Maritime Piracy Risk Assessment: A Two-step Approach

Abstract: Maritime piracy is a serious threat to the safety of cargo, ships, and crews, which can cause enormous losses to stakeholders, highlighting the significance of piracy risk assessment and prevention to the maritime industry. In this study, we propose a two-step analytical framework based on a Random Forest (RF) model, Generative Adversarial Nets (GANs), and Matrix Completion (MC) algorithm to assess the risks of successful piracy attacks. We consider different decision-makers in each step, namely, pirates first select which ships to attack and then ship operators determine the probability of a (un)successful attack. We propose different influencing factors for each step and, in the meantime, solving the problems of incomplete and imbalanced data. A case study in Southeast Asia is then conducted based on the proposed approach. The results show that, in mild wind weather, the likelihood of a piracy attack on a ship with a DWT not exceeding 50000 tons is up to 90%. Further, if the attack occurs between 0-6 a.m., the probability of success is over 90%. These results provide more specific information for ship operators and local authorities to develop efficient anti-piracy strategies and policies.

Keywords: Maritime safety; Assessment of piracy risk; Piracy prediction; Impact factors analysis; Random Forest

1 Introduction

Maritime piracy is a serious threat to the safety of cargo, ships, and crews, which can cause enormous losses to stakeholders (Jin et al., 2019). According to the International Maritime Bureau (IMB), piracy causes a loss of about $ 25 billion a year in global economic losses due to theft, ransoms, increased insurance costs, etc. Also, the occurrence of piracy has a significant deterrent effect on international trade, for more than 90% of all international trade goods are transported by sea (United Nations, 2020). For every 10 additional piracy attacks, bilateral trade decreases by 1.5% (Robitaille, 2019). Marine stakeholders have taken some measures against pirates, such as naval escorts. These measures have been effective to some extent. However, there are still a large number of piracy attacks. Taking the year 2020 as an example, a total of 195 pirate attacks were recorded worldwide, a 20% increase over 2019, and resulted in 135 crew members being kidnapped (IMO, 2020). Therefore, it is significant for marine transport and international trade to find some measures of reducing and preventing piracy by identifying the influencing factors behind attacks and estimating the likelihood of the attacks.

Traditional studies on the prediction and prevention of piracy only consider one stage of analysis, namely whether pirate attacks or whether an attack is successful (Di Salvatore, 2018; Jiang and Lu, 2020; Liu et al., 2021; Pristrom et al., 2016; Shepard and Pratson, 2020a). In fact, a successful attack is a very complicated process in practice and generally can be regarded as comprising two steps: first, pirates select which ships are more susceptible to attacks; and second, ship operators decide which

factors can effectively lower the possibility of a successful attack. Previous studies have largely ignored this process, which may result in information lost in the estimations, and thus limit the efficiency of countermeasures. On the other hand, the combined results of a two-step framework can reveal a series of high-risk scenarios that otherwise would be overlooked, and help generate more effective anti-piracy measures.

This study proposes a two-step analytical framework to assess maritime piracy risks, using Southeast Asia as an example. The piracy data from Global Integrated Shipping Information System (GISIS) is used for analysis. To achieve our research objective, two challenges need to be addressed. The first is to identify the influencing factors in each step. Although previous one-step analyses have recognized many factors and their impacts on pirate attacks (Jiang and Lu, 2020), these factors are considered to take effect simultaneously, without differentiating their difference in nature, e.g., the different decision-makers and sequences. In fact, each step is influenced by different factors (Jin et al., 2019). We have to associate these factors with the appropriate step. The second is the imbalance and missing data. According to GISIS reports, the ratio of unsuccessful attacks to successful attacks is 1 to 9. Data imbalance can lead to significant bias in econometric analysis (Longadge and Dongre, 2013). Moreover, the information in GISIS reports is incomplete, with much data missing, and ignoring this problem can introduce bias and inaccurate conclusions (Brown and Kros, 2003).

The contributions of this study are three-fold. First, we initially propose a two-step approach to assess maritime piracy risks, taking different decision-makers into consideration, which is considered more realistic in estimating the likelihood of piracy incidents. Second, we develop an analytical framework based on a Random Forest (RF) model, Generative Adversarial Nets (GANs), and Matrix Completion (MC) algorithm to solve a two-stage estimation with missing and imbalanced data. Third, the findings reveal some high-risk scenarios, which is helpful to recommend more effective and specific anti-piracy strategies for various stakeholders.

The rest of the study is organized as followed. Section 2 reviews previous studies of maritime piracy analysis. Section 3 presents the factors determining the two-step piracy risk and the models used in this study. In section 4, a case study in Southeast Asia is presented to demonstrate the application of our analytical framework and some interesting findings. Section 5 concludes the study.

2 Literature review

2.1 Analysis of maritime piracy

Maritime piracy is a global issue that directly threatens the lives of seafarers and heavily affects maritime activities, as well as the world economy and trade. Thus, increasing focuses haven been put on the analysis of maritime piracy, recently.

A careful literature review suggest that studies of piracy involve motivations for piracy (Hastings, 2009; Mo, 2002; Sumaila and Bawumia, 2014); evolution and characteristics of piracy (Hassan and Hasan, 2017); impact of piracy on trade (Flückiger and Ludwig, 2015; Morabito and Sergi, 2018; Robitaille, 2019); economic effects of piracy (Fu et al., 2010; Shepard and Pratson, 2020a; Tominaga, 2018); anti-piracy laws (Ahmad, 2020; Aziz et al., 2021; Kao, 2019); anti-piracy programs and measures (Gilmer and Kane, 2019; Gottlieb, 2013); and risk of maritime piracy (Bouejla et al., 2014; Shepard and Pratson, 2020b).

Most of the previous studies on piracy have been analyzed from governmental and legal levels (Jiang and Lu, 2020), yet a few studies have identified the factors influencing piracy attacks (Jin et al., 2019; Shane and Magnuson, 2014). In fact, analyzing the factors affecting piracy, such as ship type, is important and useful for ship operators to prevent and develop efficient anti-piracy strategies (Jin et al., 2019).

Researchers have found that the piracy attacks are not random and influenced by many factors, such as ship flag and type (Mejia Jr et al., 2009). Bateman (2010) proposed that the vulnerability of ships to piracy depends on factors such as ship type, size, speed, freeboard, and voyage. That is to say, the ship feature is a significant factor. Also, environmental condition is deemed as important. The motions of pirate vessels are usually related to environmental conditions, including season, ocean, and weather conditions (Dabrowski and de Villiers, 2015). In addition, studies have shown that the number of pirates and the weapons used by pirates are important factors (Bouejla et al., 2014; Hastings, 2020; Liu et al., 2021).

Studies have also investigated many factors that contribute to successful piracy attacks. It is commonly believed that the probability of a successful attack is higher for a vessel at berth or at anchor (Shane et al., 2018; Wong and Yip, 2012), as well as at night, in territorial waters and port areas (Jin et al., 2019). Therefore, a successful attack is influenced by ship status, time, and location. Additionally, the general perception is that crews' anti-piracy actions can reduce the risk of successful attacks (Jiang and Lu, 2020; Lewis, 2016; Shane and Magnuson, 2014). Pristrom et al. (2016) proposed that a successful attack is also related to (the lack of) naval support.

Most of these previous studies have focused on one-step analysis, either general pirate attacks or successful attacks. While, in fact, a successful pirate attack is more of a two-step, sequential process, where each step has different decision-makers. In this study, we aim to identify the influencing factors in different steps and evaluate their combined effect.

2.2 Approaches of piracy analysis

Various approaches are applied for analyzing piracy prediction and prevention, such as the logistic regression model (Jin et al., 2019), Formal Safety Assessment (FSA) (Yang et al., 2013), Analytical Hierarchy Process (AHP), and Evidential Reasoning (ER) (Hidayati et al., 2019), Bayesian Network (BN) (Jiang and Lu, 2020; Liu et al., 2021; Pristrom et al., 2016). Among them, BN is the most widely used. However, this approach relies heavily on expert knowledge; thus it is difficult to guarantee the objectivity and accuracy of the results (Jiang and Lu, 2020). The RF method, which relies on random selection instead of expert knowledge, is an objective method with high accuracy, fast learning speed, and simple computation (Breiman, 2001). With the advantages of reducing overfitting risk, determining important features, and having few parameters to tune, RF models have been successfully applied in solving a wide scope of practical problems, such as choices of travel mode (Cheng et al., 2019), price prediction (Escribano and Wang, 2021; Ghosh et al., 2021), prediction and classification of disease (Meher et al., 2014), classification and regression tasks in remote sensing (Izquierdo-Verdiguier and Zurita-Milla, 2020), as well as risk analysis (Chen et al., 2021; Keramati et al., 2020; Zhang et al., 2022). The RF model, combining the simplicity of decision trees with the flexibility and power of an ensemble model, is considered as a powerful tool for prediction and behavior analysis. In the study, we apply it for our risk analysis and prediction of maritime piracy.

