Transportation Research Part E 202 (2025) 104331

Contents lists available at ScienceDirect

Transportation Research Part E

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

Predicting out-terminals for imported containers at seaports using machine learning: Incorporating unstructured data and measuring operational costs due to misclassifications

Ying Xie [a], Dong-Ping Song [b,*], Jingxin Dong [c], Yuanjun Feng [b]

[a] Cranfield School of Management, Cranfield University, UK
[b] School of Management, University of Liverpool, UK
[c] Business School, Newcastle University, UK

ARTICLE INFO

Keywords:
Seaport
Container classification
Predictive model
Feature engineering
Explanatory machine learning
Cost analysis

ABSTRACT

Persistent bottlenecks at container ports have significantly disrupted global supply chains, necessitating more efficient operations at seaports to address yard density and port congestion. An untapped but potentially critical approach to mitigating these challenges is to leverage container characteristics and machine learning to predict the out-terminals of containers upon their discharge from vessels. The predicted results can then guide the development of a more effective container storage strategy. To formulate such a strategy, this research developed a data-enabled methodological framework that integrates four key components: 1) Utilization of structured and unstructured data to enhance prediction accuracy. 2) Practice and knowledge-informed feature engineering to construct relevant features for the machine learning models. 3) Explanatory machine learning based classification models to understand the factors influencing terminal predictions. 4) Model-induced cost analysis to capture the monetary value of the prediction model including assessing the cost implications of misclassifications. An empirical study conducted at a seaport shows that our framework yields cost savings ranging from 14.90% to 30.45% compared to the Business-as-Usual scenario. Incorporating unstructured data as an additional feature in the machine learning models improves prediction performance by up to 6%. Moreover, integrating this framework into the existing operational system poses minimal risk and can be seamlessly executed. Additionally, the proposed methodological framework and its four components has broad applications beyond the shipping industry.

1. Introduction

1.1. Motivation

Container shipping carries over 50 % of world seaborne trade by value (Lee and Song 2017) with container ports serving as pivotal connectors between seaborne and inland transport networks. However, increasing trade volume, disruptive events, and coordination challenges among stakeholders frequently lead to port congestions, disrupting global supply chains (Watkins, 2021). To address this

* Corresponding author.
E-mail addresses: Ying.Xie@cranfield.ac.uk (Y. Xie), Dongping.song@liverpool.ac.uk (D.-P. Song), Jingxin.Dong@newcastle.ac.uk (J. Dong), Yuanjun.Feng2@liverpool.ac.uk (Y. Feng).

https://doi.org/10.1016/j.tre.2025.104331
Received 28 September 2024; Received in revised form 19 February 2025; Accepted 21 June 2025
Available online 24 July 2025
1366-5545/

issue, several strategies can be employed including expanding port capacity, enhancing supply chain coordination and operational efficiency. From the perspective of port or terminal operators, one of the most readily available approaches is to improve the operational efficiency, especially in yard management (Hu et al. 2021), quayside operations (Al-Dhaheri, et al., 2016), and intermodal connectivity.

Effective yard management plays a crucial role in determining the efficiency of a seaport terminal (Caserta et al. 2020). Storage yards can easily become bottlenecks as they serve as nodes connecting seaside vessels and landside trucks and trains. Specifically, in yard management, assigning incoming containers to appropriate storage yards is a critical tactical decision, especially for multi-terminal ports (Lee et al. 2012). This allocation lays the foundation for subsequent yard block management at operational level and influences the costs and efficiencies of onward transport (Zhou et al. 2020).

After being discharged from vessels, imported containers are usually stored at storage yards for several days. Afterwards, they are collected by inland shippers and transported using road or rail via one of the multiple out-terminals. An *out-terminal for imported containers* at a seaport refers to the designated point (e.g., gate, rail terminal, or barge terminal) where containers exit the port area. A single inland transport mode (road or rail) may have multiple out-terminals (e.g. multiple rail terminals), especially for large container ports. The inland transport modes and specific out-terminals for collecting imported containers are typically unknown to port operators when the containers are discharged from vessels. The decision on when these containers will leave the port for inland destinations and whether they will travel by road or rail is made by shippers rather than the terminal operators.

The absence of information regarding imported containers' out-terminals can be attributed to the complexity of the container shipping supply chain. In the context of international shipping, the buyer, seller, or their agent establishes a haulage contract with an ocean carrier, outlining the specifics of shipment transportation. There are two types of haulage contracts—merchant haulage and carrier haulage—which define the parties' responsibility for arranging inland transport for import container collection. Under Merchant Haulage Contract, a merchant (e.g., cargo owner, freight forwarder, or agent) arranges the inland transport of import containers. The merchant has decision-making authority and determines the timing, mode of inland transport, and rail service for retrieving containers from the port. These decisions are often made after the containers have been stored at the port to accommodate the high uncertainty in international logistics. As a result, the out-terminal for import containers is often undetermined at the time of unloading from vessels. In continental Europe, merchant haulage dominates the market, accounting for over 70 % of the market share (Wagener 2014; Van den Berg & De Langen 2015; Legros et al. 2019; Gumuskaya et al. 2020). Under Carrier Haulage Contract, ocean carriers are responsible for coordinating inland transport and determining the next mode of transport. Ocean carriers often need to outsource inland transport to local operators; however, specific operational details are typically excluded from their contracts. For instance, the contracts may not specify which platform or siding a rail operator will use to deliver containers. Further, major container ports often have multiple rail (or barge) out-terminals available. These factors make it difficult for ocean carriers to specify the exact out-terminal for a given container in advance. As a result, the information about the exact out-terminal does not exist in either paper or electronic form.

It makes practical sense that, logistics decisions related to when imported containers will be picked up and from which out-terminal are made after the containers have been stored at the port to accommodate the high uncertainty in international logistics. This phenomenon reflects *delayed dynamic decision-making* on out-terminals, driven by uncertainty and fragmentation in the container transport chain (Song 2021a; Song 2021b).

This phenomenon resembles that faced by air passengers who remain unaware of their flight gates upon arrival at the airport. Flight gate information is typically announced within 30 min of the flight departure time. Similarly, rail stations (e.g. Euston Rail station and Kings Cross station in London) announce train platforms approximately 10 min before the train departure time. In our case, the specific out-terminal through which individual containers will pass remains unknown upon unloading from vessels. Cargo owners or their agents gradually reveal this information near the time of pickup. Currently, the industrial practice in container ports is to stack discharged containers in storage yards without considering their out-terminals (van Marle 2022).

Therefore, an untapped but potentially critical approach to improving yard management is to predict the out-terminals of containers upon their discharge from vessels so that they could be stacked in storage yards that are close to their out-terminals in the first place. Several benefits can be realised: (i) Reducing "rail misses". Rail misses refer to the cases that imported containers miss the rail

Table 1
A sample list of seaports with multiple rail out-terminals.

Seaports	Number of rail out-terminals
Felixstowe (UK)	3 (North, South and Central)
Southampton (UK)	2 (Maritime Terminal, Solent Rail Terminal)
Rotterdam (Netherlands)	9 (9 out of 14 container terminals have rail facilities)
Antwerp (Belgium)	6 (6 container terminals all have rail facilities)
Hamburg (Germany)	4 (4 container terminals all have rail facilities)
Valencia (Spain)	2 (2 out of 3 container terminals have rail facilities)
Bremerhaven (Germany)	2 (2 out of 3 container terminals have rail facilities; One rail terminal serves destinations throughout Germany and Hungary; the other rail terminal serves Italy and across the Alps to central Europe)
Los Angeles (US)	5 on-dock railyards and one near-dock railyard
Long Beach (US)	5 (5 out of 6 container terminals are built with on-dock rail yards)
New York and New Jersey (US)	4 rail terminals (Elizabeth, Newark, Jersey and Staten Island)
Savannah (US)	2 on-dock railyards (Chatham Intermodal Container Transfer Facility and Mason Mega Rail)

services that operate on fixed departure schedules. Rail misses are costly because they not only underutilise rail service capacity but also disrupt shippers and their customers. Rail misses are particularly problematic in container terminals with multiple rail terminals and wide storage yards. (ii) Minimising unproductive movements. By strategically placing containers near their out-terminals, we can reduce unproductive container reshuffles, internal vehicle movements, and yard crane movements. These factors significantly impact yard efficiency.

1.2. Aim and challenges

This study aims to predict the out-terminals of imported containers upon discharge from vessels and develop easy-to-implement container storage strategies to enhance terminal operations efficiency. Although the out-terminals of imported containers are uncertain, they are not entirely random. We propose a data-enabled approach that leverages machine learning to learn the pattern of out-terminal allocation. By translating these findings into actionable knowledge, we can create effective easy-to-implement strategies.

The problem encounters several challenges. **Firstly**, accurately predicting the out-terminals of imported containers poses a significant hurdle. In practice, large container seaports globally often have multiple rail terminals, as summarised in Table 1. The information of the inland transport mode **(**if known to the port operator in rare cases or obtained by prediction) does not imply the out-terminals of the imported containers especially if the port has multiple rail out-terminals.

Container ports have accumulated extensive operational and empirical data from systems like the Terminal Operating System (TOS). The first question arises: how can the data available to a port, including both structured data and unstructured data, be used to predict the out-terminals of imported containers? Although there may not be a direct causal relationship between container attributes and their out-terminals, a possible association is believed to exist. For instance, the case study carried out in the research involving a port in West Europe revealed potential dependencies between container cargo contents and the preferred mode of inland transport. Another empirical example demonstrated that imported containers with heavy and dense freight (such as furniture and home appliances) tended to use rail services in New York-New Jersey ports (Freightwaves, 2023). In addition, in the context of the Port of Bremerhaven, which has three container terminals and two rail terminals, one rail terminal facilitates shipments to Germany and Hungary, while the other serves regions in Italy and central Europe via the Alps (Eurogate, 2023). This scenario implies the choice of rail out-terminals is correlated with container destinations. Therefore, the second question emerges: how can we identify the crucial features among all available variables and enhance the interpretability of the developed predictive model, including feature contributions.

Secondly, how the results via machine learning can be turned into actionable knowledge and easy-to-implement strategy for container storage? Given the conservative and risk-averse nature of the maritime industry (Song 2021b; Chen and Shen 2022), it is essential to deploy the model and knowledge with minimal disruption to the existing operations. Ports are extremely cautious when deploying new solutions in fear of the potential disruption. For example, when the Port of Felixstowe introduced the Hutchison Ports' modular terminal-management platform 'nGen' to replace the old TOS in 2018, it caused severe delays. Truck drivers experienced tailbacks of up to 16 h (Port Technology, 2018).

Thirdly, can we quantify the impact of the data-enabled predictive model on port operations? Predicting out-terminals of imported containers upon unloading from vessels is a novel research area with no practical applications. To convince the port industry to implement predictive models, we need to validate the predictive models and evaluate the potential costs induced by prediction errors. No prediction model can achieve 100 % accuracy, so misclassifications costs (i.e. predict the imported containers' out-terminal incorrectly) must be explicitly evaluated.

1.3. Contributions of the paper

Our paper makes three primary contributions. **First**, we are the first to demonstrate the feasibility of developing a data-enabled predictive model for containers' out-terminals upon their discharge from vessels. This untapped area could lead to critical strategies for improving operational efficiency at seaports and addressing port congestion challenges. As a result, this study could open a new avenue to tackle port inefficiency and congestion problems by integrating the out-terminal predictive models with other operational decisions such as quayside, yard and landside resource allocation and scheduling. **Second**, this study develops a novel methodological framework that integrates four key components: utilising both structured and unstructured data; practice and knowledge-informed feature engineering; explanatory machine learning-based classification models; and model-induced cost analysis. Incorporating unstructured data into machine learning models improves prediction performance. The framework identifies the pivotal factors affecting out-terminal assignments, which can empower ports to invest wisely in areas that yield the greatest improvements in prediction accuracy. These important factors can also be integrated with the port's existing rules to further enhance container storage. Furthermore, the proposed methodological framework has broad applications beyond the port industry. **Third**, we design an easy-to-implement strategy to deploy our predictive model in practice, providing step-by-step guidance on container movement and storage. The empirical study conducted at the seaport shows that the predictive model can yield cost savings ranging from 14.90 % to 30.45 % compared to the Business-as-Usual scenario. Integrating the developed predictive model into the existing operational system of port poses very low risk and can be seamlessly executed.

The rest of the paper is organised as follows. Section 2 reviews relevant literature and identifies the research gaps. Section 3 outlines the problem in a real business context, discusses empirical data and constructs various business scenarios. Section 4 illustrates the methodological framework for addressing research questions. Section 5 applies machine learning-based models to predict out-terminals. Section 6 interprets these models by quantifying feature contributions. Section 7 analyses the benefits based on the prediction outcomes, including the misclassification costs. Section 8 discusses the deployment and generalisability of the methodological framework and articulates research and managerial contributions. Section 9 concludes the study and offers direction for future research.

2. Literature review

In this section, we review two streams of relevant research: one involving the application of predictive machine learning at container ports/terminals and the other concerning the flow and storage of imported containers at yards.

2.1. Predictive machine learning models developed for container ports

The data-enabled approach has enabled port/terminal operators to utilize their data to extract actionable knowledge for better operations and services. Predictive machine learning models have been developed to forecast container throughput at ports (see Filom et al. 2022) and to assist landside and seaside operations (Jahangard et al. 2025).

Landside operations: Machine learning models for landside operations can be broadly classified into models assisting with yard planning and those predicting terminal traffic flows. In yard planning, container dwell time prediction appears to be one of the most prominent research topics. Using Naïve Bayes, decision tree and the hybrid model of Naïve Baye-decision tree, Moini et al. (2012) estimated container dwell time using several factors, including container status (empty or full), container type and size, truck carrier, ocean carrier, vessel name and arrival and departure time. They also measured how changes in these factors impact dwell time, yard capacity and terminal revenue. Kourounioti et al. (2016) applied neural network to predict container dwell times. The results demonstrated that the day and month of discharge, the port of origin and container type and size are determinants of import container dwell time. In addition to dwell time, Kang et al. (2006a, 2006b) applied decision tree models to estimate container weights and to derive stacking strategies. Terminal traffic flows affect scheduling and routing of equipment such as trucks and yard cranes. Huynh and Hutson (2008) applied a decision tree technique to identify the key factors causing delays to external trucks at container terminals. Li et al. (2022) proposed a hybrid model combining gated recurrent units and fully connected neural network to predict truck arrival times, incorporating variables like weather, weekday, ship arrival, and departure times.

