IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 March 2024, accepted 14 April 2024, date of publication 16 April 2024, date of current version 1 May 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3389961*

RESEARCH ARTICLE

# A Predictive Discrete Event Simulation for Predicting Operation Times in Container Terminal

**KIKUN PARK, MINSEOP KIM, AND HYERIM BAE, (Member, IEEE)**
Department of Industrial Engineering, Pusan National University, Busan 46241, Republic of Korea

Corresponding author: Hyerim Bae (hrbae@pusan.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT)(No.RS-2023-00218913); and in part by Korea Institute of Marine Science and Technology Promotion (KIMST) funded by the Ministry of Oceans and Fisheries, South Korea, under Grant 202205382. ※ MSIT: Ministry of Science and ICT.

**ABSTRACT** Container terminals (CTs) play a crucial role within the global supply chain in the context of the global container transportation business. The primary obligation associated with these terminals involves ensuring the punctual execution of primary vessel operations, while strictly adhering to the Estimated Time of Departure (ETD) for vessels. To accomplish this goal, it is crucial to develop rigorous and exact operating strategies. Nonetheless, the accurate prediction of operation times in CT presents a significant challenge due to the concurrent involvement of various Container Handling Equipment (CHE) and the occurrence of unforeseen situations. To address this issue, this study proposes a novel approach called Predictive Discrete Event Simulation (PDES) that utilizes data collected from CT to predict the operation times. The PDES is an advanced approach that builds upon the widely used Discrete Event Simulation (DES) technique in the field of simulations. It is specifically designed to provide precise predictions of the operation times in CTs, where multiple events take place concurrently. The PDES in this paper aims to overcome the shortcomings of current simulation-based approaches for prediction in CT. These approaches often rely on predefined task sequences and assumed time for job handling time of CHE in their scenarios, which can result in reduced accuracy when predicting operation times. Through the resolution of these issues, the proposed PDES exhibits the capacity to improve predictive performance. To enhance the predictive performance of operation times in CTs through PDES, two approaches are introduced. The first approach entails the application of Support Vector Machine (SVM) algorithms for the purpose of predicting operation times. This approach is further augmented by integrating it with DES to improve the accuracy of predictive performance. The second approach involves predicting by simulating real-world operational scenarios in CTs using algorithms for CHE assignment. The predictive performance of the proposed PDES is assessed through the utilization of data gathered from Busan Port Terminal (BPT) in South Korea, demonstrating superior performance compared to alternative prediction approaches.

**INDEX TERMS** Container terminal, discrete event simulation, machine learning, prediction.

## I. INTRODUCTION

Maritime transportation is widely acknowledged as the pivotal element of global trade, responsible for managing approximately 80% of the total volume of trade worldwide [1]. The containerized cargo volume has witnessed a proportional increase alongside the expansion of the global economy. In the year 2021, following a decrease of 3.8% in the volume of international maritime containerized cargo in 2020, there was a subsequent recovery with a 3.2% rise, resulting in a cumulative quantity of 110 billion tons. The increase in maritime trade volume, the onset of the

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Ni.

COVID-19 pandemic, and the expansion of container vessels have all played a role in the increased variability in land-based container handling. As a result, there has been a rise in congestion and operational inefficiencies within CTs [2], [3]. The escalating congestion observed in CTs has a detrimental impact on productivity and the ability to adhere to schedules, consequently causing a ripple effect that disrupts the operational plans of other vessels [4]. As a result, the identification of mitigating congestion CTs as a pivotal factor is crucial in addressing the decline in productivity [5]. The resolution of congestion issues in CTs requires the development of effective strategies that consider multiple factors, such as Documentation Procedures, Ship Traffic Inputs, Port Structure, Port Operation, Management, and Government Policies [6]. This study aims to the evaluation (prediction) method of the operation plan for port operation among the above factors.

In CTs characterized by the simultaneous movement of numerous containers through berths, yards, and gates, the complexity stems from the concurrent operation of multiple CHE. CTs demonstrate a high level of complexity due to the interrelated nature of tasks within their operational processes, which in turn intensifies the intricacies associated with them. The intricate nature of this complexity presents a substantial obstacle when attempting to employ data analysis and prediction methodologies on actual data, as it gives rise to a wide range of scenarios within CTs [7]. According to experts in charge of operational planning in BPT, South Korea, the utilization of analytical and predictive technologies in terminal operations presents certain challenges. These challenges necessitate operators to heavily depend on the expertise of on-site professionals for the development and assessment of plans pertaining to primary vessel operations.

