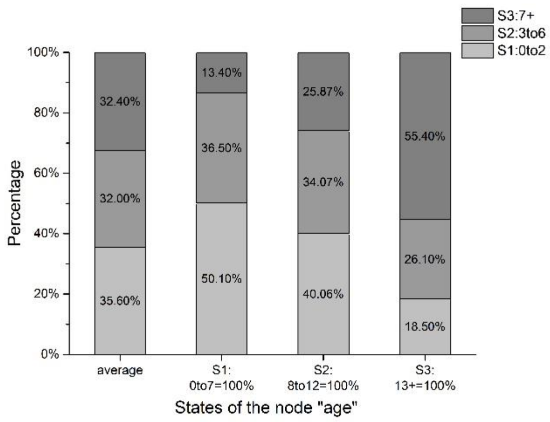
Recall the states of the variables in the TAN classifier as presented in Figure 5, in which the probability distribution of the class variable presents the proportions of the ships in the training data set belonging to the corresponding states of that variable. To identify how each state of each attribute variable will have an influence on the class variable, i.e., to identify in what states ships are more likely to have larger or smaller number of deficiencies, we assume that all the incoming ships are in one particular state of an attribute variable. To be more specific, to identify the influence of the states of "company" on the deficiency number, we can set the proportion of "S1: high", "S2: medium", "S3: low", and "S4: very_low" equal to 100% respectively, i.e., we assume that the company performance of all the incoming ships is high, medium, low and very low, respectively, and then record the proportions of the states of “deficiency_no” each time. The results are shown in the second to fifth columns in Figure 9(a). The first column in Figure 9(a) is the distribution of the variable among all the training cases, and we denote it as “average” in the horizontal ordinate. Comparing the first column with each column after the first column, if a column has the proportion of “S1: 0to2” of the class variable higher than that of the “average” column, and the proportion of “S3: 7+” of the class variable of this column is less than that of the “average” column, then it can be concluded that ships in this state of the attribute variable may have fewer deficiencies and are in better conditions than average. Conversely, if a column has the proportion of “S3: 7+” of the class variable higher than that of the “average” column and the proportion of “S1: 0to2” of the class variable of this column is lower than that of the “average” column, then the ships with this state of the attribute variable may have more deficiencies and are in worse conditions than average. The effects of different states of the states of the class variable are presented in Figure 9.

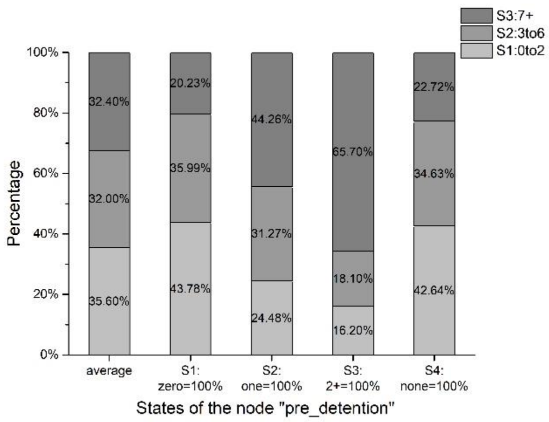


(a) Effect of different states of "company" on "deficiency_no"



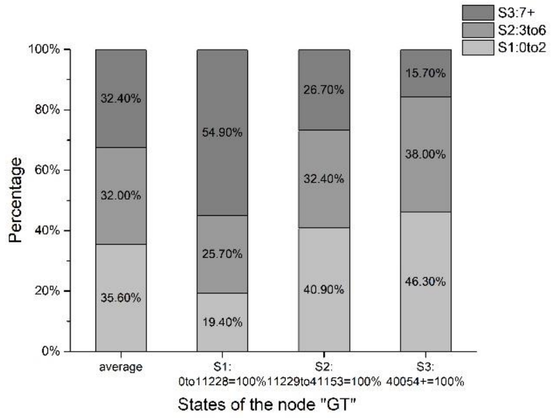
(b) Effect of different states of "last_deficiency_no" on "deficiency_no"



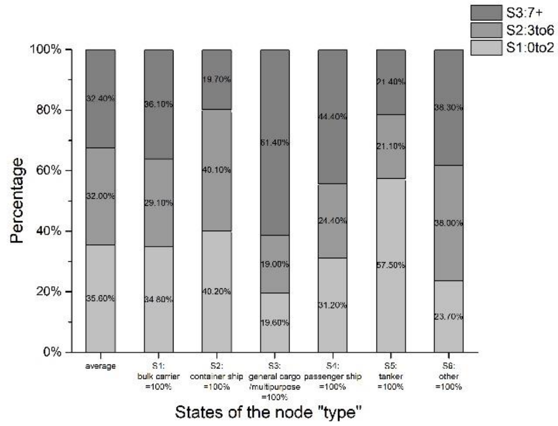


(c) Effect of different states of "age" on "deficiency_no"

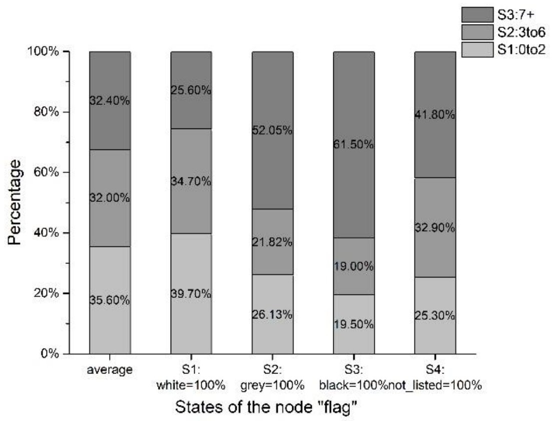
(d) Effect of different states of "pre_detention" on "deficiency_no"



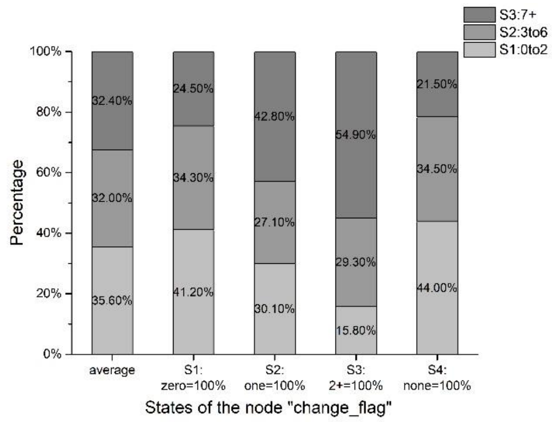
(e) Effect of different states of "GT" on "deficiency_no"