Notably, the RF model relies heavily on balanced data. In reality, piracy data is unbalanced, which can lead to significant bias in the results (Longadge and Dongre, 2013). Also, there is a large amount of missing data, e.g., more than 13% of ship speed information is missing in the relevant database, which limits the application of the RF model in risk analysis (Brown and Kros, 2003).

As for the imbalanced data problem, some approaches, such as resampling, feature selection and extraction, and cost-sensitive learning, are broadly adopted (Guo et al., 2017; Yen and Lee, 2009). Most of the resampling methods generate samples that lack diversity (Zhai et al., 2022); feature selection and extraction methods are computationally intensive (Saeys et al., 2007); for cost-sensitive learning methods, it is difficult to set values in the cost matrix. In most cases, the misclassification cost is unknown from the data and cannot be given by an expert (Guo et al., 2017). To simplify the calculation and approximate real samples, Goodfellow et al., (2014) proposed GANs, and it has been applied to many fields now, such as information security, the medical field, image processing, and computer vision (Gui et al., 2021; Ren and Xu, 2019). As an unsupervised learning method, the GANs generate artificial data that is very similar to real data and are regarded as an effective instrument to correct the class imbalance. Thus, the GANs method can deal with imbalanced data in piracy analysis.

For the missing data problem, evidential reasoning (Liu et al., 2015) and the expectation-maximization algorithm (Graham et al., 2013) have been most observed. In particular, MC was proposed to deal with the problem of a large amount of missing data (Cai et al., 2010), even with missing data rates as high as 90% (Jiang et al., 2016). This algorithm has been used in recommendation systems, risk management, and electricity distribution systems (Genes et al., 2016; Georgescu et al., 2018; Ramlatchan et al., 2018). The MC algorithm, which is commonly used to recover lost information, is considered as a promising technique and has received much attention in the past several years. The MC algorithm is used by us to solve the problem of missing data in piracy analysis.

The imbalanced and missing data problems can be easily overlooked in the analysis of maritime piracy risks, which leads to estimation bias. In this study, we propose a framework combining RF, GAN, and MC to study piracy risks, to remedy the problem.

3 Methodology

In this section, we propose a new, two-step analytical framework for assessing maritime piracy risks, combining RF, MC, and GAN, shown in Figure 1 below.

In this framework, first, the factors determining the two-step piracy risk are determined based on previous studies and the GISIS piracy reports. Second, the MC algorithm is applied to fill in the missing data in the piracy data, and the GANs method helps to generate the balanced data. Third, based on the balanced data, the RF model is used to analyze the piracy risk and the variable importance.

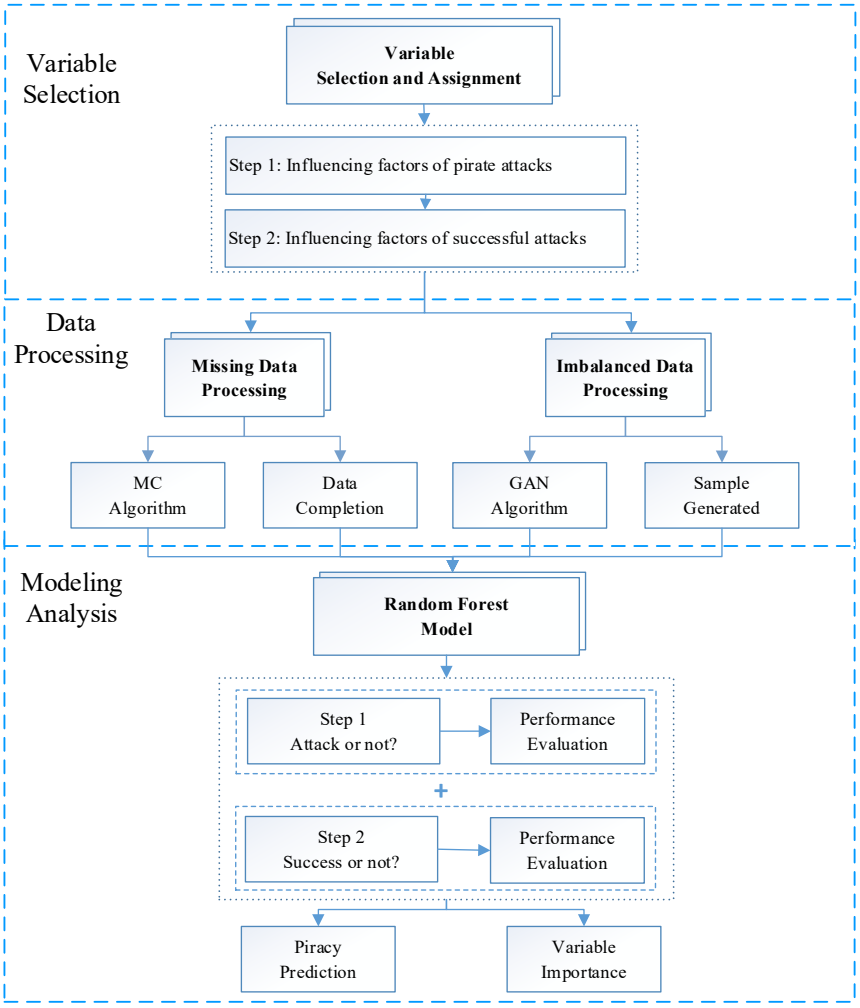

Figure 1 Framework for assessing the risk of maritime piracy

3.1 Factors determining the two-step piracy risk

The two steps of estimating piracy risks are sequential, starting from pirates who select ships to attack, ship operators then taking measures to reduce the probability of being successfully attacked. Usually, pirates consider ship characteristics before selecting a target, as some types of ships are easier to board (Jin et al., 2019). They also take into account weather conditions, such as wind, for the boats used by pirates are limited in windy weather (Jiang and Lu, 2020). In addition, the pirate ability is also an important factor (Liu et al., 2021). The next step, for ship operators, their operations choice can influence the (un)successful rate of an attack, e.g. it is safer to sail during daytime than at night (Jiang and Lu, 2020). Also, proper anti-piracy measures taken by ship operators can reduce the damage of an attack, e.g. passive resistance can decrease the rate of successful boarding (Jin et al., 2019).

In this section, we summarize and divide the influencing factors of pirate attacks from previous studies according to the aforementioned two steps. For Step 1, we propose three groups of factors, which are **ship feature**, **natural environmental condition**, and **pirate ability**. For Step 2, the factors comprise the **space-time status of a ship** and the **anti-piracy measures** taken by ship operators. We summarize

these influencing factors in Table 1, and each factor is explained in detail in the following text.

Table 1 Summary of influencing factors

Step	Factor		Literature
Step 1: Pirates select ships to attack	Ship feature	*Ship flag*	Hastings (2020); Mejia Jr et al. (2009)
		Ship type	Jin et al. (2019); Mejia Jr et al. (2009)
		Ship size	Bateman (2010); Hastings (2020)
	Natural environmental conditions	*Season*	Dabrowski and de Villiers (2015); Wong and Yip (2012)
		Wind speed	Jiang and Lu (2020)
		Visibility	Pristrom et al. (2016)
	Pirate ability	*Number of pirates*	Bouejla et al. (2014); Liu et al. (2021)
		Weapon	Hastings (2020); Tominaga (2018)
Step 2: Ship operators determine the probability of an (un)successful attack	Ship space-time status	*Speed*	Shane et al. (2018); Pristrom et al. (2016)
		Time	Jiang and Lu (2020); Shane and Magnuson (2014)
		Location	Lewis (2016); Shane and Magnuson (2014)
	Anti-piracy measures	*Avoidan*ce	Bouejla et al. (2014); Jin et al. (2019)
		Alarm	Jin et al. (2019)
		Anti-piracy watch	Pristrom et al. (2016)
		Naval support	Jiang and Lu (2020) Lewis (2016)
		ReCAAP	Jiang and Lu (2020)
		Others	Pristrom et al. (2016); Lewis (2016)

Note: *ReCAAP* (Regional Cooperation Agreement on Combating Piracy and Armed Robbery against Ships in Asia).

Ship Feature:

Ship flag shows the country where a vessel is registered. There are two categories, Flag of Convenience (FOC) and the other, with values of 1 and 2 respectively. According to GISIS piracy reports, three of the top 5 attacked ship flags are FOCs, belonging to Panama, Liberia, and the Marshall Islands (see Figure 2 (a))

Ship type includes tanker, bulk carrier, container ship, general cargo ship, ro-ro ship, tug, barge carrier, fishing boat, and so on. As shown in Figure 2 (b), in our data, the top two types of ships attacked are tanker and bulk carrier. In this study, ship type is divided into five categories (Jin et al., 2019; Mejia Jr et al., 2009), tanker, bulk carrier, container ship, general cargo ship, and other, with values of 1, 2, 3, 4 and 5 respectively.