Seaside operations: Vessel navigation, berthing and unberthing affect subsequent landside operations and their efficiency. Machine learning models have been constructed to estimate vessel arrival time (Yu et al. 2018; Abdi and Amrit 2024), vessel berthing time (Bakar et al. 2022), and vessel detention duration (Yang et al. 2023). Abdi and Amrit (2024) combined conventional neural networks, long short-term memory, and an attention mechanism while integrating Automatic Identification System data, vessel information, and weather conditions to enhance vessel arrival time prediction. Bakar et al. (2022) used multiple machine learning algorithms to predict ship berthing times and estimate cold ironing power consumption. Neural networks delivered the best performance, utilising input features such as arrival time, ship type, ship size, operation mode, and ship capacity. Yang et al. (2023) introduced a novel machine learning model to help ports estimating ship detention durations and reducing substandard vessels. The model utilizes an improved tree Augmented Naïve learning approach combined with a maximum posteriori probability Expectation Maximisation method. It serves as a tool to predict ship detention durations and identifies key risk factors and deficiencies contributing to prolonged detentions.

Integration of machine learning with prescriptive models: Beyond predictions, machine learning models have been integrated with prescriptive models to improve yard management. For instance, Maldonado et al. (2019) addressed the yard stacking problem via a two-step strategy. They predicted dwell time for each container using a Random Forest classification model. The prediction was then used as an input for a mathematical programming that minimises container rehandles heuristically. Zhang et al. (2020) integrated machine learning techniques with a branch-and-bound algorithm to address the container relocation problem at seaport terminals, demonstrating the superiority of the proposed algorithm against the existing algorithms.

Despite the proven use of machine learning models in improving port operations, no studies have focused on predicting out-terminals for containers.

2.2. Flow and storage of imported containers at shipping yards

The process of moving and storing containers to storage yards has been studied at different levels according to the storage space unit considered, including yard section, block, sub-block, bay, and slot (Jin et al. 2016).

At the bay and slot level, a specific location within a block is selected to store a container, with key criteria focusing on minimising container relocations, truck waiting time and yard crane movements (e.g., Yu and Qi, 2013; Maldonado et al. 2019; Feng et al., 2022b).

At the yard section, block, and sub-block allocation level, the flow directions/destinations of handling containers are critical for space allocation decisions, as they directly influence moving distances. At this level, key factors to be considered include the travel distances between the storage yard and quayside, yard crane workloads, block space availability, and yard density across yard area.

For yard space allocation at the block and sub-block level for import containers, several studies have considered multiple objectives to coordinate performances of various terminal resources (e.g., Zhang et al. 2003; Guldogan, 2011; Sharif and Huynh, 2013; He et al. 2022). Some research has specifically evaluated performance based on travel distance from block to gate. For instance, Rekik et al. (2018) incorporated block-to-gate distance and stack-to-gate distances into a weighted sum of four criteria to assess storage position efficiency, but they focused solely on the gate out-mode. In addition, a recent study on block allocation for inbound, outbound and transshipment containers (Feng et al., 2022a) considered multiple factors, including truck travel times from-ship-to-block and block-to-gate, as well as yard crane waiting and service times. However, none of the existing studies considered multiple out-modes or multiple out-terminals.

At the yard section level, storage space is allocated to a broader area, where a yard section comprises several neighbouring yard blocks. Research at this level has primarily focused on managing transshipment container flows, with the main objective being to minimise total moving distances and costs between the quayside and the yard (Lee et al. 2012; Lee and Jin, 2013). Especially, Lee and Jin (2013) examined terminal allocation for vessels and yard allocation and reallocation to manage inter-terminal and intra-terminal container flows. Their study considered the duration-of-stay of transshipment containers at a multi-terminal transshipment port to minimise handling costs incurred by reallocation. On the one hand, to speed up vessel loading, containers stored far from the berthing position of the outbound vessel must be moved closer to within the same terminal before loading, incurring intra-terminal handling cost. On the other hand, at a multiple-terminal transshipment port, containers often need to be moved between terminals due to mismatches in inbound and outbound vessel terminals, causing inter-terminal handling cost. A recent study explored block allocation for transshipment containers, treating future berth locations of loading vessels as known parameters (Abouelrous et al. 2025). Their goal was to balance discharge and loading times without significantly delaying either operation.

It is worth noting that the out-terminals for transshipment containers, which are seaport terminals, are determined by the visiting terminals and berthing positions of their outbound vessels. This information (Abouelrous et al. 2025) is either predefined, or can be determined by the terminal operators (Lee et al., 2012; Lee and Jin, 2013). In contrast, import containers have out-terminals such as gates or rail terminals, which remain unknown at the time of vessel discharge and are beyond the control of terminal operators.

The literature review highlights the importance of out-terminal information in container flow and storage management. However, no prior research has incorporated such information into the management of import containers at ports with multi-out-terminals. Since out-terminal information does not exist when import containers are unloaded, we propose a data-enabled approach to build a predictive model. The knowledge generated from the predictive model can then be utilized by port operators to improve yard space allocation by placing import containers close to their out-terminals. Given that most large-scale container ports around the world offer multiple out-modes and multiple rail out-terminals (see Table 1), accurately predicting these out-terminals is crucial for optimising yard operations and reducing port congestion. Furthermore, as the focus on transport decarbonisation grows, rail and barge container movements are expected to gain a higher market share in the coming years.

3. Problem description, empirical data and scenario modelling

3.1. Problem description in case context

Motivated by a Western European seaport, referred to as *Port A*, the problem is to predict out-terminals for imported containers using a substantial set of container attributes that are available when containers are discharged from vessels. Fig. 1 shows the schematic terminal map at *Port A*, illustrating two haulier out-terminals and two rail out-terminals linked to two storage yards. Haulier out-terminal H0 and rail out-terminal R0 are located close to Yard 0, while out-terminals H1 and R1 are near Yard 1. The maritime quayside, where vessels dock and containers are discharged, is located near Yard 0. Yard 1 is situated further away from the maritime quayside. In the current practice, discharged containers are primarily stored in a yard based on their proximity to the quayside berth location and yard density without considering the future out-terminals of individual containers.

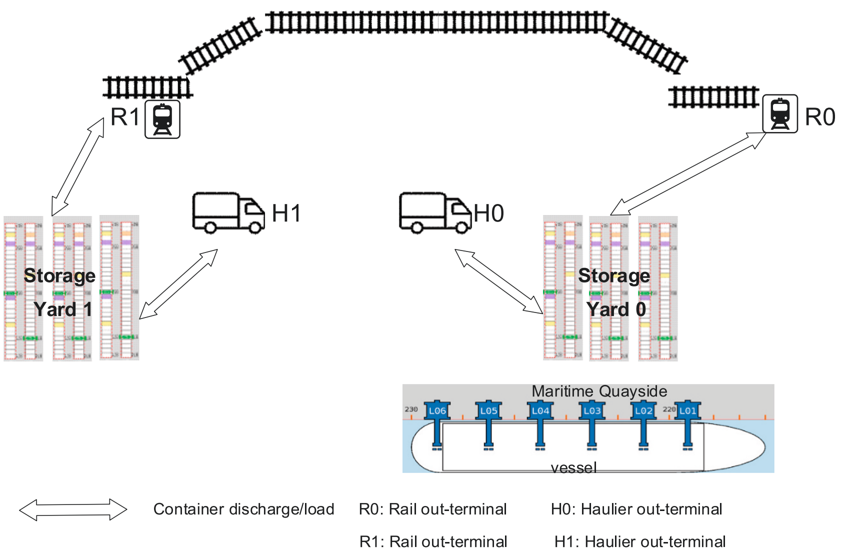

Fig. 1. The geographic location of the out-terminals in the case port.

3.2. Empirical data

Table A1 in the Appendix displays a sample of imported container dataset, encompassing their attributes, actual out-terminal (column *out_terminal*), actual out-mode (column *out_mode*), and randomly assigned storage yards (column *terminal_id*). Detailed definitions of container attributes are provided in Table 2 (also refer to Table A2 in the Appendix).

3.3. Scenario modelling

Based on the input from the operations team of *Port A*, we defined their current practices as the Business-as-Usual scenario. Fig. 2 illustrates container movements between yards and out-terminals, with notations for these movements detailed in Table 3. In Fig. 2, the arrows along with the corresponding notations, such as D_{00} and D_{01}, denote the actual number of containers transported from the quayside and stored at Yard 0 and Yard 1, respectively. Notations of D^h_{00}, D^r_{00}, and so on., represent the actual number of containers moved from yards to corresponding haulier or rail out-terminals.

Under the Business-as-Usual scenario, without container information on inland transportation mode and associated out-terminal, 70 % of the discharged containers are randomly selected and stored in the nearest Yard 0, noted as D_{00}. The remaining 30 % of containers, also considered as randomly assigned, are transported to and stored at Yard 1, denoted as D_{01}. The 70:30 split is mainly determined by the yard density and resource allocation across the container terminals. Depending on the transportation mode chosen by their owners, containers can be retrieved from one of the two haulier out-terminals H0 and H1, or collected by rail from one of the two rail terminals, R0 or R1. The determination of haulier out-terminal lies with the terminal operator, whereas the choice of rail out-terminal rests with the rail operator. That implies two key points: (i) haulier containers are generally picked up from the haulier out-terminal situated close to their storage yard; and (ii) rail containers must be retrieved from the rail out-terminal designated by the rail operator. Ideally, rail containers are stored near the intended rail out-terminal, but they may be placed in a distant yard, requiring

Table 2
Features of containers and feature selection process.

Features	Notes	Values and counts	Filter methods	Boruta algorithm	Selected features
1. cntr_height_in_feet	container height	{4.25, 6, 8, 9, 9.5, 9.54}, 6 unique values	✓	✓	✓
2. cntr_length_in_feet	container length	{20, 30, 40, 45}, 4 unique values	✓	✓	✓
3. cntr_width_in_feet	container width	{8, 8.16} 8: 601713, 8.16: 80	✖	✖	✖
4. full_empty_indr	container is full or empty	{F, E} F(Full), E(Empty), 2 unique values	✖	✖	✖
5. owner	owner of the container	$Z = [0, 1, \ldots, 20]$, 21 owners	✓	✓	✓
6. SOA	SOA, owner of the vessel	$Z = [\ [0, 1, .., 13]$, 14 codes	✖	✖	✖
7. service_code	service code	$Z = [0, 1, \ldots, 9]$, 10 service codes	✓	✓	✓
8. voyid_vessel_code	vessel code	$Z = [0, 105]$, 106 codes	✓	✓	✓
9. voyid_voyage_code	voyage code	[68444, 71025,…, 85636], 219 codes	✓	✓	✓
10. origin_port	original port	[0, 1, …, 151], 152 port names	✓	✓	✓
11. load_port	loading port	[0, 1, …, 39], 40 port names	✓	✓	✓
12. gross_weight_documented	container gross weight in kg	[1950, 2228, …, 45760], 28595values	✓	✓	✓
13. act_out_owner	actual owner of the container	[0, 1, 2, …, 20], 21 owner values	✓	✓	✓
14. gross_weight_measured_yc	container gross weight	[0, 3600, …, 30100], 86 values,	✖	✖	✖
15. gross_weight_measured_wb	information of container gross weight automatically generated by the database	[-1, 0] 2 values	✓	✖	✖
16. general_cargo_content	container content	98,271 types of content	✓	✓	✖
17. reefer	refrigerated container	{0, 1}, 0 = not refrigerated, 1 = refrigerated	✓	✓	✓
18. out_of_gauge	oversized container	{0, 1}, 0 = regular size, 1 = oversize	✖	✖	✖
19. dangerous	dangerous content	{0, 1}, 0 = not dangerous, 1 = dangerous	✓	✓	✓
20. cms_no	container number	502,852 individual numbers	✓	✓	✓
21. SITC*	SITC* classification of container content	[1, 2, …,2970], 2970 classes of content			✓

* Standard International Trade Classification

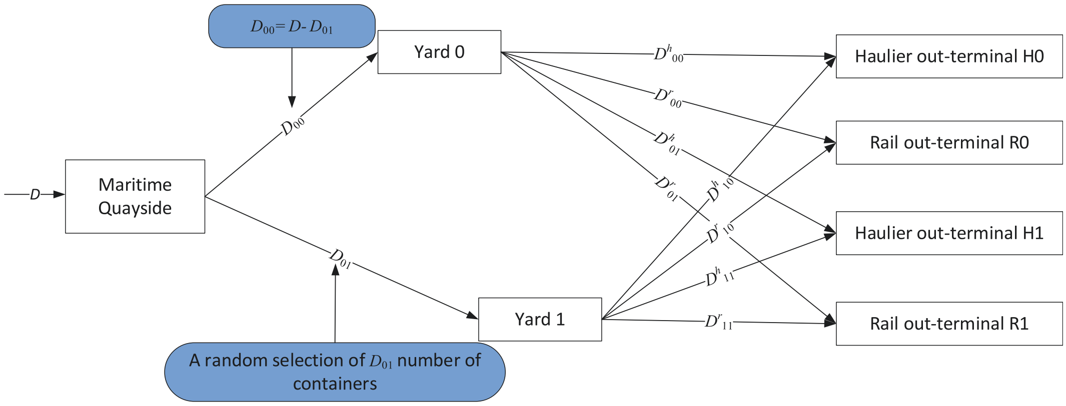

Fig. 2. Container movements under the Business-as-Usual scenario.