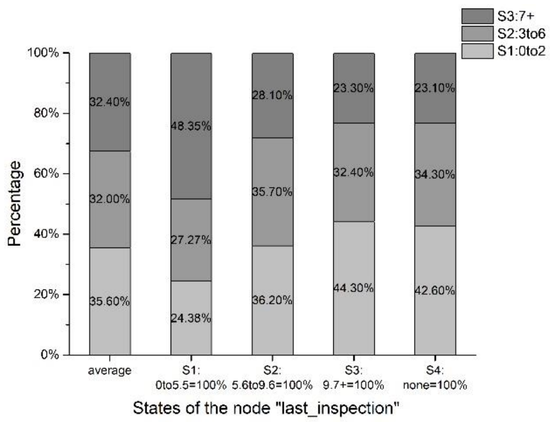
(f) Effect of different states of "type" on "deficiency_no"

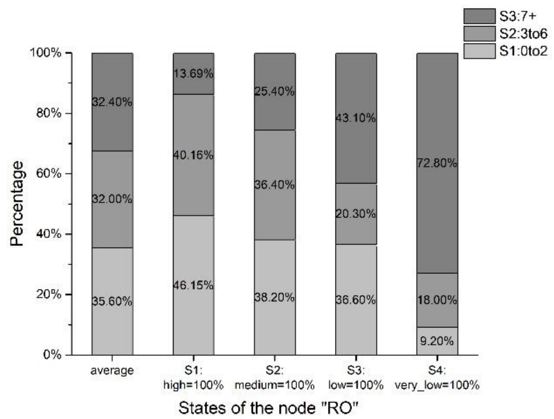


(g) Effect of different states of "flag" on "deficiency_no"



(h) Effect of different states of "change_flag" on "deficiency_no"





(i) Effect of different states of "last_inspection" on "deficiency_no"

(j) Effect of different states of "RO" on "deficiency_no"

Figure 9. Effect of different states of the attribute variables on class variable

Figure 9(a) shows that for the ship companies, the higher the company' performance is, the fewer deficiencies in PSC inspections its ships may have. Figure 9(b) indicates that, except for those ships that have no PSC inspection records, the more deficiencies there were in the last PSC inspection, the more likely it is that the ship will have more deficiencies in the next inspection. Figure 9(c) indicates that old ships may have more deficiencies than younger ships. Figure 9(d) shows that the greater the number of times a ship has been detained before, the worse performance in the latter PSC inspections it has. It is also the same for the number of times the ship changed flag, shown in Figure 9(h). As for the gross tonnage of ships, Figure 9(e) illustrates that ships with GT less than 11,228 are more likely to have more deficiencies. One of the reasons for this is that the ship's GT will be used to determine the ship's manning regulations, safety rules, registration fees, and port dues (IMO, 1969) and can thus influence the ship's conditions. Another reason is that compared to larger ships, the detention cost of smaller ships is lower, and they are more likely to have a higher number of deficiencies due to the lack of professional management of the ship companies. Figure 9(f) shows that general cargo and multipurpose ships are more likely to have a large number of deficiencies, while tankers have fewer deficiencies. Regarding the impact of ship flag performance on the number of deficiencies shown in Figure 9(g), if a ship's flag is on the white list, then it is more likely to have fewer deficiencies than ships whose flags are on the grey or black list. However, this may not be true for ships whose flags are not listed, as there are insufficient observations. However, it may be surprising that the longer the time since the last inspection, the more likely the ship is to have a smaller number of deficiencies, as indicated in Figure 9(i). That may be because ships with a lower risk profile are less frequently inspected, while ships with a worse condition are inspected more often. It is also surprising that the ships belonging to low performance ROs have fewer deficiencies in the PSC inspection than those belonging to medium performance ROs, as shown in Figure 9(j). The reason for this may be that there are only 6.21% and 5.08% ships belonging to the medium and low performance ROs respectively in the total 250 cases in the TAN classifier. The small number of cases is not typical enough to reflect the true situation.

7. Conclusion and future research

PSC inspection is viewed as an effective way to eliminate substandard shipping. One of the key issues faced by the PSC inspection authorities is how to identify high-risk incoming ships to inspect in order to find more deficiencies after inspecting a certain number of ships. To select the high-risk ships more efficiently, a data-driven Bayesian network classifier called the Tree

Augmented Naive Bayes (TAN) classifier is proposed in this paper. By using historical inspection data downloaded from the database of Tokyo MoU, which include both ship information and inspection information, the structure part and quantitative part of the TAN classifier are constructed.

The proposed model is validated by a numerical experiment based on the historic data from Hong Kong port, which shows that when the number of training cases is more than 200, the classification accuracy of the TAN model is beyond 60%. Compared with the currently used Ship Risk Profile (SRP) ship selection scheme, the TAN classifier can identify about 130.35% more deficiencies on average after inspecting the 50 ships in the testing data set. The results of the numerical experiment also show that after inspecting 10%, 20%, 30%, 40%, 50%, and 60% of the 50 total incoming ships in each testing data set, the average improvement of the TAN classifier is 348.38%, 147.23%, 108.32%, 98.29%, 70.33%, and 48.83% after inspecting 5, 10, 15, 20, 25, and 30 ships, respectively. The variable analysis shows that among all the attribute variables in the TAN classifier, the performance of the ship company and the number of deficiencies in the last PSC inspection are the dominant factors that influence the deficiency number. The results also show how the state of a specific attribute variable can have an impact on the class variable (i.e., the deficiency number). Theoretically, we propose a data equal-frequency discretization problem and present it in a mathematical and rigorous way. Then, by using dynamic programming we prove that this discretization method is bounded by $O(NV^2)$ when it is used in our model. Also, by induction, we prove that random selection of the root attribute variable of the TAN classifier will not influence the classification process of the cases in the testing data set. Practically, the proposed TAN classifier can help address the significant PSC inspection problem compared with the currently used ship selection method and the logistics regression model which is widely used in other literature on PSC inspection.

The proposed model is one of the first few data-driven models to act as a real-time predictor of the number of deficiencies of incoming ships for PSC inspection. It can predict the possible number of deficiencies of incoming foreign ships and help the PSC officers to better identify high-risk ships, as well as to make rational resource allocations.

One limitation of this research is the limited input data (i.e., the inspection records). On the one hand, some special cases may not be covered by the limited input cases. On the other hand, the CPTs may not be that accurate to reflect the real situation. In future research, more data cases, as well as more attribute variables, can be incorporated to construct the TAN model in order to further improve its prediction accuracy.

Acknowledgements

This study is supported by the National Natural Science Foundation of China (Grant Nos. 71701178 and 71831008).

References

Abuja MoU, 2017. Annual report 2016 on port state control of Abuja MoU. < http://www.abujamou.org/assets/annual_report_2016.pdf> (accessed 27.12.18).

Black Sea MoU, 2017. Annual report 2016 on port state control of Black Sea MoU. < http://www.bsmou.org/2017/06/1472/> (accessed 27.12.18).

Caribbean MoU, 2017. Annual report 2016 on port state control of Caribbean MoU. <http://www.caribbeanmou.org/sites/default/files/annual_report_2016.pdf> (accessed 27.12.18).