Busan Port Authority (BPA, South Korea) is now implementing the Smart Port development project and making progress in the construction of automated container terminals to address difficulties of CTs operation [45]. However, the container terminals that are now in operation have not yet included automation. Therefore, it is necessary to use data-driven technologies to improve operational efficiency [46].

In order to successfully reduce congestion and enhance efficiency in CT operations, it is essential to deploy approaches that actively utilize data for accurate evaluation and predicted of operating plans, instead of merely depending on expert knowledge and experience. When there is a substantial discrepancy between the anticipated outcomes of pre-determined operational schedules and the actual operational results, it becomes imperative to make necessary adjustments to the operational plans to mitigate any potential delays in the primary vessel operations at the CT. To address this issue, operators may be required to allocate supplementary resources in terms of equipment or personnel, or alternatively, redistribute equipment that was previously allocated to operations involving different vessels. The implementation of such modifications possesses the capacity to reduce the efficiency of the CT. Hence, to proactively address these concerns, it is crucial to evaluate the scheduled operational strategies and utilize precise simulations that closely correspond to the real-world operational results.

However, the methodologies suggested for the prediction and simulation of operational plans in CTs demonstrate certain constraints in accurately predicting actual operational results. There exists a prevailing inclination to make assumptions regarding the consistent or typical operational speeds of machinery. However, it has been acknowledged that the utilization of a uniform job handling time assumption for CHE is limited in its ability to adapt to situations where operation times experience significant variations. This limitation arises from the cascade effect that arises from uncertain circumstances in the CT [8].

Moreover, it has been shown that while the individual duration of a single operation may satisfy expectations, the cumulative effect of multiple operations with larger variances than the average can impact each other. This will lead to cascade effects and even make the schedule fail [8]. If there is a delay in following the planned schedule, port operators are required to create an alternate schedule, which might result in shipping delays [44].

In accordance with the analysis of the TOS at the BPT, it has been found that over 30% of changes were made to the initial scheduling plan dataset [9]. Additionally, the lack of deployed CHE has the potential to be a contributing factor to shipping delays. Currently, BPT has established regulations to manage the congestion of CTs caused by the increased amount of work from the gates [47]. This includes regulating the acceptable periods for empty container. As a result, the Quay Crane (QC) takes an average of 0.124 units of time to unloading jobs, with a standard deviation of 0.14 (Normalized values). For loading jobs, it takes an average of 0.120 units of time, with a standard deviation of 0.108. This indicates a substantial difference in the amount of time the QC operates, suggesting a lack of consistency in how tasks are carried out.

Furthermore, it is a customary practice within the simulation procedure to establish a predetermined processing sequence for containers and determine the specific quantity of equipment to be utilized. Nevertheless, it has been widely recognized that this approach is often insufficient in CTs, where a multitude of operations take place simultaneously and operational sequences frequently deviate from the original plan due to factors such as equipment arrangements and the dynamic conditions of the terminal [9].

Simulation methods commonly utilized for assessing operational plans in CTs frequently rely on assumptions regarding the sequence and job handling time of tasks. Nevertheless, these methodologies have certain limitations when it comes to effectively dealing with the inherent uncertainties associated with real-world CT operations. This study presents a novel approach called PDES that is designed to effectively handle scenarios characterized by cascade effects and variations in the sequence of operations within CT operational plans. The PDES is specifically designed to enhance predictive performance by utilizing historical data on container operations obtained from the Terminal Operating System (TOS) that is actively employed in actual CTs. This surpasses the capabilities of traditional simulation approaches [10].

The PDES is an advanced approach to DES which is widely recognized as a prominent simulation technique. PDES aims to overcome the limitations that arise when applying DES to prediction problems. The DES is recognized for its ability to accurately represent real-world phenomena and simulate dynamic processes at the level of individual events, resulting in the generation of highly detailed outcomes [11]. When the DES is utilized, it is crucial to consider the timing of event occurrences. Neglecting to properly consider the temporal aspects of events can lead to substantial inaccuracies in predictive results. The PDES proposed in this study enhances its predictive capabilities by accurately predicting the occurrence timing of events across different equipment. Furthermore, traditional methodologies employing DES frequently assume of a constant, average, or historically derived probability distribution for event (job) processing times. In instances of this nature, there exist constraints wherein simulation outcomes deviate significantly from real-world situations. This study aims to improve the predictive accuracy of CT operation times, which are equivalent to the processing time of events, by integrating a Machine Learning (ML) algorithm with the simulation process.