Ship size is defined by ship tonnage. From Figure 2 (c), smaller ships are more vulnerable to pirate attacks. In this study, the ship sizes are divided into five categories by DWT (Dead Weight Tonnage) (Pristrom et al., 2016), including (1,15 000], (15 000, 20 000], (20 000, 50 000], (50 000, 100 000], and (100 000, +∞), having values of 1, 2, 3, 4, and 5 respectively.

Natural environmental condition:

Season refers to the season in which a pirate attack occurs. As seen in Figure 2 (d), the number of piracy incidents is the highest in April and May. According to previous studies (Fu et al., 2010; Shane

and Magnuson, 2014), this variable has four states, 1st quarter, 2nd quarter, 3rd quarter, and 4th quarter, and their values are set as 1, 2, 3, and 4 respectively.

Wind speed and *Visibility* present weather conditions under which piracy incidents occur. Pirates are known to avoid hostile weather conditions, such as high winds, high waves, and poor visibility. *Wind speed* has three states, [0,5.4 m/s] (low), (5.4 m/s, 10.7 m/s] (moderate) or (10.7 m/s, +∞) (high), with values of 1, 2, and 3 respectively; *Visibility* also has three states, [0, 4 km] (bad), (4 km, 20 km) (moderate), and [20 km, +∞) (good), and the values are 1, 2, and 3, respectively.

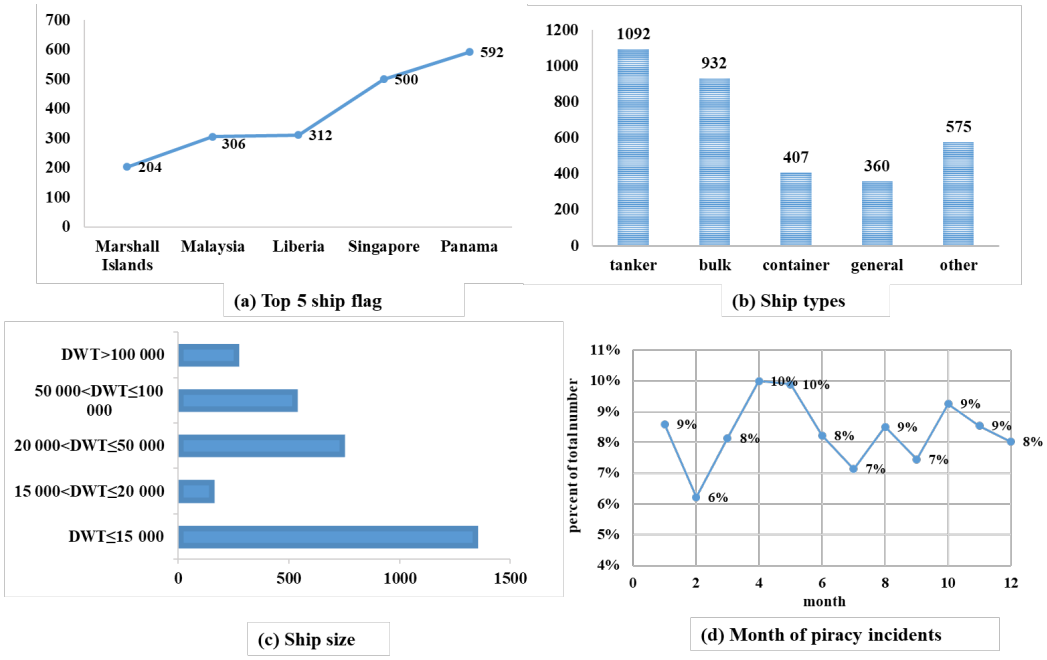

Figure 2 Features of ships attacked in Southeast Asia

Pirate ability:

Number of pirates affects piracy risks. Generally, large ships are equipped with more crew than small ships. Pirates must gather more accomplices if they want to attack large ships. In Southeast Asia, most attacks are carried out by 1-4 pirates, and most of the ships attacked are small. This variable is divided into three states, including [1, 4], [5, 9], and [10, +∞), with values of 1, 2, and 3, respectively.

Weapon refers to the weapons used by pirates, including guns, knives, etc. From GISIS, unarmed pirates are usually caught or dispersed by crew. Knives are widely used by pirates in Southeast Asia, while guns account for only 16%. This variable has three states, including guns, knives, and other, and their values are 1, 2, and 3 respectively.

Ship space-time status:

Ship speed includes three states of "At anchor", "< Fifteen knots", and "≥Fifteen knots" (Jin et al., 2019), with values of 1, 2, and 3 respectively. According to GISIS reports, more than half of attacked ships were at anchor, and the main reason is that it is easier for pirates to steal from ships at anchor.

Time refers to the time of day when a piracy event occurs. According to Figure 3 (a), the most dangerous time is between 00:00 and 01:59. This variable has four states in this study, including [00:00,06:00), [06:00-12:00), [12:00-18:00), and [18:00-24:00), whose values are 1, 2, 3, and 4

respectively.

Location indicates the location where a piracy incident occurs. From GISIS reports, most of the incidents occurred in territorial waters (see Figure 3 (b)). *Location* has three states, including in port areas, in territorial waters, or in international waters, and their values are 1, 2, and 3 respectively.

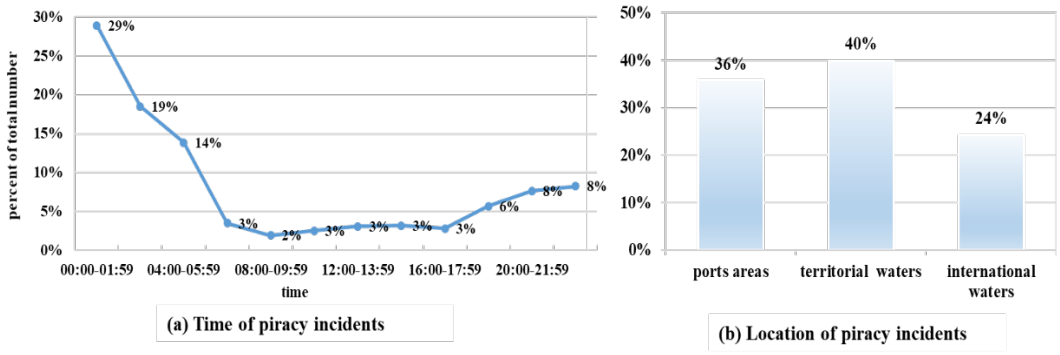

Figure 3 Ship status and environment conditions

Anti-piracy measures:

*Avoidan*ce refers to the avoidance measures taken by a crew to prevent an attack, such as conducting evasive maneuvers towards pirate boats, changing sailing course, increasing speed, and managing to outpace attackers. The variable has two states, including yes or no, and their values are 1 and 0 respectively.

Alarm is one of the warning measures taken by a crew in order to notify all crew members of a dangerous or suspicious situation. Usually, when pirates are spotted, it is a routine activity to sound the alarm and muster the crew. The variable has two states of yes and no, with values of 1 and 0 respectively.

Anti-piracy watch refers to arranging the crew to carry out an anti-piracy watch plan. An anti-piracy watch can discover pirates timely so that the master has enough time to make decisions to avoid or defend themselves against piracy. The variable has two states, including yes and no, with values of 1 and 0 respectively.

Naval support helps ships to disperse many pirates. Case studies show that, when naval forces are available within 15 min after being informed of an attempted attack, the risk of the ship being hijacked decreases to 0.71% (Pristrom et al., 2016). This variable has two states, yes and no, with values of 1 and 0 respectively.

ReCAAP is established to collaborate among the littoral states of the Straits of Malacca and Singapore and the South China Sea, and it is the first regional government-to-government agreement to promote and enhance cooperation against piracy and armed robbery against ships in Asia. The variable has two states of yes and no, and their values are 1 and 0 respectively.

Others include using fire hoses, switching on the searchlight, firing off parachute flares, and so on. The variable has two states, including yes and no, with values of 1 and 0 respectively.

3.2 Methods for handling missing and imbalanced data

For missing data, this study uses the MC algorithm (Cai et al., 2010). The algorithm first decomposes a matrix with missing values into two (or more) matrices. Then, an approximation of the original matrix

is obtained by multiplying the decomposed matrices. The new matrix is used to fill in the missing part of the original matrix. Suppose that there are m samples, and each sample has k dimensions (features). X is the piracy data matrix with missing data.

Let $X \approx \hat{X} = UW^T$. $\hat{X}$ is the estimation of X. The decomposed matrices U is a $m \times l$ matrix, and the decomposed matrices W is a $k \times l$. $l < \{m, k\}$. $rank(U) \leq l$, and $rank(W) \leq l$. In order to obtain an optimal $\hat{X}$, the loss function J is written as follows:

$$min\, J = \left\| X - \hat{X} \right\|_F^2 \tag{1}$$

Here, $\left\| X - \hat{X}^* \right\|_F^2$ is the reconstruction error; $\|\cdot\|_F$ is the Frobenius norm.