Cariou, P., Mejia, M. Q., Wolff, F. C., 2007. An econometric analysis of deficiencies noted in port state control inspections. Maritime Policy & Management 34(3), 243-258.

Cariou, P., Mejia, M. Q., Wolff, F. C., 2008. On the effectiveness of port state control inspections. Transportation Research Part E: Logistics and Transportation Review 44(3), 491-503.

Cariou, P., Mejia, M. Q., Wolff, F. C., 2009. Evidence on target factors used for port state control inspections. Marine Policy 33(5), 847-859.

Cariou, P., Wolff, F. C., 2011. Do port state control inspections influence flag-and class-hopping phenomena in shipping? Journal of Transport Economics and Policy 45(2), 155-177.

Cariou, P., Wolff, F. C., 2015. Identifying substandard vessels through port state control inspections: a new methodology for concentrated inspection campaigns. Marine Policy 60, 27-39.

Chauvin, C., Lardjane, S., Morel, G., Clostermann, J. P., Langard, B., 2013. Human and organisational factors in maritime accidents: analysis of collisions at sea using the HFACS. Accident Analysis & Prevention 59, 26-37.

Cheng, J., Greiner, R., 1999. Comparing Bayesian network classifiers. In: The Fifteenth Conference on Uncertainty in Artificial Intelligence Proceedings, 101-108.

Chow, C., Liu, C., 1968. Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information Theory 14(3), 462-467.

Cover, T. M., Thomas, J. A., 2012. Elements of Information. John Wiley & Sons, New York.

Dong, L. Y., Liu, G. Y., Yuan, S. M., Li, Y. L., Li, Z., 2007. Classifier learning algorithm based on genetic algorithms. In: Proceedings of the Second International Conference on Innovative Computing, Information and Control, 126-129.

Domingos, P., Pazzani, M., 1997. On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning 29(2-3), 103-130.

Dougherty, J., Kohavi, R., Sahami, M., 1995. Supervised and unsupervised discretization of continuous features. Machine Learning Proceedings 1995, 194-202.

Emecen Kara, E. G., 2016. Risk assessment in the Istanbul Strait using Black Sea MOU port state control inspections. Sustainability 8(4), 390-416.

Fan, L., Luo, M., Yin, J., 2014. Flag choice and port state control inspections - empirical evidence using a simultaneous model. Transport Policy 35, 350-357.

Flores, M. J., Gámez, J. A., Martínez, A. M., Puerta, J. M., 2011. Handling numeric attributes when comparing Bayesian network classifiers: does the discretization method matter? Applied Intelligence 34(3), 372-385.

Fraser, A. M., Swinney, H. L., 1986. Independent coordinates for strange attractors from mutual information. Physical Review A 33(2), 1134.

Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. Machine Learning 29 (2-3), 131-163.

Gan, X., Li, K. X., Zheng, H., 2010. Inspection policy of a Port State Control authority. In Proceedings of the International Forum on Shipping, Ports and Airports (IFSPA) 2010, 330-336.

Gao, Z., Lu, G., Liu, M., Cui, M., 2008. A novel risk assessment system for port state control inspection. In: IEEE International Conference on Intelligence and Security Informatics (ISI) Proceedings, 242-244.

Goerlandt, F., Reniers, G., 2017. An approach for reconciling different perspectives and stakeholder views on risk ranking. Journal of Cleaner Production 149, 1219-1232.

Hazelton, M. L., 2010. Bayesian inference for network-based models with a linear inverse structure. Transportation Research Part B: Methodological 44(5), 674-685.

Hänninen, M., 2014. Bayesian networks for maritime traffic accident prevention: benefits and challenges. Accident Analysis & Prevention 73, 305-312.

Hänninen, M., Kujala, P., 2014. Bayesian network modeling of port state control inspection findings and ship accident involvement. Expert Systems with Applications 41(4), 1632-1646.

Heij, C., Bijwaard, G. E., Knapp, S., 2011. Ship inspection strategies: effects on maritime safety and environmental protection. Transportation Research Part D: Transport and Environment 16(1), 42-48.

Heij, C., Knapp, S., 2018. Predictive power of inspection outcomes for future shipping accidents–an empirical appraisal with special attention for human factor aspects. Maritime Policy & Management 45 (5), 1-18.

Hruschka Jr, E. R., Ebecken, N. F., 2007. Towards efficient variables ordering for Bayesian networks classifier. Data & Knowledge Engineering 63(2), 258-269.

Huang, T. H., Huang, Y. C., Huang, C. Y., Fu, S. Y., 2016. Q-Learning approach in ship safety inspection data. Bridging the East and West, 229-235.

IMO, 2018. Port State Control. <http://www.imo.org/en/OurWork/MSAS/Pages/PortStateControl.aspx> (accessed 19.10.18).

IMO, 1969. International Convention on Tonnage Measurement of Ships. <http://www.imo.org/en/about/conventions/listofconventions/pages/international-convention-on-tonnage-measurement-of-ships.aspx > (accessed 13.2.2019)

Indian Ocean MoU, 2017. Annual report 2016 on port state control of Indian Ocean MoU. < http://www.iomou.org/armain.htm> (accessed 27.12.18).

Kasoulides, G. C., 1993. Port State Control and Jurisdiction: Evolution of the Port State Regime. Kluwer Academic Publishers, Berlin.

Knapp, S., 2007. The econometrics of maritime safety: "recommendations to enhance safety at sea" (No. 96). ERIM Ph.D. series research in management.

Knapp, S., Franses, P. H., 2007. Econometric analysis on the effect of port state control inspections on the probability of casualty: can targeting of substandard ships for inspections be improved? Marine Policy 31(4), 550-563.

Knapp, S., Bijwaard, G., Heij, C., 2011. Estimated incident cost savings in shipping due to inspections. Accident Analysis & Prevention 43(4), 1532-1539.

Knapp, S., Van de Velden, M., 2009. Visualization of differences in treatment of safety inspections across port state control regimes: a case for increased harmonization efforts. Transport Reviews 29(4), 499-514.

Li, F., Montewka, J., Goerlandt, F., Kujala, P., 2017. A probabilistic model of ship performance in ice based on full-scale data. In: IEEE 4th International Conference on Transportation Information and Safety (ICTIS) Proceedings, 242-244.

Li, K. X., Wonham, J., 1999. Who is safe and who is at risk: a study of 20-year-record on accident total loss in different flags. Maritime Policy & Management 26(2), 137-144.

Li, K. X., Yin, J., Fan, L., 2014. Ship safety index. Transportation research part A: Policy and Practice (66), 75-87.

Li, K. X., Yin, J., Bang, H. S., Yang, Z., Wang, J., 2014. Bayesian network with quantitative input for maritime risk analysis. Transportmetrica A: Transport Science 10(2), 89-118.

Li, K. X., Zheng, H., 2008. Enforcement of law by the port state control (PSC). Maritime Policy & Management 35(1), 61-71.