Table 3
Container movement notations.

Notations	Descriptions
D	the total number of containers unloaded at maritime quayside
D_{00}	the number of containers moved from maritime quayside and stored at Yard 0
D_{01}	the number of containers moved from maritime quayside and stored at Yard 1
D_{00}^h	the number of containers moved from Yard 0 to haulier out-terminal H0
D_{00}^r	the number of containers moved from Yard 0 to rail out-terminal R0
D_{01}^r	the number of containers moved from Yard 0 to rail out-terminal R1
D_{01}^h	the number of containers moved from Yard 0 to haulier out-terminal H1
D_{10}^h	the number of containers moved from Yard 1 to haulier out-terminal H0
D_{10}^r	the number of containers moved from Yard 1 to rail out-terminal R0
D_{11}^r	the number of containers moved from Yard 1 to rail out-terminal R1
D_{11}^h	the number of containers moved from Yard 1 to haulier out-terminal H1

transportation over longer distances. For example, if rail containers discharged at the maritime quayside are misplaced in Yard 1 but are supposed to be collected from rail out-terminal R0, moving these containers from Yard 1 to R0 would incur cross-yard transportation costs.

Although the terminal operator decides on the haulier out-terminals for imported containers once informed about the road mode of inland transport, cross-yard transportation of haulier containers still occur in current practices at *Port A*. In 2020, approximately 0.2 % of imported containers initially stored at Yard 1 were later moved to H0 and picked up by the haulier. Conversely, nearly 2 % of imported containers initially stored at Yard 0, were transported to H1 and collected by the haulier. This unnecessary behaviour may be attributed to road hauliers mistakenly accessing the port via incorrect gates, leading to cross-yard transportation for haulier containers. We will later show that eliminating cross-yard transportation for haulier containers would result in approximately a 1 % cost-savings (as detailed at the end of Section 7.1).

These current practices have imposed two direct challenges on the port operations: 1) The substantial distance between the two storage yards and cross-yard movements increase the chances of missing rail services with fixed departure schedules, commonly referred to as “rail misses”. 2) Complicated container yard management arises from unproductive container reshuffles, internal vehicle movements, and yard crane movements. Classifying containers by out-terminals and ideally storing containers near their designated out-terminals upon discharge from vessels can reduce unproductive crane movements and vehicle travel. This strategy improves yard management efficiency and streamlines the internal movement of the discharged containers.

Predicting the target out-terminals of imported containers was formulated as a traditional multiclass classification task, where each out-terminal was treated as a class label. The dataset of 668,080 imported containers recorded in the year of 2020 was retrieved from

Table 4
Characteristics of datasets.

Dataset	Number of containers per class
Raw dataset 668,080	**Haulier:** 492,500, including **H0:** 339,600, **H1:** 152,900 **Rail:** 175,580, including **R0:** 36,293, **R1:** 139,287
Pre-processed dataset 628,114	**Haulier:** 460,985, including **H0:** 317,892 **H1:** 143,093 **Rail:** 167,129, including **R0:** 34,272 and **R1:** 132,857
Dataset with SITC classification similarity over 0: 627,886	**Haulier:** 460,832, including **H0:** 317,784, **H1:** 143,048 **Rail:** 167,054, **R0:** 34,253, **R1:** 132,801

the TOS at *Port A*. As shown in Table 2, container features are in both structured and unstructured formats. The structured features represent container size, weight, origin port, shipping line, owner, and out-terminal, while the unstructured data, presented in text format, describes over 98,000 types of container content.

The data science team at the port advised that nearly 70 % of containers are transported by road, whereas the rest 30 % are moved by rail. This is evidenced in the raw dataset that contains 492,500 haulier observations, and 175,580 rail observations (see Table 4). Among the rail containers, there were 139,287 containers collected from R1 and 36,293 collected from R0. Since the terminal operator can choose the haulier out-terminals, the predictive task is essentially to classify the imported containers into three classes, H (haulier), R0, and R1. However, the imbalance of data objects among the out-terminals poses a challenge for predictive classification, as most machine learning algorithms perform optimally on data with an equal distribution of classes. The abundance of data from the majority class can hinder machine learning models' ability to effectively learn the characteristics of the minority class and differentiate it from the majority class.

4. The methodological framework

A methodological framework is proposed in Fig. 3 to address the research challenges. While businesses have traditionally relied on structured data, typically stored in formatted relational databases as numerical and categorical data, the effective utilisation of unstructured data, such as text, audio, and video files, can offer a competitive advantage (Harbert 2021). In the following subsections, the key steps and activities outlined in Fig. 3 are discussed in detail.

4.1. Pre-processing structured and unstructured data

We pre-processed the raw data by eliminating rows with missing, meaningless, and repeated data, and by removing features that contain repeated information or non-valuable information. For example, features like *cntr_id* and *bill_of_lading_no* possess high cardinality, assigning a unique ID number to each container. Another feature, *cntr_status*, takes a single value for all the containers and carries zero-variance. These features were deemed non-contributory to the training of the machine learning models and were consequently excluded from consideration. The data pre-processing steps eliminated 39,966 rows of container data, resulting in a dataset comprising 628,114 individual container records (see Table 4). After cross checking the two sources of information, a total of 628,089 containers with both berthing terminal and storage yard information were retained in the dataset. Following the classification of container content using Standard International Trading Classification (SITC) codes through feature engineering (as detailed in Section 4.3), an additional 203 containers were removed due to meaningless container entries (for example, entries of telephone numbers), leading to a training and testing dataset containing 627,886 containers. Our attempts to rectify the container data imbalance by oversampling the minority class and undersampling the majority class (Fernández et al., 2018) failed to show appreciable improvement in classification outcomes. Consequently, we opted not to pursue further corrective action to balance the dataset. In subsequent research, we have conducted threshold tuning to investigate different probability thresholds in classification models and determine the one that maximizes precision and recall while addressing class imbalances. But this is beyond the scope of this research.

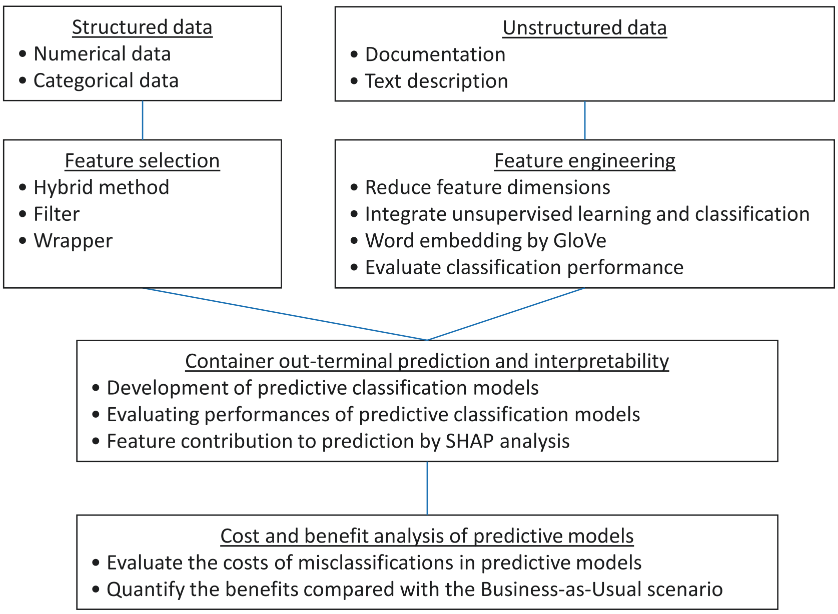

Fig. 3. Methodological framework for container out-terminal prediction.

4.2. Feature selection of structured data

Feature selection reduces computation time in training machine learning models, reduces overfitting risks, enhances precision accuracy by selecting critical features and eliminating redundant and irrelevant ones, and improves model interpretability. Without feature selection, the training time using the same dataset is 25 % higher. Furthermore, if the feature selection step is omitted, it would lead to the inclusion of redundant features with many missing values or irrelevant information. This inclusion can result in failures during machine learning model training.

In this section, features that highly differentiate among prediction outcomes and contribute to the reduction of prediction errors are selected (Ghaddar and Naoum-Sawaya, 2018). Two types of methods were employed for feature selection: filter methods and wrapper methods (Chandrashekar et al. 2014). Commonly used filter methods are Pearson's correlation coefficient, ANOVA correlation coefficient, Kendall's rank coefficient, and Chi-squared tests. In wrapper methods, supervised learning algorithms are iteratively fitted to extract correlations and dependencies between the features and exclude less significant features. Wrapper methods encompass techniques such as forward selection, backward elimination, Boruta algorithm, and Recursive Feature Elimination. Filter methods operate independently of classification algorithms, and they are computationally efficient, whereas wrapper methods are computationally effective (Chen et al. 2020). Wrapper methods excel in identifying the best subset of features, potentially leading to improved classification outcomes (Hsu et al. 2011). However, in the wrapper methods, the selected features can be influenced by the classifiers used for prediction. As both methodologies exhibit unique advantages and different shortcomings, a hybrid feature selection method is adopted in this research by combining the filter and wrapper methods (see Section A2 in Appendix).

As a result, a total of 14 features out of the initial set of 20 features were retained, denoted by a "✓" symbol present in both the "Filter methods" column and the "Boruta algorithm" column in Table 2. The feature of '*general_cargo_content*' was replaced by '*SITC*' (see section 4.3). *Port A* is an import port, with 99.8 % of imported containers being full, which means the feature of "*full_empty_indr*" carries almost zero variance and is deemed as unimportant in the process of feature selection.

4.3. Feature engineering of unstructured data

The unstructured data, denoted as "*general_cargo_content*," represents 98,271 unique container content names. High dimensionality in data can significantly affect machine learning performance, a challenge referred to as the "curse of dimensionality" (Aremu et al., 2020). To mitigate this issue, we performed feature engineering by transforming the original high-dimensional features into new low-dimensional ones (Bocca and Rodrigues, 2016). Specifically, we reduced the feature dimensionality of "*general_cargo_content*" through unsupervised classification. This approach aims to enhance the interpretability of machine learning models while preserving prediction accuracy.

The data science team at the port believes a relationship exists between container content and out-terminals, as some specific containers were transferred directly to rail services, and some were moved to shunting areas before being loaded onto trains. From an operational standpoint, enhanced interpretability helps the operations team identify influential container attributes on target out-terminals, which in turn, reveals associations between these attributes and improves decision making of container movement.

SITC is a product classification of the United Nations used for international trade. SITC Level 5 contains a list of 2970 predefined headings for goods, and this list includes almost all the internationally traded goods transported by containers. We adopted the SITC Level 5 structure to classify the 98,271 types of container content. Text classification presents a unique challenge in that creating labelled training documents can be labour-intensive and intellectually demanding, particularly when dealing with a large volume of text data. Even when subject knowledge is applied to label text documents, there tends to be a high degree of variability. This is because classification depends on personal opinions and is open to different interpretations (Shafiabady et al. 2016). Therefore, unsupervised learning methods have often been used for text classification (Ko and Seo 2000), including self-organising maps and k-means to measure data similarities (Li and Huang 2009). In the absence of a labelled training dataset, we leveraged GloVe word embeddings and Cosine similarity to perform unsupervised text classifications. GloVe word embeddings facilitated the creation of vector representations of words within the container content and the SITC headings. Subsequently, pairwise Cosine similarity was calculated between the matrix representation of SITC headings and the container content. The SITC heading that achieved the maximum similarity with the given instance of container content was identified, and its corresponding code was assigned as the class label to the container content. The detailed algorithms of the unsupervised classification are presented in Section A.3 in Appendix. Unsupervised classification performances were compared with the classification results obtained from a manual labelling process. The comparison achieved an accuracy score of 80.0 %, demonstrating that unsupervised text classification offers a practical solution for categorizing container contents when a labelled training dataset is unavailable.

After feature engineering, unsupervised classification assigns a SITC code to each of the 98,271 container content types, resulting in a new feature called "*SITC*" with reduced dimensionality. Following this feature engineering, the original "*general_cargo_content*" feature is replaced by the newly created "*SITC*" feature, as shown in Table 2.

4.4. Development of multiclassification models

Several machine learning methods can be used to build classification models. Deep learning, including various neural network models, delivers higher accuracy but requires a large volume of labelled data, often in the millions, and substantial computing power. Conversely, tree-based models consistently outperform standard deep learning models on tabular-style datasets, where features are individually meaningful and lack strong multiscale temporal or spatial structures (Lundberg et al. 2020). Given that the container dataset is in tabular format, the selection of tree-based models was deemed appropriate for the classification prediction task. Specifically, XGBoost and Random Forest were selected to construct classifiers. Random Forest, a bagging ensemble machine learning model, is designed to reduce variance, while XGboost, a gradient-boosting ensemble model, works to reduce both variance and bias.

After performing feature selection and feature engineering, the dataset with the chosen features was split into training and test data using a 75:25 ratio based on the original dataset. Under the haulier mode, the default collection location is the out-terminal adjacent to its storage yard, for example, H0 for Yard 0, or H1 for Yard 1. Hauliers receive directions from the terminal operator to collect containers from either H0 or H1. Under the rail mode, containers are collected from R0 or R1. For the multi-class classifier, we considered three out-terminal classes: H, R0, and R1, where H encompasses both H0 and H1, reflecting the operator's discretion in determining haulier out-terminals. The optimal hyperparameters were tuned to configure the machine learning classifiers (see Section A.4 in Appendix for more details). Since the distribution of target classes, i.e., H, R0 and R1, is severely imbalanced, class weights were set to "balanced" mode. This mode automatically adjusts weights inversely proportional to class frequencies in the input data.

4.5. Evaluating performances of predictive classification models

As with any predictive classification model, choosing the most appropriate and adequate metrics is essential in guiding the classifier modelling and assessing the classification performance. Given the rail network's stringent fixed departure and arrival schedules, correctly identifying the minority class (rail class) is more important than the majority (haulier class). This ensures that containers are efficiently loaded onto trains within the fixed time window. Hence, the problem is more sensitive to classification errors for the rail class than for the haulier class.