The imbalanced data problem is solved by a GAN (Goodfellow et al., 2014). In a GAN, the generative model produces fake samples, and uses these fake samples without detection. While the discriminative model learns to determine whether a sample is from a model distribution or a data distribution. Through this game of competition, the two models are both improved. Let G be the generative model, and D the discriminative model. $D(x)$ represents the probability that x comes from data set V, and $G(z)$ is the data generated via inputting noise variable z. We train D to maximize the probability of assigning the correct labels to both training examples and samples from G. We simultaneously train G to minimize $log(\,1 - D(G(z)))$. In other words, D and G play the following two-player minimax game with a value function $F(D, G)$:

$$\min_G \max_D F(D, G) = E[log(\,D(x)] + E[log(\,1 - D(G(z)))] \tag{2}$$

3.3 The Random Forest model

RF is an ensemble method based on machine learning theory and is used to solve prediction problems. RF was proposed by Breiman (2001) and has proved to be a powerful and effective tool in prediction across a number of industries (Chen et al., 2021; Cheng et al., 2019; Escribano and Wang, 2021).

Suppose that there are m samples in piracy data set V. The input variable has k dimensions (features), recorded as $X = (x_1, x_2, \cdots, x_k)$, and $x_i = (x_{1i}, x_{2i}, \cdots, x_{mi})$, $i = 1,2, \cdots, k$. The output variable is Y, $Y = (y_1, y_2, \cdots, y_m)$. In Step 1, if pirates attack a ship, the output variable $y_j = 1$, or $y_j = 0$, $j = 1,2, \cdots, m$; in Step 2, if the attack is successful, the output variable $y'_j = 1$, or $y'_j = 0$, $j = 1,2, \cdots, m$. The process of the RF model involves three steps, which presented in Figure 4. The first step is bootstrap sampling, and n bootstrap sets $V_1^*, V_2^*, \cdots, V_n^*$ are obtained. The second step is to grow the decision trees. In the bootstrap set V_1^*, k^* features are selected randomly from k features, and then the tree is grown to its maximum size with CART (Classification and Regression Trees) method. The last step is to aggregate the results of all decision trees.

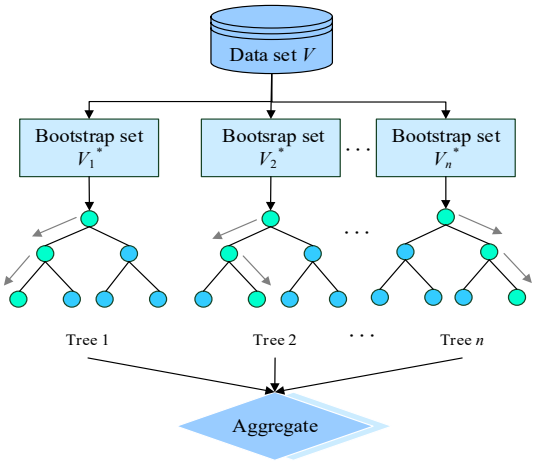

Figure 4 RF process

Out-of-bag (OOB) estimate is used to monitor error, and its three steps are as follows (see Figure 5). The first step is to take the OOB estimates as inputs to calculate the results of the decision trees. Next is to aggregate the results of the OOB estimates. The last step is to compute the OOB error. According to these steps, we will obtain a better RF with a lower generalization error, which has a lower correlation between classifiers and higher strength.

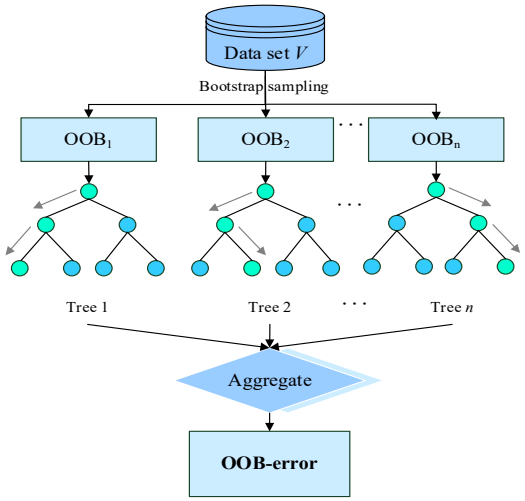

Figure 5 OOB estimates

According to (Lu et al., 2022; Makariou et al., 2020), we set the parameters as follows: the number of trees is 100; the criteria for decision trees splitting is the mean squared error; the maximum number of features selected when building trees is “None”; and bootstrap sampling is used when building the tree. In addition, this study splits the training set and testing set according to the ratio of 7:3. Also, some primary judgement criteria are needed, including accuracy, precision, recall, and F1-sore. Before explaining these criteria, we make the following definitions:

Positive: in Step 1, “attack” is a positive class; in Step 2, “success” is a positive class.

Negative: in Step 1, “not attack” is a positive class; in Step 2, “unsuccess” is a positive class.

True positives (TP): an outcome where the model correctly predicts the positive class.

True negatives (TN): an outcome where the model correctly predicts the negative class.

False positives (FP): an outcome where the model incorrectly predicts the positive class.

False negatives (FN): an outcome where the model incorrectly predicts the negative class.

Figure 6 takes Step 1 as an example to illustrate the definitions above.

		Actual	
		Positive "attack"	**Negative** "not attack"
Predicted	**Positive** "*attack*"	**True positives(TP)**	**False Positive(FP)**
	Negative "*not attack*"	**False Negative(FN)**	**True Negative(TN)**

Figure 6 The confusion matrix of Step 1

Next, we introduce the four criteria. *Accuracy* is the fraction of accurate predictions and is defined as:

$$Accuracy = \frac{Number\ of\ predictions}{Total\ number\ of\ predictions} = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Precision is the proportion of positive identifications being correctly predicted:

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

Recall is the proportion of actual positives being identified correctly:

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

To fully evaluate the effectiveness of a model, both precision and recall must be examined. Unfortunately, the two are often in competition. That is, improving precision typically reduces recall and vice versa. To produce a model that balances both, we can use F1-score, the last criterion.

F1-score is the harmonic average of precision and recall. A good F1-score means there are low FP and low FN. F1-score is defined as follows:

$$F1 - score = \frac{2\times Recall\times Precision}{Recall+Precision} \tag{6}$$

To analyze piracy risks, it is critical to understand the interaction between variables that are affecting predictive accuracy. To solve this problem, Breiman (2001) proposed an OOB estimate method to measure variable importance, and the steps are as follows: First, calculate the error of the original OOB estimates, recorded as $ErrOOB_0$. Second, permute the values of the j^{th} variable. Third, compute a new OOB error, recorded as $ErrOOB_j$. Fourth, measure the importance of the j^{th} variable, as $Err_j = ErrOOB_j - ErrOOB_0$. If $Err_j > 0$, the accuracy of RF decreases, the larger the Err_j the stronger the ability of x_j to predict the results. On the contrary, if $Err_j \leq 0$, x_j has no contribution to the prediction results. According to the four steps, we can identify the important factors influencing the two-

step piracy risk.

4 Empirical research

According to GISIS, from 1995 to 2020, there were more than 8,000 piracy attacks in the world, and these attacks are spatially clustered, mainly in four regions, including Southeast Asia, West Africa, East Africa, and other regions (see Figure 7). This study selects the area of Southeast Asia as an empirical example, which accounts for more than 40% of all piracy incidents worldwide.