Lu, L., Goerlandt, F., Banda, O. A. V., Kujala, P., Höglund, A., Arneborg, L., 2019. A Bayesian network risk model for assessing oil spill recovery effectiveness in the ice-covered Northern Baltic Sea. Marine Pollution Bulletin 139, 440-458.

Mediterranean MoU, 2017. Annual report 2016 on port state control of Mediterranean MoU. < htt p://www.medmou.org/ > (accessed 27.12.18).

Menard, S., 2002. Applied Logistic Regression Analysis (106). Sage, New York.

Paris MoU, 2013. Criteria for responsibility assessment of recognized organizations (RO). <https:/ /www.parismou.org/sites/default/files/RO%20responsibility%20rev11_0.pdf> (accessed 20.1 0.18).

Paris MoU, 2014. Annex 8, inspection and selection scheme. < https://www.parismou.org/inspecti ons-risk/library-faq/selection-scheme> (accessed 10.12.18).

Paris MoU, 2017. Annual report 2016 on port state control of Paris MoU. < https://www.parismou. org/2017-paris-mou-annual-report-%E2%80%9Csafeguarding-responsible-and-sustainable-sh ipping%E2%80%9D> (accessed 27.12.18).

Paris MoU, 2019. Organization of Paris MoU. < https://www.parismou.org/about-us/organisation> (accessed 13.04.19).

Pernkopf, F., 2005. Bayesian network classifiers versus selective k-NN classifier. Pattern Recognition 38(1), 1-10.

Riyadh MoU, 2017. Annual report 2016 on port state control of Riyadh MoU. < https://www.riya dhmou.org/public_html/assets/uploads/images/d7521229df2bbccf3c0880b413f9a588.pdf > (a ccessed 27.12.18).

Sun, X. T., Chung, S. H., Chan, F. T., Wang, Z., 2018. The impact of liner shipping unreliability on the production–distribution scheduling of a decentralized manufacturing system. Transportation Research Part E: Logistics and Transportation Review 114, 242-269.

Tan, Z., Meng, Q., Wang, F., Kuang, H. B., 2018. Strategic integration of the inland port and shipping service for the ocean carrier. Transportation Research Part E: Logistics and Transportation Review 110, 90-109.

Teye, C., Bell, M. G., Bliemer, M. C., 2017. Locating urban and regional container terminals in a competitive environment: an entropy maximising approach. Transportation Research Part B: Methodological 117, 971-985.

Titz, M. A., 1989. Port state control versus marine environmental pollution. Maritime Policy & Management 16(3), 189-211.

Tokyo MoU, 2014. Information sheet of the new inspection regime (NIR). < http://www.tokyo-mou.org/doc/NIR-information%20sheet-r.pdf> (accessed 20.11.18).

Tokyo MoU, 2017a. Annual report 2016 on port state control in the Asia-Pacific region. < http://www.tokyo-mou.org/doc/ANN16.pdf> (accessed 27.12.18).

Tokyo MoU, 2017b. Black – grey – white lists. <http://www.tokyo-mou.org/doc/Flag%20performance%20list%202017.pdf> (accessed 25.10.18).

Tokyo MoU, 2017c. Memorandum of Understanding on port state control in the Asia-Pacific region. < http://www.tokyo-mou.org/doc/Memorandum%20rev17.pdf> (accessed 29.3.19).

Tokyo MoU, 2018. Annual report on port state control in the Asia-Pacific Region 2017. < http://www.tokyo-mou.org/doc/ANN17.pdf> (accessed 28.10.18).

Trucco, P., Cagno, E., Ruggeri, F., Grande, O., 2008. A Bayesian Belief network modelling of organisational factors in risk analysis: a case study in maritime transportation. Reliability Engineering & System Safety 93(6), 845-856.

Tsou, M. C., 2018. Big data analysis of port state control ship detention database. Journal of Marine Engineering & Technology 17, 1-9.

UNCTAD, 2017. Review of Maritime Transportation 2017 <https://unctad.org/en/PublicationsLibrary/rmt2017_en.pdf> (accessed 5.11.18).

Viña del Mar Agreement, 2017. Annual report of Viña del Mar 2016. < http://alvm.prefecturanaval.gob.ar/cs/Satellite?c=Page&cid=1434394760856&pagename=CIALA%2FPage%2FtemplateImageToDoc > (accessed 27.12.18).

Wang, Y., and Vassileva, J., 2003. Bayesian network-based trust model. In: IEEE International Conference on Web Intelligence Proceedings, 372-378.

Warfield, S. K., Kaus, M., Jolesz, F. A., Kikinis, R., 2000. Adaptive, template moderated, spatially varying statistical classification. Medical Image Analysis 4(1), 43-55.

Wróbel, K., Krata, P., Montewka, J., Hinz, T., 2016. Towards the development of a risk model for unmanned vessels design and operations. TransNav: International Journal on Marine Navigation and Safety of Sea Transportation 10 (2), 267-274.

Wyner, A. D., 1978. A definition of conditional mutual information for arbitrary ensembles. Information and Control 38(1), 51-59.

Xu, R., Lu, Q., Li, W. J., Li, K. X., Zheng, H. S., 2007a. A risk assessment system for improving port state control inspection. In: International Conference on Machine Learning and Cybernetics Proceedings, 818-823.

Xu, R., Lu, Q., Li, K. X., and Li, W., 2007b. Web mining for improving risk assessment in port state control inspection. In: International Conference on Natural Language Processing and Knowledge Engineering Proceedings, 427-434.

Yang, Z., Yang, Z., Yin, J., 2018a. Realising advanced risk-based port state control inspection using data-driven Bayesian networks. Transportation research part A: Policy and Practice 110, 38-56.

Yang, Z., Yang, Z., Yin, J., Qu, Z., 2018b. A risk-based game model for rational inspections in port state control. Transportation Research Part E: Logistics and Transportation Review 118, 477-495.

Yu, J., Goos, P., Vandebroek, M., 2012. A comparison of different Bayesian design criteria for setting up stated preference studies. Transportation Research Part B: Methodological 46(7), 789-807.

Zheng, J., Qi, J., Sun, Z., Li, F., 2018. Community structure based global hub location problem in liner shipping. Transportation Research Part E: Logistics and Transportation Review 118, 1-19.

Zheng, W., Li, B., Song, D., 2017. Effects of risk-aversion on competing shipping lines' pricing strategies with uncertain demands. Transportation Research Part B: Methodological 104, 337-356.

Zhang, D., Yan, X., Yang, Z., Wall, A., Wang, J., 2013. Incorporation of formal safety assessment and Bayesian network in navigational risk estimation of the Yangtze River. Reliability Engineering & System Safety 118, 93-105.

Zhang, G., Thai, V. V., 2016. Expert elicitation and Bayesian Network modeling for shipping accidents: a literature review. Safety Science 87, 53-62.

