Hindawi
Journal of Advanced Transportation
Volume 2023, Article ID 2144733, 17 pages
https://doi.org/10.1155/2023/2144733

Research Article

Machine Learning for Promoting Environmental Sustainability in Ports

Meead Mansoursamaei,[1,2] **Mahmoud Moradi**,[1] **Rosa G. González-Ramírez**,[3] **and Eduardo Lalla-Ruiz**[2]

[1]*University of Guilan, Rasht, Iran*
[2]*University of Twente, Enschede, Netherlands*
[3]*Facultad de Ingeniería y Ciencias Aplicadas, Universidad de los Andes, Santiago, Chile*

Correspondence should be addressed to Meead Mansoursamaei; m.mansoursamaei@utwente.nl

Received 19 May 2022; Revised 12 September 2022; Accepted 10 February 2023; Published 3 March 2023

Academic Editor: Dongjoo Park

Maritime transportation is one of the essential drivers of the global economy as it enables both lower transportation costs and intermodal operations across multiple forms of transportation. Maritime ports are essential interfaces that support cargo handling between sea and hinterland transportation. Besides, in this area, environmental protection is becoming extremely important. Global warming, air pollution, and greenhouse gas emissions are all having a detrimental influence on the environment and will most likely continue to do so for future generations. Hence, there is a growing need to promote environmental sustainability in maritime-based transportation. The application of machine learning (ML), as one of the main subdomains of artificial intelligence (AI), can be considered a component within the process of digital transformation to advance green activities in maritime port logistics. Thus, this article presents the results of a systematic literature review of the recent literature on machine learning for promoting environmentally sustainable maritime ports. It collects and analyses the articles whose contributions lie in the interplay between three main dimensions, i.e., machine learning, port-related operations, and environmental sustainability. Throughout a review protocol, this research is constituted on the major focuses of impact, problems, and techniques to discern the current state of the art as well as research directions. The research findings indicate that the articles using polynomial regression models are dominant in the literature, and the recurrent neural network (RNN) and long short-term memory (LSTM) are the most recent approaches. Moreover, in terms of environmental sustainability, emissions and energy consumption are the most studied problems. mAccording to the research gaps observed in the review, two broad directions for future research are identified: (i) altering attention on a greater diversity of machine learning approaches for promoting environmental sustainability in ports and (ii) leveraging new outlooks to perform more green practical works on port-related operations.

1. Introduction

Maritime freight transportation is one of the vital drivers of the global economy as it enables both lower transportation costs and faster intermodal operations across multiple forms of transportation [1, 2]. Indeed, maritime ports are the essential interfaces that support cargo handling between sea and hinterland. Besides, environmental sustainability has become one of the important foundations on the agenda of many maritime ports due to the challenges of climate change as well as the growing demands of the logistics and transportation sectors [3]. Expanding maritime transportation activities has enabled urban economies to prosper to some extent; however, they also have caused resource waste and environmental pollution. To achieve the sustainable growth of ports and cities, energy-saving and emission-reduction methods have to be used [4, 5]. The numerous sources and types of port-related emissions, such as those from maritime vessels, trucks, and cargo-handling machines, have a considerable impact on the environment [6]. Moreover, a significant proportion of emissions in ports are due to interterminal transportation and container drayage operations [7]. Therefore, the research on environmental enhancements in maritime ports as main transportation nodes has grown in importance as they allow for improvements in a variety of areas of environmental sustainability that governments and entrepreneurs are attempting to address [3].

Artificial intelligence (AI) approaches and subdisciplines can be divided into two main categories: (i) reasoning and decision making and (ii) learning [8]. While the first set is focused on decision making concerning planning, solution searching, and optimization, the second set relates to learning, communication, and perception. In this sense, learning refers to the ability of systems to automatically learn, determine, predict, adapt, and react to changes without being explicitly programmed [9]. The techniques related to learning include neural networks, deep learning, reinforcement learning, and decision trees, among others [8, 10], and the three main subdomains are supervised, unsupervised, and reinforcement learning [11].

In this research, the application of ML as one of the main subdomains of AI to advance green activities in port logistics is studied. A supervised learning model as a subdomain of ML can learn and check the plausibility of schedules and predict the energy consumption of battery-electricauto-guided vehicles (AGVs) in horizontal transportation areas [12]. Neural networks are being used in the ports of Hamburg, Rotterdam, and Singapore. For instance, Hamburger Hafen und Logistik AG (HHLA) has become one of the first ports in the world to implement ML solutions for its Hamburg container terminals to predict the dwell time of a container at the yard of the terminal [13]. Regarding environmental concerns and emissions from maritime traffics at the port of Rotterdam, machine learning techniques have been developed to predict the estimated time of arrival (ETA) for containerships to reduce waiting time and thus emissions [14]. The SAFER project of the maritime and port authority (MPA) of Singapore and IBM have implemented seven ML-based modules within Singapore port waters in order to predict arrival times of vessels and potential congestion areas as well as to detect offending ships to environmental regulations [15]. Therefore, an ever-increasing demand for ML-based technologies to support green ports is arising. Besides, several relevant works have been published recently in journals and proceedings. Hence, a systematic literature review (SLR) that analyses the articles investigating green activities in maritime ports using ML applications becomes necessary.

This study with the purpose of investigating the applications of machine learning techniques to stimulate environmental actions in maritime ports seeks to explore the state-of-the-art research within the interplay of machine learning, environmental sustainability, and maritime ports. Thus, the current research contributes to the literature by identifying research categories based on a systematic literature review (SLR), i.e., impacts, problems, and techniques.

Most recently, the sustainability of ports is receiving more attention. The number of research on the topic has increasingly expanded and been published in some literature reviews. However, port sustainability in terms of environmental impacts is currently experiencing new challenges; hence, the literature needs to be updated. In addition to the previous reviews, the studies that have used machine learning approaches as solutions to cut down on port-related pollution have been considered in this research. Concerns among stakeholders, including port officials, policymakers, users, and citizens have grown over the environmental sustainability of ports and machine learning techniques as technological solutions. Therefore, to direct the future growth of this topic, a timely literature review is essential.

The article is organized into four sections. Section 2 presents the research methodology by describing the data collection and analysis methods along with initial statistics regarding recent publication trends in ML applications in green-port operations. Section 3 presents the initial bibliometric analysis. Then, relying on the bibliometric analysis, the evolution of ML in port logistics is studied. First, the evolving process of ML applied to aid sustainable maritime port operations based on the growing volume of publications over time is examined. Second, the impact and citation patterns to characterize the selected articles are presented. Then, the most cited articles and researchers with a higher number of contributions within the scope of the review up to 15th October 2021 are presented. Finally, the evolution and interaction of topics, techniques, and problems are scrutinized. The article concludes with a discussion on research limitations and potential research directions.

2. Research Methodology

There are several reporting guidelines to perform a systematic literature review [16–18]. Such reporting guidelines, e.g., SRQR (standards for reporting qualitative research), ENTEQ (enhancing transparency in reporting the synthesis of qualitative research), eMERGe (meta-ethnography reporting guidelines), and RAMESES (realist and meta-narrative evidence syntheses: evolving standards) are suggested for qualitative reviews [19]. Likewise, the PRISMA (preferred reporting items for systematic reviews and meta-analyses) proposes a standardized method for ensuring transparent and thorough systematic review reporting [20]. PRISMA is a systematic review protocol that includes a 27-item checklist and a four-phase flow diagram [21] (see Figure 1). In this study, PRISMA was selected as the methodology for the systematic review over other existing protocols because of its comprehensiveness, its use in a variety of disciplines, and its potential to improve consistency across reviews. This protocol involves (i) the definition of research questions and (ii) the identification of the search string and source selection.

2.1. Definition of Research Questions. The first step of this review is the definition of the research questions (RQs) related to the main research question of this work, i.e., how machine learning has been used to promote environmentally sustainable port operations. This way, the following research questions are addressed:

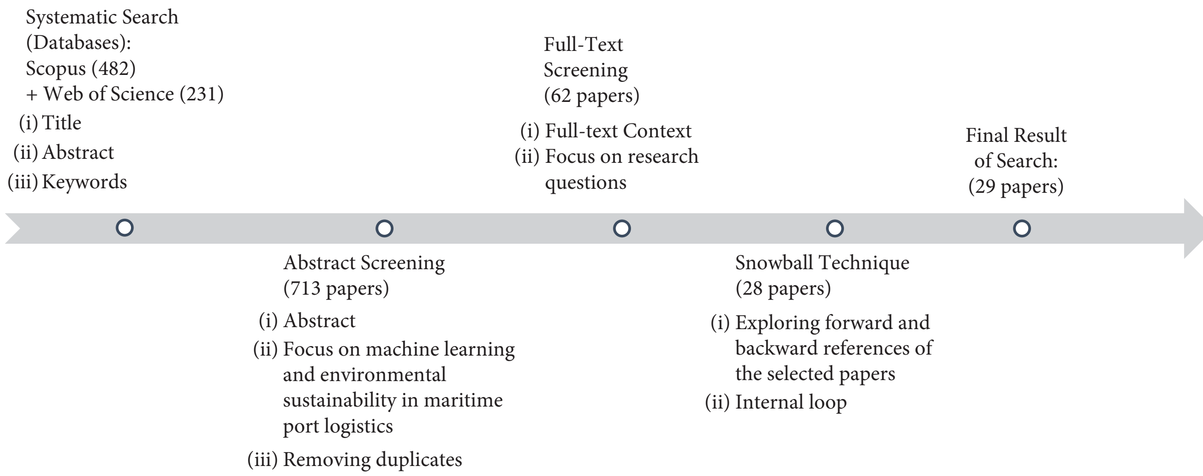

Figure 1: The overall procedure for selecting and filtering the articles [21].