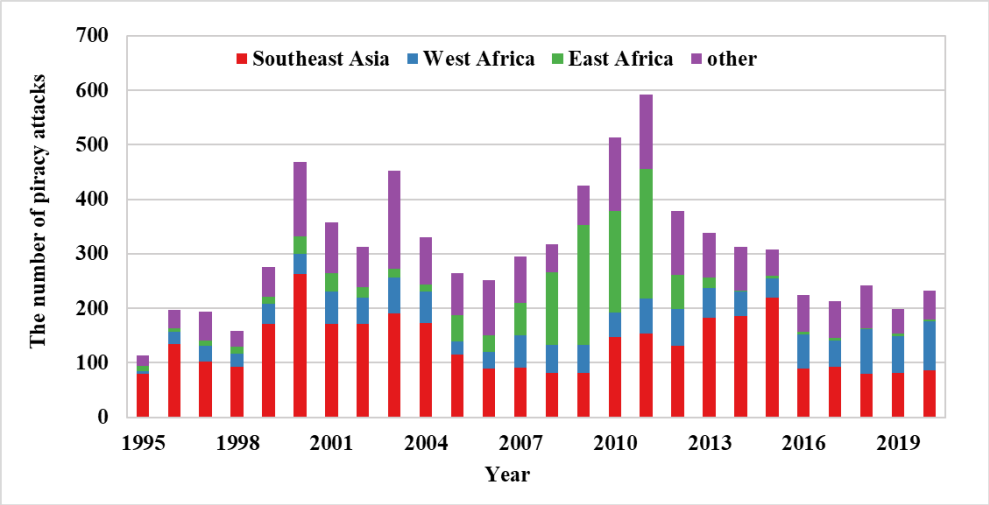

Figure 7 The number of piracy incidents from 1995 to 2020

4.1 Data collection and processing

We collected 3,475 reports of piracy incidents in Southeast Asia from January 1, 1995, to June 16, 2021. Among these reports, 837 reports have incomplete information. The main missing information includes the ship flag, vessel type, vessel tonnage, and ship status. In addition, geographic coordinates are provided for only 1,661 of the attacks in our collection. The following Figure 8 shows the locations of piracy attacks. According to their geographical distribution, we found that piracy clusters spacially and occurs usually close to coastlines. The areas near and in the Singapore Strait which is part of the Malacca Strait, are especially vulnerable. The Malacca Strait (excluding the Singapore Strait), waterways between the Sulu Sea and the Celebes Sea, and the Sunda Strait are also high-incident areas.

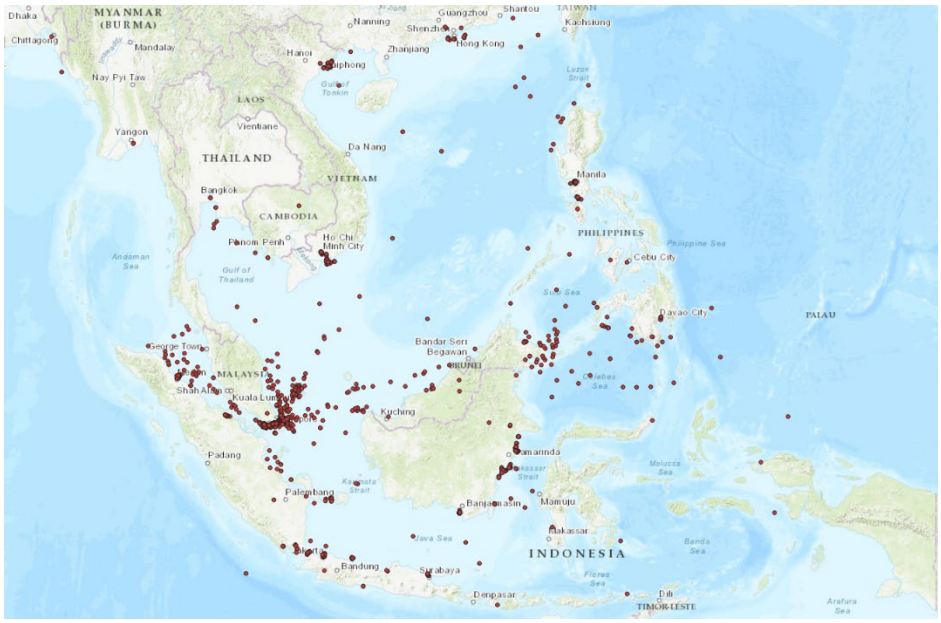

Figure 8 Piracy attacks in Southeast Asia

According to the analysis of factors determining the two-step piracy risk, we selected seventeen factors and two outcome variables for our model (shown in Table 2). Their data sources are as follows:

- *Ship type, status, season, time, location, pirates' number, weapon, avoidance, alarm, anti-piracy watch, naval support, ReCAAP, other measures*, *Attack*, and *Success* are from the Piracy and Armed Robbery database of GISIS (Global Integrated Shipping Information System, https://gisis.imo.org/).
- The data on *ship flag* and *size* are from the website of Shipping Online (http://www.sol.com.cn/).
- The data on maritime *wind speed* and *visibility* come from ICOADS (International Comprehensive Ocean-Atmosphere Data Set, https://icoads.noaa.gov/).

According to Table 2, Some preliminary findings can be observed. First, the standard deviation values of *Ship type*, *Ship size*, *Season,* and *Time* are more than one, which means that most values of these four variables differ significantly from their mean values. Second, some variables, including *Flag, Ship type, Ship size,* and *Status*, have missing values. Third, outcome variables Attack and Success have imbalanced data. According to GISIS, all reports collected record details of ships being attacked, and 88% of these attacks were successful. Therefore, the value of Attack for all reports is one, and the value of Success for 3,062 reports is one.

Table 3 shows the statistics of data processed by the MC and GANs. According to the results, the mean and SD value of the data processed by the MC is similar to the original data. While the data is balanced using the GANs.

Table 2 Variable, description, and statistics

Variable	Description	No. of data	Value	Statistic			
				min	max	mean	SD
Step 1*Flag*	Flag flown by the ship	3230	1,2	1.000	2.000	1.555	0.497

	Ship type	Type of the ship	3366	1, 2, 3, 4, 5	1.000	5.000	2.522	1.462
	Ship size	Classification by DWT	3082	1, 2, 3, 4, 5	1.000	5.000	2.423	1.417
	Season	Season when piracy occurs	3475	1, 2, 3, 4	1.000	4.000	2.516	1.106
	Wind speed	Wind speed at the incident location	3475	1, 2, 3	1.000	3.000	1.019	0.142
	Visibility	Visibility at the incident location	3475	1, 2, 3	1.000	3.000	2.700	0.582
	No. of pirates	Number of pirates	3475	1, 2, 3	1.000	3.000	1.374	0.558
	Weapon	Weapon used by pirates	3475	1, 2, 3	1.000	3.000	2.513	0.761
	Attack	Result of attack or not. State 1: yes; State 2: no	3475	1	1.000	1.000	1.000	0.000
	Speed	Speed of the ship	3019	1, 2, 3	1.000	3.000	1.466	0.506
	Time	Time of day when piracy occurs	3475	1, 2, 3, 4	1.000	4.000	1.850	1.219
	Location	Location of piracy incident	3475	1, 2, 3	1.000	3.000	1.885	0.766
	Avoidance	Evasive measure taken	3475	0, 1	0.000	1.000	0.167	0.373
	Alarm	Alert measure taken	3475	0, 1	0.000	1.000	0.564	0.496
Step 2	*Anti-piracy watch*	The watch measure taken	3475	0, 1	0.000	1.000	0.101	0.301
	Naval support	Report to navy	3475	0, 1	0.000	1.000	0.145	0.352
	ReCAAP	Report to the ReCAAP	3475	0, 1	0.000	1.000	0.003	0.051
	Other	Other measures taken	3475	0, 1	0.000	1.000	0.125	0.331
	Success	Result of success or not. State 1: yes; State 2: no	3475	0, 1	0.000	1.000	0.881	0.324

Notes: SD is standard deviation.

Table 3 Statistics of variables after the MC and GAN

		After the MC					After the GAN				
	Variable	**No. of**	**Statistic**				**No. of**	**Statistic**			
		data	**min**	**max**	**mean**	**SD**	**data**	**min**	**max**	**mean**	**SD**
	Flag	3475	1.000	2.000	1.565	0.496	6950	1.000	2.000	1.536	0.499
	Ship type	3475	1.000	5.000	2.525	1.442	6950	1.000	5.000	2.796	1.239
	Ship size	3475	1.000	5.000	2.402	1.343	6950	1.000	5.000	2.748	1.187
	Season	3475	1.000	4.000	2.516	1.106	6950	1.000	4.000	2.510	0.923
Step 1	*Wind speed*	3475	1.000	3.000	1.019	0.142	6950	1.000	3.000	1.485	0.637
	Visibility	3475	1.000	3.000	2.700	0.582	6950	1.000	3.000	2.333	0.695
	No. of pirates	3475	1.000	3.000	1.374	0.558	6950	1.000	3.000	1.666	0.652
	Weapon	3475	1.000	3.000	2.513	0.761	6950	1.000	3.000	2.245	0.711
	Attack	3475	1.000	1.000	1.000	0.000	6950	0.000	1.000	0.500	0.500
	Speed	3475	1.000	3.000	1.475	0.506	6124	1.000	3.000	1.701	0.598
	Time	3475	1.000	4.000	1.850	1.219	6124	1.000	4.000	2.161	1.099
	Location	3475	1.000	3.000	1.885	0.766	6124	1.000	3.000	1.946	0.694
Step 2	*Avoidance*	3475	0.000	1.000	0.167	0.373	6124	0.000	1.000	0.575	0.494
	Alarm	3475	0.000	1.000	0.564	0.496	6124	0.000	1.000	0.531	0.499
	Anti-piracy watch	3475	0.000	1.000	0.101	0.301	6124	0.000	1.000	0.274	0.446
	Naval support	3475	0.000	1.000	0.145	0.352	6124	0.000	1.000	0.293	0.445

ReCAAP	3475	0.000	1.000	0.003	0.051	6124	0.000	1.000	0.220	0.415
Other	3475	0.000	1.000	0.125	0.331	6124	0.000	1.000	0.285	0.451
Success	3475	0.000	1.000	0.881	0.324	6124	0.000	1.000	0.500	0.500

4.2 Model performance

In order to demonstrate the effectiveness of the data processing, this study compared the RF performance of different data sets. Table 4 gives the results. The RF model with data processed by the MC and GANs performs best. Therefore, we adopted the processed data by the MC and GANs in the following study.