Zhang, J., Teixeira, Â. P., Guedes Soares, C., Yan, X., Liu, K., 2016. Maritime transportation risk assessment of Tianjin Port with Bayesian belief networks. Risk Analysis 36(6), 1171-1187.

Zhang, X., Lam, J. S. L., 2018. Shipping mode choice in cold chain from a value-based management perspective. Transportation Research Part E: Logistics and Transportation Review 110, 147-167.

Zhou, C., Sun, J., 2010. Automatically optimized and self-evolutional ship targeting system for port state control. In: IEEE International Conference on Systems Man and Cybernetics (SMC) Proceedings, 791-795.

Appendix A. Proof of Theorem 1

The problem can be solved by dynamic programming. The dynamic programming approach has N stages. The state ω of a stage $s=2,...,N$ means that the categories $\omega+1,...,V$ belong to stages $s,...,N$ and that the categories $1,...,\omega$ belong to stages $1,...,s-1$ and stage $s=1$ has only one state $\omega=0$. The set of possible states of a stage s is denoted by $\Omega_s=\{s-1,...,V-(N-s+1)\}$. At state ω of stage s, the immediate decision is the number of categories that are incorporated in state s. That is, if the immediate decision is d, then categories $\omega+1,...,\omega+d$ belong to stage s and the resulting state of stage s is $\omega+d$. The set of possible immediate decisions is $D(s,\omega)=\{1,...,V-\omega-(N-s)\}$. Let $u(s,\omega)$ be the minimum sum of squared errors over stages $s,...,N$ when the system is at state ω of stage s. The recursive relation is:

$$u(s,\omega)=\min_{d\in D(s,\omega)}\left(\frac{\sum_{v=\omega+1}^{\omega+d}\theta_v}{K}-\frac{1}{N}\right)^2+u(s+1,\omega+d), s=1,...,N-1, \omega\in\Omega_s \tag{A1}$$

and the boundary conditions are

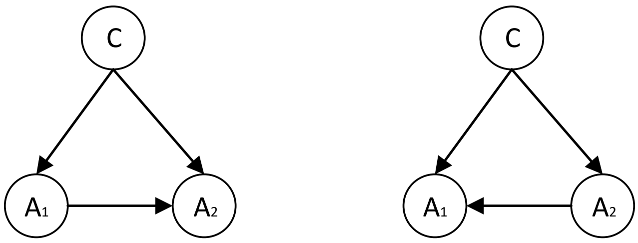
$$u(N,\omega)=\left(\frac{\sum_{v=\omega+1}^{V}\theta_v}{K}-\frac{1}{N}\right)^2, \omega\in\Omega_N\,. \tag{A2}$$

The optimal solution can be obtained by solving $u(1,0)$. Since the dynamic programming approach has N stages, each stage has at most V states, at each state of each stage, there are at most V decisions, and the time required to evaluate a decision is bounded by $O(1)$, the problem can be solved in time bounded by $O(NV^2)$.□

Appendix B. Proof of Theorem 2

To prove the theorem, we will prove that, for a TAN classifier with I attribute variables, $I \geq 2$, different choices of root attribute variable node all have the same value $\tilde{P}^I(a_{1,s^{(1)}},...,a_{I,s^{(I)}},c_{\bar{s}})$ in Eq. (7) for a particular combination of $\bar{s}=1,...,N_C$, $j=1,...,I$, $s^{(j)}=1,...,N_j$. We will prove this conclusion by induction. That is, we first prove that this conclusion is true for a TAN classifier with two attribute variables; we then prove that if this conclusion is true for a TAN classifier with I attribute variables, $I \geq 2$, it will also be true for a TAN classifier with $I+1$ attribute variables.

First, consider a TAN classifier with two attribute variables $A=(A_1,A_2)$ and one class variable C. The two structures of the TAN classifier are shown in Figure B1.



(a) A_1 is the parent variable of A_2 (b) A_2 is the parent variable of A_1

Figure B1. TAN classifier with two attribute variables

For any case k with states of the attribute variables $ATT_k=(a_{1,s'},a_{2,s''})$, $s'=1,...,N_1$, $s''=1,...,N_2$, we can use either the TAN classifier in Figure B1(a) (referred to hereafter as the left classifier, or "L" for short) or the TAN classifier in Figure B1(b) (the right classifier, or "R" for short) to calculate the values in Eq. (7). If we use the left TAN classifier, we have

$$\begin{aligned}
&\tilde{P}^L(a_{1,s'},a_{2,s''},c_{\bar{s}}) \\
&= P^L(c_{\bar{s}})\times P^L(a_{1,s'}\,|\,c_{\bar{s}})\times P^L(a_{2,s''}\,|\,a_{1,s'},c_{\bar{s}}) \\
&= P^L(c_{\bar{s}})\times P^L(a_{1,s'}\,|\,c_{\bar{s}})\times \frac{P^L(a_{2,s''},a_{1,s'}\,|\,c_{\bar{s}})}{P^L(a_{1,s'}\,|\,c_{\bar{s}})}, \bar{s}=1,...,N_C
\end{aligned} \tag{B1}$$

If we use the right TAN classifier, we have

$$\begin{aligned}
&\tilde{P}^R(a_{1,s'},a_{2,s''},c_{\bar{s}}) \\
&= P^R(c_{\bar{s}})\times P^R(a_{2,s''}\,|\,c_{\bar{s}})\times P^R(a_{1,s'}\,|\,a_{2,s''},c_{\bar{s}}) \\
&= P^R(c_{\bar{s}})\times P^R(a_{2,s''}\,|\,c_{\bar{s}})\times \frac{P^R(a_{2,s''},a_{1,s'}\,|\,c_{\bar{s}})}{P^R(a_{2,s''}\,|\,c_{\bar{s}})}, \bar{s}=1,...,N_C.
\end{aligned} \tag{B2}$$

Note that in Eqs. (B1) and (B2), both $P^L(c_{\bar{s}})$ and $P^R(c_{\bar{s}})$ refer to the proportion of cases in the data set whose class state is $c_{\bar{s}}$, and both $P^L(a_{2,s''}, a_{1,s'} \mid c_{\bar{s}})$ and $P^R(a_{2,s''}, a_{1,s'} \mid c_{\bar{s}})$ refer to the proportion of cases with $a_{1,s'}$ as the state of attribute variable A_1 and $a_{2,s''}$ as the state of attribute variable A_2 among cases in the data set with class state $c_{\bar{s}}$. Therefore,