While classification accuracy stands as the most used metric in classification tasks, it is worth noting that in the case of imbalanced classification, relying solely on common metrics like accuracy can lead to sub-optimal classification models. This is because these metrics tend to be insensitive to situations where the data distribution is skewed (Branco et al. 2016). The preferences for specific metrics vary with application context and different objectives. Threshold metrics are one family of evaluation metrics that quantify the classification prediction errors. These metrics are designed to summarise the fraction, ratio, or rate at which predicted classes do not match the expected class in a holdout dataset. In this research, the threshold metrics are used to measure the predictive classification performances, including confusion matrix, precision rate per class, recall rate per class, F1 score per class, overall accuracy, overall F1 score, and Receiver Operating Characteristic (ROC) Area Under the Curve (AUC) score using One vs Rest (OvR) method. Using the OvR method, a single class is considered the positive class, while the rest are regarded as the negative class. ROC AUC scores are then calculated for each specific positive class. In Section 5, the performances of two classifier models: Random Forest and XGBoost, are compared and discussed.

4.6. Feature contribution to prediction by SHAP analysis

Sophisticated machine learning algorithms usually can produce accurate predictions, but their "black box" nature hinders the adoption of the machine learning model. To help the port better understand why a container is classified into an out-terminal and how the container attributes influence the classification, we applied model-agnostic interpretation techniques to interpret the predictive classification models. Feature importance is a measure that quantifies the significance of each feature using a single value. In machine learning models, various methods exist for ranking feature importance (Bommert et al. 2020), including: decision tree-based feature importance, permutation importance, and SHAP (SHapley Additive exPlanations) analysis (Lundberg and Lee, 2017). Good feature attribution methods should follow two properties: consistency and accuracy (Lundberg and Lee, 2017). Consistency means that the change in feature importance aligns with the direction of change in the feature's marginal contribution. Accuracy requires that the sum of all the features' importance adds up to the total importance of the model. The importance of the model can be measured as the probabilities belonging to classes, Log odds in classification problems, or R^2 for regression problems. The SHAP values proved to be unique, consistent, and locally accurate attribution values (Lundberg and Lee, 2017). SHAP analysis helps discern patterns in the data and understand the impacts of features on predictions through data visualization and model interpretability. Section 6 provides in-depth discussions on how to interpret machine learning models for multi-classification using SHAP analysis.

4.7. Cost and benefit analysis of predictive models

The added values of the operational decisions obtained from the predictive classification models are calculated and discussed. Of particular importance is the explicit quantification of the costs associated with decision errors induced by misclassifications. These errors entail incorrectly predicting classes and consequently storing containers in the wrong storage yards. While often overlooked in the literature that applies data analysis and machine learning to operations management, such cost considerations are critical to sound business decision making. Type I and II error costs are often investigated as misclassification costs for binary classification problems. Despite many successful applications of machine learning models in international maritime freight transportation management (Barua et al. 2020), there has been limited research that analyses misclassification costs in the context of the shipping industry or port management. In this research, we extend the concept of Type I and Type II costs to address multiclass classification problems, explicitly quantifying the costs incurred by the misclassifications generated by multiclass classification predictive models. This approach allows us to measure the benefits of employing predictive models and inform the yard management's decision making in the international shipping port.

We evaluated the performance of the predictive models by drawing fixed-sized samples from the test data using bootstrap sampling.

This assessment involved analysing confusion matrices and quantifying misclassification costs using the error rates within those matrices, as shown in Section 7.

5. Evaluating predictive model performances using empirical data

The methodological framework was applied to predict out-terminals for containers imported in the year of 2020, using the data in Table 4. The performances of various machine learning classifiers are compared in this section. The tuned classifiers, Random Forest and XGBoost, were used to classify containers using two distinct feature sets: one set contains the new feature “SITC” and another that excludes it. The classification performances for both cases are compared in Table 5. It is apparent that both Random Forest and XGBoost classifiers with the feature “SITC” outperform the classifiers without it, which verifies the importance of using the unstructured data (i.e. the container cargo contents) in classifying the out-terminals of containers. Due to the severe imbalance in class distribution, the aggregate standard metrics, such as weighted F1 scores and Accuracy, become unreliable or even misleading. Random Forest and XGBoost classifiers deliver superior results in classifying containers to the haulier gate H, as evidenced by higher Precision rates, Recall Rates, and F1 scores of this class. Setting class weights to be “balanced” increases the recall rates of R0 and R1 while maintaining the precision rates of all the three classes of H, R0 and R1. When SITC feature was included, both Random Forest and XGBoost achieved an overall accuracy of 80 % and an average ROC AUC of 86 %. The Recall rates and precision scores for out-terminals R0 and R1 remain lower in comparison to those associated with the H out-terminal.

High Recall rates of rail out-terminals mean that models are sensitive enough to capture containers destined for Rail. Nevertheless, the recall scores in our case are somewhat misleading because we have three class labels. Lower Recall rates of rail out-terminals do not necessarily incur higher transportation or storage costs. For example, misclassifying containers destined for R1 as H (or R0 as H) does not incur significant additional costs compared to the Business-as-Usual scenario, because these containers have similar probabilities of being stored in Yard 0 and Yard 1 in both scenarios. However, if containers destined for R0 are misclassified as R1, or vice versa, the additional costs could be more significant because cross-yard transportation during the container collection stage will increase the probability of rail misses (which is highly costly). Therefore, the confusion matrix in Table 6 is more appropriate to measure the accuracy of the multi-classification model, which will be discussed in Section 7. For example, from the confusion matrix in Table 6, we can observe that a_{12} and a_{21} (representing misclassifications of R0 as R1, or R1 as R0) are very small compared to other misclassification errors. This indicates that the model performs well.

Future research is required to further enhance the prediction performances for rail out-terminals. However, the focus of this research is to investigate whether a reasonably accurate predictive model can be developed from the currently available data, and whether the predictive model can improve the cost-effectiveness of container stacking at *Port A* when compared with the current practices. Our method is demonstrated to be effective in Section 7 even though the recall scores are relatively low.

6. Machine learning interpretability using SHAP analysis

In this section, we interpret the importance of the features utilised in constructing the predictive models. SHAP analysis was performed on the XGBoost model and baseline SHAP values were obtained. The SHAP summary plots in Fig. 4 offer the global overview of feature importance by sorting the sum of absolute SHAP values over all samples. The plots further show the global positive and negative relationships of the features with the predicted class. The plots also provide a way to demonstrate local explanation of a feature, by showing the effect of a single feature across the entire dataset, and the distribution of the impacts of each feature on the predicted output. The dotted violin plots correspond to individual observations in the dataset. The dot’s position on the X-axis shows the impact that a feature has on the prediction, the log odds (used in XGBoost) of belonging to that class. The colour of each dot signifies the value of that feature for the respective container, and multiple dots accumulate at the X-axis to show density.

Table 5
Performances of classifiers with feature selection.

Feature set	Classifiers	Class	Precision score	Recall score	F1 score	ROC AUC OvR	Weighted F1	Accuracy	Average ROC AUC
With “*SITC*”	Random Forest	R1	0.68	0.48	0.56	0.85	0.79	0.81	0.86
		R0	0.67	0.37	0.47	0.88			
		H	0.83	0.93	0.88	0.84			
Without “*SITC*”	Random Forest	R1	0.57	0.44	0.50	0.80	0.76	0.77	0.81
		R0	0.54	0.31	0.40	0.84			
		H	0.82	0.90	0.86	0.79			
With “*SITC*”	XGBoost	R1	0.60	0.59	0.60	0.85	0.79	0.79	0.86
		R0	0.54	0.46	0.50	0.88			
		H	0.86	0.94	0.87	0.85			
Without “*SITC*”	XGBoost	R1	0.48	0.52	0.50	0.78	0.73	0.73	0.80
		R0	0.38	0.37	0.37	0.83			
		H	0.83	0.82	0.82	0.77			

Table 6
Confusion matrix of multiclass classification using XGBoost.

	Actual R1	Actual R0	Actual H
Predicted R1	a_{11}(mean = 0.123, std = 0.005)	a_{12}(mean = 0.007, std = 0.001)	a_{13}(mean = 0.073, std = 0.004)
Predicted R0	a_{21}(mean = 0.006, std = 0.001)	a_{22}(mean = 0.025, std = 0.002)	a_{23}(mean = 0.015, std = 0.002)
Predicted H	a_{31}(mean = 0.082, std = 0.004)	a_{32}(mean = 0.023, std = 0.002)	a_{33}(mean = 0.646, std = 0.007)

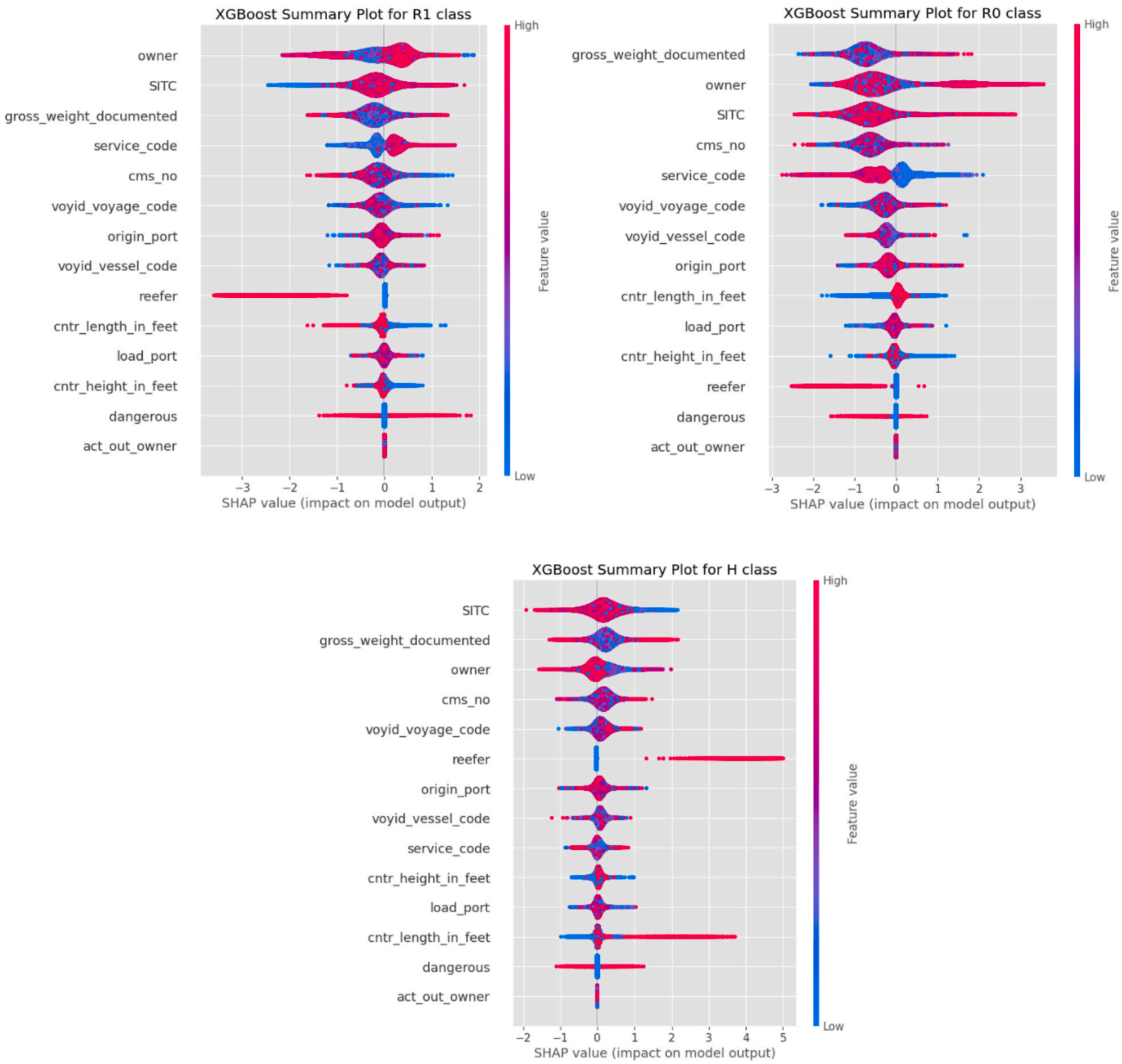

Fig. 4. The SHAP summary plots for XGBoost classification.

6.1. *SHAP robustness analysis*

To evaluate the consistency and robustness of SHAP analysis results, we applied bootstrap undersampling techniques to create subsets comprising 80 % of the original training dataset, while retaining its essential characteristics. Ten subsets were generated, and the XGBoost model was retrained on the subsets. SHAP analysis was performed using each subset. The SHAP values obtained from these models using the resampled data were compared with the baseline SHAP values, and the Symmetric Mean Absolute Percentage Error (SMAPE) was calculated using equation (1) to assess each feature's robustness to data resampling. The robustness of SHAP analysis signifies the stability and reliability of SHAP values in the face of training dataset variations and resulting differences in trained machine learning models.

$$SMAPE_j = \left(\frac{1}{n}\right) * \sum_{i=1}^{n} \frac{\left|S_j(baseline) - S_j(sample_i)\right|}{\frac{\left|S_j(baseline)\right| + \left|S_j(sample_i)\right|}{2}} \tag{1}$$

j: the j-th feature, $j = 1, 2,..,14$ in this study;
i: the i-th resampled training dataset, $i = 1, 2,..,10$ in this study;
$SMAPE_j$: Symmetric Mean Absolute Percentage Error for feature j;
n: total number of resampled training datasets, $n = 10$ in this study;
$S_j(baseline)$: baseline SHAP values for feature j;
$S_j(sample_i)$: SHAP values for feature j using resampled training dataset i.