(i) *RQ1.* What is the environmental impact of ML approaches on ports operations? (Impact related)
(ii) *RQ2.* Which environment-related port operations have ML techniques being used for? (Problem related)
(iii) *RQ3.* How ML techniques have been used to promote environmentally sustainable port operations? (Technique related)

2.2. Identification of Search String and Source Selection. To achieve the final result of the search, several steps including the systematic search, abstract screening, full-text screening, and snowball technique were applied. The Scopus and Web of Science (WoS) databases were utilized to collect and select related references systematically. The query used in this work is composed of a list of keywords distributed into three main pillars (i.e., machine learning, maritime ports, and environmental sustainability) as shown in Figure 2. The AND operator delimited each group, while the OR operator bundled all keywords within each group. The AND NOT operator eliminated irrelevant subject areas. After that, the query was used in the abovementioned databases to find those articles considering the selected keywords within their titles, abstracts, or keywords from 2017 to 2021. By the initial search, 482 articles from Scopus and 231 articles from WoS were discovered from the databases; however, English articles were included as a limitation and irrelevant subject areas such as agriculture, medicine, and chemical engineering were excluded. Moreover, since this review is focused on ports, maritime shipping-related keywords were excluded. The last search was run on 15th October, 2021.

Information of articles including the title, abstract, publication year, and sources were screened to determine whether a article is included or not. Based on the three pillars of our scope, there was one reason to discard articles, namely, the articles whose contribution did not lie within the interplay between machine learning, maritime port operations, and sustainability were removed. After the abstract and full-text screening, a total of 28 articles were collected. During the selection process, the forward and backward snowballing technique was applied to all the selected articles to minimize the risk of missing relevant studies [22]. This resulted in addition of one more article leading to a total of 29 articles that were analyzed in this SLR. The procedure was carried out by two of the authors of this manuscript who independently screened the selected articles. Finally, there was a consensus among all authors about the articles that were excluded [23]. Figure 1 indicates a flow diagram of the search procedure and the results obtained at each stage.

3. Discussion and Results

In order to answer the research questions raised in Section 2.1, three subsections, each corresponding to a research question, are provided. First, the impact of selected articles is discussed, and the most cited works and authors are listed. Second, based on the physical area of port-related operations, the environmental problems of the selected articles are identified and discussed. Third, the machine learning-based techniques that have been proposed in the literature to address the environmental problems are outlined.

3.1. Impact Analysis (RQ1). With the aim of answering RQ1 involving impact, the number of works published so far, the number of citations, and data collection methods are investigated.

The number of articles published annually from 2017 to 2021 by the source type is shown in Figure 3. Based on the data, it is observed that the majority of works (25 articles) were published in journals, while the other 4 articles were published in conference proceedings. The Journal of Cleaner Production with four articles and Transportation Research Part D with three articles are the journals where more of the selected articles were published, while the other works (76%) are from various sources.

To illustrate the citation rate up to 15th October, 2021, Figure 4 displays a histogram. It can be observed that 25 of the 29 collected articles have been cited so far. The most cited article, [24] with 44 citations was published in the Transport Policy. In this article, the performance of 17 ports in China

Maritime ports logistics

(i) ("maritime" AND ("logistic*" OR "transport*" OR "port")) OR (("container" OR "cargo" OR "roro") AND ("terminal" OR "port" OR "harbor" OR "yard" OR "staking" OR "dispatching" OR "drayage" OR "railhead" OR "truck" OR "interchange area" OR "freight station")) OR ("yard" AND ("terminal" OR "operations" OR "storage" OR "truck" OR "management" OR "area")) OR "inter-terminal transport" OR "intermodal drayage" OR "seaport" OR "sea port" OR "bulk port" OR "landside" OR "seaside" OR "berth*" OR "crane" OR "stowage planning" OR "hinterland" OR "internal vehicles" OR "straddle carrier" OR "gantry carrier" OR "terminal gate" OR "gate area" OR "quay" OR "transport area" OR "horizontal transport system" OR "truck area" OR "train area" OR "transport area" OR "gate operations" OR "container storage"

Machine learning

(i) "machine learning" OR "active learning" OR "feature extraction" OR "adaptive learning" OR "generative adversarial network" OR "adversarial machine learning" OR "generative model" OR "adversarial network" OR "multi-task learning" OR "anomaly detection" OR "neural network" OR "artificial neural network" OR "pattern recognition" OR "automated machine learning" OR "probabilistic learning" OR "automatic classification" OR "probabilistic model" OR "automatic recognition" OR "recommender system" OR "bagging" OR "recurrent neural network" OR "bayesian modelling" OR "recursive neural network" OR "gaussian process" OR "reinforcement learning" OR "classification" OR "semi-supervised learning" OR "clustering" OR "statistical learning" OR "collaborative filtering" OR "statistical relational learning" OR "content-based filtering" OR "supervised learning" OR "unsupervised learning" OR "convolutional neural network" OR "support vector machine" OR "data mining" OR (data?driven) OR "data pre-processing" OR "transfer learning" OR "deep learning" OR "unstructured data" OR "deep neural network" OR "ensemble method" OR "regression" OR "dimen* reduc*" OR "K-Means" OR "principal component analysis" OR "kernel" OR "decision tree" OR "K-Nearest" OR "random forest" OR "Q-Learning" OR "proximal policy"

Environmental sustainability

(i) "green" OR "emission*" OR "sustainab*" OR "environment*" OR "noise" OR "waste" OR "energy*" OR "pollution" OR "biofuel*" OR "biomass" OR "battery" OR "decarboni?ation" OR "greenhouse" OR "renewable" OR "solar" OR "recycle" OR "carbon" OR "footprint" OR "contamination" OR "ecolog*" OR "spill" OR "electr*" OR "smart grid*"

AND NOT

(i) "fisher*" OR "algae" OR "trout" OR "salmon" OR "mussel" OR "molecular" OR "herring" OR "larva" OR "oyster" OR ("sea AND bass") OR "aerosol" OR "atmospheric chemistry" OR "atmospheric deposition" OR "troposphere" OR "sewage" OR "ozone" OR "microb*" OR "biology" OR "organism" OR "animal" OR "plants" OR "electron" OR "species" OR "eutrophication" OR "odor" OR "house" OR "building" OR "retinal" OR "steel" OR "fishing" OR "ore"

Figure 2: The used keywords based on the three pillars: maritime ports, machine learning, and environmental sustainability.

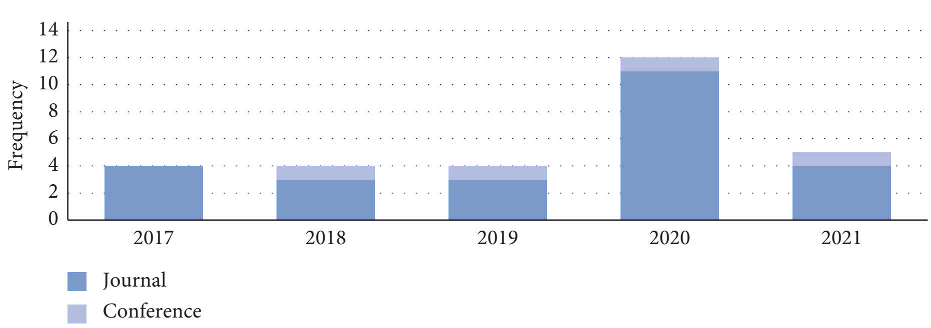

Figure 3: The number of articles published annually from 2017 to 2021 by source.

under environmental concerns using regression models was evaluated.

In terms of data collection methods and data sources, 21 out of 29 articles (72%) used secondary data sources from earlier works, research institutes, or governments. However, in 7 of the articles (24%), researchers gathered data by observations, measurements, and experiments as primary data sources. Nonetheless, three of the articles (10%) utilized interviews and surveys solely or as a part of their data collection method.

The environmental indicators used in port logistics applications include emissions, water pollution, noise

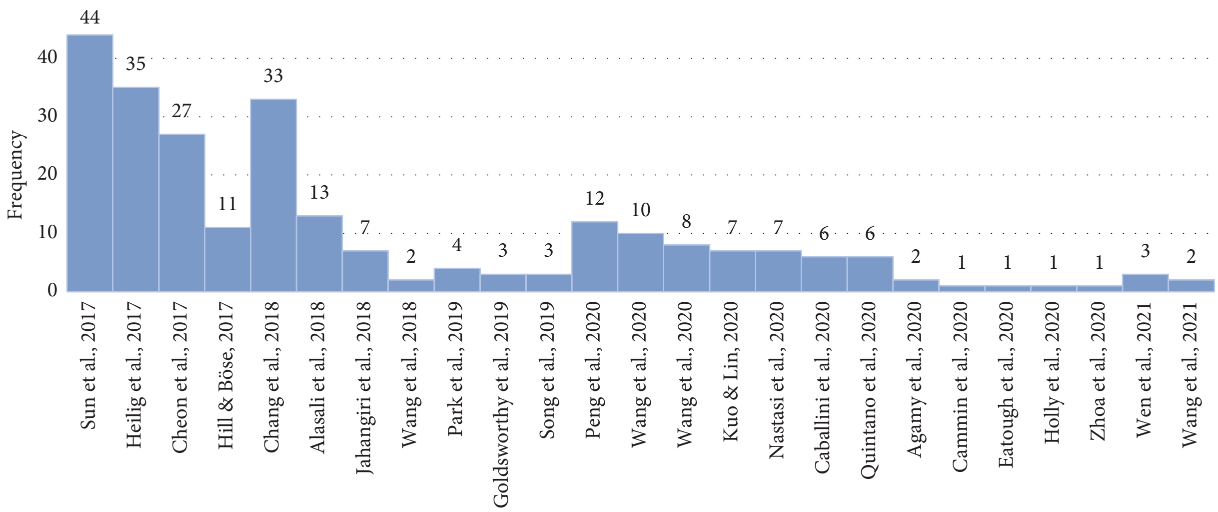

FIGURE 4: Citations of the selected articles (the last update was on 15th October, 2021).

pollution, solid waste, energy-saving, and renewable energy were reported in the collected articles. Therefore, among them, 25 articles (86%) raised emissions, 13 articles (45%) considered energy-saving, and 6 articles (21%) used water pollution as the environmental indicator of their work. Nonetheless, noise pollution, renewable energy, and solid waste, each with only one article (3%), are the smallest used environmental indicators. Figure 5 presents the contribution of the selected articles based on environmental sustainability indicators.

