Ocean Engineering 334 (2025) 121614

Contents lists available at ScienceDirect

Ocean Engineering

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

Research paper

Machine learning approaches for identifying substandard ships in port state control inspections with imbalanced data

Likun Wang [a], Jinhui Wang [b], Yizhang Hua [c], Wenming Shi [d,*], Zaili Yang [e], Mei Sha [a]

[a] College of Transport and Communication, Shanghai Maritime University, Shanghai, 201306, China
[b] College of Ocean Science and Engineering, Shanghai Maritime University, Shanghai, 201306, China
[c] Innovation and Development Department, Shanghai Lingang Sci-Tech City Co, Ltd, Shanghai, 201306, China
[d] Center for Maritime and Logistics Management, Australian Maritime College, University of Tasmania, Launceston, TAS, 7250, Australia
[e] Liverpool Logistics, Offshore and Marine Research Institute, Liverpool John Moores University, Liverpool, UK

ARTICLE INFO

Keywords:
Port state control
Substandard ship identification
Imbalanced data
Machine learning
SVM-DBN model

ABSTRACT

To effectively identify substandard ships in port state control (PSC) inspections, this study proposes a novel machine learning (ML) approach that integrates support vector machines (SVMs) and deep belief networks (DBNs). A comprehensive dataset sourced from the Paris Memorandum of Understanding (MoU) inspection database is used, covering the years 2018–2020. The proposed ML approach effectively addresses the imbalanced data problem in PSC inspections, where the probability of substandard ships is typically no more than 5 %. The results of prediction accuracy indicate that the ensembled SVM-DBN model outperforms the standalone SVM model on both the original and oversampling datasets in all scenarios, except for a slightly lower precision rate on the oversampling dataset. In contrast, the SVM-DBN model underperforms relative to the standalone SVM model on the under-sampling dataset. These results have practical implications for both shipowners and port authorities. Shipowners can use the results to proactively assess and improve ship conditions, thereby mitigating potential detention risks, and port authorities can use them to improve risk-based inspection strategies, thereby optimizing inspection resources and increasing the effectiveness of substandard ship identification.

1. Introduction

Substandard ships are widely recognized as a significant threat to maritime safety, contributing to various types of maritime accidents, such as collisions, fires, and explosions, which result in casualties and property losses (Wang et al., 2022a, 2022b). These ships also aggravate the challenges of global maritime transport and trade, severely affecting economic prosperity worldwide (Wang et al., 2021; Wu et al., 2022). The International Maritime Organization (IMO) defines substandard ships as those with significant deficiencies in equipment, machinery, hull, and/or operational safety that deviate from international standards. In practice, port state control (PSC) plays a crucial role in identifying substandard ships, ensuring compliance with flag state regulations, enhancing maritime safety, and protecting the marine environment (Liu et al., 2022).

However, PSC inspections must strike a balance between ensuring that as many ships as possible are inspected and preventing unnecessary delays caused by excessive inspections. This poses a significant challenge for PSC authorities, who must effectively identify high-risk substandard ships for inspection (Wang et al., 2019). To address this, various decision support systems have been implemented through several Memoranda of Understandings (MoUs) (Yan et al., 2021). For example, the Paris and Tokyo MoUs introduced a New Inspection Regime (NIR), while the United States Coast Guard (USCG) developed the Maritime Transportation Security Act (MTSA) compliance targeting matrix, which considers factors such as ship age, company performance, and previous detentions. In summary, existing decision support systems assign specific values to different risk factors based on expert judgment, which are then aggregated to assign a risk score for each ship. The higher the score, the more likely the ship is to be selected for inspection. Unfortunately, these systems have faced increasing criticism for their subjectivity, as they rely heavily on the empirical judgment of experts. As a result, recent studies have shifted toward using machine learning (ML) techniques, such as support vector machine (SVM), random forest

* Corresponding author.
E-mail addresses: 77912463@qq.com (L. Wang), sunrise@mail.ustc.edu.cn (J. Wang), lzshmm@hotmail.com (Y. Hua), Wenming.Shi@utas.edu.au (W. Shi), z.yang@ljmu.ac.uk (Z. Yang), meisha@shmtu.edu.cn (M. Sha).

https://doi.org/10.1016/j.oceaneng.2025.121614
Received 4 January 2025; Received in revised form 17 May 2025; Accepted 18 May 2025
Available online 24 May 2025

(RF), and Bayesian network (BN), which provide a more objective approach to identifying substandard ships through data mining.

When mining PSC datasets, the problem of data imbalance between ships with and without detentions often arises, complicating the identification of substandard ships. More specifically, ships without detentions significantly outnumber those with detentions. As shown in Fig. 1, the Paris MoU annual report highlights that the detention rate has consistently remained below 5 % from 2008 to 2021. In such imbalanced PSC datasets, conventional ML techniques tend to favor the classes with more non-detentions, resulting in low prediction accuracy for the classes with fewer detained ships (Yan et al., 2021). Consequently, PSC authorities are increasingly focusing on improving the accuracy of ship risk prediction by paying more attention to the "minority" class of substandard ships, thereby enhancing the effectiveness of substandard ship identification in PSC inspections.

To address the above concerns, this study proposes a novel ML approach that integrates SVMs and deep belief networks (DBNs) to overcome the challenges of subjectivity and data imbalance in existing PSC inspection datasets. In doing so, this study advances previous studies in the following ways: First, it shifts the focus to detained ships and their influencing factors, thereby enhancing our understanding of PSC detention data and laying the groundwork for more effective substandard ship identification. Second, resampling techniques are used to create more balanced (i.e., oversampling and under-sampling) datasets. These resampled datasets, along with the original dataset, are used to validate the predictive performance of the SVM-DBN model. This approach enables a more objective and scientifically sound assessment of ship detention risk in PSC inspections. Third, the constructed SVM-DBN model contributes to the application of ML approaches to substandard ship identification by leveraging SVM's ability to prevent overfitting and DBN's ability to uncover hidden relationships from imbalanced PSC data.

The remainder of this study is structured as follows: Section 2 thoroughly reviews previous studies on risk-based PSC inspection factors, ship selection methods for PSC inspections, and imbalanced data handling. Section 3 presents the process of handling imbalanced data using resampling techniques and constructing the SVM-DBN model. Section 4 reports the results of validating and comparing the predictive performance of the base SVM and the SVM-DBN models. Section 5 summarizes the main findings and implications and suggests possible directions for future research.

2. Literature review

Given the critical role of PSC inspections in identifying substandard ships, considerable efforts have been dedicated to analyzing the factors influencing these inspections and the methods used for ship selection (Yan and Wang, 2019; Yan et al., 2021; Yang et al., 2023; Zhang et al., 2024). This section provides a comprehensive review of these factors and approaches. It also examines existing solutions to address data imbalance, offering valuable insights for the present study.

2.1. Risk-based PSC inspection factors

As previously noted, decision support systems in PSC inspections predominantly rely on expert judgment to provide a risk score for each ship. According to the PMoU's guidance for the ship risk calculator (http://parismou.org/PMoU-Procedures/ship-risk-calculator), both generic parameters (e.g., ship type, ship age, flag, recognized organization, and ISM company performance) and historic parameters from the last 36 months (e.g., the number of inspection, the number of deficiencies, and the number of detentions) are used to generate a ship's risk profile. In line with these guidelines, several risk-based PSC inspection factors have been identified in previous studies, mainly including ship type, age, gross tonnage, flag performance, recognized organization performance, type of inspection, inspection country, and recorded deficiencies. These factors are employed in this study and summarized in Table 1.