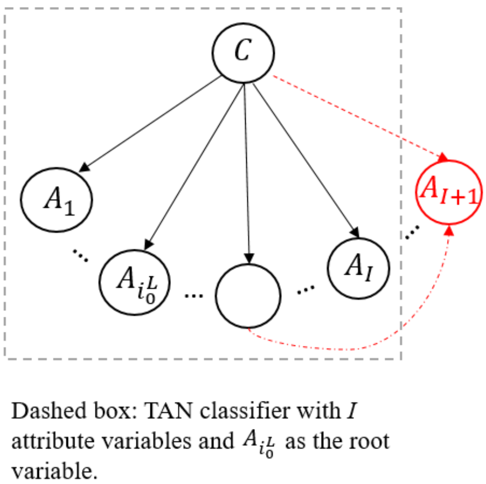
$$\tilde{P}^L(c_{\bar{s}} \mid a_{1,s'}, a_{2,s''}) = \tilde{P}^R(c_{\bar{s}} \mid a_{1,s'}, a_{2,s''}), \quad \bar{s} = 1, \ldots, N_C. \tag{B3}$$

For a TAN classifier with I attribute variables, we have (the superscript "I" means the TAN classifier has I attribute variables)

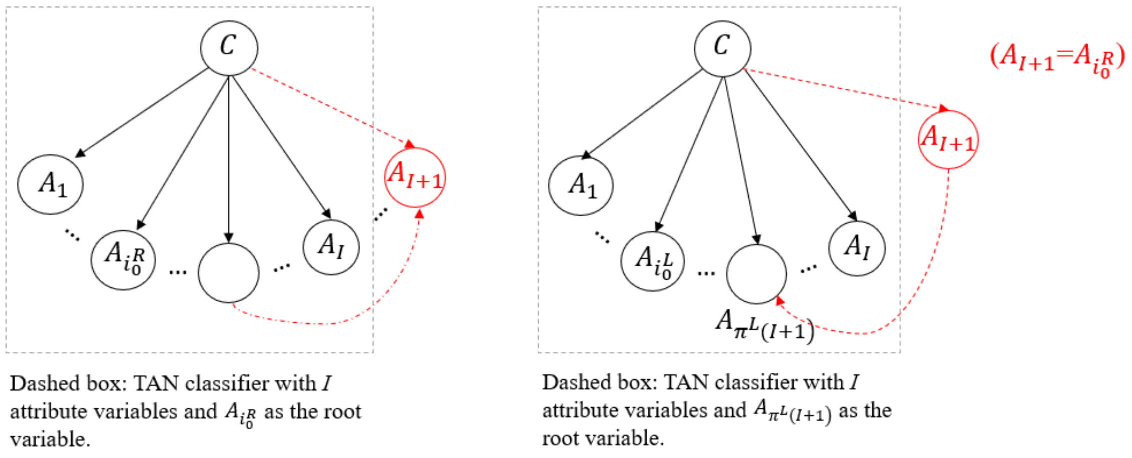
$$\begin{aligned}&\tilde{P}^I(a_{1,s^{(1)}}, \ldots, a_{I,s^{(I)}}, c_{\bar{s}}) \\ &= P^I(c_{\bar{s}}) \times P^I(a_{i_0,s^{(i_0)}} \mid c_{\bar{s}}) \times \prod_{i=1, i \neq i_0}^{I} \frac{P^I(a_{i,s^{(i)}}, a_{\pi(i),s^{(\pi(i))}} \mid c_{\bar{s}})}{P^I(a_{\pi(i),s^{(\pi(i))}} \mid c_{\bar{s}})}, \bar{s} = 1, \ldots, N_C, j = 1, \ldots, I, s^{(j)} = 1, \ldots, N_j\end{aligned} \tag{B4}$$

Suppose that for a TAN classifier with I attribute variables, $I \geq 2$, different choices of root attribute variable node all have the same value of $\tilde{P}^I(a_{1,s^{(1)}}, \ldots, a_{I,s^{(I)}}, c_{\bar{s}})$ in Eq. (B4). Next, we prove that for a TAN classifier with $I+1$ attribute variables and with a given maximum spanning tree, different choices of root attribute variable node all have the same value of $\tilde{P}^{I+1}(a_{1,s^{(1)}}, \ldots, a_{I,s^{(I)}}, a_{I+1,s^{(I+1)}}, c_{\bar{s}})$.

Consider two TAN classifiers with the same maximum spanning tree of $I+1$ attribute variables, and one classifier (left classifier, or "L") has root attribute variable $A_{i_0^L}$ and the other (right classifier, or "R") has root attribute variable $A_{i_0^R}$, $i_0^R \neq i_0^L$, as shown in Figure B2. $A_{\pi^L(i)}$ is the unique parent attribute variable of attribute variable A_i, $i = 1, \ldots, I, i \neq i_0^L$, in the left classifier and $A_{\pi^R(i)}$ is the unique parent attribute variable of attribute variable A_i in the right classifier.



(a) The left classifier with *I+1* as the root variable



(b) The right classifier with $I+1$ as the root variable (case i)

(c) The right classifier with $I+1$ as the root variable (case ii)

Figure B2. The structures of left classifier and right classifier

In a maximum spanning tree with at least two nodes, there exist at least two nodes, each of which is connected to exactly one other node in the tree. Therefore, in the left classifier, we can find a node that is not the root attribute variable and that is connected to exactly one other node in the tree. Without loss of generality, we assume that this node is the attribute variable A_{I+1} (otherwise we just swap its sequence with the sequence of A_{I+1} in vector A).

Then, in the left classifier,

$$\begin{aligned}
&\tilde{P}^{I+1,L}(a_{1,s^{(1)}},\dots,a_{I,s^{(I)}},a_{I+1,s^{(I+1)}},c_{\bar{s}}) \\
&= P^{I+1,L}(c_{\bar{s}})\times P^{I+1}(a_{i_0^L,s^{(i_0^L)}}\,|\,c_{\bar{s}})\times\prod_{i=1,i\neq i_0^L}^{I}P^{I+1,L}(a_{i,s^{(i)}}\,|\,a_{\pi^L(i),s^{(\pi^L(i))}},c_{\bar{s}})\times P^{I+1,L}(a_{I+1,s^{(I+1)}}\,|\,a_{\pi^L(I+1),s^{(\pi^L(I+1))}},c_{\bar{s}}) \\
&= \underbrace{P^{I+1,L}(c_{\bar{s}})\times P^{I+1,L}(a_{i_0^L,s^{(i_0^L)}}\,|\,c_{\bar{s}})\times\prod_{i=1,i\neq i_0}^{I}\frac{P^{I+1,L}(a_{i,s^{(i)}},a_{\pi^L(i),s^{(\pi^L(i))}}\,|\,c_{\bar{s}})}{P^{I+1,L}(a_{\pi^L(i),s^{(\pi^L(i))}}\,|\,c_{\bar{s}})}}_{AA}\times\frac{P^{I+1,L}(a_{I+1,s^{(I+1)}},a_{\pi^L(I+1),s^{(\pi^L(I+1))}}\,|\,c_{\bar{s}})}{P^{I+1,L}(a_{\pi^L(I+1),s^{(\pi^L(I+1))}}\,|\,c_{\bar{s}})}, \\
&\bar{s}=1,\dots,N_C
\end{aligned} \tag{B5}$$

It should be noted that AA in Eq. (B5) is actually the value $\tilde{P}(a_{1,s^{(1)}},\dots,a_{I,s^{(I)}},c_{\bar{s}})$ for the TAN classifier with I attribute variables and root attribute variable $A_{i_0^L}$ in Figure B2(a).