3.2. Port-Related Problems (RQ2). To answer RQ2, the works were categorized in terms of the port-related operations and their application areas in the port (i.e., seaside, yard, landside, and overall port areas). The specific machine learning techniques of those works are later discussed in Section 3.3. Therefore, based on the area within ports, the articles have been distributed according to the location of the addressed problems, which resulted in seaside operations with 28%, yard operations with 12%, and landside operations with 16%. Nonetheless, 44% of the articles have considered the environmental sustainability indicators for benchmarking, performance evaluation, and air quality prediction not being specifically focused on an area but the overall port. Hence, these contributions are categorized as the "overall port" in Figure 6.

3.2.1. Overall Port Areas. Several articles benchmarked and evaluated ports in terms of environmental efficiency and did not consider a specific area within port (i.e., seaside, yard, and landside), the reason for which they are classified as "overall port areas." In this regard, the authors of reference [24] evaluated 17 Chinese ports in terms of NOx emissions and energy savings. Similarly, the authors of reference [25] investigated those environmental problems in 18 ports of China. Using a different machine learning technique, the authors of reference [26] benchmarked 15 seaports in China in terms of wasted water treatment as well as air quality evaluation and energy-saving expenditures. Moreover, the authors of reference [27] evaluated the environmental efficiency of the Kaohsiung container port in Taiwan, considering the same problems. In the same way, the authors of reference [28] benchmarked the top 10 ports in the US and considered water pollution and emissions within port areas. Furthermore, the authors of reference [29] benchmarked operating practices of 20 ports in the US for enhancing environmental efficiency in terms of greenhouse gas emissions, oil spill prevention, and energy efficiency of port operations. Similarly, the authors of reference [30] benchmarked 24 container ports in Europe regarding emissions and energy-saving problems.

To evaluate ports environmental efficiency in emission control areas (ECAs) (emission control areas (ECAs), as outlined by Annex VI of the 1997 MARPOL protocol, are sea areas where regulations have been implemented to prevent emissions from ships), 23 ports in the Baltic and the North Sea and 25 non-ECA ports in Europe were investigated by [31] to examine the impact of ECA regulations on reducing emissions in European ports. Furthermore, using an Intergovernmental Panel on Climate Change (IPCC) method, the authors of reference [32] measured CO_2 emissions from the port container distribution (PCD) to evaluate the sustainable development ability of 30 ports in China. Likewise, the authors of reference [33] predicted air quality, fine particulate composition, and mass in the area of Long Beach port in California. In addition, the authors of reference [34, 35] predicted air quality and emissions in 4 ports in Turkey (Ambarli, Izmir, Mersin, and Kocaeli ports) and Busan Port in Korea, respectively. Moreover, the authors of reference [36] simulated the indoor air quality of roll-on/roll-off (RORO) ships and predicted pollution emitted from cars in maritime ports. Furthermore, the authors of reference [37] developed a container terminal logistics generalized computing architecture (CTL-GCA) for planning, scheduling and decision making to establish a better connection among liners and rubber-tired gantry cranes (RTGCs) and block community to reduce carbon emissions.

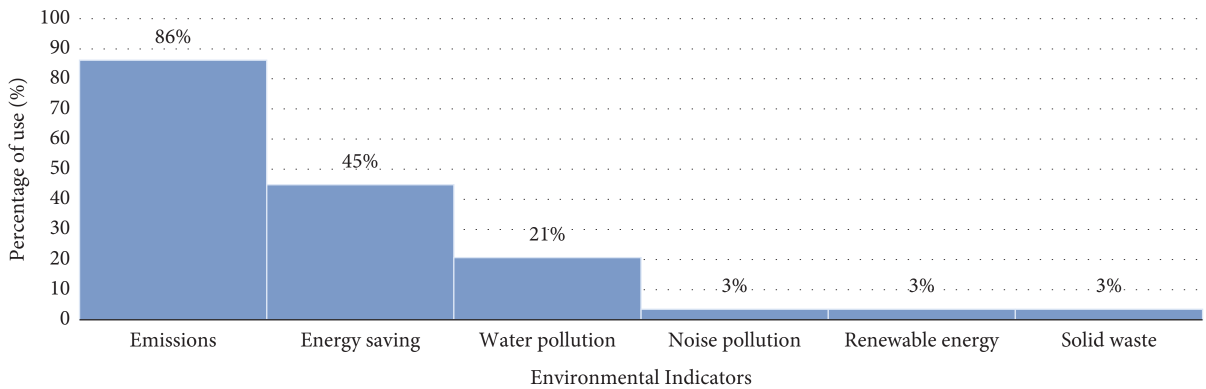

Figure 5: ML-based articles in terms of environmental sustainability in maritime port logistics.

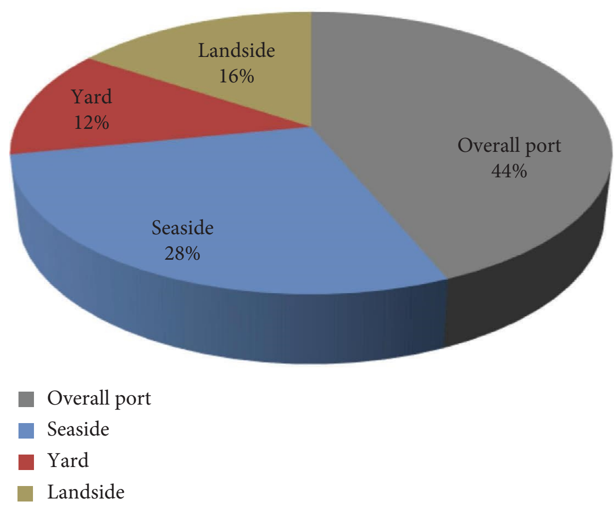

Figure 6: Distribution of selected articles based on application areas within the port.

3.2.2. Seaside Area. The seaside operations that received more attention are those regarding berth allocation planning. For instance, to manage real-time data and air emissions reduction in maritime ports due to the berth operations, [38] developed a predictive system for vessel arrivals, considering ship features and expanding estimated time of arrival (ETA) features to date, time, and weekday, based on the previous model presented by the authors of reference [39]. The authors of reference [40] simulated berth planning problems and predicted the arrival time of vessels using machine learning techniques. In the same way, the authors of references [41–43] and [44] used case studies to solve berth planning problems at different ports that are presented in detail in Table 1. From the other perspective, noise emissions by ships around the port areas are one of the important issues for port cities. This has been studied in [45] where the authors with machine learning techniques identified the affecting parameters of noise emitted by ships in the industrial port of Livorno, Italy.

3.2.3. Yard Area. Several contributions related to container and cargo handling in the yard area to mitigate greenhouse gas emissions, energy-saving, and promoting the use of renewable energy. For instance, the authors of reference [12] presented a study on how to use battery-electricauto-guided vehicles (AGVs) in the yard for handling containers in the port of Hamburg (Germany). They utilized a synthetic case by generating data for checking the plausibility of schedules and predicting energy consumption. Container cranes are also one of the main sources of energy consumption and pollution in the yard. In this regard, the authors of references [46, 47], and [48] considered environmental problems (i.e., energy consumption and emissions) of the rubber-tyred gantry (RTG) in the port of Felixstowe (UK) and port of Casablanca (Morocco) and a synthetic case, respectively.

3.2.4. Landside Area. Trucks are the main source of emissions in the landside area. Hence, several studies have paid attention to the environmental problems caused by trucks in the area. For instance, the authors of reference [49] proposed a forecasting engine for truck arrivals to logistics nodes, i.e., empty container depots, packing facilities, or terminals, to mitigate greenhouse gas emissions from truck congestions beyond the gates of an empty container depot in northern Germany. Based on the proposed model, companies can adjust their route planning to minimize truck waiting times.

Table 1: Problems and ML techniques for promoting environmental sustainability in maritime port logistics.