Using a dataset of 4080 PSC inspections reported by the Swedish

Table 1
Factors identified in existing literature.

Variables	References
Ship type	(Cariou et al., 2007; Xu et al., 2007a; Heij et al., 2011; Bateman 2012; Wang et al., 2019)
Ship age	(Degré, 2007; Cariou et al., 2007; Heij and Knapp, 2018; Heij and Knapp, 2018; Hänninen and Kujala, 2014; Yan et al., 2021)
Ship gross tonnage	(Cariou et al., 2007; Degré, 2007; R.-F. Xu et al., 2007; Cariou and Wolff, 2015; Heij and Knapp, 2018)
Ship flag performance	(Xu et al., 2007; Cariou et al., 2007; Degré, 2007; Batista et al., 2004; Cariou and Wolff, 2015; Heij and Knapp, 2018; Yan et al., 2021)
Ship recognized organization performance	(Xu et al., 2007; Degré, 2007; Cariou and Wolff, 2015; Cariou et al., 2007; Heij and Knapp, 2018; Yang et al., 2018a; Yan et al., 2021)
Type of inspection	(Hänninen and Kujala, 2014; Wang et al., 2019; Yan et al., 2021; Wang et al., 2021)
Inspection country	(Cariou and Wolff, 2015; Yang et al., 2018a)
Deficiencies	(Xu et al., 2007; Akyuz and Celik, 2014)

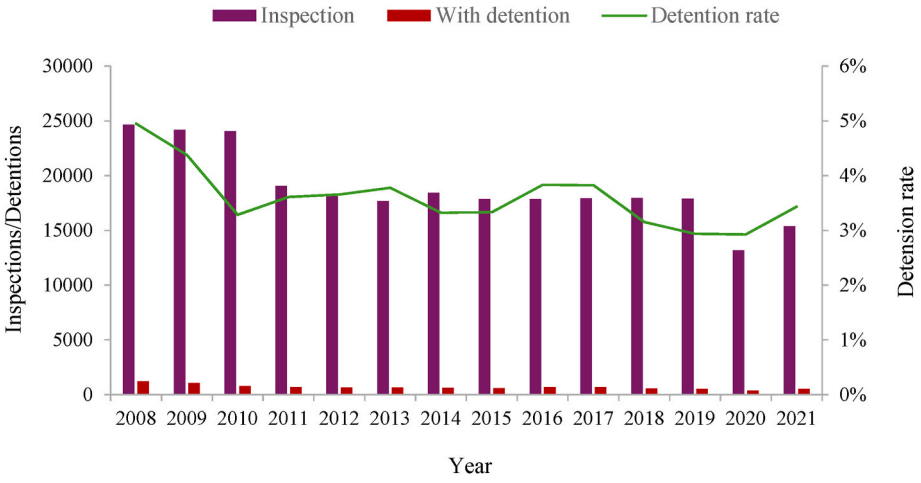

Fig. 1. Number of ship detentions in Port State from 2008 to 2021. *Notes*: The horizontal axis represents the year, the left vertical axis indicates the number of inspections and detentions, and the right vertical axis shows the detention rate.

Maritime Administration from 1996 to 2001, Cariou et al. (2007) concluded that factors such as ship age, type, and the flag of registry are crucial determinants in ship selection for PSC inspections. They also highlighted the variability in the relationship between ship age and the frequency of detected deficiencies across different ship types. Based on data from ships passing the Istanbul Strait, Emecen Kara (2016) argued that the effectiveness of flags within the Black Sea MoU is a key indicator, directly correlating with their associated risk levels. Yang et al. (2018) analyzed over 80,000 inspection records from seven major European countries between 2005 and 2008, using a data-driven BN to identify the primary factors contributing to bulk carrier detentions, including inspection type, number of deficiencies, ship age, and recognized organization. Wang et al. (2021) categorized commonly recognized risk-based PSC inspection factors into three groups: 1) ship-related factors, such as gross tonnage, type, and age; 2) inspection-related factors, such as inspection type and country; and 3) deficiency-related factors, such as the number of deficiencies.

Since it is widely recognized that the factors identified above are correlated with predefined safety standards in PSC inspections, this study thoroughly investigates the dependencies among all identified factors and introduces detailed deficiency information into the substandard ship identification model.

2.2. Ship selection methods for PSC inspections

Apart from identifying risk-based PSC inspection factors, port authorities are also concerned with effectively selecting high-risk substandard ships for inspection. This approach aims to ensure maritime safety while optimizing inspection resources (Degré, 2007). Guided by different concepts and theories of PSC inspections, two main methods have been proposed to improve the efficiency of substandard ship selection. First, different statistical models and mathematical analyses are widely utilized. For example, Cariou and Wolff (2015) used count and probit data models to estimate the probability of encountering numerous deficiencies of specific types, aiming to enhance the selection process and supplement current detention-focused procedures. Şanlıer (2020) conducted a basic statistical analysis of the Black Sea MoU inspection database from 2012 to 2017, providing insights into how specific risk factors influence ship detentions and assisting PSC authorities in identifying high-risk ships.

Second, with the rapid development of computer technology, researchers are increasingly focusing on developing intelligent ship selection models using ML techniques. For example, Xu et al. (2007a) introduced an SVM-based risk assessment system to classify incoming ships according to target factors. Xu et al. (2007b) combined network mining techniques to enhance classification accuracy, achieving an increase of about 20 %. Gao et al. (2021) integrated K-nearest neighbors (KNN) and SVM techniques to filter out noisy training examples, thereby enhancing the accuracy of ship selection. This refinement led to a prediction accuracy of more than 20 %. Xu et al. (2007a) presented an SVM risk assessment system that categorizes ships into high-risk or low-risk groups based on target factors. Wang et al. (2019) developed a data-driven Bayesian network (BN) classifier to identify high-risk foreign vessels that warrant the attention of PSC inspection authorities. Other relevant studies include those by Akyuz and Celik (2014), Hänninen and Kujala (2014), and Heij and Knapp (2018). Recently, Wang et al. (2019) introduced a data-driven BN classifier called the Tree Augmented Naive Bayes (TAN) classifier to identify high-risk foreign vessels. They found that the TAN-based BN approach can more efficiently and accurately identify substandard ships. Yang et al. (2023) developed a novel model which is trained via the incorporation of an Improved Tree Augmented Naïve (ITAN) learning approach and a maximum a posteriori probability (MAP) of Expectation Maximization (EM) approach, which could be used as a prediction tool to determine rational durations for detained vessels.

Although the existing ML models, to a certain extent, avoid overfitting, their predictive performance is highly influenced by both the size of the dataset and its ratio to the influential factors trained within the model. Classification algorithms (e.g., KNN and BN) often exhibit a bias towards the categories with large numbers (e.g., non-detained ships), resulting in predictions that favor the more prevalent categories. Then, the characteristics of minority classes (e.g., detained ships) are frequently regarded as noise, misleading predictions when identifying detained ships.