To model piracy risks in Southeast Asia, we compared the performance of the RF model with multiple machine learning models in both of the two steps. Table 5 shows the performance of the classification results of these models. Among the seven models, the RF model shows the best performance, K-Nearest Neighbor (KNN), Decision Tree (DT), and Support Vector Machines (SVM) generate relatively good results. Naïve Bayes (NB), Logistic Regression (LR), and Support Vector Machines Cross Validation (SVMCV) show the worse performance compared to the rest models. One possible reason is that RF is an ensemble method based on bagging that can obtain a lower variance than any single method. In addition, the introduction of randomness makes the RF model less prone to overfitting and has good noise immunity.

Table 4 Model performance comparison with different data

		RF+MC+GAN	RF+GAN	RF only
Step 1	**Accuracy**	0.967	0.959	0.563
	Precision	0.968	0.967	0.563
	Recall	0.967	0.951	1.000
	F1-score	0.967	0.959	0.720
Step 2	**Accuracy**	0.935	0.928	0.746
	Precision	0.907	0.891	0.749
	Recall	0.975	0.974	0.994
	F1-score	0.940	0.931	0.854

Table 5 Model performance comparison after the MC and GAN

		RF	NB	KNN	LR	DT	SVM	SVMCV
Step 1	**Accuracy**	0.967	0.908	0.967	0.926	0.959	0.966	0.917
	Precision	0.968	0.882	0.967	0.914	0.957	0.953	0.875
	Recall	0.967	0.944	0.969	0.941	0.961	0.979	0.982
	F1-score	0.967	0.912	0.968	0.927	0.959	0.966	0.925
Step 2	**Accuracy**	0.935	0.842	0.932	0.899	0.932	0.922	0.917
	Precision	0.907	0.850	0.897	0.873	0.904	0.891	0.875
	Recall	0.975	0.847	0.981	0.944	0.972	0.969	0.982
	F1-score	0.940	0.848	0.937	0.907	0.937	0.928	0.925

4.3 Analysis of the importance of factors

In this section, we illustrated the importance of the influencing factors to piracy risks in the two-step process. The results are shown in Figure 9. The numbers in this figure indicate the importance of factors. The larger the number, the more important the factor is. Further, to determine the most dangerous state of each of the most influential factors (the results of feature importance not less than 0.03 in Figure 9), we created a total of 10 groups of 28 scenarios in both two steps, each group, respectively, investigating the impact of a different state of an influential factor, with the other factors' states unchanged. The piracy risks in these scenarios are shown in Table 6.

According to Figure 9 and Table 6, some interesting findings are identified. In Step 1, *wind speed*, *ship size*, *visibility*, *weapon*, *ship type*, *No. of pirates*, and *season* have an impact on the risk of piracy, while ship flag has little effect. Among all, *wind speed*, *ship size*, and *visibility* significantly affect pirate behaviors, and pirates are likely to choose small ships to attack, when there are mild winds and good visibility. Generally, small ships have low freeboards, and are thus easier to be boarded by pirates (Jin et al., 2019). In Southeast Asia, armed pirates usually approach ships in high-speed boats with the ultimate goal of stealing valuables (Jiang and Lu, 2020). The use of these high-speed boats is limited by wind and visibility. Generally, when the average wind speed is higher than 10.7 m/s, there is almost no pirate activity (Jiang and Lu, 2020). Similarly, when visibility is limited due to fog, snow, or heavy rainfall, the lower probability of piracy would be (Liu et al., 2021). On the contrary, *ship flag* has little impact, which seems to be inconsistent with some previous literature (Jin et al., 2019). In reality, pirates focus more on a particular type of ship, such as tankers that have a low freeboard (Mejia Jr et al., 2009).

In Step 2, *ReCAAP*, *avoidance*, *time*, *alarm*, *anti-piracy watch*, *navy support*, *other strategy,* and *location* influence the success of piracy, while *ship speed* has minor impact. Successful attacks are mostly influenced by *ReCAAP* and *time*. Ship operators reporting to the ReCAAP can decrease the probability of a successful attack. A possible reason is that the ReCAAP releases information on attacks through the Information Sharing Center (ISC) after receiving reports from the crew, which can help the nearest liaison office of a member country to rescue or support the attacked vessel in a timely manner. Besides, choosing to sail or anchor at night (between 0 to 6 a.m.) can lead to different risks. At night, the crews are perhaps slack, and their vigilance may be reduced. In addition, it is difficult for crews to spot the hidden pirates at night. Numerous cases also have shown that successful attacks usually occur at night in Southeast Asia. In contrast, *ship speed* only has a little impact, which is inconsistent with findings from Jiang and Lu (2020). According to an interview with a captain, if the ship is steaming, it may speed up and/or change direction when encountering pirates, and thus easier to get away from pirates. Therefore, in historical cases, there were fewer attacked ships steaming than at anchor. In fact, it is not the status of steaming itself that reduces the likelihood of a successful attack, but rather the counter-piracy measures at work, such as acceleration and redirection.

Through observing the findings from two steps analysis, the high-risk scenarios can be identified. In mild wind weather, the likelihood of a piracy attack on a ship with a DWT not exceeding 50,000 tons is up to 90%. Further, if the attack occurs between 0-6 a.m., the probability of success is over 90%. In reality, in these scenarios, most piracy incidents are thefts according to GISIS reports. Previous studies have shown that it is safe for ships to sail during the daytime in Southeast Asia (Jiang and Lu, 2020), we've found that, if ships are attacked by pirates, the likelihood of success in the daytime remains at a

high level of about 70%. Moreover, although large ships are less vulnerable to attack than small ships, the risk of piracy attacks on large ships exceeds 0.5.

Notably, the success of a pirate attack depends, to a large extent, on crew actions. For example, although small ships are more vulnerable, their crews can significantly reduce the piracy risk by taking anti-piracy measures, such as *ReCAAP* and *anti-piracy watch*. Furthermore, we find that the piracy risk is nearly zero when the crews adopt a comprehensive strategy, that includes *avoidance, alarm, anti-piracy watch, naval support, ReCAAP,* and *other measures*.