Articles	*Areas* Overall port	Seaside	Yard	Landside	Research scope	*Environmental problems* Emission	Water pollution	Noise pollution	Energy saving	Renewable energy	Solid waste	*Machine learning* Input	Techniques (tools)	Output	Data collection methods	Case studies
[24]	●				Port performance evaluation	●						Port assets, berth quantity, and geographical location	PR	Net profit, cargo throughput, and NOx emissions	Secondary data sources	17 Chinese ports
[28]	●				Port performance evaluation	●	●					Facility, vessel and other pollution incidents	KDE	A smoothed graph for the distribution of pollution incidents probability density	Secondary data sources	10 American ports
[49]				●	Truck scheduling	●						Historical data, truck arrival time, administrative waiting start and end time; intermediate waiting start and end time; node-specific forecasting parameters, e.g., dispatching modes and storage policies; and external forecasting parameters, weather and traffic information	NN (BP)	Waiting time, arrival rates that translates into a reduction of traffic congestion and air pollution	Secondary data sources	An empty container depot in Northern Germany
[50]				●	Truck scheduling	●						Number of clusters and the archive containing *n* solutions	K-means	Cluster centroids	Secondary data sources	Port of Hamburg (Germany)
[31]	●				Port performance evaluation	●						Berth length, the number of cranes, terminal area for the efficiency estimation. City gross domestic product, variance inflation factors, and emissions control regulations	PR	TEUs handled and the impact of emissions control regulations	Secondary data sources	48 ports in Europe
[46]			●		Yard crane demand				●			The average of the previous day load, the average of the previous week load, the same hour load for previous day, and the previous hour load	NN (BP) and SVM	RTG crane demand of one hour	Primary data sources	Port of Felixstowe in the UK
[40]		●			Berthing	●						Maximum continuous rate (MCR) measured by megawatt, shaft speed	PR	Emissions (NO_x, SO_x, CO_2, and CO)	Primary data sources	Two ocean-going vessels in Australia
[48]		●	●		Antiswing crane	●	●					Trolley position, trolley speed, loading angle, and angular velocity	NN (ANFIS)	The driving force of the trolley	Secondary data sources	—
[29]	●				Port performance evaluation	●	●		●			Number of quay crane, acres, berth and depth, undesirable output (CO_2), and desirable outputs (calls, throughput, and deadweight tonnage)	SOM	Clusters of decision-making units (DMUs)	Secondary data sources	20 American container ports
[41]	●				Berthing	●						Ship identification, position, speed, course, heading and navigational status, and timestamp	PR	Emissions (NO_x, SO_2, $PM_{2.5}$, VOC, CO, NH_3, CO, N_2O, and CH_4)	Primary data sources	Ports of Newcastle, Jackson, Botany, and Kembla in Australia
[44]		●			Berthing	●	●					Observed video	PCA	The background with low-rank property and the foreground with sparse property	Primary data sources	Unknown
[37]	●				Port performance evaluation	●			●			RTGC number, block number, handling container specification, stevedoring full or empty category, handling volume for a task, and the number of clusters	K-means	Resource allocation for container terminals	Secondary data sources	A container terminal on the east coast of China
[42]		●			Berthing				●			The net tonnage, deadweight tonnage, actual handling volume, and efficiency of facilities	GBoost, RF, NN (BP), PR, and KNN	Energy consumption	Secondary data sources	Jingtang port (China)
[25]	●			●	Port performance evaluation	●			●			CO_2 emission driver factors of the city where the port is located are gross domestic product, total resident population, the number of port berth, total imports, total exports, the first industrial value, the secondary industrial value, the primary industrial value, gross industrial production, fixed assets investment in the tertiary industry, per capita income, railway freight volume, highway freight volume, and waterway freight volume	Spatial clustering	Clusters of similar ports in terms of environmental sustainability (LISA cluster maps of PCD carbon emissions)	Secondary data sources	30 Chinese container ports

Table 1: Continued.

Articles	Areas				Research scope	Environmental problems						Machine learning			Data collection methods	Case studies
	Overall port	Seaside	Yard	Landside		Emission	Water pollution	Noise pollution	Energy saving	Renewable energy	Solid waste	Input	Techniques (tools)	Output		
[32]	●				Port performance evaluation	●			●		●	Number of berths, the length of the terminal, the number of staff, and the total fixed assets	PR	Cargo throughput, NO_X emissions, SO_X emissions, and solid waste containers	Secondary data sources	18 Chinese ports
[27]	●				Port performance evaluation	●	●		●			Highly correlated input variables	PCA	Determinant factors of the survey are lean management, green operational practices, green behavior (green participation and green compliance), and green climate	Survey	Kaohsiung container port (Taiwan)
												Determinant factors of the survey are lean management, green operational practices, green behavior (green participation, green compliance), and green climate	PR	Green performance (financial and nonfinancial)		
[45]		●			Noise of moving ships in port areas			●				Draught, speed, and ship-to-microphone distance	PR	Sound emitted	Primary data sources	Industrial port of Livorno (Italy)
[51]				●	Truck scheduling	●						Container features are cycle, type, weight, special (e.g., hazard shipping), agreement (between stakeholders), vessel departure time, distance (of two containers in the yard), customs clearance, dwell time, and final destination	Hierarchical clustering	Container groups	Secondary data sources, survey	Port of Altamira (Mexico) and Port of Genoa (Italy)
[30]	●				Port performance evaluation	●			●			Energy consumption and number of employees	Hierarchical clustering, PR	Total gross weight of goods, air pollutant emissions, and the rank of ports in terms of eco-efficiency	Secondary data sources	24 European container ports
[36]		●			Indoor air quality prediction (RORO)	●			●			CO concentration and load (number of cars)	NN (BP)	The reference flow rate of the ventilation system	Secondary data sources	A liner between Egypt and Saudi Arabia ports
[38]		●			Berthing	●						ETA features (date, time, and weekday) and ship features (ship type and length)	SVM	Arrival time of vessels	Secondary data sources	—
[33]	●				Air quality prediction	●						Fine particulate mass and fine particulate composition	PR	Air quality	Primary data sources	Long Beach (US)
[12]			●		AGV				●	●		Scheduled arrival, departure, and load/unload start time, planned berthing place, planned position of front and rear of the ship, and number of containers to load and unload	NN (BP)	Availability of AGVs	Primary data sources	Hamburg container terminal (Germany)
[52]				●	Container truck emissions	●						Highly correlated data of traffic and particle number concentrations (PNC)	PCA	Principal components (container truck volume, other vehicles volume, and PNC data)	Secondary data sources	Waigaoqiao port (China)
[34]	●				Air quality prediction	●						Type of pollutant, the operating mode, and gross tonnage of ships	PR	Emissions (SO_2, NO_x, CO_2, VOC, PM, and CO)	Secondary data sources	Ports of Ambarlı, Izmir, Mersin, and Kocaeli (Turkey)
[47]		●	●		RTG crane	●			●			Energy consumption of hoist, gantry, and trolley	PR	General energy consumption of RTG	Secondary data sources	Casablanca port (Morocco)
[35]	●				Air quality prediction	●						Meteorological data, air quality data, and shipping activity data	RNN and LSTM	Emissions ($PM_{2.5}$, PM_{10}, SO_2, O_3, NO_2, CO)	Secondary data sources	Busan port (Korea)
[43]		●			Berthing	●			●			Hourly data of energy (electricity) prices and load demands	LSTM, NN (BP), Elman, RBF	Day-ahead prices of energy	Secondary data sources	A navigation route in Australia
[26]	●				Port performance evaluation	●	●		●			Highly correlated input variables	PCA	Air quality, rate of treatment of wastewater, standard-reaching rate of nearshore water, green coverage rate in developed areas, and expenditure on energy-saving investments per capita	Secondary data sources, survey	15 Chinese seaports
												Air quality, rate of treatment of wastewater, standard-reaching rate of nearshore water, green coverage rate in developed areas, and expenditure on energy-saving investments per capita	Hierarchical clustering	The rank of ports based on environmental sustainability features		

Furthermore, considering truck emissions in the port of Hamburg, the authors of reference [50] developed a multi-objective model for interterminal truck routing problems and utilized a machine learning technique as part of the decision support system. Moreover, using two real container terminals, i.e., the port of Altamira (Mexico) and the port of Genoa (Italy) as case studies, the authors of reference [51] proposed a methodological framework to reduce empty truck trips to minimize the deviation from their preferred time slots and turnaround times in container terminals and reduce emissions. The authors of reference [52] studied the relationship between traffic volume and the particle number concentrations (PNC) caused by emissions of container trucks in the port of Waigaoqiao (China). For this, they combined a machine learning technique with statistical methods to characterize the variation of particles in the port area.

3.3. ML Techniques to Promote Green Port Operations (RQ3). Researchers or practitioners who seek to apply ML in maritime port operations should possess the fundamental competency of selecting an algorithm that is appropriate for a given task or problem. However, conceptualizing a way toward using ML to improve the performances of port operations is challenging in the absence of expertise or prior research of a similar type, especially when taking into account the numerous algorithms that have been offered in the technical literature.

ML is mainly classified into three different types, i.e., supervised learning, unsupervised learning, and reinforcement learning [53]. Given that division, a systematic literature review by [54] illustrated the machine learning techniques used in industrial applications so far which are organized in Table 2. The different tools related to (i) supervised learning with classification and regression algorithms, (ii) unsupervised learning with clustering and dimensionality reduction algorithms, and (iii) reinforcement learning, are presented in the table. In order to highlight current and emerging trends and, more importantly, to guide researchers or practitioners in the selection of ML techniques, the table demonstrating the subclassification of ML algorithms is used in this review to map techniques when analyzing the collected articles. For further information on the tools, see the reference of the table.

To provide a better vision of the problems (RQ2) and the techniques (RQ3) discussed in this SLR, Table 1 is presented. Table 1 summarizes the main characteristics of the reviewed articles based on the following categories: the application area within the port (i.e., seaside, yard, landside, and overall port), the research scope, environmental problems (i.e., emissions, water pollution, noise pollution, energy saving, renewable energy, and solid waste), the machine learning technique, involved factors (input and output), the data collection method, and the case study. The information provided in this table is discussed in Sections 3.1, 3.2, and 3.3.

Considering the ML techniques used in the collected articles, a hierarchical categorization including the ML techniques, the subdomains, and the tools is shown in Figure 7. As can be observed in the figure, supervised learning with 70% is the most used type of technique while unsupervised learning is the second option with 30% of collected articles. Particularly, polynomial regression (PR) with 30% and neural networks (NN) with 27.5% are the most used tools. Therefore, according to the main categorization of ML techniques discussed in Section 3.3, there is no article using classification nor reinforcement learning among the collected articles.

3.3.1. Supervised Learning. Supervised learning, commonly called predictive learning, is used for labelled datasets in which the response of a scenario or example is known [53]. It enables several regression tools (e.g., polynomial regression, neural networks, and k-nearest neighbour) for predicting the behaviour of a dataset. Moreover, some classification tools (e.g., Bayesian network, logistic regression, and decision tree) are other applications of supervised learning when the output is categorical [54]. As seen in Table 1, only regression-related algorithms have been developed in the scope of this review.