2.3. Approaches for addressing data imbalance

According to a recent study by Yan et al. (2021), imbalanced class distribution in datasets poses a significant challenge for many classifier learning algorithms, which typically assume or rely on a roughly balanced class distribution, often with a roughly 1:1 class ratio. To mitigate the impact of data imbalance, several approaches have been explored in various domains, such as medical diagnosis (Ahmed and Rashid, 2024), financial fraud detection (Shahana et al., 2023), and image classification (Islam et al., 2023). These efforts generally fall into two main categories:

First, data resampling techniques based on evaluation functions aim to achieve a balanced class distribution, typically through over-sampling or under-sampling methods (Batista et al., 2004). For example, García et al. (2009) proposed an enhanced decision tree algorithm for training set selection, demonstrating superior accuracy compared to over-sampling methods and outperforming standard under-sampling techniques. Gao et al. (2021) integrated various “rebalancing” heuristics into SVM modeling, with numerical results demonstrating its ability to match or exceed the performance of previously recognized top-performing algorithms across various datasets. Second, ensemble classification approaches combine two or more methods, often leading to significantly higher predictive accuracy than any single method alone. Galar et al. (2012) proposed a taxonomy of ensemble-based approaches and demonstrated their superiority over the use of preprocessing techniques alone to deal with class imbalance. They also highlighted the beneficial synergy between sampling techniques and bagging.

As observed in the PSC inspection dataset in Fig. 1, data imbalance is a common issue, with ships without detentions significantly outnumbering ships with detentions. Despite this, few studies have specifically addressed the issue of data imbalance in PSC datasets. A notable exception is the work of Yan et al. (2021), who employed an ensemble model proposed by Yu et al. (2018) to develop a classification model called Balanced Random Forest (BRF) to predict ship detentions, making a pioneering contribution to address data imbalance in PSC inspections. In summary, the widespread use of ML techniques in various domains and concerns about data imbalance provide valuable insights for this study. In the context of imbalanced PSC data, this study uses two ML methods to improve the identification of detained ships. One approach is to generate more balanced (i.e., oversampling and under-sampling) datasets through resampling techniques. The other focuses on constructing an ensembled SVM-DBN model that exploits the ability of SVM to avoid overfitting and the ability of DBN to discover hidden relationships in imbalanced PSC data. The resampled datasets, along with the original dataset, are used to assess the predictive performance of the SVM-DBN model. Results indicate that the ensembled SVM-DBN model outperforms the standalone SVM model on both the original and oversampling datasets in all scenarios, except for a slightly lower precision rate on the oversampling dataset. Additionally, the proposed model-building approach does not require subjective variables or manual adjustments to the network structure, thus proposing a new training process using gradient descent and an improved prediction method for ship detentions.

3. Methodology

As shown in Fig. 2, the construction of the SVM-DBN model for substandard ship identification in PSC inspection involves several steps: risk factor determination, data acquisition, data processing, model training and model evaluation, all of which are discussed in this section.

3.1. Risk factor determination

A comprehensive review of the risk-based PSC inspection factors in Table 1 (Section 2.1) identifies eight key risk factors, based on existing literature and shipping industry expertise, that are highly correlated with ship detentions. These factors include ship-related factors (i.e., ship type, age, gross tonnage, flag performance), inspection-related factors (i.e., type of inspection and inspection country), and defect-related factors (i.e., previous deficiencies and ship recognized organization performance). This study incorporates these eight factors to predict the probability of ship detention during PSC inspections.

3.2. Data acquisition

This study collects PSC inspection data from the Paris MoU (https://parismou.org/) for the period from January 1, 2018, to December 31, 2020, covering all 27 member states. After excluding inspection records with missing values (not exceeding 5 %), a comprehensive dataset of 46,658 records was created, consisting of 1363 detentions and 45,295 non-detentions. For classification feature coding, one-hot encoding is applied to the decision attribute (i.e., y), while label encoding is used for all risk factors x_i (i = 1, 2, …, 8). Table 2 presents the notations, descriptions, values and occurrence frequencies for the decision attribute and the eight risk factors.

3.3. Data processing

The original dataset was randomly partitioned into an 80 %–20 % split, resulting in a training set and a test set. Table 3 provides information about each set. As observed, the training set contains 37,326 records with 1091 detentions, while the test set consists of 9332 records with 272 detentions. The proposed substandard ship identification model is trained on the training set and its performance is evaluated using the test set.

A close examination of Table 3 reveals that PSC detentions account for approximately 2.9 % of all inspection records, while non-detentions account for about 97.1 %, exhibiting a significant class imbalance. As suggested by Yu et al. (2018), to improve classification accuracy, it is effective to adjust class weights by oversampling the minority class and under-sampling the majority class to rebalance the data. Specifically, oversampling randomly duplicates instances of the minority class, while under-sampling removes instances from the majority class, effectively increasing the weight of "detained vessels". These resampling techniques can address data imbalance and mitigate the lower accuracy for minority classes in the original PSC inspection dataset. While resampling techniques are valuable for addressing data imbalance and improving statistical inference, they can introduce certain biases. For example, resampling can unintentionally select certain groups more or less frequently than others, leading to selection bias and an unrepresentative sample. Splitting the data into training and test sets can reduce the model's ability to capture complex patterns, especially if the test set is too small, potentially resulting in over-smoothing or underfitting. In addition, after resampling, the test set may no longer reflect the

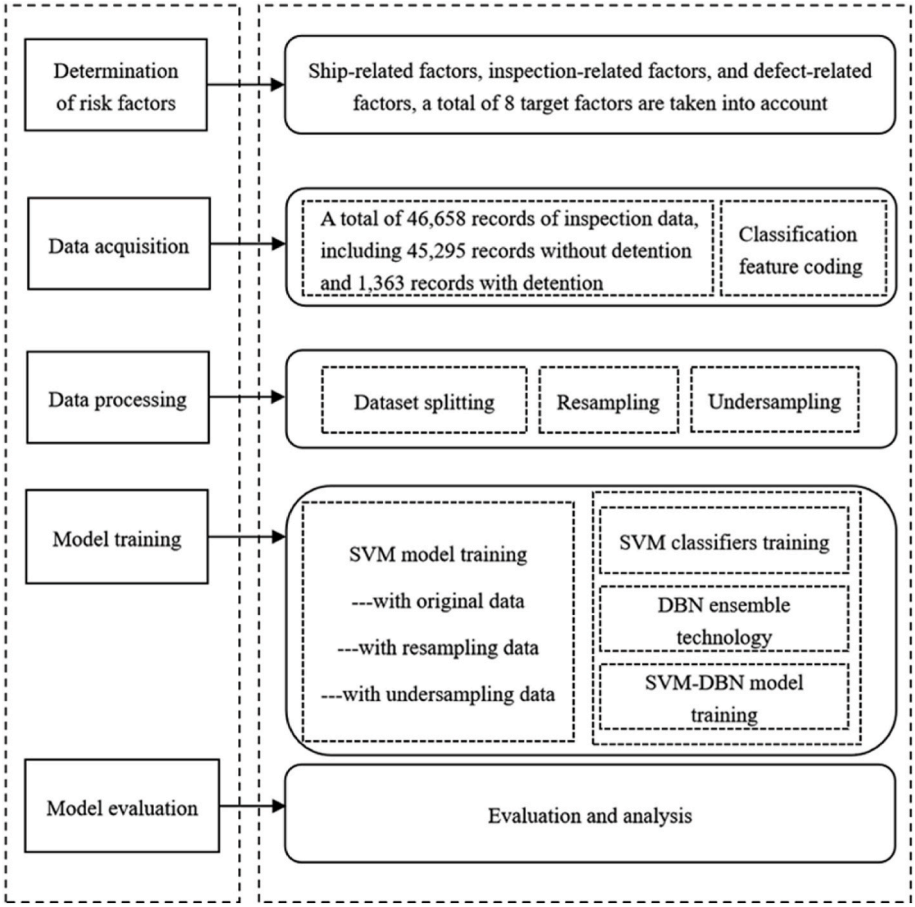

Fig. 2. Process of ship detention prediction model.