There are two cases of the right classifier. In Case (i), as shown in Figure B2(b), the root node $A_{i_0^R}$ is not A_{I+1}. Then, similar to Eq. (B5), we have

$$\begin{aligned}
&\tilde{P}^{I+1,R}(c_{\bar{s}}\,|\,a_{1,s^{(1)}},\dots,a_{I,s^{(I)}},a_{I,s^{(I+1)}}) \\
&= \underbrace{P^{I+1,R}(c_{\bar{s}})\times P^{I+1,R}(a_{i_0^R,s^{(i_0^R)}}\,|\,c_{\bar{s}})\times\prod_{i=1,i\neq i_0}^{I}\frac{P^{I+1,R}(a_{i,s^{(i)}},a_{\pi^R(i),s^{(\pi^R(i))}}\,|\,c_{\bar{s}})}{P^{I+1,R}(a_{\pi^R(i),s^{(\pi^R(i))}}\,|\,c_{\bar{s}})}}_{BB}\times\frac{P^{I+1,R}(a_{I+1,s^{(I+1)}},a_{\pi^R(I+1),s^{(\pi^R(I+1))}}\,|\,c_{\bar{s}})}{P^{I+1,R}(a_{\pi^R(I+1),s^{(\pi^R(I+1))}}\,|\,c_{\bar{s}})}, \\
&\bar{s}=1,\dots,N_C
\end{aligned} \tag{B6}$$

Note that BB in Eq. (B6) is actually the value $\tilde{P}(a_{1,s^{(1)}},...,a_{I,s^{(I)}},c_{\bar{s}})$ for the TAN classifier with I attribute variables and root attribute variable $A_{i_0^R}$ in Figure B2(b). Based on the precondition of the induction, we have $AA=BB$. Since A_{I+1} is connected to exactly one other node in the tree, we have $\pi^L(I+1)=\pi^R(I+1)$ and therefore

$$\frac{P^{I+1,L}(a_{I+1,s^{(I+1)}},a_{\pi^L(I+1),s^{(\pi^L(I+1))}}\,|\,c_{\bar{s}})}{P^{I+1,L}(a_{\pi^L(I+1),s^{(\pi^L(I+1))}}\,|\,c_{\bar{s}})}=\frac{P^{I+1,R}(a_{I+1,s^{(I+1)}},a_{\pi^R(I+1),s^{(\pi^R(I+1))}}\,|\,c_{\bar{s}})}{P^{I+1,R}(a_{\pi^R(I+1),s^{(\pi^R(I+1))}}\,|\,c_{\bar{s}})}. \quad \text{(B7)}$$

Hence, for Case (i),

$$\tilde{P}^{I+1,L}(c_{\bar{s}}\,|\,a_{1,s^{(1)}},...,a_{I,s^{(I)}},a_{I,s^{(I+1)}})=\tilde{P}^{I+1,R}(c_{\bar{s}}\,|\,a_{1,s^{(1)}},...,a_{I,s^{(I)}},a_{I,s^{(I+1)}}). \quad \text{(B8)}$$

In Case (ii), as shown in Figure B2(c), the root node $A_{i_0^R}$ is A_{I+1}. Then since A_{I+1} is connected to exactly one other node in the tree, we have $\pi^R(\pi^L(I+1))=I+1$, that is, the parent attribute variable of $A_{\pi^L(I+1)}$ as the parent of A_{I+1}. Moreover, $\pi^R(i)\neq I+1, i=1,...,I, i\neq\pi^L(I+1)$, that is, no attribute variable other than $A_{\pi^L(I+1)}$ has parent A_{I+1}. Therefore, in the right classifier,

$$\begin{aligned}
&\tilde{P}^{I+1,R}(a_{1,s^{(1)}},...,a_{I,s^{(I)}},a_{I+1,s^{(I+1)}},c_{\bar{s}})\\
&=P^{I+1,R}(c_{\bar{s}})\times P^{I+1,R}(a_{I+1,s^{(I+1)}}\,|\,c_{\bar{s}})\times\prod_{i=1,I\neq\pi^L(I+1)}^{I}P^{I+1,R}(a_{i,s^{(i)}}\,|\,a_{\pi^R(i),s^{(\pi^R(i))}},c_{\bar{s}})\times P^{I+1,R}(a_{\pi^L(I+1),s^{(\pi^L(I+1))}}\,|\,a_{I+1,s^{(I+1)}},c_{\bar{s}})\\
&=P^{I+1,R}(c_{\bar{s}})\times P^{I+1,R}(a_{I+1,s^{(I+1)}}\,|\,c_{\bar{s}})\times\prod_{i=1,I\neq\pi^L(I+1)}^{I}\frac{P^{I+1,R}(a_{i,s^{(i)}},a_{\pi^R(i),s^{(\pi^R(i))}}\,|\,c_{\bar{s}})}{P^{I+1,R}(a_{\pi^R(i),s^{(\pi^R(i))}}\,|\,c_{\bar{s}})}\times\frac{P^{I+1,R}(a_{\pi^L(I+1),s^{(\pi^L(I+1))}},a_{I+1,s^{(I+1)}}\,|\,c_{\bar{s}})}{P^{I+1,R}(a_{I+1,s^{(I+1)}}\,|\,c_{\bar{s}})}\\
&=P^{I+1,R}(c_{\bar{s}})\times\prod_{i=1,I\neq\pi^L(I+1)}^{I}\frac{P^{I+1,R}(a_{i,s^{(i)}},a_{\pi^R(i),s^{(\pi^R(i))}}\,|\,c_{\bar{s}})}{P^{I+1,R}(a_{\pi^R(i),s^{(\pi^R(i))}}\,|\,c_{\bar{s}})}\times P^{I+1,R}(a_{\pi^L(I+1),s^{(\pi^L(I+1))}},a_{I+1,s^{(I+1)}}\,|\,c_{\bar{s}})\\
&=\underbrace{P^{I+1,R}(c_{\bar{s}})\times P^{I+1,R}(a_{\pi^L(I+1),s^{(\pi^L(I+1))}}\,|\,c_{\bar{s}})\times\prod_{i=1,I\neq\pi^L(I+1)}^{I}\frac{P^{I+1,R}(a_{i,s^{(i)}},a_{\pi^R(i),s^{(\pi^R(i))}}\,|\,c_{\bar{s}})}{P^{I+1,R}(a_{\pi^R(i),s^{(\pi^R(i))}}\,|\,c_{\bar{s}})}}_{CC}\times\frac{P^{I+1,R}(a_{\pi^L(I+1),s^{(\pi^L(I+1))}},a_{I+1,s^{(I+1)}}\,|\,c_{\bar{s}})}{P^{I+1,R}(a_{\pi^L(I+1),s^{(\pi^L(I+1))}}\,|\,c_{\bar{s}})}\\
&\bar{s}=1,...,N_C
\end{aligned}$$