Regression is a supervised learning technique that aims to identify the correlation between variables and predict the continuous output variables based on one or more predictor variables. In this regard, to evaluate port efficiency in terms of environmental problems, [24] used polynomial regression for predicting the amount of NOx emission based on port assets, berth quantity, and geographical location of 17 port enterprises in China. The authors showed a benefit of the polynomial regression model for ports performance evaluation and found that the medium-sized and large-scale ports should focus on emissions reduction compared to small-sized ports that should focus on improving the service level and full resource utilization. Similarly, [25] considered the number of berths, the length of the terminal, the number of staff, and the total fixed assets of 18 ports in China as the input variables of their regression model to predict NOx and SOx emissions as well as solid waste and energy consumption in the selected ports. Based on the results of a regression model, the authors found that economic development positively impacts green efficiency. Reference [31] used berth length, the number of cranes, terminal area, and amount of cargo handled as the independent variables of their model for predicting emissions as well as port performance evaluation. The authors found, by applying a regression model, that although ECA regulation reduces emissions, it significantly harms port productivity due to losing cargoes.

For evaluating the port of Kaohsiung in Taiwan in terms of environmental sustainability, [27] used several input variables including lean management, green operational practices, green behaviour (green participation and green compliance), and green climate to predict green performance (financial and nonfinancial). Using the results of a regression model, the authors concluded that lean management positively impacted green operations and green behaviour. Green operational practices had a positive influence on both green behaviour and green performance.

Table 2: Categorization and used machine learning techniques in industrial applications [54].

ML domain	ML subdomains	Algorithms	Tools		
Machine learning	Unsupervised learning	Clustering	(i) K-means (ii) K-median (iii) Hierarchical clustering (HC)	(i) Spatial cluster (SC) (ii) Gaussian mixture model (GMM)	(i) Local outlier factor (LOF) (ii) Neighbour-based clustering (NBC) (iii) Parzen windows (PW)
		Dimensionality reduction	(i) Principal component analysis (PCA) (ii) Linear discriminant analysis (LDA) (iii) Kernel density estimator (KDE)	(i) Kernel principal component analysis (K-PCA) (ii) Singular value decomposition (SVD)	(i) t-distributed stochastic neighbour embedding (t-SNE) (ii) Uniform manifold approx. and projection (UMAP) (iii) Self-organizing maps (SOM)
	Supervised learning	Regression	(i) Neural networks (NN) (ii) NN, back propagation (BP) (iii) NN, convolutional neural network (CNN) (iv) NN, extreme learning machine (ELM) (v) NN, long-short term memory (LSTM) (vi) NN, deep learning (DL) (vii) NN, adaptive neuro-fuzzy inference system (ANFIS)	(i) NN, multilayer perception (MLP) (ii) NN, radial basis function (RBF) (iii) NN, recurrent neural network (RNN) (iv) Linear regression (LR) (v) Polynomial regression (PR) (vi) Fuzzy regression (FR) (vii) Bayesian regression (BR) (viii) Lasso regression (LASSO)	(i) Locally weighted regression (LWR) (ii) Support vector machine (SVM)- regressor (iii) Gradient boosting (GBoost) (iv) Random forest (RF)- regressor (v) K-nearest neighbor (KNN)-regressor (vi) Gaussian process regression (GPR)
		Classification	(i) Decision tree (DT) (ii) Gradient boosting (GBoost) (iii) Naive bayes (NB) (iv) Bayesian network (BN) (v) Kernel method (KM) (vi) Support vector machine (SVM)	(i) Adaptive support vector machine (ASVM) (ii) Learning vector quantization (LVQ) (iii) Linear discriminant analysis (LDA) (iv) Stochastic gradient descent (SGD)	(i) K-nearest neighbor (KNN) (ii) Quadratic discriminant analysis (QDA) (iii) Random forest (RF) (iv) Logistic regression (LogR) (v) Pattern recognition (PattR) (vi) Multi-layer perception (MLP)
	Reinforcement learning		(i) Adaptive heuristic critic (AHC) (ii) Deep deterministic policy gradient (DDPG) (iii) Q-learning (QL)	(i) Approximate dynamic programming (ADP) (ii) Proximal policy optimization (PPO) (iii) Deep Q- learning (DQL)	(i) State-action-reward-state-action (SARSA) (ii) Temporal difference learning (TD) (iii) Trust region policy optimization (TRPO)

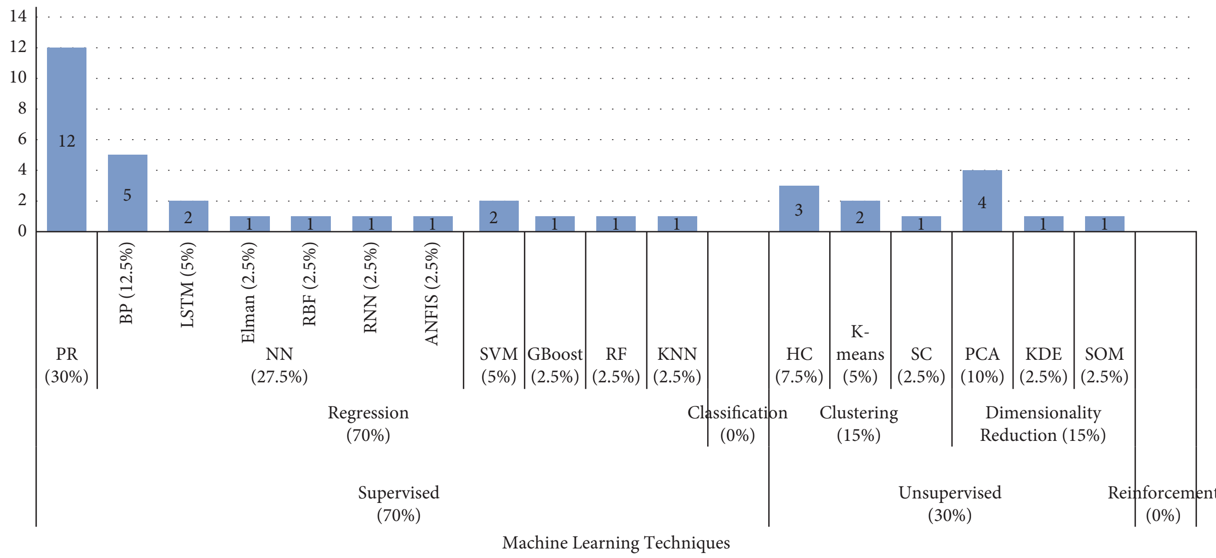

Figure 7: Categorization of ML techniques (subdomains, algorithms, and tools) for promoting environmental sustainability in port logistics.

Reference [30] used the total energy consumption of ports and the number of employees as the input variables of their model to predict the total gross weight of goods, air pollutant emissions, and the eco-efficiency rank of ports. With a regression model, they revealed that the energy consumption variable had a significant diverse correlation with the eco-efficiency of ports. Moreover, [33] used the fine particulate mass and the fine particulate composition as the input variables of their model to predict air quality. The authors concluded that polynomial regression models provided useful analysis for air quality management.

The authors of reference [34] considered the type of pollutant, the operating mode, and the gross tonnage of ships to predict the amount of emission. Based on the regression analysis, they found that innovative methods proposed by the International Maritime Organization (IMO) such as carbon capture and storage systems, increasing energy efficiency, and emissions converting technologies had a significant impact on emissions reduction. The authors of reference [35], by using long short-term memory (LSTM) and the recurrent neural network (RNN), used meteorological data, air quality data, and shipping activity data as the input variables to predict emissions in ports. The authors indicated that besides meteorological data and air quality data, ship activities, as one of the main sources of emissions in port areas, should be considered in the prediction model to enhance the performance. Using neural networks, the authors of reference [36] developed a predictive model for controlling the CO concentration in RORO ships indoors. They considered CO concentration, load (number of cars), and the reference flow rate of the ventilation system as the input variables of their model. The authors concluded that neural networks models combined with other methodologies such as fuzzy controlling and particle swamp optimization significantly guarantee the robustness of the indoors CO concentration reduction in RORO ships to an allowable limit. The authors of reference [49] developed a forecasting engine for truck arrivals to logistics nodes, i.e., empty container depots and packing facilities or terminals that mitigate greenhouse gas emissions from truck congestions in the landside. In doing so, they proposed a neural network model by taking historical data of truck arrival time, administrative wait time, intermediate wait time, node-specific forecasting parameters (e.g., dispatching modes and storage policies), and external forecasting parameters (e.g., weather information and traffic information) as the inputs. The authors showed the benefit of neural networks in the smoothed peak workloads at the nodes due to adaptive truck routing and reduced waiting times.

Intending to manage the energy consumption of RTG cranes, the authors of reference [46] utilized neural networks and a support vector machine and considered the average of the previous day load, the average of the previous week load, the same hour load for the previous day, and the previous hour load as the input variables of their model to predict RTG crane demand of one hour. They revealed that the effectiveness of the neural networks model was significantly high when the estimation of the number of crane moves and container gross weight was accurate. Furthermore, to predict the general energy consumption of RTGs, the authors of reference [47] proposed a regression model based on the energy consumption of hoist RTG, gantry RTG, and trolley RTG. The authors showed huge air pollution decrease and cost-saving on energy by the forecasting model. In other work, through an adaptive neuro-fuzzy inference system (ANFIS), the authors of reference [48] developed a model to minimize swings of RTG during loading/unloading of containers and cargo in the yard and seaside area which leads to prevent emission of hazardous materials into the air and water. They considered trolley position, trolley speed, loading angle, angular velocity, and the driving force of the trolley as the input variables of their neural network model. The authors showed that the ANFIS control method was robust and quick-response under different rope lengths and working conditions, but not reliable enough when the noise was strong. In the same area, to predict the demand for battery-electric AGVs, the authors of reference [12] took the scheduled arrival, departure and work started, planned berthing place, the planned position of front and rear of the ship, and the number of containers to load and unload as the input variables of their neural networks model. The authors reported a benefit of the neural networks model for checking the availability of AGVs in the horizontal transportation area of ports.