Table 2
Description of the variables used.

Variable	Notation	Description	Values	Occurrence frequencies (%)
Detention	y	Detention, without detention	0, 1	97.08 %, 2.92 %
Ship type	x_1	Bulk carrier, Chemical tanker, General cargo/ multipurpose, Oil tanker, Container, Other	1, 2, 3, 4, 5, 6	21.79 %, 9.23 %, 27.88 %, 9.21 % 10.79 %, 21.09 %
Ship age	x_2	0 to 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25, 26 to 30, more than 30	1, 2, 3, 4, 5, 6, 7	15.92 %, 24.69 %, 23.51 %, 11.52 %, 8.62 %, 5.75 %, 15.74 %
Ship gross tonnage	x_3	0 to 4999, 5000 to 9999, 10000 to 14999, 15000 to 19999, 20000 to 24999, 25000 to 29999, 30000 to 35000, more than 35000	1, 2, 3, 4, 5, 6, 7, 8	31.69 %, 12.20 %, 6.00 %, 5.15 %, 7.52 %, 6.35 %, 5.88 %, 25.20 %
Ship flag performance	x_4	White, Grey, Black	1, 2, 3	91.76 %, 2.57 %, 5.67 %
Ship RO performance	x_5	High, Medium, Low, Very low	1, 2, 3, 4	94.32 %, 3.65 %, 1.67 %, 0.36 %
Type of inspection	x_6	PSC-Initial inspection, PSC-Expanded inspection, PSC-More detailed inspection	1, 2, 3	39.23 %, 18.78 %, 41.99 %
Inspection country	x_7	Malta, Netherlands, Cyprus, Norway, Denmark, Russian, Federation, Portugal, Italy, United Kingdom, Greece, Germany, Finland, Sweden, France, Belgium, Spain, Ireland, Lithuania, Croatia, Latvia, Estonia, Poland, Bulgaria, Canada, Slovenia, Romania, Iceland	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27	1.10 %, 6.70 %, 0.52 %, 3.01 %, 2.98 %, 6.60 %, 2.88 %, 8.27 %, 7.69 %, 5.43 %, 5.66 %, 1.58 %, 2.61 %, 5.77 %, 5.42 %, 8.94 %, 1.58 %, 1.81 %, 1.76 %, 1.68 %, 1.22 %, 3.01 %, 2.09 %, 7.17 %, 0.87 %, 3.24 %, 0.39 %
Deficiencies	x_8	0, 1 to 5, 6 to 10, 11 to 15, 16 to 20, 21 to 25, 26 to 30, more than 30	1, 2, 3, 4, 5, 6, 7, 8	48.87 %, 39.05 %, 8.71 %, 2.18 %, 0.72 %, 0.24 %, 0.12 %, 0.11 %

Table 3
Information about training and test sets.

	Data size	Number of detentions	Number of non-detentions	Imbalance rate
Training set (80 %)	37,326	1091	36,235	2.9 %
Test set (20 %)	9332	272	9060	2.9 %

real-world distribution of the data (Sauerbrei, 1999; Molinaro et al., 2005). To mitigate these potential biases, this study incorporates careful data cleaning, transformation, and feature engineering, and employs random sampling along with repeated resampling procedures. Consequently, both oversampling and under-sampling methods are employed, with the results summarized in Table 4.

3.4. Construction of the SVM-DBN model

3.4.1. SVM classifier training

Due to its high accuracy and low susceptibility to overfitting in binary classification, the classic SVM is commonly used as the base classifier (Cortes and Vapnik, 1995). Since ship detention is a nonlinear problem, this study employs the Gaussian radial basis kernel function: $K(x, y) = exp(-gamma \cdot \|x-y\|^2)$. To further mitigate the risk of overfitting, parameter tuning and model selection are performed using exhaustive search during training to optimize classification performance. Additionally, a random selection process creates 20 sub-datasets, each containing 70 % of the training data. The 20 individual SVM classifiers are trained on these sub-datasets to reduce the impact of noise on model performance. These classifiers then classify the test ensemble input, producing 20 different prediction sets, alongside the true label, denoted as $[Y_{svm1}, \cdots, Y_{svm20}; Y_{ture}]$. Finally, to address potential issues such as poor performance due to parameter initialization or dataset inconsistencies, an ensemble learning algorithm that combines multiple SVM learners is employed. This method has been shown to outperform individual SVM learners and has been widely applied in financial fraud detection and image classification (Yu et al., 2018).

Table 4
Information about training set after resampling.

Training set	Data size	Number of detentions	Number of non-detentions
Training set (80 %)	37,326	1091	36,235
Oversampling	72,470	36,235	36,235
Under-sampling	2182	1091	1091

3.4.2. DBN ensemble training

A DBN-based ensemble strategy is employed for its deep structure and powerful learning capabilities, which enable effective feature extraction from unlabeled data. Consisting of multiple Restricted Boltzmann Machines (RBMs) and a Back Propagation (BP) Neural Network, the DBN-based ensemble approach in this study excels in mining joint distribution information to refine weak classifier predictions (Yu et al., 2018). Notably, the DBN is only responsible for result integration and does not participate in the initial classification phase.

The results from the 20 SVM classifiers are then consolidated by the DBN ensemble learner to generate the final classification results. These sets $[Y_{svm1}, \cdots, Y_{svm20}; Y_{ture}]$ serve as the training data for the DBN, allowing both pre-training and fine-tuning of the network. This process results in an ensemble learner refined by DBN learning.

3.4.3. SVM-DBN model construction

In the final classification stage, the test set data is classified by the 20 trained SVM classifiers, producing 20 sets of classification results denoted as $[Y'_{svm1} \cdots, Y'_{svm20}]$. These results are then consolidated by the DBN ensemble learner to produce the final classification results. This study adopts SVM as the base classifier and DBN as the ensemble method to integrate the predictions of multiple SVMs, thereby enhancing the classification accuracy and generalization ability of the model. The 20 SVM models are trained on different training subsets randomly sampled from data, which reduces the risk of overfitting. The schematic of the integrated model structure is shown in Fig. 3, and the data flow processing workflow is shown in Fig. 4.

3.5. Model evaluation

In binary classification, evaluating the performance of a classifier using a confusion matrix is the most straightforward method, as shown in Table 5 (Batista et al., 2004). This study utilizes three widely accepted metrics (i.e., precision, recall, and F1-score) to assess the performance of the proposed classifiers (Yan et al., 2021; Hocek et al., 2025). These metrics are explained in detail in Table 6.