In Fig. 5, the SMAPE values between the baseline SHAP values and those obtained from retrained models depict features' sensitivity. A high SMAPE score indicates a sensitive feature, meaning variations in training data could lead to significant changes in the model's output and ranking of feature importance. Overall, the maximum SMAPE values are below 0.15 (15 %) for R1 class, 0.10 (10 %) for H class and 0.20 (20 %) for R0 class, which indicates that none of the features is highly sensitive to resampled training sets. There exists a correlation between features with high SHAP values and those with high SMAPE metrics. This suggests the XGBoost model under investigation remains robust to resampled data, and the ranking of feature importance based on SHAP values remains consistent, despite minor disparities in the ranking of more sensitive features, such as *service code*, *cntr_length_in_feet* and *reefer*.

6.2. Feature importance based on SHAP analysis

When interpreting machine learning classifiers, the focus is on quantifying the contribution of each feature to the predicted label. However, it is important to note that a unit change in categorical variables does not have a meaningful interpretation. For categorical variables, the interpretation is how each categorical value affects the prediction. The summary plots for the three classes predicted by XGBoost are shown in Fig. 4, and several observations can be drawn from the plots:

(1). **Feature importance ranking:** When predicting H, R0, and R1 classes, the top three important features are *SITC*, *owner*, and *gross_weight_documented*. Regarding the feature *SITC*, transforming the high-dimensional container content into a new low-dimensional feature improves the interpretability of how it contributes to the prediction results. As for the *owner* feature, it is notable that containers' out-terminals are associated with *owner* information, suggesting that owners may exhibit preferences for either inland Rail or Haulier transport. Additionally, the container weight, *gross_weight_documented*, also influences the choice of inland transportation mode and ultimately the corresponding out-terminal at the port.

(2). Impact and correlation:
The plot in Fig. 4 reveals the direction of effects. For example, a higher value of *gross_weight_documented* is associated with a higher probability of prediction of H Class, implying that heavy containers are more likely to be transported by road haulage from *Port A* to their destinations. While this observation may appear counterintuitive, it is due to the age of the rail track and weight limits on bridges along the rail routes, as verified by *Port A*. It is noted that *SITC* code and *Owner* are categorical variables, thus, an increment or decrement in the values of these variables do not necessarily reveal an association between them and the prediction. In Section A.5 in Appendix, the impacts of categorical variables are quantified based on categorical values and some interesting results were revealed: i) containers from *Owner* 13 tend to be transported by rail from out-terminal R0, while containers from *Owner* 2 are more likely to be collected from out-terminal R1. ii) Content characterised by SITC codes 1291 (*Meat and edible meat offal of rabbits or hares*) and 4812 (*Cereal other than maize*) exhibit a preference for being collected from R1 out-terminal. Conversely, containers having content corresponding to SITC codes 1211 (*Meat of sheep*) and 5429 (*Medicaments*) are more likely to be retrieved from R0 out-terminal.

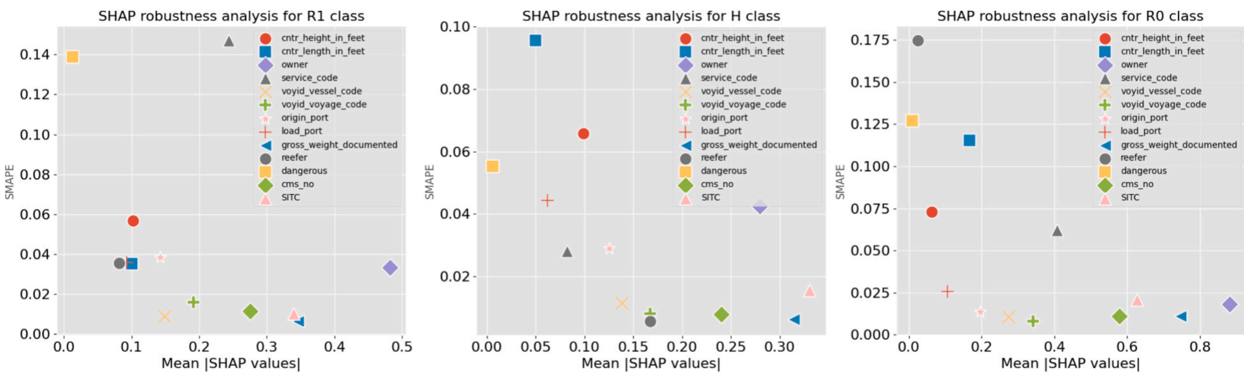

Fig. 5. Mean absolute SHAP values v.s. SMAPE for XGBoost classification model.

The plot also reveals the distribution of effect sizes. The long right (or left) tails mean that features with low global importance can be extremely important for specific classifications of individual containers. For example, high value of *reefer* stretches to the left for both rail classes, implying that refrigerated containers ($reefer = 1$) tend to be transported by hauliers. High values of *cntr_length_in feet* stretch to the left for R1 class and stretch to the right for H class, showing that longer containers tend to be collected from H out-terminal; low values of *cntr_height_in feet* increase SHAP values of R0 and R1, implying that lower containers are more likely to be transported by rail while higher containers are more likely to be picked up by haulier. One explanation given by the port is that a loading gauge is in place defining the maximum height, width, and loads of containers transported by rail. Containers over 8.5 feet in height tend to be transported by hauliers.

SHAP, as a tool to interpret complex machine learning models, helps the port in identifying and verifying the relationships between containers' attributes and the out-terminals. Some of these results concur with the port operations team's experiential judgement, such as the relationships between container height, refrigerated containers, and container owners with out-terminals.

7. Evaluating costs arising from predictive models

In this section, we assess the advantages of utilizing predictive models by comparing the costs of the Business-as-Usual scenario with the costs under two scenarios that integrate the prediction results while accounting for misclassification errors.

By employing bootstrap sampling, we drew 1000 samples, each containing 5000 instances of container data from the test dataset. A confusion matrix is produced based on the XGBoost classification results, as shown in Table 6. This matrix shows the mean values and standard deviations of the normalized confusion matrix elements, with normalization achieved by dividing each element by the total sum of all matrix values.

Rather than focusing on Type I and II errors for each class, the error rates in the confusion matrix in Table 6 were used to calculate the transportation-related costs resulting from misclassifications. The minimum error rate would effectively reduce unnecessary transportation costs. Furthermore, it is important to note that misclassifying haulier containers does not typically lead to missed lorries or substantial penalty costs, as lorries do not adhere to scheduled timetables. However, if rail containers are incorrectly predicted to the wrong out-terminals, it could result in significant penalties due to rail misses at their intended out-terminal.

To capture both direct and indirect costs of container movements at ports, the operations team at the port consolidated these cost elements into a single cost element called *transportation-related cost*. Table 7 reports the estimated coefficients for these costs. It should be noted that these cost coefficients were estimated by normalising the transport-related cost between a yard and its associated haulier out-terminal to a value of 1. The relative values of these cost elements are representative of the true costs and reflect the ordinal relationships among them. Due to the high penalties incurred when missing the loading containers to a train service, the transportation cost of moving containers from a yard to a rail out-terminal is higher than moving them from yard to haulier out-terminal.

Transporting containers from the maritime quayside to Yard 0 and Yard 1 was treated as an exogenous input variable, taking precedence over container movements between yards and out-terminals. Consequently, the costs associated with container movements from the maritime quayside to the two yards were fixed and omitted in our scenario analysis. Yard 0 is approximately 1 mile from Yard 1, with H0 and H1 located within 200–300 yards of Yard 0 and 1, respectively. The within-yard transportation costs between Yard 0 (1) and haulier gates H0 (1) are minimal, normalised to 1, and denoted as C_{00}^{yh} and C_{11}^{yh}. Based on a travelling distance of 1 mile, the cross-yard transportation costs of moving containers between the yard and the haulier out-terminals, noted as C_{01}^{yh} and C_{10}^{yh}, are estimated to be 5. The rail terminals R0 and R1 are located about 2 miles from Yards 0 and 1, respectively. The within-yard transportation costs between the yards and their respective rail out-terminal, denoted as C_{00}^{yr} and C_{11}^{yr}, are estimated to be 10. The most significant costs are attributed to cross-yard transportation costs between the yard and the rail out-terminal, noted as C_{01}^{yr} and C_{10}^{yr}, which could range 2–6 times of the within yard transportation costs between yards and haulier gates C_{00}^{yh} and C_{11}^{yh}, due to penalties imposed by rail misses and waiting time of RMG (Rail Mounted Gantry) cranes. Therefore, the order of cost coefficients in terms of their values is: $C_{01}^{yr} = C_{10}^{yr} > C_{00}^{yr} = C_{11}^{yr} > C_{01}^{yh} = C_{10}^{yh} > C_{00}^{yh} = C_{11}^{yh} = 1$;

Table 7
Estimated cost coefficients.

Cost coefficients descriptions	Notation	Value (£)
from yard to rail		
transport-related cost from Yard 0 to R1, incl. miss penalty and RMG* waiting	C_{01}^{yr}	[20,60]
transport-related cost from Yard 1 to R0, incl. miss penalty and RMG* waiting	C_{10}^{yr}	[20,60]
transport-related cost from Yard 0 to R0, incl. miss penalty and RMG* waiting	C_{00}^{yr}	10
transport-related cost from Yard 1 to R1, incl. miss penalty and RMG* waiting	C_{11}^{yr}	10
from yard to haulier		
transport-related cost from Yard 0 to H1	C_{01}^{yh}	5
transport-related cost from Yard 1 to H0	C_{10}^{yh}	5
transport-related cost from Yard 0 to H0	C_{00}^{yh}	1
transport-related cost from Yard 1 to H1	C_{11}^{yh}	1

* RMG: Rail Mounted Gantry cranes.

Three business scenarios were constructed to compare the costs incurred under different yard operation strategies:

Business as Usual scenario: Yard management followed current practices described in Section 3.3.

Predictive scenario: In this scenario, an easy-to-implement strategy is designed to deploy the predictive model in TOS with minimum effort and disruption to existing operations. Under this strategy, several rules are defined for container movement and storage in specific yards, based on the results from the predictive mode. These rules can be adapted and adjusted to suit various real world business cases:

1. The predictive model determines the containers' destinations. First, all containers predicted for R1 are moved from the maritime quayside to Yard 1, while all containers predicted for R0 are transferred from the maritime quayside to Yard 0.
2. Alongside the containers predicted for R1, a portion of containers predicted for the haulier are randomly selected and moved from maritime quayside to Yard 1. This step is implemented to maintain consistent volume distribution in container movement operations between the maritime quayside and the two yards in both scenarios, i.e. to keep D_{00} and D_{01} having the same values as the Business-as-Usual scenario, which minimises the disruptions to existing operations (such as resource assignment and workload distribution).
3. Haulier containers utilise the out-terminal located within their respective stored yard terminals. The rule is implemented based on the rationale that the terminal operator holds the responsibility of specifying the haulier out-terminal for lorries to pick up containers.

Ideal Predictive scenario: this scenario assumes that the predictive model achieves 100 % accuracy, which is equivalent to the scenario in which we know the actual out-terminals of imported containers when they are unloaded from the vessel. All containers destined for R1 are stored at Yard 1, all containers destined for R0 are stored at Yard 0, and haulier containers are stored either at Yard 0 or Yard 1 to ensure D_{00} and D_{01} having the same values as the Business-as-Usual scenario. Haulier containers will be collected from their respective storage yards. There will be no cross-yard movements from yards to out-terminals. The transport operations in the Ideal predictive scenario are identical to that in the Predictive scenario except the predictive accuracy.

7.1. Costs under the business-as-usual scenario

The relevant total costs of transporting containers under the Business-as-Usual scenario were incurred by the following activities. A total number of D containers were unloaded at maritime quayside, D_{00} were stored at Yard 0 and D_{01} are stored at Yard 1, then moved to one of the out-terminals: H0, H1, R0, and R1. The costs can be formulated as follows:

$$\text{Transport} - \text{related costs of containers at yard } 0 = D^h_{00} \times C^{yh}_{00} + D^r_{00} \times C^{yr}_{00} + D^r_{01} \times C^{yr}_{01} + D^h_{01} \times C^{yh}_{01} \tag{2}$$

$$\text{Transport} - \text{related costs of containers at yard } 1 = D^h_{10} \times C^{yh}_{10} + D^r_{10} \times C^{yr}_{10} + D^r_{11} \times C^{yr}_{11} + D^h_{11} \times C^{yh}_{11} \tag{3}$$

The sum of (2) and (3) gives the total relevant cost in the Business-as-Usual scenario, i.e.

$$\text{Total relevant cost under Business} - \text{as} - \text{Usual scenario} = D^h_{00} \times C^{yh}_{00} + D^r_{00} \times C^{yr}_{00} + D^r_{01} \times C^{yr}_{01} + D^h_{01} \times C^{yh}_{01} + D^h_{10} \times C^{yh}_{10} + D^r_{10} \times C^{yr}_{10} + D^r_{11} \times C^{yr}_{11} + D^h_{11} \times C^{yh}_{11} \tag{4}$$

According to the historical data recorded in the Business-as-Usual scenario, a small number of haulier containers were still moved between yards despite the terminal operator's ability to assign haulier out-terminals, which could be attributed to road hauliers accessing incorrect gates by mistake. Eliminating cross-yard movements for haulier containers can improve Business-as-Usual scenario

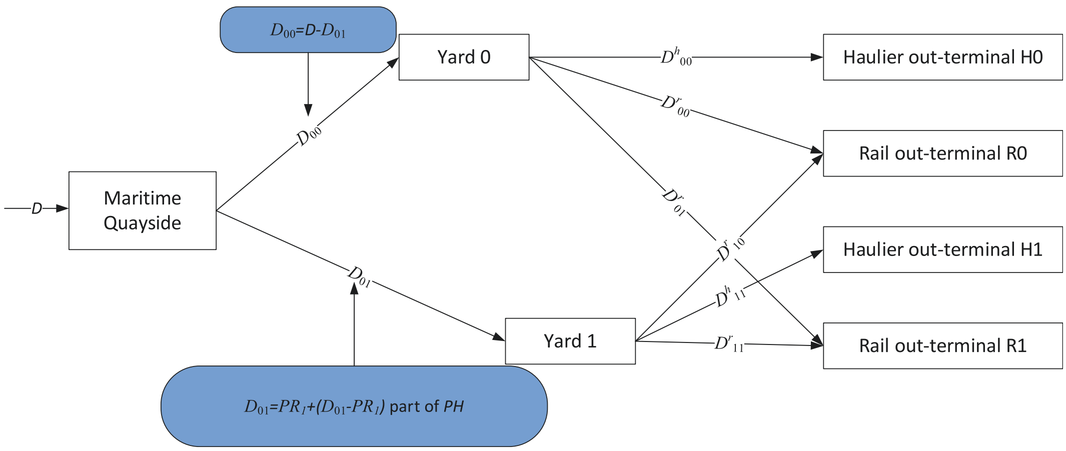

Fig. 6. Container movements under Predictive scenario.

outcomes. This can be achieved by merging D_{01}^h into D_{00}^h and D_{10}^h into D_{11}^h, resulting in cost saving of approximately 1 %. In Section 7.4, we will demonstrate that this cost saving is significantly lower than the cost saving achieved by the predictive model.