In the seaside area, the authors of reference [40] predicted the engine exhaust emissions using the polynomial regression based on the maximum power output and shaft speed of ships during berthing operations. They proposed a forecasting model that was significantly accurate for different engine types at berth, manoeuvring, and sea. Moreover, to develop the berth allocation planning to manage real-time data and air emissions reduction, [38] used ETA features (date, time, and weekday) and ship features (ship type and length) as the inputs of their support vector machine model to build a predicting system for vessel arrivals. The authors concluded that the use of additional features (e.g., weekday) and discarding irrelevant inputs (e.g., the shipping line) have a positive influence on the performance of the SVM model. With the same purpose, using a regression model, [41] utilized ship identification, position, speed, course, timestamp, heading, and navigational status of ships. They reported the benefit of regression analysis to model the spatial extent (the active area) of the emissions at different temporal resolutions (hourly and daily). In a similar work, using the net tonnage, deadweight tonnage, actual handling volume, time of ships arrival, and efficiency of facilities as the input variables, [42] used five regression tools of the machine learning technique including gradient boost (i.e., Gboost), backpropagation neural network, linear regression, k-nearest neighbour, and random forest to predict energy consumption and emissions from ships during berthing operation. The authors found that the time of ships arrival without influencing the performance of the model could be eliminated to reduce the difficulty of data collection. They also concluded that when the efficiency of facilities was doubled, the energy consumption of ships was reduced by 34.17% at berth and 8.41% in overall port areas.

To manage the energy consumption of an all-electric ship (AES), the authors of reference [43] proposed several neural network-based models including the Elman, backpropagation (BP), the radial basis function (RBF), and long short-term memory (LSTM). They considered hourly data of electricity prices and load demands as the input variables. The authors revealed that the LSTM method could predict the hourly price of electricity onshore accurately. As a result, the method combined with an optimization model resulted in the minimum cost and emission of the AES.

Ship sources, one of the main sources of noise emissions in the port area (i.e., roads, railways, ships, port activities, and industrial plants), are from all the activities related to the movement and stationing of ships. Engines, funnels, and ventilation, as well as transit in port regimes, are the main sources of noise from ships [55]. For this sake, the authors of reference [45] proposed a regression model based on draught, distance from a recording microphone, and speed to evaluate the correlations among the variables and predict sound emitted from moving ships in port areas. The authors indicated that ship draught was not an influencing parameter for noise emissions. Also, the authors concluded that for the noise assessment in port areas, the right placement of the noise source that provides precise input data plays an essential role in improving the output of an acoustic model.

3.3.2. Unsupervised Learning. Unsupervised learning or descriptive learning is used for unlabeled datasets that the response to a scenario or example is unknown [53]. It enables several clustering tools (e.g., hierarchical clustering, k-means, and fuzzy c-means) for recognizing the behaviour of a dataset that there is no historical output. Moreover, some dimensionality reduction tools (e.g., principal component analysis and self-organizing maps) are known as other applications of unsupervised learning that minimize the volume of datasets to an efficient computation process.

(1) Clustering. Clustering is an unsupervised learning technique that considers a set of selected features to group objects with similar attributes. The purpose of the clustering technique is to construct clusters where data objects within the same cluster are similar to one anther but different from the objects in other clusters. In this regard, for benchmarking ports in terms of environmental sustainability, hierarchical clustering is used to create dendrograms or cluster trees. For instance, the authors of references [26, 30] evaluated and benchmarked several ports in Europe and China considering energy consumption, rate of wastewater treatment, standard-reaching rate of nearshore water, the green coverage rate in developed areas, and expenditure on energy-saving investments per capita. Based on the revealed clusters, they both concluded that the ports in the same cluster with the best performance in terms of technical efficiency showed a better eco-efficiency performance than other clusters.

Spatial clustering by splitting spatial data into a series of meaningful subclasses aims to consider the selected features to group spatial objects in the same cluster that are similar to each other and dissimilar to those in different clusters. In this regard, to investigate the spatial characteristics of emissions from the port container distribution (PCD), the authors of reference [32] used this type of clustering and considered several parameters such as CO_2 emission driver factors of the city where the port is located, gross domestic product, total resident population, the number of port berths, total imports, total exports, the first industrial value, the secondary industrial value, the primary industrial value, gross industrial production, fixed assets investment in the tertiary industry, per capita income, railway freight volume, highway freight volume, and waterway freight volume. With the spatial clustering technique, they reported that ports with similar geographical locations showed a similar pattern of PCD carbon emissions.

In the yard area, to manage trucks operations in container terminals and reduce empty truck trips, it is important to identify container features. In doing so, the authors of reference [51] performed a dendrogram and clustered containers based on several input variables including the cycle of import/export, the ISO type of the container, the weight of the container, special (e.g., hazard shipping), the agreement between stakeholders, vessel departure time, the distance of two containers in the yard, customs clearance for import/export, the dwell time of the container, and the final destination of container in the hinterland. The authors showed the benefit of container clustering in terms of reducing the number of trucks for moving the same number of containers. In the same area, the authors of reference [37] developed a k-means model based on the number of RTGs, blocks, handling container specifications, stevedoring full or empty category, handling volume for a task, and the number of clusters to manage the relationship between RTG crane teams and the given block sets. The authors concluded that the k-means model was an efficient tool for clustering block communities and dispatching RTG cranes in the yard area. Furthermore, the authors of reference [50] utilized a k-means as part of a decision support system to provide representative solutions for a multiobjective interterminal truck routing problem. They used the number of clusters and a solution archive as the inputs and the cluster centroids as the output of the model. The authors showed that using k-means inside their multiobjective algorithm was a suitable clustering approach for reducing the set of solutions and, thus, making the decision process more manageable.

(2) Dimensionality Reduction. Dimensionality reduction in machine learning is a data preprocessing technique that refers to reducing the number of input variables in a dataset to minimize computational costs and increase speed [53]. It enables several tools (e.g., principal component analysis (PCA), kernel density estimator (KDE), and self-organizing map (SOM)) for dataset volume reduction [54].

Regarding benchmarking and evaluating ports, [29] used SOM combined with a data envelopment analysis (DEA) to propose similar groups of ports in terms of environmental efficiency. In doing so, they took the number of quay cranes, acres, berth, depth, the number of calls, throughput, deadweight tonnage, and CO_2 emissions as well as the inputs of the SOM. The authors concluded that the SOM was a suitable tool for reducing the dimensions of the features to a simple visualized map. For the same purpose, the authors of reference [28] implemented the KDE tool to measure pollution incidents intensity among the ports. The KDE selected a common distribution and estimated the parameters for the density (intensity) function from the data sample. The authors indicated that KDE was an accurate tool for generating a smooth surface variation and mapping the spatial distribution of ports pollution probability density of a geographic area. Reference [27] applied PCA for reducing the dimension of highly correlated input variables of a survey to a few independent factors including lean management, green operational practices, green behaviour (green participation and green compliance), and green climate to evaluate ports in terms of green performance (financial and nonfinancial). They showed the benefit of PCA as a data preprocessing tool for finding the relationship between principal components of an equation in polynomial regression models. Moreover, for ranking ports based on several environmental sustainability factors, [26] utilized PCA to identify the independent variables. They proposed the rate of treatment of wastewater, the standard-reaching rate of nearshore water, the green coverage rate in developed areas, and expenditure on energy-saving investments per capita as the input variables. The work concluded that PCA helped in reducing the dimension of indicators when combined with a hierarchical clustering model. The authors of reference [44] developed a computer vision-based model to detect ships entering into the imaging area at the seaside and help them with automatic berthing. They considered the observed video as the input variable of the PCA and separated the foreground object (ship) from the background scene of each video frame as the outputs. The authors reported the benefit of PCA for reducing the dimensions of image features. Moreover, to identify the relationship between the traffic and particle number concentrations (PNC) data from container truck emissions in the yard, the authors of reference [52] applied the PCA and proposed container truck volume, other vehicles' volume, and PNC data as uncorrelated variables for characterizing the variation of particles. They found that the method had a high performance in dimensionality reduction when combined with a Pearson correlation analysis. They also concluded that dimensionality reduction significantly reduced the computation cost and data collection difficulties.

3.4. Advantages and Disadvantages. Regarding Table 1 that summarized the main characteristics of the reviewed articles based on the following categories: the application area within the port (i.e., seaside, yard, landside, and overall port), the research scope, environmental problems (i.e., emissions, water pollution, noise pollution, energy saving, renewable energy, and solid waste), the machine learning technique, involved factors (input and output), the data collection method, and the case study, Table 3 reports the advantages and disadvantages of the used ML techniques in the collected articles.

3.5. Open Issues and Future Works. This subsection discusses some important future research directions and open viewpoints toward the application of ML techniques in the environmental sustainability of maritime ports. Future research may concentrate on analyzing the effectiveness of various machine learning (ML) algorithms for reducing the impact of environmental concerns related to solid wastes or noise pollution near port cities, as there are numerous pollutants but few solutions for those issues as of yet. The use of ML approaches to help renewable energy technology is another important study area that can be tackled by future academics. The use of optimization techniques like metaheuristics, mathematical programming, and heuristic approaches to aid decision making for planning and carrying

TABLE 3: Advantages and disadvantages of the used ML techniques for promoting environmental sustainability in maritime port logistics.