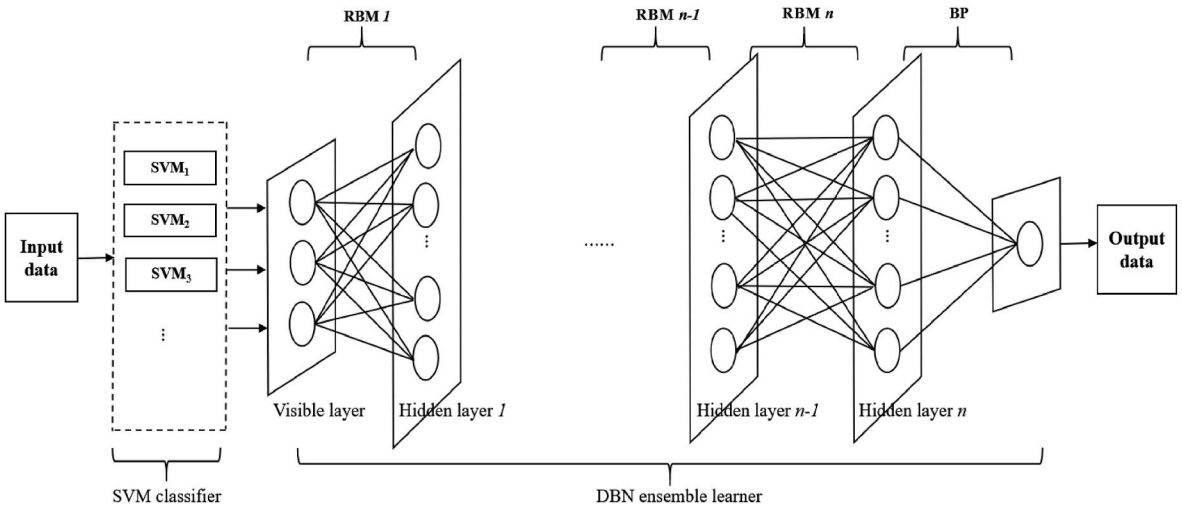

Fig. 3. Detailed structure of the SVM-DBN model.

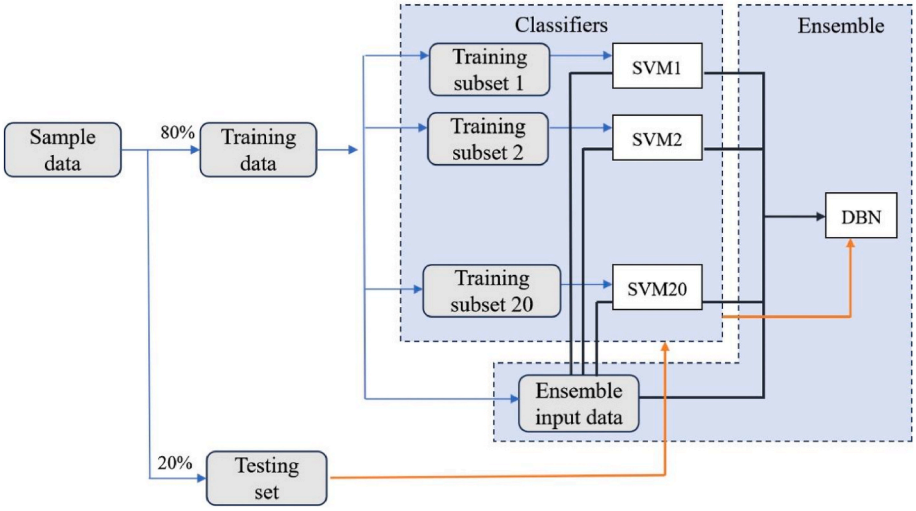

Fig. 4. Data flow processing workflow of the SVM-DBN model.

Table 5
Confusion matrix for binary classification problem.

Reality	Prediction	
	Class "1" (with detention)	Class "2" (without detention)
Class "1" (with detention)	True positives (TP)	False negative (FN)
Class "2" (without detention)	False positive (FP)	True negative (TN)

Table 6
A detailed explanation of evaluation indicators.

Metric	Definition	Explanation
Precision	$Precision = \frac{TP}{TP + FP}$	It indicates the proportion of all predicted detentions that correspond to actual detentions.
Recall	$Recall = \frac{TP}{TP + FN}$	It indicates the percentage of detained samples that are correctly predicted as detentions.
F1-score	$F1 - score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 \times Precision \times Recall}{Precision + Recall}$	The harmonic mean of precision and recall shows the accuracy and robustness of the classifier.

4. Results

4.1. Model learning and validation

Through repeated experiments, the optimal parameters for the SVM model are determined as the Gaussian radial basis kernel function with the regularization parameter $C = 0.1$ and gamma $\gamma = 0.1$. In the DBN ensemble model, after iterative trials, a learning rate of 0.1 and a momentum rate of 0.9 are chosen to achieve the optimal classification performance. A random greedy algorithm is then employed for both pre-training and fine-tuning the backpropagation algorithm. Specifically, two RBMs with 128 and 64 neurons, respectively, are trained over 10 iterations. During the fine-tuning phase, a stochastic gradient descent (SGD) optimizer with a batch size of 64, a learning rate of 0.1, and a momentum rate of 0.9 is used for 3 training iterations. Additionally, the Rectified Linear Unit (ReLU) activation function and the cross-entropy loss function are employed for model training. In particular, the selection of the ReLU activation function is mainly due to its computational

efficiency, requiring only one max operation, and its superior convergence and generalization performance compared to other commonly used activation functions such as sigmoid and tanh, thus mitigating the vanishing gradient problem. These pre-training and fine-tuning steps improve classification performance, effectively addressing the data imbalance concern.

4.2. Prediction accuracy

4.2.1. Comparisons between different datasets using the SVM model

The SVM model is performed for the original, oversampling, and under-sampling PSC inspection datasets. The prediction accuracy metrics (i.e., precision, recall and F1-score) are shown in Fig. 5.

First, the precision rate measures the accuracy of detention predictions for each dataset using the SVM model, reflecting its ability to avoid falsely labeling non-detentions as detentions. As shown in Fig. 5, the SVM models trained on the oversampling (0.973) and under-sampling (0.970) datasets demonstrate slightly higher average precision rates than the model trained on the original (0.969) dataset, demonstrating an increase of 0.09 % and 1.5 %, respectively. Second, the recall rate represents the proportion of actual detentions correctly identified by the model. A higher recall rate indicates better detection of detentions, with fewer missed cases. From Fig. 5, the SVM model trained on the under-sampling (0.899) dataset achieves the highest average recall rate, followed by the original (0.876) and oversampling (0.875) datasets. In other words, compared to the original dataset, the SVM model trained on the under-sampling dataset shows a 2.63 % increase in the recall rate, while the model trained on the oversampling dataset experiences a slight decline of 0.11 %. The lower recall rate for the oversampling dataset may be due to the fact that the oversampling reduces the generalization ability of the SVM model during training, leading to overfitting. Third, the F1-score, a comprehensive metric that combines precision and recall, provides a holistic assessment of model performance. Its value ranges from 0 to 1, with a higher value indicating superior model performance. Compared to the F1-score for the original dataset (0.913), it is 0.914 for the oversampling dataset, showing a 0.11 % increase, and 0.927 for the under-sampling dataset, showing a 1.53 % increase. That is, the SVM model trained on the under-sampling dataset exhibits the highest overall performance for both detained and non-detained samples, as evidenced by its highest F1-score. Overall, the SVM models trained on the oversampling and under-sampling datasets outperform the model trained on the original dataset in all scenarios, with the exception that the recall rate for the original dataset is slightly higher than that for the oversampling dataset.