7.2. Costs under the predictive scenario

In the predictive scenario, the container movement follows rules 1)-3) defined for this scenario, and the movement is depicted in Fig. 6. Instead of random selection and transportation to Yard 1 as in the Business-as-Usual scenario, containers were selected and transported based on the prediction outcomes generated by the classification model. Using the percentage values in the confusion matrix presented in Table 6, we can work out the predicted numbers of containers destined for H, R0, and R1, denoted as PH, PR_0, and PR_1, respectively:

$$PH = D \times (a_{31} + a_{32} + a_{33}) \tag{5}$$

$$PR_0 = D \times (a_{21} + a_{22} + a_{23}) \tag{6}$$

$$PR_1 = D \times (a_{11} + a_{12} + a_{13}) \tag{7}$$

In our specific case context, the total volume of containers to be moved from maritime quayside to Yard 1, denoted as D_{01}, always exceeded the value of PR_1. Consequently, the remaining volume, calculated as $(D_{01} - PR_1)$, was filled by a random selection of containers from the PH category. The remaining containers, $D_{00} = D - D_{01}$, were stored in Yard 0. Since the classification model may incur misclassification, to work out the actual transportation-related costs, it is necessary to determine the actual movement of containers, such as D_{00}^h, D_{00}^r, D_{01}^r, and so on, using the information provided in the confusion matrix presented in Table 6. Rule 3) indicates that $D_{01}^h = D_{10}^h = 0$. Following rules 1)-3), we can calculate the corresponding total movement costs using the formulas (2) and (3).

Firstly, we worked out the movements of containers stored at Yard 0. By utilising the predictions of PH, PR_0 and PR_1, along with the classification rates provided in the confusion matrix in Table 6, we obtained D_{00}^h, D_{00}^r and D_{01}^r using the following equations:

$$D_{00}^h = PR_0 \times a_{23}/(a_{21} + a_{22} + a_{23}) + (PH - D_{01} + PR_1) \times a_{33}/(a_{31} + a_{32} + a_{33}) \tag{8}$$

$$D_{00}^r = PR_0 \times a_{22}/(a_{21} + a_{22} + a_{23}) + (PH - D_{01} + PR_1) \times a_{32}/(a_{31} + a_{32} + a_{33}) \tag{9}$$

$$D_{01}^r = PR_0 \times a_{21}/(a_{21} + a_{22} + a_{23}) + (PH - D_{01} + PR_1) \times a_{31}/(a_{31} + a_{32} + a_{33}) \tag{10}$$

The first part of D_{00}^h, i.e., $PR_0 \times a_{23}/(a_{21} + a_{22} + a_{23})$, are the actual haulier containers incorrectly predicted as R0 containers. These containers are stored in the default Yard 0 and transported to H0. The second part of D_{00}^h i.e., $(PH - D_{01} + PR_1) \times a_{33}/(a_{31} + a_{32} + a_{33})$, represents the accurately predicted haulier containers residing in Yard 0, to be sent to H0. The numbers of containers required to be transported from Yard 0 to rail out-terminals R0 and R1, noted as D_{00}^r and D_{01}^r, can be similarly worked out. The transport-related costs associated with the containers stored at Yard 0 is given by:

$$\textit{Transport} - \textit{related costs of containers at yard } 0 = D_{00}^h \times C_{00}^{yh} + D_{00}^r \times C_{00}^{yr} + D_{01}^r \times C_{01}^{yr} \tag{11}$$

Secondly, we worked out the movements of the containers stored at Yard 1. The costs of transporting containers from Yard 1 to out-terminals H1, R1 and R0 can then be calculated:

$$D_{11}^h = PR_1 \times a_{13}/(a_{11} + a_{12} + a_{13}) + (D_{01} - PR_1) \times a_{33}/(a_{31} + a_{32} + a_{33}) \tag{12}$$

$$D_{11}^r = PR_1 \times a_{11}/(a_{11} + a_{12} + a_{13}) + (D_{01} - PR_1) \times a_{31}/(a_{31} + a_{32} + a_{33}) \tag{13}$$

$$D_{10}^r = PR_1 \times a_{12}/(a_{11} + a_{12} + a_{13}) + (D_{01} - PR_1) \times a_{32}/(a_{31} + a_{32} + a_{33}) \tag{14}$$

$$\textit{Transport} - \textit{related costs of containers at yard } 1 = D_{10}^r \times C_{10}^{yr} + D_{11}^r \times C_{11}^{yr} + D_{11}^h \times C_{11}^{yh} \tag{15}$$

The sum of (11) and (15) gives the total relevant cost in the Predictive scenario, i.e.

$$\textit{Total relevant cost under predictive scenario} = D_{00}^h \times C_{00}^{yh} + D_{00}^r \times C_{00}^{yr} + D_{01}^r \times C_{01}^{yr} + D_{10}^r \times C_{10}^{yr} + D_{11}^r \times C_{11}^{yr} + D_{11}^h \times C_{11}^{yh} \tag{16}$$

7.3. Costs under the Ideal predictive scenario

This scenario was only used for comparison purpose because it is impossible to achieve 100 % prediction accuracy in reality. The costs can be formulated as follows:

$$\textit{Transport} - \textit{related costs of containers at yard } 0 = D_{00}^r \times C_{00}^{yr} + \left(D_{00} - D_{00}^r\right) \times C_{00}^{yh} \tag{17}$$

$$\textit{Transport} - \textit{related costs of containers at yard } 1 = D_{11}^r \times C_{11}^{yr} + \left(D_{01} - D_{11}^r\right) \times C_{11}^{yh} \tag{18}$$

In the above two equations, the first term represents the costs incurred by transporting containers from the corresponding yard to the nearby rail terminal, while the second term signifies the costs associated with transferring containers from the corresponding yard to the nearby haulier gate.

The sum of (17) and (18) gives the total relevant cost in the Ideal predictive scenario, i.e.

$$\textit{Total relevant cost under Ideal} - \textit{predictive} = D^r_{00} \times C^{yr}_{00} + \left(D_{00} - D^r_{00}\right) \times C^{yh}_{00} + D^r_{11} \times C^{yr}_{11} + \left(D_{01} - D^r_{11}\right) \times C^{yh}_{11} \tag{19}$$

7.4. Cost comparison across the three scenarios

In this section, the cost elements among three scenarios were compared while maintaining a fixed parameter, C^{yr}_{01}=£20, using XGBoost. Next, the overall costs between the Predictive scenario and the Business-as-Usual scenario were evaluated, with varying parameter C^{yr}_{01} from £20 to £60. Finally, the misclassification costs associated with the predictive model were discussed.

7.4.1. Comparison of the three scenarios

Transportation-related costs achieved under the three scenarios are compared in Table 8. The bold numbers in Table 8 highlight the cross-yard transport costs to rail out-terminals. A noticeable difference can be observed when comparing Predictive and Ideal scenarios with the Business-as-Usual scenario. Specially, the predictive model can reduce the transportation-related cost from Yard 0 to rail out-terminal R1, reducing it from £1747340 to £972060 (a 44.37 % decrease). Similarly, it can decrease the cost from Yard 1 to rail out-terminal R0 from £182840 to £125214 (a 31.38 % reduction). In the Ideal scenario, the cross-yard transport cost to rail out-terminals is reduced to zero; nevertheless, the costs from the yard to its nearby rail out-terminal are higher compared to the other two scenarios. Under the condition that D_{00} and D_{01} are divided between the two yards, the Ideal scenario sets a low bound for the total relevant costs, amounting to £2043575. The parameter C^{yr}_{01} represents the unit transport-related cost from Yard 0 to R1, including penalties for rail misses and crane waiting cost (similarly for C^{yr}_{10}). In subsequent analysis, the value of C^{yr}_{01} was set to vary in the range of £20 to £60, which was in line with the estimation given by the operations team at *Port A*. It should be noted that the parameter C^{yr}_{01} does not affect the Ideal scenario since cross-yard transportation to rail out-terminals does not exist in the Ideal scenario.

7.4.2. Sensitivity analysis of cost coefficients

In Table 9, the relevant total costs generated in the Business-as-Usual scenario and Predative scenario are compared. The results demonstrate that yard management in Predictive scenario outperforms Business-as-usual scenario, irrespective of the values assigned to C^{yr}_{01} and C^{yr}_{10}. In other words, using the predictive classification model could lead to significant cost savings in handling imported containers. The cost reductions attained by the predictive classification model increase as the value of C^{yr}_{01} and C^{yr}_{10} increase, as shown in Table 9. With C^{yr}_{01} and C^{yr}_{10} ranging from £20 to £60, Predictive scenario achieved total cost reductions of £469,280 ~ £2,134,371 when compared with the Business-as-Usual scenario, corresponding to a percentage decrease of 14.90 % ~30.45 %. It is noteworthy that the cost coefficients have been normalised, with the transport-related cost between a yard and its associated haulier out-terminal to be £1. Therefore, the actual cost reduction would be even higher than the values reported in Table 9, and the percentage cost reduction holds greater significance.

7.4.3. Sensitivity analysis of predication error rates

Misclassification errors are inevitable with virtually any classification model but increasing classification accuracies per class could offset the costs incurred by misclassification errors; and the implications of error rate on costs vary with the class label and the type of errors (Type I or II error per class). Misclassifying containers intended for rail out-terminals incurs higher costs compared to misclassifying containers destined for haulier out-terminals. Furthermore, mistakenly storing containers in a yard that is farther from the destination rail out-terminal would increase the risk of missing scheduled trains and incurring substantial penalty costs.

The confusion matrix in Table 6 provides insights into the error rates associated with different misclassifications. Specifically, the error rate a_{12} represents the misclassification of actual R0 containers as R1, while a_{21} indicates the rate of misclassifying actual R1 containers as R0. On the other hand, the error rates a_{31} and a_{32} determine the number of containers that are incorrectly predicted as H. In terms of importance, the error rates a_{12} and a_{21} hold greater significance compared to a_{31} and a_{32}. This is because misclassifying rail containers to a different rail out-terminal results in cross yard transportation costs and possible penalty charges, whereas misclassifying rail containers as haulier containers does not necessarily lead to cross yard transportation. If actual R0 or R1 containers are misclassified to haulier containers, they could be stored in either Yard 0 or Yard 1, depending on their inclusion in the randomly selected

Table 8
Transport-related costs (£) from storage yards to out-terminals underC^{yr}_{01} = 20 (£).

Scenario	Out-terminal	H0	H1	R0	R1	Total costs
Business as Usual	Yard 0	315,815	56,005	251,290	**1,747,340**	**3,148,962**
	Yard 1	9845	131,847	**182,480**	454,340	
Predictive	Yard 0	362,636	0	282,730	**972,060**	**2,679,682**
	Yard 1	0	98,232	**125,214**	838,810	
Ideal	Yard 0	388,814	0	324,830	**0**	**2,043,575**
	Yard 1	0	52,781	**0**	1,277,150	

Table 9
Relevant total costs incurred in the two scenarios and cost reduction.

$C_{01}^{yr} = C_{10}^{yr}$	20	30	40	50	60
Business-as-Usual (£)	3,148,962	4,113,872	5,078,782	6,043,692	7,008,602
Predictive (£)	2,679,682	3,228,319	3,776,957	4,325,594	4,874,231
Cost reduction in Predictive (£)	469,280	885,553	1,301,825	1,718,098	2,134,371
Cost reduction in Predictive (%)	14.90	21.53	25.63	28.43	30.45

portion of haulier containers (H) that are to be transported to Yard 1. Regarding a_{13} and a_{23}, misclassifying actual haulier containers (H) as rail containers, whether R1 or R0, does not incur any cross-yard transportation costs or penalty costs associated with missed scheduled lorries. This is because the terminal operator has control over the haulier out-terminal selection, and the haulier collection does not follow scheduled timetables. In terms of the importance ranking of the error rates, our judgement concurs with the opinions of the operations team at the port, which is as follows: $a_{12} \approx a_{21} > a_{31} \approx a_{32} > a_{13} \approx a_{23}$. The results in Table 6 show that the top two most important error rates, a_{12} and a_{21}, are relatively small, indicating that the costs associated with such classification errors are modest compared to the model's ability to accurately detect containers destined to each out-terminal.

Comparing the cost reductions achieved by XGBoost reveals that improved accuracies in predicting rail classes (R0 and R1), represented by increased rates of a_{11} and a_{22} in Table 6, leads to a substantial decrease in cross yard transportation and its associated costs. However, there is still potential for further cost reduction. Incorporating additional features, such as information on consignor, consignee, and destinations of the containers would be valuable in enhancing the prediction accuracy rates of a_{11}, a_{22} and a_{33}, and reducing the top important error rates of a_{12} and a_{21}.