ML techniques	Advantages	Disadvantages	Related articles
Polynomial regression (PR)	(i) Works on any size of the dataset (ii) Gives information about the relevance of features	(i) The correct polynomial degree should be chosen for a good bias	[24, 25, 27, 30, 31, 33, 34, 40–42, 45, 47]
Neural networks (NN)	(i) Efficiency (ii) Continuous learning (iii) Data retrieval (iv) Multitasking	(i) Hardware dependent (ii) Complex algorithms (iii) Black-box nature (iv) Approximate results (v) Data-dependency	[12, 35, 36, 42, 43, 46, 48, 49]
Support vector machine (SVM)	(i) Works very well on nonlinear problems (ii) Easily adaptable (iii) Ignores outliers	(i) Requires feature scaling	[38, 46]
Gradient boost (GBoost)	(i) Performs very well on medium and small datasets (ii) Easy to interpret (iii) Prevents overfitting (iv) A great approach for enhancing classification and regression solutions	(i) Sensitive to outliers (ii) Hardly scalable (iii) Poor results on unstructured data	[42]
Random forest (RF)	(i) Accurate (ii) Powerful (iii) Works very well on linear/nonlinear problems	(i) The number of trees should be chosen (ii) No interpretability	[42]
K-nearest neighbor (KNN)	(i) Fast (ii) Easy to implement	(i) Need feature scaling (ii) Data should be cleaned	[42]
Hierarchical clustering (HC)	(i) Easy to implement (ii) Does not require the number of clusters	(i) Long runtime	[26, 30, 51]
K-means	(i) Easy to implement (ii) Large datasets (iii) Convergence (iv) Different shapes and sizes with generalization	(i) Requires the number of clusters (ii) Requires initial values (iii) Requires dimensionality reduction tools if the number of dimensions is high (iv) Does not ignore outliers	[37, 50]
Spatial clustering (SC)	(i) Does not require the number of clusters (ii) Different shaped clusters (iii) Ignores outliers	(i) Poor results for datasets with various densities (ii) Poor results for unstructured data (iii) Not deterministic	[32]
Principal component analysis (PCA)	(i) Eliminates correlated features (ii) Enhances algorithm performance (iii) Reduces overfitting (iv) Enhances visualization	(i) Less interpretable (ii) Data has to be uniformed (iii) Loss of information	[26, 27, 44, 52]
Kernel density estimator (KDE)	(i) Smooth visualization (ii) Works with various shapes and sizes	(i) Biased at the boundaries (ii) Information loss by oversmoothing	[28]
Self-organizing maps (SOM)	(i) Interpretable (ii) Applicable for large datasets	(i) Requires initial weights	[29]

out port operations that incorporate environmental aspects is also something we propose as an extension given that the focus of this SLR was on the implementation of ML techniques to promote environmentally sustainable practices in ports.

4. Conclusions

This article presented the results of a systematic literature review of the recent literature on machine learning for promoting environmental sustainability in ports. The review explored contributions of five years (2017–2021) with the aim of capturing the most recent approaches to foster environmental sustainable port operations. It categorized the identified articles based on their application area within the port as well as the application of the ML approach. Using the PRISMA protocol and bibliometric tools, the research framework was constituted on the major considerations of impacts, techniques, and problems. In general, the challenges and barriers of machine learning to aid decision making and promote more sustainable green practices at ports were discussed. By analyzing the 29 identified articles with their methodological approaches, this review summarized the academic contributions considering the three main dimensions including machine learning, port logistics, and environmental sustainability. The articles that used regression models were dominant in the literature, while LSTM and RNN were the most recent approaches. Also, in terms of environmental indicators, investigations on emissions and energy consumption were predominant among collected articles.

Concerning the type of approaches, supervised learning was the most used type of technique while unsupervised learning was the second option with about 30% of collected articles. Dimensionality reduction tools as data preprocessing techniques reduced computation costs as they minimized the volume of data in databases. Clustering techniques were beneficial to constructing clusters where data objects within the same cluster were similar to one anther but different from the objects in other clusters. Moreover, the information on ML-based prediction algorithms such as polynomial regression and neural networks showed a high contribution to promoting environmental indicators including greenhouse gas emissions, energy consumption, and water pollution.

The outcomes of this SLR can be used as a starting point for investigations regarding the application of machine learning techniques for enhancing environmental sustainability in ports and provide a guideline for researchers to find research gaps in the literature. In the light of this review, future studies may focus on examining the performance of different ML algorithms for mitigating the impact of environmental concerns related to solid wastes or noise pollution near port cities, as there are several pollutants, but there are only a few contributions for solving those problems yet. Another relevant research direction is the one related to supporting renewable energy technologies based on ML techniques that can be addressed by future researchers. Moreover, given that the scope of this SLR was on the implementation of ML techniques to promote environmentally sustainable practices at ports, we propose as an extension the consideration of optimization techniques such as metaheuristics, mathematical programming, and heuristic approaches to aid decision making for planning and executing port operations incorporating environmental aspects. Finally, a complementary SLR on green maritime shipping can be a relevant research direction to provide insights and analyse the current contributions of ML in that application domain.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The third author of the manuscript acknowledges the support of the National Agency of Research and Development (ANID), Chile, through the grant no. FONDECYT N.1210735, the program STIC-AMSUD-22-STIC-09, and the grant no. FOVI220133 “Digital Ports of Next Generation”.

References

[1] E. Lalla-Ruiz, L. Heilig, and S. Voß, “Environmental sustainability in ports,” *Sustainable Transportation and Smart Logistics: Decision-Making Models and Solutions*, Vol. 65–89, Elsevier, , Amsterdam, Netherlands, 2018.

[2] A. Vega-Muñoz, G. Salazar-Sepulveda, J. F. Espinosa-Cristia, and J. Sanhueza-Vergara, “How to measure environmental performance in ports,” *Sustainability*, vol. 13, no. 7, p. 4035, 2021.

[3] H. Davarzani, B. Fahimnia, M. Bell, and J. Sarkis, “Greening ports and maritime logistics: a review,” *Transportation Research Part D: Transport and Environment*, vol. 48, pp. 473–487, 2016.

[4] M. Ghobaei-Arani, M. Shamsi, and A. A. Rahmanian, “An efficient approach for improving virtual machine placement in cloud computing environment,” *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 29, no. 6, pp. 1149–1171, 2017.

[5] Y. Zheng, J. Zhao, and G. Shao, “Port city sustainability: a review of its research trends,” *Sustainability*, vol. 12, no. 20, p. 8355, 2020.

[6] H. Geerlings and T. Vellinga, “Sustainability,” in *Ports and Networks*, pp. 296–314, Routledge, Oxfordshire, UK, 1 edition, 2018.

[7] F. Schulte, E. Lalla-Ruiz, R. G. González-Ramírez, and S. Voß, “Reducing port-related empty truck emissions: a mathematical approach for truck appointments with collaboration,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 105, pp. 195–212, 2017.

[8] HLEG, “Independent High-Level Expert Group on Artificial Intelligence Set up by the European Commission,” 2019, https://www.aepd.es/sites/default/files/2019-12/ai-definition.pdf.

[9] A. Shakarami, A. Shahidinejad, and M. Ghobaei-Arani, "An autonomous computation offloading strategy in Mobile Edge Computing: a deep learning-based hybrid approach," *Journal of Network and Computer Applications*, vol. 178, Article ID 102974, 2021.

[10] F. Jazayeri, A. Shahidinejad, and M. Ghobaei-Arani, "Autonomous computation offloading and auto-scaling the in the mobile fog computing: a deep reinforcement learning-based approach," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 8, pp. 8265–8284, 2021.

[11] S. Samoili, M. López Cobo, E. Gómez, G. De Prato, F. Martínez-Plumed, and B. Delipetrev, "AI Watch - defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence," in *Joint Research Centre*European Commission, Brussels, Belgium, 2020.

[12] S. Holly, A. Nieße, M. Tröschel et al., "Flexibility management and provision of balancing services with battery-electric automated guided vehicles in the Hamburg container terminal Altenwerder," *Energy Inform*, vol. 3, no. 1, p. 26, 2020.

[13] N. H. Prevljak, "HHLA container terminals use machine learning technology to boost productivity," Offshore Energy, Broussard, Louisiana, 2020.

[14] I. de la Peña Zarzuelo, M. J. Freire Soeane, and B. López Bermúdez, "Industry 4.0 in the port and maritime industry: a literature review," *Journal of Industrial Information Integration*, vol. 20, Article ID 100173, 2020.

[15] G. Yeo, S. H. Lim, L. Wynter, and H. Hassan, "MPA-IBM project SAFER: sense-making analytics for maritime event recognition," *INFORMS Journal on Applied Analytics*, vol. 49, no. 4, pp. 269–280, 2019.

[16] E. F. France, N. Ring, J. Noyes et al., "Protocol-developingmeta-ethnography reporting guidelines (eMERGe)," *BMC Medical Research Methodology*, vol. 15, no. 1, pp. 103–114, 2015.

[17] A. Tong, K. Flemming, E. McInnes, S. Oliver, and J. Craig, "Enhancing transparency in reporting the synthesis of qualitative research: entreq," *BMC Medical Research Methodology*, vol. 12, no. 1, pp. 181–188, 2012.

[18] B. C. O'Brien, I. B. Harris, T. J. Beckman, D. A. Reed, and D. A. Cook, "Standards for reporting qualitative research: a synthesis of recommendations," *Academic Medicine*, vol. 89, no. 9, pp. 1245–1251, 2014.

[19] K. Flemming, A. Booth, K. Hannes, M. Cargo, and J. Noyes, "Cochrane Qualitative and Implementation Methods Group guidance series—article 6: reporting guidelines for qualitative, implementation, and process evaluation evidence syntheses," *Journal of Clinical Epidemiology*, vol. 97, pp. 79–85, 2018.

[20] A. Liberati, D. G. Altman, J. Tetzlaff et al., "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration," *BMJ*, vol. 339, 2009.

[21] D. Moher, A. Liberati, J. Tetzlaff et al., "Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement," *PLoS Medicine*, vol. 6, no. 7, Article ID e1000097, 2009.