4.2.2. Comparison between the SVM and SVM-DBN models

Using the original PSC inspection dataset as the test set, both the standalone SVM model and the ensembled SVM-DBN are evaluated using prediction accuracy metrics (i.e., precision, recall and F1-score), as presented in Fig. 6. The ensembled SVM-DBN model demonstrates superior performance over the standalone SVM model in most scenarios. Specifically, it achieves a higher average precision rate (0.972), recall rate (0.973), and F1-score (0.962) on the original dataset. On the oversampling dataset, it maintains a higher recall rate (0.943) and F1-score (0.964), although its precision rate (0.931) is slightly lower than that of the standalone SVM model. Conversely, on the under-sampling dataset, the ensembled SVM-DBN model underperforms compared to the standalone SVM model, achieving the same precision rate (0.970) but with a lower recall rate (0.829) and F1-score (0.884). Despite the higher computational cost and time requirements, the ensembled SVM-DBN model can make incremental improvements on both the original and oversampling datasets in most scenarios. These findings support the practical application of the resampling techniques, particularly the oversampling method, and their integration with other approaches (e.g., SVM and DBNs), in PSC inspections.

4.2.3. Comparison between the SVM-DBN ensemble model and alternative standalone ML models

Benefiting from the strengths of ML techniques, a range of standalone ML predictive models beyond SVM have been proposed, including complement naïve Bayes (cNB), logistic regression (Lo), adaptive boosting (AdaBoost), neural network classifier (nn_clf), k-nearest neighbors (kn), decision tree (DT), and AdaBoostSVM. This study evaluates the performance of these models across the original, oversampling, and under-sampling PSC inspection datasets, with detailed results presented in Appendix Tables A, B, and C, respectively. On the test set, these models show relatively similar prediction performance in terms of the precision rate, the recall rate, and F1-score. Notably, these standalone ML models inherently mitigate the effects of data imbalance, particularly the AdaBoost model, which demonstrates the best overall performance. This suggests its potential for integration with other models to enhance predictive accuracy. While the SVM model performs comparably to AdaBoost, this study makes a novel attempt to integrate resampling techniques with SVM and DBN models for PSC inspections. The results demonstrate that the SVM-DBN ensemble model outperforms the standalone SVM model, offering incremental improvements. This finding provides a valuable reference for future research on the integration of ML models with other techniques in PSC inspections (Yu et al., 2018).

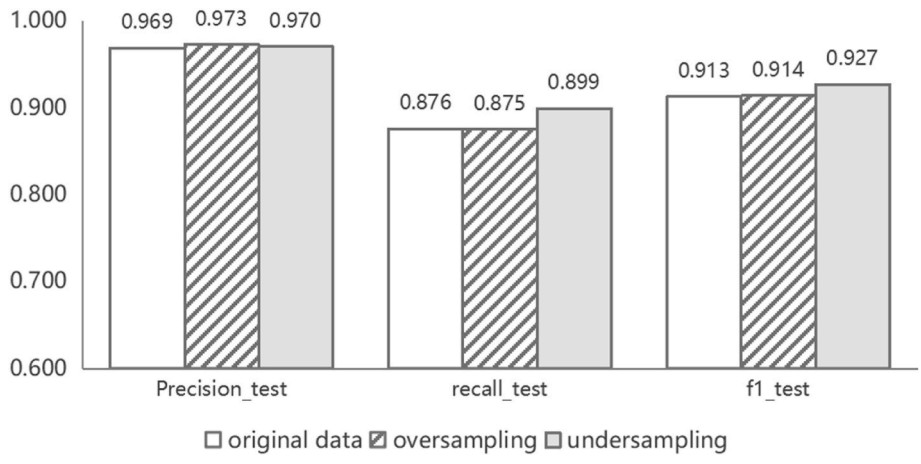

Fig. 5. Prediction accuracy of different test datasets using the SVM model. N*otes*: The horizontal axis displays the prediction accuracy metrics (i.e., precision, recall and F1-score) for the original, oversampling, and under-sampling PSC inspection datasets. The vertical axis represents the values of these prediction accuracy metrics.

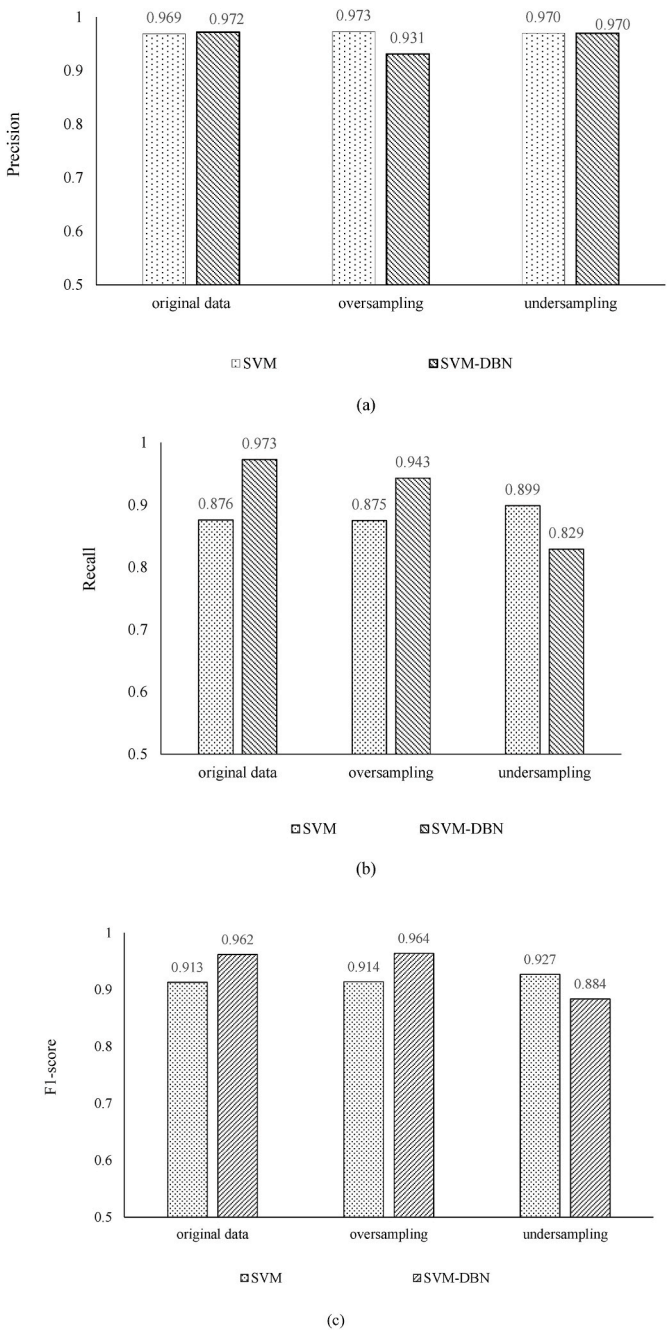

Fig. 6. Prediction accuracy of SVM and SVM-DBN on the original, oversampling, and under-sampling datasets. *Notes*: The horizontal axis displays the original, oversampling, and under-sampling PSC inspection datasets. The vertical axis represents the values of precision in Panel (a), recall in Panel (b), and F1-score in Panel (c).