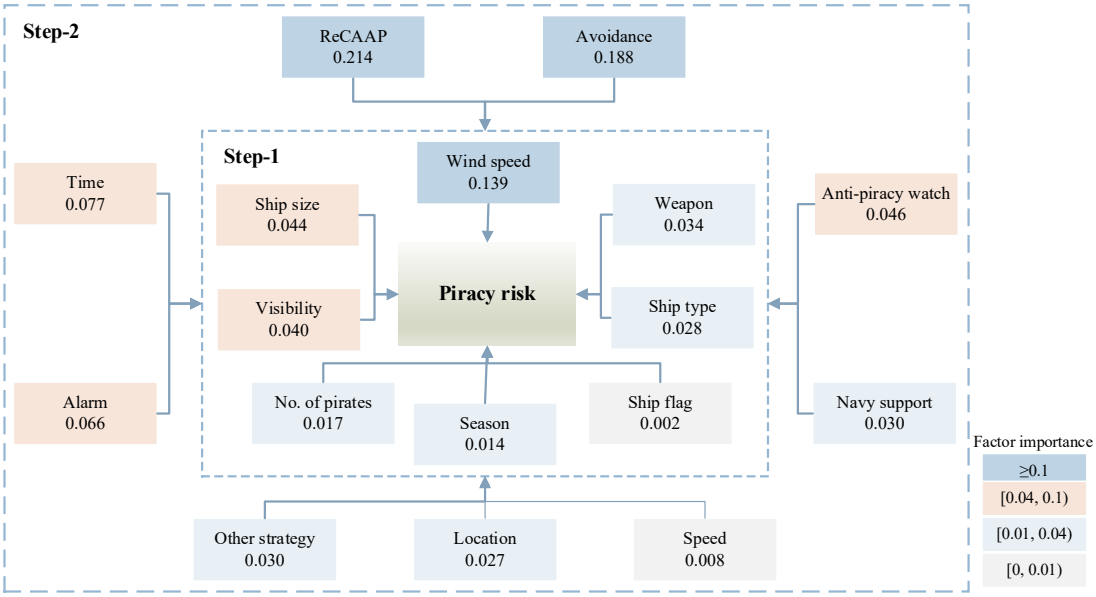

Figure 9 The results of factor importance to piracy risks

Table 6 The probabilities associated with different scenarios in Step 1 and 2

Step 1				Step 2			
Factor	State	Scenario	Probability	Factor	State	Scenario	Probability
Wind speed	Low	PS1	0.894	*ReCAAP*	Yes	SS1	0.013
	Moderate	PS2	0.010		No	SS2	0.536
	High	PS3	0.010	*Avoidance*	Yes	SS3	0.931
Ship size	(0, 15 000]	PS4	0.981		No	SS4	0.979
	(15 000, 20 000]	PS5	0.970	*Time*	[00:00,06:00)	SS5	0.931
	(20 000, 50 000]	PS6	0.993		[06:00-12:00)	SS6	0.691
	(50 000, 100 000]	PS7	0.156		[12:00-18:00)	SS7	0.708
	(100 000, +∞)	PS8	0.560		[18:00-24:00)	SS8	0.851
Visibility	Bad	PS9	0.364	*Alarm*	Yes	SS9	0.823
	Moderate	PS10	0.334		No	SS10	0.979
	Good	PS11	0.558	*Anti-piracy watch*	Yes	SS11	0.931
Weapon	Gun	PS12	0.990		No	SS12	0.946
	Knife	PS13	0.622	*Navy support*	Yes	SS13	0.817
	Other	PS14	0.989		No	SS14	0.945

4.4 Discussion and countermeasure suggestions

The findings of this study provide important implications for ship owners in preventing and resisting piracy. First, we suggest that the owners of small ships who regularly operate in Southeast Asia strengthen anti-piracy equipment in the ships and improve the anti-piracy consciousness of the crew, as these types of ships are most vulnerable to pirate attacks. The protection work, such as in the bulwarks, pilots, and living areas, should be well deployed (Jiang and Lu, 2020). Second, we recommend that the owners are better to pass through the waters of Southeast Asia during the daytime, as the success of piracy is much higher at night. Third, we advise the owners to organize anti-piracy drills to familiarize the crews with anti-piracy measures, such as avoidance, alarm, or navy support, as these measures can largely reduce the likelihood of a successful attack. Finally, based on our two-step risk estimation, various attentions should be paid for ships in different size under certain circumstance. We suggest that the owners of ships with a DWT not exceeding 50 000 tons should strengthen anti-piracy watch and surveillance capabilities when their ships sail in mild winds and with good visibility weather; during nighttime. As an efficient countermeasure, two crew groups can be arranged as anti-piracy patrols, one group at the bow and the other at the stern. Another group is assigned to conduct an anti-piracy watch. Notably, unlike tankers and bulk carriers, fully loaded container ships have large blind spots, such as the stern; thus, anti-piracy spotters are also needed in these areas. For the owners of ships with a DWT exceeding 100 000 tons, we recommend them to maintain anti-piracy watch in order to keep distance from suspicious boats and persons when in the wild wind and good visibility, as the risk of piracy attacks on these ships exceeds 0.5.

The findings also provide a reference for local authorities to formulate efficient policies. On the one hand, we suggest that they issue warnings to tankers that are anchoring in ports or sailing at night in Southeast Asian waters to enhance prevention, as these scenarios are subject to a high risk of piracy. On the other hand, we recommend that local authorities strengthen mutual cooperation and jointly combat piracy by participating in the ReCAAP, as the ReCAAP plays an important role in decreasing the success rate of attacks. Also, ReCAAP members should conduct joint exercises in dangerous waters to improve their joint search, rescue, and escort capabilities. Meanwhile, in order to combat piracy, we advise the local authorities to strengthen maritime escorts and patrols in hot spots frequented by pirates.

Finally, we believe more interesting findings and implications can be achieved if the two-step estimation method is applied to cases in other regions.

5 Conclusion and further research

This study presented a two-step framework for maritime piracy risk assessment, with Step 1 examining which ships are more likely to be attacked from the pirates' perspective and Step 2 analyzing which factors can lower the possibility of a successful attack, from ship operators' perspective. The combined risk from the two steps provides more specific implications to relative stakeholders in the shipping industry. We also address the problem of missing and imbalanced data when modeling this risk. The framework is then applied to the historical maritime piracy data in Southeast Asia. The findings suggest that tankers are more vulnerable to attack than container ships, and a successful attack is significantly influenced by time and anti-piracy measures. The combined probability of the two-step process identifies many high-risk scenarios, such as tankers being at ports in mild wind weather. Some of the

scenarios that were previously deemed safe are proven to still need further attention. This study provides more comprehensive information for ship operators to better manage piracy risks, and can also be helpful for local authorities to develop effective anti-piracy policies.

Finally, although our study provides an extensive understanding of piracy risks, still, there are some limitations. First, this study ignores the time-varying characteristics of piracy behaviors. For example, in the 1990s, most of the piracy incidents occurred in the South China Sea; in the past decade, however, the number of such incidents in the Malacca Strait has significantly increased. Therefore, piracy risk analysis taking into account the time-varying features of piracy behaviors, is worth exploring. Second, as the features of pirate attacks vary across regions, it is always a challenge to measure the applicability of a risk model, tweaked with the specific data of one area, to another region. This limitation in regional data is a common difficulty in pirate risk analysis. We expect to see more studies on the effect of piracy data aggregation in the future.

References

Ahmad, M., 2020. Maritime piracy operations: Some legal issues. Journal of International Maritime Safety, Environmental Affairs, and Shipping 4, 62–69. https://doi.org/10.1080/25725084.2020.1788200

Aziz, S.N.B.A., Kumar, A., Maurya, D., Khobragade, J.W., 2021. The Anti-Maritime Piracy Law in India and Malaysia: An Analytical Study. Journal of International Maritime Safety, Environmental Affairs, and Shipping 5. https://doi.org/10.1080/25725084.2021.2006462

Bateman, S., 2010. Maritime piracy in the Indo-Pacific region–ship vulnerability issues. Maritime Policy & Management 37, 737–751. https://doi.org/10.1080/03088839.2010.524739

Bouejla, A., Chaze, X., Guarnieri, F., Napoli, A., 2014. A Bayesian network to manage risks of maritime piracy against offshore oil fields. Safety Science 68, 222–230. https://doi.org/10.1016/j.ssci.2014.04.010

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brown, M.L., Kros, J.F., 2003. Data mining and the impact of missing data. Industrial Management & Data Systems 103, 611–621. https://doi.org/10.1108/02635570310497657

Cai, J.-F., Candès, E.J., Shen, Z., 2010. A Singular Value Thresholding Algorithm for Matrix Completion. SIAM J. Optim. 20, 1956–1982. https://doi.org/10.1137/080738970

Chen, Y., Zheng, W., Li, W., Huang, Y., 2021. Large group activity security risk assessment and risk early warning based on random forest algorithm. Pattern Recognition Letters 144, 1–5. https://doi.org/10.1016/j.patrec.2021.01.008

Cheng, L., Chen, X., De Vos, J., Lai, X., Witlox, F., 2019. Applying a random forest method approach to model travel mode choice behavior. Travel Behaviour and Society 14, 1–10. https://doi.org/10.1016/j.tbs.2018.09.002

Dabrowski, J.J., de Villiers, J.P., 2015. Maritime piracy situation modelling with dynamic Bayesian networks. Information Fusion 23, 116–130. https://doi.org/10.1016/j.inffus.2014.07.001

Di Salvatore, J., 2018. Does criminal violence spread? Contagion and counter-contagion mechanisms of piracy. Political Geography 66, 14–33. https://doi.org/10.1016/j.polgeo.2018.07.004

Escribano, Á., Wang, D., 2021. Mixed random forest, cointegration, and forecasting gasoline prices. International Journal of Forecasting 37, 1442–1462. https://doi.org/10.1016/j.ijforecast.2020.12.008

Flückiger, M., Ludwig, M., 2015. Economic shocks in the fisheries sector and maritime piracy. Journal of Development Economics 114, 107–125. https://doi.org/10.1016/j.jdeveco.2014.12.003

Fu, X., Ng, A.K.Y., Lau, Y.-Y., 2010. The impacts of maritime piracy on global economic development: the case of

Somalia. Maritime Policy & Management 37, 677–697. https://doi.org/10.1080/03088839.2010.524736

Ghosh, P., Neufeld, A., Sahoo, J.K., 2021. Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. arXiv:2004.10178 [cs, q-fin, stat].

Gilmer, B.V., Kane, B.E., 2019. Counter piracy programming and potential barriers to reintegrating Somali pirates: An African restorative justice critique. International Journal of Law, Crime and Justice 58, 12–21. https://doi.org/10.1016/j.ijlcj.2019.07.002

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets, in: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2, NIPS'14. MIT Press, Cambridge, MA, USA, pp. 2672–2680.

Gottlieb, Y., 2013. Combating Maritime Piracy: Inter-Disciplinary Cooperation and Information Sharing. Case W. Res. J. Int'l L. 46, 303.