[22] V. Garousi, K. Petersen, and B. Ozkan, "Challenges and best practices in industry-academia collaborations in software engineering: a systematic literature review," *Information and Software Technology*, vol. 79, pp. 106–127, 2016.

[23] S. Pahlevan-Sharif, P. Mura, and S. N. R. Wijesinghe, "A systematic review of systematic reviews in tourism," *Journal of Hospitality and Tourism Management*, vol. 39, pp. 158–165, 2019.

[24] J. Sun, Y. Yuan, R. Yang, X. Ji, and J. Wu, "Performance evaluation of Chinese port enterprises under significant environmental concerns: an extended DEA-based analysis," *Transport Policy*, vol. 60, pp. 75–86, 2017.

[25] L. Wang, C. Peng, W. Shi, and M. Zhu, "Carbon dioxide emissions from port container distribution: spatial characteristics and driving factors," *Transportation Research Part D: Transport and Environment*, vol. 82, Article ID 102318, 2020.

[26] C. Wang, H. Haralambides, and L. Zhang, "Sustainable port development: the role of Chinese seaports in the 21st century Maritime Silk Road," *International Journal of Shipping and Transport Logistics*, vol. 13, no. 1/2, p. 205, 2021.

[27] S. Y. Kuo and P. C. Lin, "Determinants of green performance in container terminal operations: a lean management," *Journal of Cleaner Production*, vol. 275, Article ID 123105, 2020.

[28] S. Cheon, A. Maltz, and K. Dooley, "The link between economic and environmental performance of the top 10 U.S. ports," *Maritime Policy and Management*, vol. 44, no. 2, pp. 227–247, 2017.

[29] Y. S. Park, N. M. A. Mohamed Abdul Ghani, F. Gebremikael, and G. Egilmez, "Benchmarking environmental efficiency of ports using data mining and RDEA: the case of a U.S. container ports," *International Journal of Logistics Research and Applications*, vol. 22, no. 2, pp. 172–187, 2019.

[30] C. Quintano, P. Mazzocchi, and A. Rocca, "Examining eco-efficiency in the port sector via non-radial data envelopment analysis and the response based procedure for detecting unit segments," *Journal of Cleaner Production*, vol. 259, Article ID 120979, 2020.

[31] Y. T. Chang, H. K. Park, S. Lee, and E. Kim, "Have emission control areas (ECAs) harmed port efficiency in Europe?" *Transportation Research Part D: Transport and Environment*, vol. 58, pp. 39–53, 2018.

[32] L. Wang, Z. Zhou, Y. Yang, and J. Wu, "Green efficiency evaluation and improvement of Chinese ports: a cross-efficiency model," *Transportation Research Part D: Transport and Environment*, vol. 88, Article ID 102590, 2020.

[33] D. J. Eatough, P. Cropper, W. Keeton et al., "Apportionment of PM2.5 adjacent to the I-710 harbor freeway in long Beach, CA," *Journal of the Air and Waste Management Association*, vol. 70, no. 3, pp. 260–282, 2020.

[34] A. Ekmekçioğlu, K. Ünlügençoğlu, and U. B. Çelebi, "Container ship emission estimation model for the concept of green port in Turkey," *Journal of Engineering for the Maritime Environment*, vol. 236, no. 2, pp. 504–518, 2021.

[35] H. Hong, H. Jeon, C. Youn, and H. Kim, "Incorporation of shipping activity data in recurrent neural networks and long short-term memory models to improve air quality predictions around busan port," *Atmosphere*, vol. 12, no. 9, p. 1172, 2021.

[36] H. Agamy, M. Abdelgeliel, M. Mosleh, K. Elserafy, and N. A. N. Mohamed, "Neural fuzzy control of the indoor air quality onboard a RO–RO ship garage," *International Journal of Fuzzy Systems*, vol. 22, no. 3, pp. 1020–1035, 2020.

[37] G. Song, B. Li, and Y. He, "Container terminal logistics generalized computing architecture and green initiative computational pattern performance evaluation," *Information*, vol. 10, no. 12, pp. 383–424, 2019.

[38] P. Cammin, M. Sarhani, L. Heilig, and S. Voß, "Applications of real-time data to reduce air emissions in maritime ports," in *Design, User Experience, and Usability. Case Studies in Public and Personal Interactive Systems. HCII 2020*, A. Marcus and E. Rosenzweig, Eds., Vol. 12202, Springer Cham, New York, NY, USA, 2020.

[39] J. Yu, G. Tang, X. Song et al., “Ship arrival prediction and its value on daily container terminal operation,” *Ocean Engineering*, vol. 157, pp. 73–86, 2018.

[40] S. Jahangiri, N. Nikolova, and K. Tenekedjiev, “An improved emission inventory method for estimating engine exhaust emissions from ships,” *Sustainable Environment Research*, vol. 28, no. 6, pp. 374–381, 2018.

[41] B. Goldsworthy, H. Enshaei, and S. Jayasinghe, “Comparison of large-scale ship exhaust emissions across multiple resolutions: from annual to hourly data,” *Atmospheric Environment*, vol. 214, Article ID 116829, 2019.

[42] Y. Peng, H. Liu, X. Li, J. Huang, and W. Wang, “Machine learning method for energy consumption prediction of ships in port considering green ports,” *Journal of Cleaner Production*, vol. 264, Article ID 121564, 2020.

[43] S. Wen, T. Zhao, Y. Tang et al., “Coordinated optimal energy management and voyage scheduling for all-electric ships based on predicted shore-side electricity price,” *IEEE Transactions on Industry Applications*, vol. 57, no. 1, pp. 139–148, 2021.

[44] Y. Zhan and T. Liu, “Weighted RPCA based background subtraction for automatic berthing,” in *Proceedings of the Chinese Control Conference (CCC)*, pp. 3519–3524, Guangzhou, China, July 2019.

[45] M. Nastasi, L. Fredianelli, M. Bernardini, L. Teti, F. Fidecaro, and G. Licitra, “Parameters affecting noise emitted by ships moving in port areas,” *Sustainability*, vol. 12, no. 20, p. 8742, 2020.

[46] F. Alasali, S. Haben, V. Becerra, and W. Holderbaum, “Day-ahead industrial load forecasting for electric RTG cranes,” *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 2, pp. 223–234, 2018.

[47] S. Fahdi, M. Elkhechafi, and H. Hachimi, “Machine learning for cleaner production in port of Casablanca,” *Journal of Cleaner Production*, vol. 294, Article ID 126269, 2021.

[48] Z. Wang, Y. Shi, and S. Li, “Research on anti-swaying of crane based on t-s type adaptive neural fuzzy control,” in *Proceedings of the Chinese Control and Decision Conference (CCDC)*, pp. 5503–5508, Shenyang, China, June 2018.

[49] A. Hill and J. W. Böse, “A decision support system for improved resource planning and truck routing at logistic nodes,” *Information Technology and Management*, vol. 18, no. 3, pp. 241–251, 2017.

[50] L. Heilig, E. Lalla-Ruiz, and S. Voß, “Multi-objectiveinter-terminal truck routing,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 106, pp. 178–202, 2017.

[51] C. Caballini, M. D. Gracia, J. Mar-Ortiz, and S. Sacone, “A combined data mining – optimization approach to manage trucks operations in container terminals with the use of a TAS: application to an Italian and a Mexican port,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 142, Article ID 102054, 2020.

[52] H. Zhao, H. He, W. Lu, and Y. Hao, “Characterizing the variation of particles in varied sizes from a container truck in a port area,” *Environmental Monitoring and Assessment*, vol. 192, no. 12, p. 787, 2020.

[53] K. Murphy, *Machine Learning: A Probabilistic Perspective*The MIT Press, Cambridge, MA, USA, 2012.

[54] M. Bertolini, D. Mezzogori, M. Neroni, and F. Zammori, “Machine Learning for industrial applications: a comprehensive literature review,” *Expert Systems with Applications*, vol. 175, Article ID 114820, 2021.

[55] L. Fredianelli, M. Bolognese, F. Fidecaro, and G. Licitra, “Classification of noise sources for port area noise mapping,” *Environments*, vol. 8, no. 2, p. 12, 2021.

[56] H. Du and Y. Jiang, “Backup or reliability improvement strategy for a manufacturer facing heterogeneous consumers in a dynamic supply chain,” *IEEE Access*, vol. 7, Article ID 50419, 2019.

[57] A. Kara and I. Dogan, “Reinforcement learning approaches for specifying ordering policies of perishable inventory systems,” *Expert Systems with Applications*, vol. 91, pp. 150–158, 2018.

[58] A. Kuhnle, J. Jakubik, and G. Lanza, “Reinforcement learning for opportunistic maintenance optimization,” *Production Engineering*, vol. 13, no. 1, pp. 33–41, 2018.

[59] M. F. Monaco, L. Moccia, and M. Sammarra, “Operations Research for the management of a transhipment container terminal: the Gioia Tauro case,” *Maritime Economics and Logistics*, vol. 11, no. 1, pp. 7–35, 2009.

[60] Q. Tan, Y. Tong, S. Wu, and D. Li, “Modeling, planning, and scheduling of shop-floor assembly process with dynamic cyber-physical interactions: a case study for CPS-based smart industrial robot production,” *International Journal of Advanced Manufacturing Technology*, vol. 105, no. 9, pp. 3979–3989, 2019.

[61] G. Wong, T. Greenhalgh, G. Westhorp, J. Buckingham, and R. Pawson, “RAMESES publication standards: meta-narrative reviews,” *Journal of Advanced Nursing*, vol. 69, no. 5, pp. 987–1004, 2013.

[62] Z. Zhu, N. Anwer, Q. Huang, and L. Mathieu, “Machine learning in tolerancing for additive manufacturing,” *CIRP Annals*, vol. 67, no. 1, pp. 157–160, 2018.