5. Conclusion

Effective identification of substandard ships in PSC inspections not only optimizes port authorities' inspection resources but also ensures maritime safety. This study proposes an ensembled SVM-DBN model that integrates resampling techniques, SVM and DBN to produce a more balanced dataset, avoid overfitting, and discover hidden relationships among inspection risk factors. The findings from a comprehensive dataset sourced from the Paris MoU inspection database are as follows: First, eight key risk factors are identified to predict the probability of ship detention in PSC inspections, including ship type, age, gross tonnage, flag performance, type of inspection, inspection country, ship recognized organization performance, and previous deficiencies. Second, the ML approach overcomes the subjectivity inherent in existing decision support systems, offering a scientific and objective data-driven method for PSC ship selection. Third, the resampling techniques effectively address the data imbalance issue in PSC inspections, resulting in more balanced datasets. Finally, the SVM-DBN model outperforms the standalone SVM model on both the original and oversampling datasets in most scenarios, suggesting its practical application to PSC inspections.

The findings of this study have important implications for both shipowners and port authorities. They allow shipowners to proactively assess ship conditions and mitigate potential detention risks. For example, by using the identified factors and the constructed model, shipowners can predict the probability of detention, enabling them to prioritize ship maintenance efforts. For port authorities, as the number of ships requiring PSC inspections increases, these findings offer a rapid, effective, and objective method to identify substandard ships, optimize inspection resources, avoid delays, and ensure maritime safety.

Despite these implications, there are opportunities to further extend this study in different ways. First, additional risk factors excluded in the current PSC database, such as personnel-related factors, shipowner-specific factors, and the ship's operational history, can be explored through data mining from the Lloyd's Register database and the Global Integrated Shipping Information System database. Second, another potential extension is to integrate text mining of ship detention cases, expert domain knowledge, and the PSC database to conduct a comprehensive analysis of the underlying reasons for ship detentions during PSC inspections. This could include reasons such as insufficient shore-based support, non-compliance with maritime conventions, and inadequate ship equipment. Meanwhile, integrating other standalone ML models (e.g., AdaBoost) with complementary techniques presents a promising direction for enhancing the predictive accuracy of ship detentions. Third, to increase its practical adoption by PSC authorities, a platform like the existing "Ship risk profile calculator" used by the Paris MoU could be developed, incorporating the SVM-DBN model to facilitate more efficient and convenient inspections by PSC officers.

CRediT authorship contribution statement

Likun Wang: Writing – original draft, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jinhui Wang:** Writing – review & editing, Supervision, Funding acquisition. **Yizhang Hua:** Writing – review & editing, Supervision, Data curation. **Wenming Shi:** Writing – review & editing, Supervision, Project administration, Formal analysis. **Zaili Yang:** Writing – review & editing, Supervision, Investigation. **Mei Sha:** Writing – review & editing, Supervision, Investigation, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

This research was funded by the National Natural Science Foundation of China (Grant NOs. 71904117 by **Likun Wang**, 71972128 by **Mei Sha**), Humanities and Social Science Fund of Ministry of Education of China (NO. 24YJAZH145 by **Jinhui Wang**).

Appendix

Table A
Prediction accuracy of different models on the original PSC inspection dataset

Original	SVM	cNB	Lo	AdaBoost	nn_clf	kn	DT	AdaBoostSVM	SVM-DBN
pre_train	0.976206	0.963497	0.974406	0.942448	0.985268	0.972632	0.988557	0.868400	0.966954
rec_train	0.889648	0.920190	0.977897	0.970798	0.986149	0.975218	0.988667	0.868056	0.972432
f1_train	0.922311	0.938408	0.974097	0.956413	0.984739	0.967164	0.987446	0.868025	0 961285
pre_test	0.968779	0.961744	0.971301	0.942347	0.969244	0.959994	0.965659	0.970394	0.971594
rec_test	0.876018	0.920381	0.975782	0.970746	0.973746	0.970746	0.970210	0.890913	0.972782
f1_test	0.912941	0.937941	0.971386	0.956336	0.970609	0.959838	0.967469	0.921866	0.961849

Notes: Pre_train, rec_train, and f1_train represent the precision rate, recall rate, and F1-score for the training set, respectively, while their values for the test set are denoted as pre_test, rec_test, and f1_test, respectively.

Table B
Prediction accuracy of different models on the oversampling PSC inspection dataset

Oversampling	SVM	cNB	Lo	AdaBoost	nn_clf	kn	DT	AdaBoostSVM	SVM-DBN
pre_train	0.891696	0.841819	0.880687	0.877554	0.960208	0.951863	0.973218	0.690671	0.966954
rec_train	0.891351	0.841803	0.880680	0.877550	0.958241	0.949377	0.972124	0.639193	0.972432
f1_train	0.891327	0.841801	0.880679	0.877550	0.958197	0.949307	0.972108	0.613078	0.934134
pre_test	0.972988	0.964694	0.967129	0.967212	0.962797	0.959842	0.956475	0.958047	0.931253
rec_test	0.875482	0.841406	0.873875	0.875589	0.907415	0.881054	0.918667	0.888234	0.942755
f1_test	0.913744	0.889704	0.910115	0.911183	0.929804	0.913308	0.934911	0.919485	0.964119

Notes: The prediction accuracy metrics and models are defined in the same way as before.

Table C
Prediction accuracy of different models on the under-sampling PSC inspection dataset

Under-sampling	SVM	cNB	Lo	AdaBoost	nn_clf	kn	DT	AdaBoostSVM	SVM-DBN
pre_train	0.854620	0.833561	0.876732	0.872325	0.971733	0.874444	0.972427	0.953240	0.942857
rec_train	0.853625	0.833483	0.876571	0.871634	0.971724	0.873429	0.971724	0.951613	0.941014
f1_train	0.853522	0.833473	0.876558	0.871574	0.971723	0.873343	0.971713	0.951569	0.940952
pre_test	0.970191	0.970929	0.972938	0.972821	0.969983	0.972010	0.969524	0.969792	0.970263
rec_test	0.899057	0.840763	0.872375	0.860051	0.836048	0.815045	0.849443	0.831976	0.828654
f1_test	0.926673	0.892819	0.912329	0.904821	0.889834	0.876881	0.898007	0.885670	0.883622

Notes: The prediction accuracy metrics and models are defined in the same way as before.

References

Ahmed, M., Rashid, A.N.M.B., 2024. EDSUCh: a robust ensemble data summarization method for effective medical diagnosis. Digital Communications and Networks 10, 182–189. https://doi.org/10.1016/j.dcan.2022.07.007.

Akyuz, E., Celik, M., 2014. A hybrid decision-making approach to measure effectiveness of safety management system implementations on-board ships. Saf. Sci. 68, 169–179. https://doi.org/10.1016/j.ssci.2014.04.003.

Batista, G.E.A.P.A., Prati, R.C., Monard, M.C., 2004. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor. Newsl. 6, 20–29. https://doi.org/10.1145/1007730.1007735.

Cariou, P., Mejia, M.Q., Wolff, F.-C., 2007. An econometric analysis of deficiencies noted in port state control inspections. Marit. Pol. Manag. 34, 243–258. https://doi.org/10.1080/03088830701343047.

Cariou, P., Wolff, F.-C., 2015. Identifying substandard vessels through port state control inspections: a new methodology for concentrated inspection campaigns. Mar. Pol. 60, 27–39. https://doi.org/10.1016/j.marpol.2015.05.013.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297. https://doi.org/10.1007/BF00994018.

Degré, T., 2007. The use of risk concept to characterize and select high risk vessels for ship inspections. WMU J Marit Affairs 6, 37–49. https://doi.org/10.1007/BF03195088.