To further evaluate the robustness of the results and the model's sensitivity to the misclassification errors, we hypothetically introduced a 20 % increase in the six specific misclassification errors, including a_{21}, a_{31}, a_{12}, a_{32}, a_{13}, and a_{23}. For instance, when a_{21} is increased by 20 %, an equal amount is decreased from a_{11}, while the other a_{ij} values remain unchanged. The results are given in Table 10. The column 'a_{ij}(%)' in Table 10 represents the cost reduction percentage in the cases where the misclassification error a_{ij} was increased by 20 % and the correct prediction ratio a_{jj} equally decreased. The second column "*Pred(%)*" presents the cost reduction percentage achieved in the Predictive scenario as shown in Table 9. Comparing the results in the column "*Pred(%)*" with other columns, it is evident that even with a 20 % increase in misclassification errors under the fixed C_{01}^{yr}, the predictive model continues to deliver substantial cost savings. This shows the robustness of our predictive model with respect to the misclassification errors. Notably, a_{12} and a_{21}, which are the most critical misclassification errors, are very small in our predictive model. In Table 10, a_{31} appears to be the misclassification error that is most sensitive to the 20 % increment, which is due to its considerably high absolute value (see Table 6). A 20 % increase of a_{31} would significantly decrease the value of a_{11} (the correct prediction rate of the actual R1 containers), which leads to reduced cost savings percentages. One counter-intuitive observation is that an increase in misclassification error a_{23} and decrease in the correct prediction ratio a_{33} slightly enhance cost savings. This can be explained as follows: 1) an increment in a_{23} causes more authentic H (haulier) containers to be predicted as R0 containers and are stored in Yard 0, while a decrement in a_{33} leads to fewer correctly predicted haulier containers; 2) since a portion of predicted haulier containers must be chosen for transportation to Yard 1, a greater number of authentic haulier containers that are predicted as R0 containers will remain at Yard 0, reducing the number of authentic haulier containers being transferred to Yard 1. However, this also increases the chance of actual R1 containers, which are incorrectly classified as haulier containers, being chosen and sent to Yard 1, ultimately resulting in higher cost savings.

To have a clearer view of the relationship between the predictive model classification errors and the cost saving percentage, the parameter C_{01}^{yr} was fixed at 40, and the misclassification error was increased by 5 %, 10 %, 15 %, 20 % and 25 % respectively. From Fig. 7, it can be observed that (i) for five types of misclassification errors (a_{21}, a_{12}, a_{32}, a_{13} and a_{23}), the cost saving percentages do not change significantly as the misclassification error increases by up to 25 %; (ii) In the worst case, as the misclassification error a_{31} is increased by 25 %, the cost saving achieved is still close to 20 %. This confirms the effectiveness and robustness of the predictive model.

8. Discussion

In this section, we first discuss the deployment of the results and the generalisability of the methodological framework. Then, we summarise the main contributions of this work in terms of academic research and managerial insights.

Table 10
The cost savings achieved as misclassification errors increased by 20%.

C_{01}^{yr} (£)	Pred (%)	a_{21} (%)	a_{31} (%)	a_{12} (%)	a_{32} (%)	a_{13} (%)	a_{23} (%)
20	14.90	14.68	12.33	14.60	14.78	14.70	14.91
30	21.53	21.19	17.59	21.07	21.33	21.22	21.54
40	25.63	25.22	20.86	25.07	25.40	25.25	25.64
50	28.43	27.97	23.08	27.80	28.16	28.00	28.44
60	30.45	29.96	24.68	29.78	30.17	30.00	30.47

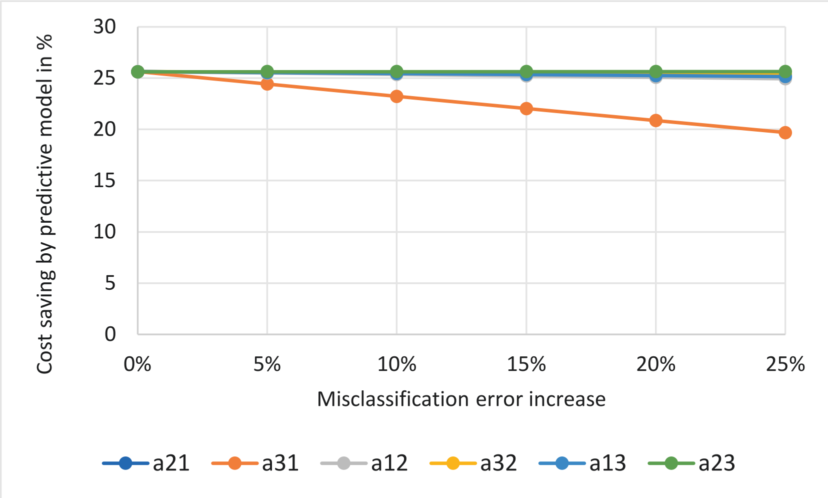

Fig. 7. Cost saving percentage vs predictive model classification errors increasement.

8.1. Deployment and generalisability

The data science team at *Port A* believes that the easy-to-implement strategy would enable them to incorporate the predictive models and cost analysis into the current TOS with minimal efforts and disruptions. The deployment essentially replaces random selection operations with predictive model-based selection operations as described in the predictive scenario. It does not change the imported containers' volume distributions across storage yards and causes minimal disruption to the allocation of existing resources. When vessels dock at the quayside, the predictive model records containers in the TOS, predicts the out-terminal, and facilitates container movement to a nearby storage yard aligned with the predicted out-terminal. Our study shows that deploying our predictive model as a straightforward add-in function to the existing yard block assignment strategies can yield a significant efficiency improvement at container ports. In addition, with the increasing demand for transportation decarbonisation, many countries (e.g. US, UK and European continent countries) are promoting modal shift from road to rail in terms of container movements. This indicates the great deployment market for the proposed predictive models.

In this research, the predictive model is trained using container data in 2020. In the actual deployment, the model can be regularly updated using new container information when they are collected from the out-terminals. The frequency of model updates can be adjusted to align with different work shifts (the day shifts are busier than night shifts), or seasonal container throughput. For example, the classification model can be updated after every 1000 additional containers are collected. There is also flexibility in the size of training data. It can be fixed for a specific duration or a fixed volume of data, or can dynamically increase along with the arrival of new containers. The choice depends on factors like server storage capacity, computing power, and the desired prediction model accuracy. To further improve the classification performance, the port advises that it would be beneficial to retrieve additional information from shipping lines, such as consignor and consignee information. The consignees' information can also be utilised to achieve smart stacking at container yards to further improve the yard operation efficiency (Feng et al., 2022b).

Fragmented decision-making, uncertainty and complexity pose challenges within global container shipping supply chains. *Port A* is just one example of the numerous major ports grappling with these issues. Many leading ports in both the United States and Europe feature multiple container terminals and multiple rail terminals (see Table 1). The predictive and cost analysis models developed in this research are applicable to most container ports worldwide that operate multiple out-terminals using different modes of inland transport.

Furthermore, the proposed methodological framework can be applied to predictive data analytics in various industries, comprising four key components essential for constructing predictive models and conducting cost analysis. Firstly, combining structured and unstructured data (such as text, audio, video files) is crucial for deriving substantial value to businesses (Harbert 2021). Secondly, feature engineering, informed by either practical insights or established theory, proves invaluable for optimal use of the datasets. In this research, the unstructured container content was transformed into a new feature using feature engineering and incorporated into the machine learning models to improve prediction performances for container out-terminals. Yoganarasimhan (2020) developed feature generation and feature categorization by capturing both global and individual-level attributes to improve prediction accuracy for personalized marketing. Chuang et al. (2021) applied operations management theory informed feature engineering to improve predictive accuracy of demand forecasting. Thirdly, quantification of feature contribution in machine learning enhances result interpretability, as demonstrated in Chuang et al. (2021). Lastly, evaluation of misclassification costs and quantifying the benefits of predictive model can convince industrial managers and facilitate model deployment. In this regard, Khan et al. (2014) and Tiwari et al. (2020) investigated misclassification costs for production quality inspection errors.

8.2. Research contributions

A primary objective of operations management is to conduct theoretically generalisable research that can aid the operations team to improve their decision making. Empirical research has enhanced operations management in two important ways: providing valuable inputs to models that improve operational decisions and identifying evidence of a phenomenon that happens in practice (Fisher et al. 2020). This research adopts a data-enabled approach and develops a predictive model that automates yard decision-making in a shipping port by integrating container information from various sources. This model offers a new avenue to address the challenges of container yard management and port congestion by predicting the out-terminals of import containers and utilising this knowledge to develop container storage strategy. The predictive models have the potential to integrate data from multiple departments or businesses. The significant cost reductions derived from deploying the predictive model can serve as financial incentives to promote cross-functional collaboration and information integration. A new research paradigm may be opened if the out-terminal predictive model could be integrated with other port operations, e.g. slot assignment, yard container housekeeping, and vehicle and crane scheduling. Development of these models constitutes our generalisable contributions to the empirical research methodology for improving operational decisions.

The novelty of the methodological framework lies in several aspects, including the extraction of valuable insights from unstructured data to enrich training dataset for improved prediction accuracy, the provision of global and local interpretability of predictive models and the measurement of added values in operational decisions through scenario modelling. Of particular significance is the explicit quantification of the costs associated with misclassifications. Such costs are critical to business decision making, but they have seldom been investigated in literature focused on applying data analytics and machine learning in an operations management context.

8.3. Managerial contributions

The lack of information on discharged containers' inland transport such as when the import containers will be picked up and which out-terminal these containers will go through is a common phenomenon. Such a phenomenon is unlikely to change due to individual cargo owners or their agents' delayed dynamic decision-making behaviour. This research attempted to adopt a data-enabled approach to processing heterogeneous datasets and gaining insights into container information, and then utilise these insights to predict out-terminals of the import containers. For *Port A*, the developed predictive tool effectively separated rail containers from haulier containers with a precision rate of 60 %. At the same time, the model can capture circa 35–60 % rail containers and store them near their target out-terminals, and therefore reduce the unproductive crane movement, vehicle travelling distance, and the possible rail misses.

Practice-informed feature selection and knowledge-informed feature engineering of unstructured data improve machine learning model prediction accuracy and interpretability. For *Port A*, the predicted out-terminals and discharged container attributes will be synchronised with yard information, crane information, and tractor information in real time, to assist long term tactical yard planning, while balancing with short term workload decisions.

This research can support shipping ports in exploiting the operational data held in-house and generating real-time insights and predictions to improve efficiency of yard operations and reduce costs. The substantial cost savings resulting from the implementation of the predictive model can act as a catalyst for a financial motivation to encourage collaboration across shipping supply chain stakeholders and information integration (Simatupang and Sridharan, 2008), which will lead to new disruptive supply chain business models characterised by dynamic allocation of processes and dynamic supply chain structures (Ivanov et al. 2021).

Integrating proposed predictive models into existing operations would cause minimal disruption, which holds particular significance for the port industry, known for its conservative and risk-averse nature (Chen and Shen 2022). The predictive model has the capability to provide real-time forecasts and streamline the handling of routine and repetitive decisions. This shift toward real-time predictive operations management decision-making represents a transformative development that will reshape the landscape of strategic, tactical, and operational decision-making processes (Ivanov et al. 2021).

9. Conclusions and future work

The growing demand to tackle high yard density and port congestion in the global supply chain requires more efficient operations of container movements at seaports. This motivates terminal operators to seek potential ways to improve port operations. One untapped research area is to incorporate the out-terminal information of import containers into yard management. Because the out-terminal information does not exist at the time when containers are unloaded from vessels, we developed a machine learning predictive tool to forecast the out-terminals of imported containers when they are discharged from vessels, and conducted cost analysis of misclassifying containers. The predictive tool and the cost analysis tool have the following attributes: 1) informed by the operation practice at the port, feature selection and feature engineering processes were performed to select container attributes to develop the prediction tool for container out-terminals; 2) combining embeddings with Cosine similarity, the prediction tool conducted unsupervised classification of unstructured data to improve prediction performance; 3) SHAP analysis was conducted to interpret contributions of container attributes and identify the key attributes that distinguish container out-terminals; and 4) misclassification costs were evaluated based on the multiclass confusion matrix (implies Type I and Type II errors), and the added values of the predictive models were measured and compared to the Business-as-Usual scenario. The predictive tool effectively classifies containers to their out-terminals and reduces the total transportation-related costs, albeit with misclassification costs. The SHAP shows improved better consistency with human intuition.

This study initiates a new avenue to address yard management and port congestion. It explores how advanced data analytics can yield valuable insights, aiding operational decision-making and enhancing overall efficiency. In the context of a shipping port, the predictive model offers the ability to make real-time prediction and streamline routine decisions, particularly concerning out-terminals. The proposed predictive model can be synchronized with port infrastructure and resource data to facilitate tactical yard planning. This extension of data analytics and machine learning from the operational level to tactical decision-making marks a significant advancement in accelerating digitisation in seaport operations. The transition to real-time predictive operations management signifies a transformative shift that will reshape strategic, tactical, and operational decision-making processes. The substantial cost saving derived from the data analytics and predictive models can assist in designing effective incentives to encourage collaboration and information sharing among supply chain participants. This, in turn, could result in more aligned decision-making, improved operational efficiency, and increased overall profits throughout the supply chain.

Several intriguing avenues for further research can be pursued. **Firstly**, many classification methods, like logistic regression or Random Forest, do not directly classify observations but provide a probability estimate of class membership. Typically, a probability threshold of 0.5 is used; observations with a probability above 0.5 are classified as the target class, otherwise as the default class. However, this default threshold may not be optimal for imbalanced class distributions. To address this, future research will perform threshold tuning to explore various probability thresholds and find the one that optimizes precision and recall while accounting for class imbalances (H, R0, and R1). **Secondly**, future research will consider synchronising the results from the predictive models with real time yard, crane and tractor information, to support long term tactical yard planning while also considering short-term workload decisions. The incorporation of predictive models into yard housekeeping and container pre-staging optimisation (Xie and Song 2018) warrants exploration. Additionally, the predictive models can be integrated with intricate process of container stacking, reshuffling, and retrieval to reduce unproductive reshufflings (Feng et al. 2020). **Thirdly**, it is desirable to obtain additional container information from shipping lines or freight forwarders to enrich the features in the training dataset, explore the impacts of the new features, and improve the prediction accuracy. Incentive alignment mechanisms will be a focal point, to foster collaboration and promote information integration across supply chains.