Graham, J.W., Cumsille, P.E., Shevock, A.E., 2013. Methods for handling missing data, in: Handbook of Psychology: Research Methods in Psychology, Vol. 2, 2nd Ed. John Wiley & Sons, Inc., Hoboken, NJ, US, pp. 109–141.

Gui, J., Sun, Z., Wen, Y., Tao, D., Ye, J., 2021. A Review on Generative Adversarial Networks: Algorithms, Theory, and Applications. IEEE Transactions on Knowledge and Data Engineering 1–1. https://doi.org/10.1109/TKDE.2021.3130191

Guo, H., Li, Y., Shang, J., Gu, M., Huang, Y., Gong, B., 2017. Learning from class-imbalanced data: Review of methods and applications. Expert Systems with Applications 73, 220–239. https://doi.org/10.1016/j.eswa.2016.12.035

Hassan, D., Hasan, S.M., 2017. Origion, development and evolution of maritime piracy: A historical analysis. International Journal of Law, Crime and Justice 49, 1–9. https://doi.org/10.1016/j.ijlcj.2017.01.001

Hastings, J., 2020. The Return of Sophisticated Maritime Piracy to Southeast Asia. Pacific Affairs 93, 5–30. https://doi.org/10.5509/20209315

Hastings, J.V., 2009. Geographies of state failure and sophistication in maritime piracy hijackings. Political Geography 28, 213–223. https://doi.org/10.1016/j.polgeo.2009.05.006

Hidayati, N., Salleh, N.H.M., Harun, M., 2019. Maritime Terrorisms and Navigational Security in Sulu Sea ☆ 12, 42–054.

IMO, 2020. Global Integrated Shipping Information System (GISIS) [WWW Document]. URL https://gisis.imo.org (accessed 11.30.21).

Izquierdo-Verdiguier, E., Zurita-Milla, R., 2020. An evaluation of Guided Regularized Random Forest for classification and regression tasks in remote sensing. International Journal of Applied Earth Observation and Geoinformation 88, 102051. https://doi.org/10.1016/j.jag.2020.102051

Jiang, B., Ma, S., Causey, J., Qiao, L., Hardin, M.P., Bitts, I., Johnson, D., Zhang, S., Huang, X., 2016. SparRec: An effective matrix completion framework of missing data imputation for GWAS. Sci Rep 6, 35534. https://doi.org/10.1038/srep35534

Jiang, M., Lu, J., 2020. The analysis of maritime piracy occurred in Southeast Asia by using Bayesian network. Transportation Research Part E: Logistics and Transportation Review 139, 101965. https://doi.org/10.1016/j.tre.2020.101965

Jin, M., Shi, W., Lin, K.-C., Li, K.X., 2019. Marine piracy prediction and prevention: Policy implications. Marine Policy 108, 103528. https://doi.org/10.1016/j.marpol.2019.103528

Kao, M., 2019. Assessing Maritime Piracy in American Law: A Century-old Punishment for an Evolving Crime. The International Journal of Marine and Coastal Law 34, 1–23. https://doi.org/10.1163/15718085-23441084

Keramati, A., Lu, P., Iranitalab, A., Pan, D., Huang, Y., 2020. A crash severity analysis at highway-rail grade crossings:

The random survival forest method. Accident Analysis & Prevention 144, 105683. https://doi.org/10.1016/j.aap.2020.105683

Lewis, J.S., 2016. Maritime piracy confrontations across the globe: Can crew action shape the outcomes? Marine Policy 64, 116–122. https://doi.org/10.1016/j.marpol.2015.11.012

Liu, K., Yang, L., Li, M., 2021. Application of Cloud Model and Bayesian Network to Piracy Risk Assessment. Mathematical Problems in Engineering 2021, e6610339. https://doi.org/10.1155/2021/6610339

Liu, Z.-G., Pan, Q., Mercier, G., Dezert, J., 2015. A New Incomplete Pattern Classification Method Based on Evidential Reasoning. IEEE Transactions on Cybernetics 45, 635–646. https://doi.org/10.1109/TCYB.2014.2332037

Longadge, R., Dongre, S., 2013. Class Imbalance Problem in Data Mining Review. International Journal of Computer Science & Network 2.

Lu, J., Su, W., Jiang, M., Ji, Y., 2022. Severity prediction and risk assessment for non-traditional safety events in sea lanes based on a random forest approach. Ocean & Coastal Management 225, 106202. https://doi.org/10.1016/j.ocecoaman.2022.106202

Makariou, D., Barrieu, P., Chen, Y., 2020. A random forest based approach for predicting spreads in the primary catastrophe bond market. Papers, Papers.

Meher, P.K., Rao, A.R., Wahi, S.D., Thelma, B.K., 2014. An approach using random forest methodology for disease risk prediction using imbalanced case–control data in GWAS. Current Medicine Research and Practice 4, 289–294. https://doi.org/10.1016/j.cmrp.2014.11.011

Mejia Jr, M.Q., Cariou, P., Wolff, F.-C., 2009. Is maritime piracy random? Applied Economics Letters 16, 891–895. https://doi.org/10.1080/13504850701222186

Mo, J., 2002. Options to Combat Maritime Piracy in Southeast Asia. Ocean Development and International Law - OCEAN DEV INT LAW 33. https://doi.org/10.1080/00908320290054819

Morabito, G., Sergi, B.S., 2018. HOW DID MARITIME PIRACY AFFECT TRADE IN SOUTHEAST ASIA? Journal of East Asian Studies 18, 255–265. https://doi.org/10.1017/jea.2018.5

Pristrom, S., Yang, Z., Wang, J., Yan, X., 2016. A novel flexible model for piracy and robbery assessment of merchant ship operations. Reliability Engineering & System Safety 155, 196–211. https://doi.org/10.1016/j.ress.2016.07.001

Ren, C., Xu, Y., 2019. A Fully Data-Driven Method Based on Generative Adversarial Networks for Power System Dynamic Security Assessment With Missing Data. IEEE Transactions on Power Systems 34, 5044–5052. https://doi.org/10.1109/TPWRS.2019.2922671

Robitaille, M.-C., 2019. Maritime Piracy and International Trade. Defence and Peace Economics 31, 1–18. https://doi.org/10.1080/10242694.2019.1627511

Saeys, Y., Inza, I., Larrañaga, P., 2007. A review of feature selection techniques in bioinformatics. Bioinformatics 23, 2507–2517. https://doi.org/10.1093/bioinformatics/btm344

Shane, J., Magnuson, S., 2014. Successful and Unsuccessful Pirate Attacks Worldwide: A Situational Analysis. Justice Quarterly 33, 1–26. https://doi.org/10.1080/07418825.2014.958187

Shane, J., Piza, E., Silva, J., 2018. Piracy for ransom: the implications for situational crime prevention. Security Journal 31. https://doi.org/10.1057/s41284-017-0115-0

Shepard, J.U., Pratson, L.F., 2020a. Maritime piracy in the Strait of Hormuz and implications of energy export security. Energy Policy 140, 111379. https://doi.org/10.1016/j.enpol.2020.111379

Shepard, J.U., Pratson, L.F., 2020b. Maritime piracy in the Strait of Hormuz and implications of energy export security. Energy Policy 140, 111379. https://doi.org/10.1016/j.enpol.2020.111379

Sumaila, U.R., Bawumia, M., 2014. Fisheries, ecosystem justice and piracy: A case study of Somalia. Fisheries Research 157, 154–163. https://doi.org/10.1016/j.fishres.2014.04.009

Tominaga, Y., 2018. Exploring the economic motivation of maritime piracy. Defence and Peace Economics 29, 383–406.

United Nations, 2020. World Economic Situation and Prospects 2019 Report.

Wong, M., Yip, T.L., 2012. Maritime piracy: An analysis of attacks and violence. Int. J. of Shipping and Transport Logistics 4, ng and Transport Logistics. https://doi.org/10.1504/IJSTL.2012.049315

Yang, Z.L., Wang, J., Li, K.X., 2013. Maritime safety analysis in retrospect. Maritime Policy & Management 40, 261–277. https://doi.org/10.1080/03088839.2013.782952

Yen, S.-J., Lee, Y.-S., 2009. Cluster-based under-sampling approaches for imbalanced data distributions. Expert Systems with Applications 36, 5718–5727. https://doi.org/10.1016/j.eswa.2008.06.108

Zhai, J., Qi, J., Shen, C., 2022. Binary imbalanced data classification based on diversity oversampling by generative models. Information Sciences 585, 313–343. https://doi.org/10.1016/j.ins.2021.11.058

Zhang, Y., Li, H., Ren, G., 2022. Estimating heterogeneous treatment effects in road safety analysis using generalized random forests. Accident Analysis & Prevention 165, 106507. https://doi.org/10.1016/j.aap.2021.106507