(B9)

Then, CC in Eq. (B9) is actually the value $\tilde{P}(a_{1,s^{(1)}},...,a_{I,s^{(I)}},c_{\bar{s}})$ for the TAN classifier with I attribute variables and $A_{\pi^L(I+1)}$ as the root attribute variable, as shown in Figure B2(c). Based on the precondition of the induction, we have $AA=CC$. Therefore, for Case (ii),

$$\tilde{P}^{I+1,L}(c_{\bar{s}}\,|\,a_{1,s^{(1)}},...,a_{I,s^{(I)}},a_{I,s^{(I+1)}})=\tilde{P}^{I+1,R}(c_{\bar{s}}\,|\,a_{1,s^{(1)}},...,a_{I,s^{(I)}},a_{I,s^{(I+1)}}). \quad \text{(B10)}$$

This concludes the proof of the theorem. □

Appendix C. Method used to calculate the CPTs

The CPT of a variable in the BN contains the probabilities of each state of the variable under the condition of the states of its parent variables. For the class variable (i.e., the node "deficiency_no"), the CPT is reduced to the prior probability distribution of its states as it has no parent variable, as is shown in Table C1.

Table C1 CPT of deficiency_no

deficiency_no	prior probability
S1:0to2	35.6%
S2:3to6	32.0%
S3:7+	32.4%

For an attribute variable, the CPT is dependent on the states of its parent variables, which include the class variable and/or another attribute variable. For the root attribute variable "age", whose parent only contains the class variable, the conditions in its CPT only contain three states of the variable "deficiency_no", and the probabilities of different states of "age" under the condition of a specific state of "deficiency_no" are the probabilities of the cases belonging to that state of "deficiency_no" and the state of "age" in the training data set. The sum of each column of the CPT is equal to 100%. The CPT of the root attribute variable "age" is shown in Table C2.

Table C2 CPT of age

age	deficiency_no = S1:0to3	deficiency_no = S2:3to6	deficiency_no = S3:7+
S1:0to7	44.56%	36.15%	13.10%
S2:8to12	36.96%	34.94%	26.19%
S3:13+	18.48%	28.91%	60.71%

For the non-root attribute variables, whose parent variables contain the class variable and another attribute variable, the conditions in CPT are the combination of one state of the class variable and one state of the parent attribute variable. An example of the CPT of node "RO" is shown in Table C3.

Table C3 CPT of RO

RO(%)	flag											
	S1: white	S2: grey	S3: black	S4: not _listed	S1: white	S2: grey	S3: white	S4: not _listed	S1: white	S2: grey	S3: black	S4: not_ listed
	deficiency_no											
	S1: 0to2	S1: 0to2	S1: 0to2	S1: 0to2	S2: 3to6	S2: 3to6	S2: 3to6	S2: 3to6	S3: 7+	S3: 7+	S3: 7+	S3: 7+
S1:high	93.5	40.0	25.0	25.0	93.9	25.0	25.0	50.0	93.3	66.7	58.9	12.5
S2: medium	1.1	20.0	25.0	25.0	1.2	25.0	25.0	16.7	1.7	16.7	11.7	37.5
S3:low	1.1	20.0	25.0	25.0	1.2	25.0	25.0	16.7	1.7	8.3	11.7	12.5
S4:not_ listed	4.3	20.0	25.0	25.0	3.7	25.0	25.0	16.7	3.3	8.3	17.7	37.5

Appendix D. Procedure 2: Selection of n ships by the SRP selection scheme

Procedure 2. *Selection of n ships by the SRP selection scheme.*

Step 1: Divide the ships in Ψ' into four categories in sequence: ships without any PSC inspections before, ships whose inspection time windows are closed, ships within the inspection time window, and ships out of (not entering) the time window. Ships in the first category are considered to have equal priority. The priority of ships in the first category is higher than ships in the second, followed by ships in the third and fourth categories. Different ships in the second category have different priorities, so do ships in the third and fourth categories. The priorities of ships in the second, third, and fourth categories are determined in Step 2.

Step 2: Calculate the risk index RI of each ship in Ψ'. Denote the last inspection time for ship i, $i = 1,...,50$, as L_i. The risk index RI is used to indicate the relative risk ranking of the ships in their corresponding categories. The method to calculate the ship risk index RI is in Table C1.

Step 3: Sort the ships in Ψ' to generate the sequence of the inspection list. The sequence of ships is: ships in the first category are randomly sequenced, followed by ships in the second category in descending order of RI, followed by ships in the third category in descending order of RI, and followed by ships in the fourth category in descending order of RI. The first n ships in the inspection list are selected

Table D1 Calculation of ship risk index

Ship risk profile	Time window (months)	State of time window		
		out of time window	within time window	time window closed
LRS	9 to 18	$RI = \frac{L_i}{9}$	$RI = \frac{L_i - 9}{18-9}$	$RI = \frac{L_i}{18}$
SRS	5 to 8	$RI = \frac{L_i}{5}$	$RI = \frac{L_i - 5}{8-5}$	$RI = \frac{L_i}{8}$
HRS	2 to 4	$RI = \frac{L_i}{2}$	$RI = \frac{L_i - 2}{4-2}$	$RI = \frac{L_i}{4}$

Appendix E. Procedure 3: Selection of n ships by the TAN classifier

Procedure 3: *Selection of n ships by the TAN classifier.*

Step 1: Train the TAN classifier using data set Ψ. The class variable "deficiency_no" has three states: S1:0to2, S2:3to6 and S3:7+. Calculate the average number of deficiencies of each state of deficiency_no in the 250 cases in Ψ. The results are: ships with 0 to 2 deficiencies on average have 1.00 deficiency, ships with 3 to 6 deficiencies on average have 3.85 deficiencies, and ships with 7+ deficiencies on average have 10.07 deficiencies.

Step 2: Input the states of each ship in Ψ' into the TAN classifier and the probability distribution of deficiency number is shown in the states of deficiency_no. Denote the probability for a ship to have 0 to 2, 3 to 6, or 7+ deficiencies by D_{0to2}, D_{3to6} and D_{7+} respectively.

Step 3: Use the average number of deficiencies of each state of deficiency_no in the 250 cases in Ψ to denote the expected number of deficiencies of that state, and calculate the expected number of deficiencies for every ship in Ψ' by $E(deficiency\_no) = 1.00 \times D_{0to2} + 3.85 \times D_{3to6} + 10.07 \times D_{7+}$.

Step 4: Sort the 50 ships in Ψ' in descending $E(deficiency\_no)$ to generate the sequence of inspection list. The first n ships in the inspection list are selected.