The objective of this study is to predict the makespan of QC and vessels. To enhance the accuracy of predicting the period of job completion and the occurrence time of jobs, ML algorithms are employed. The heuristic method is employed to predict the duration of the Internal Truck (IT) movement, while the probability distribution is utilized to estimate the time it takes for the Yard Crane (YC) to operate. The TOS data is utilized to construct prognostic models for equipment uptime and scheduling of tasks.

The examination of data obtained from the TOS of CTs indicates that the times of tasks performed in CTs exhibit a notable level of variability because of cascade effects. Moreover, the prediction task is challenging as all features available for prediction are categorical. Furthermore, as the number of variables considered in the simulation increases, the computational complexity rises, necessitating a predictive model for operational time that takes cardinality into account. The application of PDES involves the utilization of feature selection using Bayesian Network (BN) and the implementation of a prediction model using SVM to accurately predict the time of events. This approach exhibits enhanced predictive performance in comparison to methods that employ conventional simulation approaches.

Furthermore, the assessment and projection of operational plans for CTs necessitate the careful consideration of the equipment allocation strategy employed for each individual operation. The objective of equipment allocation in CTs is to assign equipment in a manner that minimizes operational time [18]. Nevertheless, it is widely acknowledged that the optimization of equipment allocation in real CTs falls short of achieving optimality due to the inherent stochasticity of terminal processes, resulting in suboptimal outcomes [38]. In order to address this problem, it is necessary to employ dynamic equipment allocation algorithms that are specifically designed for the operational context. These algorithms should also integrate prior information obtained from real CTs into the equipment allocation process. This study presents a novel approach aimed at enhancing prediction accuracy through the simulation of equipment allocation strategies that closely resemble those observed in real-world scenarios.

The contributions of this study are as follows. Firstly, this study presents the approach for predicting the operation times of CT's integrating the ML algorithm and simulation techniques. To predict the operation times of CT's, this paper proposes three approaches. The first step involves introducing a SVM predictive model application approach designed to precisely predict the occurrence time and process time of events (jobs). Secondly, addressing the constraints inherent in DES which relies on the timing of events to drive the simulation, a predictive simulation model is developed that advances events based on prediction outcomes. Thirdly, by emulating actual CHE allocation, a strategy for assigning equipment is devised, considering the container operational conditions. This strategy exhibits an improved level of precision in predicting the operation times of the CT. The prediction model was trained using container operation history data collected from BPT located in South Korea. The efficacy of the proposed approach is substantiated through comparative experiments.

The remainder of this paper is organized as follows. Section II introduces the related studies on prediction of CT, simulation for CT and the scope of DES applications. Section III provides an overview of the problems addressed in this study and introduces the data used for analysis. In Section IV, PDES for predicting the operational time of the CT is presented. Section V summarizes the experimental results, and the study concludes in the final section.

## II. RELATED WORKS

### A. OPERATIONAL PROBLEM IN CONTAINER TERMINAL

CT utilize a range of equipment, such as QC, YC, IT, and External Truck (ET) [12]. The QC is tasked with the responsibility of facilitating the transfer of containers between the vessel and the berth, encompassing the loading, and unloading processes. The responsibility of YC involves the relocation and placement of containers within the yard, ensuring that they are positioned in their assigned locations. IT plays a crucial role in enabling the movement of containers from the berth to the yard, whereas ET is responsible for the transportation of containers between the landside and the yard.