CRediT authorship contribution statement

Ying Xie: Writing – original draft, Project administration, Visualization, Data curation, Methodology, Funding acquisition, Validation, Software, Conceptualization, Resources, Writing – review & editing, Supervision, Formal analysis, Investigation. **Dong-Ping Song:** Software, Visualization, Project administration, Resources, Investigation, Data curation, Writing – original draft, Supervision, Funding acquisition, Writing – review & editing, Validation, Methodology, Formal analysis. **Jingxin Dong:** Investigation, Writing – review & editing, Methodology, Visualization, Data curation, Validation, Formal analysis, Conceptualization. **Yuanjun Feng:** Visualization, Data curation, Methodology, Formal analysis, Validation, Conceptualization, Writing – review & editing, Investigation.

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.

Acknowledgement

This work was partially supported by the UK Engineering and Physical Sciences Research Council (EPSRC) [grant number: EP/W028492/1; EP/Y024605/1].

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tre.2025.104331.

References

Abdi, A., Amrit, C., 2024. Enhancing vessel arrival time prediction: a fusion-based deep learning approach. Expert Syst. Appl. 252, 123988.
Abouelrous, A., Dekker, R., Bliek, L., Zhang, Y., 2025. Real-time policy for yard allocation of transshipment containers in a terminal. Transp. Res. B Methodol. 192, 103138.
Al-Dhaheri, N., Jebali, A., Diabat, A., 2016. A simulation-based Genetic Algorithm approach for the quay crane scheduling under uncertainty. Simul. Model. Pract. Theory 66, 122–138.
Aremu, O.O., Hyland-Wood, D., McAree, P.R., 2020. A machine learning approach to circumventing the curse of dimensionality in discontinuous time series machine data. Reliab. Eng. Syst. Saf. 195, 106706.
Bakar, N.N.A., Bazmohammadi, N., Çimen, H., Uyanik, T., Vasquez, J.C., Guerrero, J.M., 2022. Data-driven ship berthing forecasting for cold ironing in maritime transportation. Appl. Energy 326, 119947.
Barua, L., Zou, B., Zhou, Y., 2020. Machine learning for international freight transportation management: a comprehensive review. Res. Transp. Bus. Manag. 34, 100453.
Bocca, F.F., Rodrigues, L.H.A., 2016. The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. Comput. Electron. Agric. 128, 67–76.

Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., Lang, M., 2020. Benchmark for filter methods for feature selection in high-dimensional classification data. Comput. Stat. Data Anal. 143, 106839.
Branco, P., Torgo, L., Ribeiro, R.P., 2016. A survey of predictive modeling on imbalanced domains. ACM Computing Surveys (CSUR) 49, 1–50.
Caserta, M., Schwarze, S., Voß, S., 2020. Container rehandling at maritime container terminals: a literature update. In: Böse, J.W. (Ed.), *Handbook of Terminal Planning*, 2nd ed. Springer, Cham, pp. 343–382.
Chandrashekar, G., Sahin, F., 2014. A survey on feature selection methods. Comput. Electr. Eng. 40 (1), 16–28.
Chen, X. and Shen, C. (2022). Port digitalisation progressing, but challenges remain, *Lloyds List*, 15 Jul 2022.
Chen, C.W., Tsai, Y.H., Chang, F.R., Lin, W.C., 2020. Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results. Expert. Syst. 37 (5), e12553.
Chuang, H.H.C., Chou, Y.C., Oliva, R., 2021. Cross-item learning for volatile demand forecasting: an intervention with predictive analytics. J. Oper. Manag. 67, 828–852.
Eurogate (2023), Eurogate Intermodal transport, accessed on 31st Aug, 2023.
Feng, Y., Song, D., Li, D., Zeng, Q., 2020. The stochastic container relocation problem with flexible service policies. Transp. Res. B Methodol. 141, 116–163.
Feng, X., He, Y., Kim, K.H., 2022a. Space planning considering congestion in container terminal yards. Transp. Res. B Methodol. 158, 52–77.
Feng, Y., Song, D.P., Li, D., 2022b. Smart stacking for imported containers using customer information at automated container terminals. Eur. J. Oper. Res. 301 (2), 502–522.
Fernández, A., García, S., Galar, M., Prati, R.C., Krawczyk, B., Herrera, F., 2018. Learning from imbalanced data sets. Springer, Cham.
Filom, S., Amiri, A.M., Razavi, S., 2022. Applications of machine learning methods in port operations–a systematic literature review. Transport. Res. Part E Logist. Transport. Rev. 161, 102722.
Fisher, M., Olivares, M., Staats, B.R., 2020. Why empirical research is good for operations management, and what is good empirical operations management? Manuf. Serv. Oper. Manag. 22 (1), 170–178.
Freightwaves (2023), Intermodal rail taking growing share of containers out of NY-NJ Ports - FreightWaves, accessed on 31st Aug, 2023.
Ghaddar, B., Naoum-Sawaya, J., 2018. High dimensional data classification and feature selection using support vector machines. Eur. J. Oper. Res. 265, 993–1004.
Guldogan, E.U., 2011. Simulation-based analysis for hierarchical storage assignment policies in a container terminal. SIMULATION 87 (6), 523–537.
Gumuskaya, V., van Jaarsveld, W., Dijkman, R., Grefen, P., abd Veenstra, A., 2020. A framework for modelling and analysing coordination challenges in hinterland transport systems. Maritime Econom. Logist. 22, 124–145.
Harbert, T. (2021). Tapping the power of unstructured data – *MIT Sloan School of Management*, Feb 1, 2021.
He, J., Xiao, X., Yu, H., Zhang, Z., 2022. Dynamic yard allocation for automated container terminal. Ann. Oper. Res. 1–22.
Hsu, H.H., Hsieh, C.W., Lu, M.D., 2011. Hybrid feature selection by combining filters and wrappers. Expert Syst. Appl. 38 (7), 8144–8150.
Hu, X., Liang, C., Chang, D., Zhang, Y., 2021. Container storage space assignment problem in two terminals with the consideration of yard sharing. Adv. Eng. Inf. 47, 101224.
Huynh, N., Hutson, N., 2008. Mining the sources of delay for dray trucks at container terminals. Transp. Res. Rec. 2066 (1), 41–49.
Ivanov, D., Tang, C.S., Dolgui, A., Battini, D., Das, A., 2021. Researchers' perspectives on Industry 4.0: multi-disciplinary analysis and opportunities for operations management. Int. J. Prod. Res. 59 (7), 2055–2078.
Jahangard, M., Xie, Y., Feng, Y., 2025. Leveraging machine learning and optimization models for enhanced seaport efficiency. Maritime Econom. Logist. 1–42.
Jin, J.G., Lee, D.H., Cao, J.X., 2016. Storage yard management in maritime container terminals. Transp. Sci. 50 (4), 1300–1313.
Kang J., Ryu K.R. and Kim K.H. (2006a). determination of storage locations for incoming containers of uncertain weight. In: Ali M., Dapoigny R. (eds) *Advances in Applied Artificial Intelligence*. IEA/AIE 2006. Lecture Notes in Computer Science, vol 4031. Springer, Berlin, Heidelberg, 1159-1168.
Kang, J., Ryu, K.R., Kim, K.H., 2006b. Deriving stacking strategies for export containers with uncertain weight information. J. Intell. Manuf. 17, 399–410.
Khan, M., Jaber, M.Y., Ahmad, A.R., 2014. An integrated supply chain model with errors in quality inspection and learning in production. Omega 42 (1), 16–24.
Ko, Y., Seo, J., 2000. Automatic text categorization by unsupervised learning. *In COLING 2000 Volume 1: the 18th International Conference on Computational Linguistics*.
Kourounioti, I., Polydoropoulou, A., Tsiklidis, C., 2016. Development of Models predicting Dwell Time of import Containers in Port Container Terminals – an Artificial Neural Networks Application. Transp. Res. Procedia 14, 243–252.
Lee, D.H., Jin, J.G., 2013. Feeder vessel management at container transshipment terminals. Transport. Res. Part E: Logist. Transport. Rev. 49 (1), 201–216.
Lee, C.Y., Song, D.P., 2017. Ocean container transport in global supply chains: overview and research opportunities. Transp. Res. B Methodol. 95, 442–474.
Lee, D.H., Jin, J.G., Chen, J.H., 2012. Terminal and yard allocation problem for a container transshipment hub with multiple terminals. Transport. Res. Part E: Logist. Transport. Rev. 48 (2), 516–528.
Legros, B., Bouchery, Y., Fransoo, J., 2019. A time-based policy for empty container management by consignees. Prod. Oper. Manag. 28 (6), 1503–1527.
Li, T.S., Huang, C.L., 2009. Defect spatial pattern recognition using a hybrid SOM–SVM approach in semiconductor manufacturing. Expert. Syst.. Appl. 36 (1), 374–385.
Li, N., Sheng, H., Wang, P., Jia, Y., Yang, Z., Jin, Z., 2022. Modeling categorized truck arrivals at ports: big data for traffic prediction. IEEE Trans. Intell. Transp. Syst. 24 (3), 2772–2788.
Lundberg, S., Lee, S.I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Proces. Syst.
Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.I., 2020. From local explanations to global understanding with explainable AI for trees. Nat. Mach. Intell. 2 (1), 56–67.
Maldonado, S., González-Ramírez, R.G., Quijada, F., Ramírez-Nafarrate, A., 2019. Analytics meets port logistics: a decision support system for container stacking operations. Decis. Support Syst. 121, 84–93.
Moini, N., Boile, M., Theofanis, S., Laventhal, W., 2012. Estimating the determinant factors of container dwell times at seaports. Maritime Econom. Logist. 14, 162–177.
Port Technology, 2018. New Port of Felixstowe System Causes Delays. https://www.porttechnology.org/news/new_port_of_felixstowe_system_causes_delays/. accessed on 21st July, 2025.
Rekik, I., Elkosantini, S., Chabchoub, H., 2018. A case based heuristic for container stacking in seaport terminals. Adv. Eng. Inf. 38, 658–669.
Shafiabady, N., Lee, L., Rajkumar, R., Kallimani, V., Akram, N.A., Isa, D., 2016. Using unsupervised clustering approach to train the support vector machine for text classification. Neurocomputing 211, 4–10.
Sharif, O., Huynh, N., 2013. Storage space allocation at marine container terminals using ant-based control. Expert Syst. Appl. 40 (6), 2323–2330.
Simatupang, T.M., Sridharan, R., 2008. Design for supply chain collaboration. Bus. Process. Manag. J. 14 (3), 401–418.
Song, D.P., 2021a. A Literature Review, Container Shipping Supply Chain: Planning Problems and Research Opportunities. Logistics 5 (2), 41.
Song, D.P., 2021b. *Container Logistics and Maritime Transport*. Routledge, London.
Tiwari, S., Kazemi, N., Modak, N.M., Cárdenas-Barrón, L.E., Sarkar, S., 2020. The effect of human errors on an integrated stochastic supply chain model with setup cost reduction and backorder price discount. Int. J. Prod. Econ. 226, 107643.
Van den Berg, R., De Langen, P.W., 2015. Towards an 'inland terminal centred' value proposition. Marit. Policy Manag. 42 (5), 499–515.
van Marle, G. (2022). *Port sector players get together again – but their views on congestion are apart*, The Loadstar, 2022.
Wagener, N., 2014. Intermodal transport in Europe—Opportunities through innovation. LogForum 10 (4), 371–382.
Watkins, E. (2021). *Dwell times hit 'alarming' highs in Southern California ports*, Lloyds List, 2021.
Xie, Y., Song, D.P., 2018. Optimal planning for container prestaging, discharging, and loading processes at seaport rail terminals with uncertainty. Transport. Res. Part E: Logist. Transport. Rev. 119, 88–109.
Yang, Z., Wan, C., Yu, Q., Yin, J., Yang, Z., 2023. A machine learning-based Bayesian model for predicting the duration of ship detention in PSC inspection. Transport. Res. Part E Logist. Transport. Rev. 180, 103331.
Yoganarasimhan, H., 2020. Search personalization using machine learning. Management Science. 66 (3), 1045–1070.

Yu, M., Qi, X., 2013. Storage space allocation models for inbound containers in an automatic container terminal. Eur. J. Oper. Res. 226 (1), 32–45.
Yu, J., Tang, G., Song, X., Yu, X., Qi, Y., Li, D., Zhang, Y., 2018. Ship arrival prediction and its value on daily container terminal operation. Ocean Eng. 157, 73–86.
Zhang, C., Liu, J., Wan, Y.W., Murty, K.G., Linn, R.J., 2003. Storage space allocation in container terminals. Transp. Res. B Methodol. 37 (10), 883–903.
Zhang, C.R., Guan, H., Yuan, Y.F., Chen, W.W., Wu, T., 2020. Machine learning-driven algorithms for the container relocation problem. Transp. Res. B Methodol. 139, 102–131.
Zhou, C., Wang, W., Li, H., 2020. Container reshuffling considered space allocation problem in container terminals. Transport. Res. Part E: Logist. Transport. Rev. 136, 101869.

Predicting out-terminals for imported containers at seaports using machine learning: Incorporating unstructured data and measuring operational costs due to misclassifications

Xie, Ying

2025-10

Xie Y, Song D-P, Dong J, Feng Y. (2025) Predicting out-terminals for imported containers at seaports using machine learning: Incorporating unstructured data and measuring operational costs due to misclassifications. Transportation Research Part E: Logistics and Transportation Review, Volume 202, October 2025, Article number 104331

https://doi.org/10.1016/j.tre.2025.104331