Emecen Kara, E., 2016. Risk assessment in the istanbul strait using black sea MOU port state control inspections. Sustainability 8, 390. https://doi.org/10.3390/su8040390.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F., 2012. A review on ensembles for the class imbalance problem: bagging-, Boosting-, and hybrid-based approaches. IEEE Trans. Syst., Man, Cybern. C 42, 463–484. https://doi.org/10.1109/TSMCC.2011.2161285.

Gao, J., Liu, K., Wang, B., Wang, D., Hong, Q., 2021. An improved deep forest for alleviating the data imbalance problem. Soft Comput. 25, 2085–2101. https://doi.org/10.1007/s00500-020-05279-8.

García, S., Fernández, A., Herrera, F., 2009. Enhancing the effectiveness and interpretability of decision tree and rule induction classifiers with evolutionary training set selection over imbalanced problems. Appl. Soft Comput. 9, 1304–1314. https://doi.org/10.1016/j.asoc.2009.04.004.

Hänninen, M., Kujala, P., 2014. Bayesian network modeling of port state control inspection findings and ship accident involvement. Expert Syst. Appl. 41, 1632–1646. https://doi.org/10.1016/j.eswa.2013.08.060.

Heij, C., Knapp, S., 2018. Predictive power of inspection outcomes for future shipping accidents – an empirical appraisal with special attention for human factor aspects. Marit. Pol. Manag. 45, 604–621. https://doi.org/10.1080/03088839.2018.1440441.

Hocek, H., Yay, S., Yazir, D., 2025. Comprehensive analysis of ship detention probabilities using binary logistic regression method with machine learning. Ocean Eng. 315, 119889. https://doi.org/10.1016/j.oceaneng.2024.119889.

Islam, M.A., Rashid, S.I., Hossain, N.U.I., Fleming, R., Sokolov, A., 2023. An integrated convolutional neural network and sorting algorithm for image classification for efficient flood disaster management. Decision Anal. J. 7, 100225. https://doi.org/10.1016/j.dajour.2023.100225.

Liu, K., Yu, Q., Yang, Zhisen, Wan, C., Yang, Zaili, 2022. BN-based port state control inspection for paris MoU: new risk factors and probability training using big data. Reliab. Eng. Syst. Saf. 224, 108530. https://doi.org/10.1016/j.ress.2022.108530.

Molinaro, A.M., Simon, R., Pfeiffer, R.M., 2005. Prediction error estimation: a comparison of resampling methods. Bioinformatics 21 (15), 3301–3307.

Şanlıer, Ş., 2020. Analysis of port state control inspection data: the black sea region. Mar. Pol. 112, 103757. https://doi.org/10.1016/j.marpol.2019.103757.

Shahana, T., Lavanya, V., Bhat, A.R., 2023. State of the art in financial statement fraud detection: a systematic review. Technol. Forecast. Soc. Change 192, 122527. https://doi.org/10.1016/j.techfore.2023.122527.

Sauerbrei, W., 1999. The use of resampling methods to simplify regression models in medical statistics. J. Roy. Stat. Soc. C Appl. Stat. 48 (3), 313–329.

Wang, J., Zhou, Y., Zhang, S., Zhuang, L., Shi, L., Chen, J., Hu, D., 2022a. Societal risk acceptance criteria of the global general cargo ships. Ocean Eng. 261, 112162. https://doi.org/10.1016/j.oceaneng.2022.112162.

Wang, J., Zhou, Y., Zhuang, L., Shi, L., Zhang, S., 2022b. Study on the critical factors and hot spots of crude oil tanker accidents. Ocean Coast Manag. 217, 106010. https://doi.org/10.1016/j.ocecoaman.2021.106010.

Wang, L., Wang, J., Shi, M., Fu, S., Zhu, M., 2021. Critical risk factors in ship fire accidents. Marit. Pol. Manag. 48, 895–913. https://doi.org/10.1080/03088839.2020.1821110.

Wang, S., Yan, R., Qu, X., 2019. Development of a non-parametric classifier: effective identification, algorithm, and applications in port state control for maritime transportation. Transp. Res. Part B Methodol. 128, 129–157. https://doi.org/10.1016/j.trb.2019.07.017.

Wang, Y., Zhang, F., Yang, Zhisen, Yang, Zaili, 2021. Incorporation of deficiency data into the analysis of the dependency and interdependency among the risk factors influencing port state control inspection. Reliab. Eng. Syst. Saf. 206, 107277. https://doi.org/10.1016/j.ress.2020.107277.

Wu, B., Yip, T.L., Yan, X., Guedes Soares, C., 2022. Review of techniques and challenges of human and organizational factors analysis in maritime transportation. Reliab. Eng. Syst. Saf. 219, 108249. https://doi.org/10.1016/j.ress.2021.108249.

Xu, R., Lu, Q., Li, K.X., Li, W., 2007. Web mining for improving risk assessment in port state control inspection. In: 2007 International Conference on Natural Language Processing and Knowledge Engineering. Presented at the 2007 International Conference on Natural Language Processing and Knowledge Engineering. IEEE, Beijing, China, pp. 427–434. https://doi.org/10.1109/NLPKE.2007.4368066.

Xu, R.-F., Lu, Q., Li, W.-J., Li, K.X., Zheng, H.-S., 2007. A risk assessment system for improving port state control inspection. In: 2007 International Conference on Machine Learning and Cybernetics. Presented at the 2007 International Conference on Machine Learning and Cybernetics, IEEE, Hong Kong, China, pp. 818–823. https://doi.org/10.1109/ICMLC.2007.4370255.

Yan, R., Wang, S., 2019. Ship inspection by port state control—review of current research. In: Qu, X., Zhen, L., Howlett, R.J., Jain, L.C. (Eds.), Smart Transportation Systems 2019, Smart Innovation, Systems and Technologies. Springer Singapore, Singapore, pp. 233–241. https://doi.org/10.1007/978-981-13-8683-1_24.

Yan, R., Wang, S., Peng, C., 2021. An artificial intelligence model considering data imbalance for ship selection in port state control based on detention probabilities. J. Comput. Sci. 48, 101257. https://doi.org/10.1016/j.jocs.2020.101257.

Yang, Zhisen, Wan, C., Yu, Q., Yin, J., Yang, Zaili, 2023. A machine learning-based Bayesian model for predicting the duration of ship detention in PSC inspection. Transport. Res. E Logist. Transport. Rev. 180, 103331. https://doi.org/10.1016/j.tre.2023.103331.

Yang, Zhisen, Yang, Zaili, Yin, J., 2018. Realising advanced risk-based port state control inspection using data-driven Bayesian networks. Transport. Res. Pol. Pract. 110, 38–56. https://doi.org/10.1016/j.tra.2018.01.033.

Yu, L., Zhou, R., Tang, L., Chen, R., 2018. A DBN-Based resampling SVM ensemble learning paradigm for credit classification with imbalanced data. Appl. Soft Comput. 69, 192–202. https://doi.org/10.1016/j.asoc.2018.04.049.

Zhang, X., Liu, C., Xu, Y., Ye, B., Gan, L., Shu, Yaqing, 2024. A knowledge graph-based inspection items recommendation method for port state control inspection of LNG carriers. Ocean Eng. 313, 119434. https://doi.org/10.1016/j.oceaneng.2024.119434.