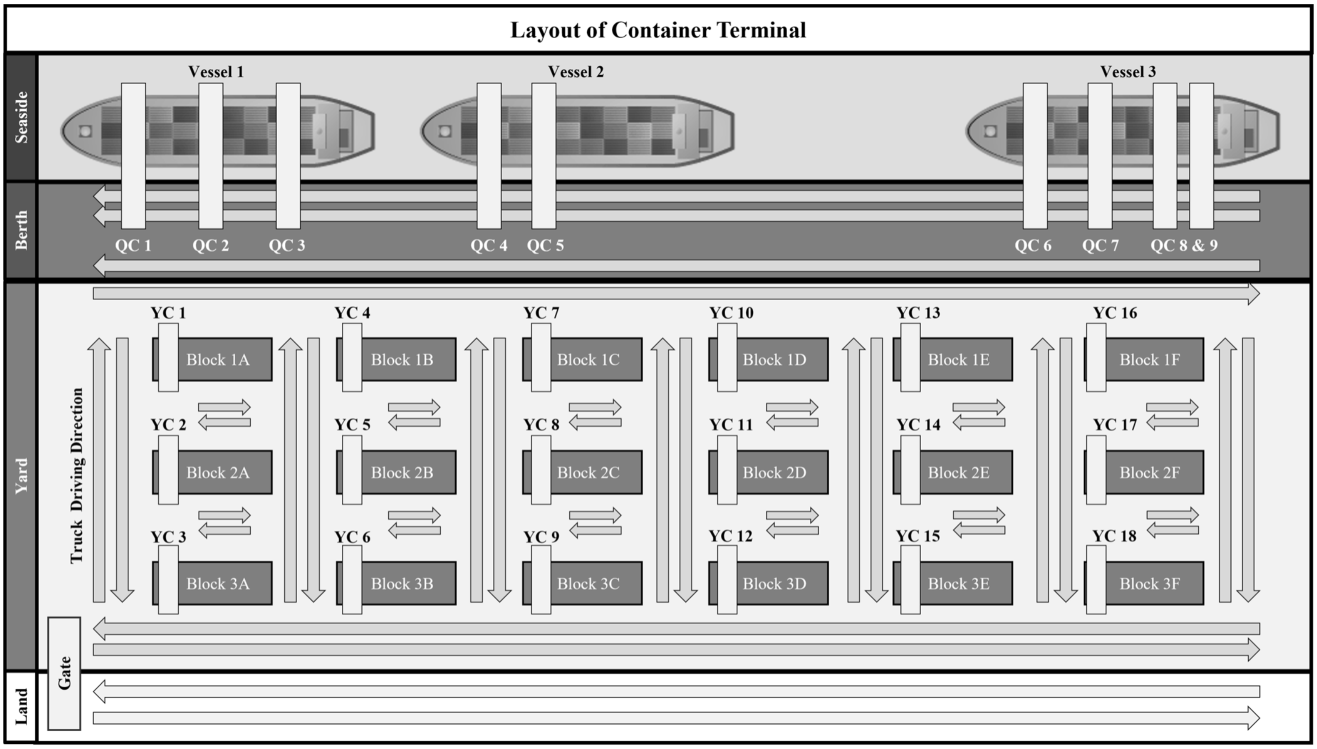



**FIGURE 1. The layout of the container terminal (CT).**

The geographical positions of each equipment unit can be observed in Figure 1. In the first stage, container vessels navigate towards the berth located at the coastline in order to carry out mainline operations. QCs are strategically positioned at the berth to facilitate mainline operations, where each crane is exclusively assigned to a particular vessel. The regions denoted by arrows in the figure correspond to the roadways employed by IT or ET. It is imperative to acknowledge that every CT conforms to specific road regulations, necessitating trucks to adhere to designated traffic directions. Within the yard, containers are loaded, stored, and organized. YCs are strategically positioned in each block to carry out these operations. Lastly, at the land side, gates are designated for the entry and exit of ET responsible for importing or exporting containers.

The operational activities performed in a CT can be classified into three main categories: discharging, which involves the unloading of containers from vessels; loading, which entails the loading of containers onto vessels; and the operational processes associated with containers entering and exiting through the In-Gate and Out-Gate at the terminal gate [13]. Furthermore, the incorporation of twin operations is a direct result of the specifications of container equipment and operational plans, enabling the concurrent handling of two containers [14].

According to Figure 2, during the discharge process, which involves the unloading of containers from the vessel, the QC commences the operation from the berth. Concurrently, it is essential for the IT to be strategically positioned at the designated site of the QC to efficiently receive the containers that are being unloaded. After the QC delivers the container, the IT proceeds to transport it to the yard. Subsequently, the IT awaits the YC's operation. The discharge procedure is completed by transferring the container, which is being transported by the IT to the yard. In contrast, the loading procedure, which entails the placement of containers into the vessel, begins with the YC initiating the process by transferring containers from the yard onto the vessel. The IT, strategically stationed in the yard, takes delivery of the loaded containers from the YC. The IT then moves to the QC, and the loading operation is completed as the QC loads the containers transported by the IT to the yard. In contrast, the loading procedure, which entails the placement of containers into the vessel, begins with the YC initiating the process by transferring containers from the yard onto the vessel. The IT, strategically stationed in the yard, takes delivery of the loaded containers from the YC. The IT then moves to the QC, and the loading operation is completed as the QC loads the containers transported by the IT onto the vessel. The Gate In procedure entails the arrival of the ET with containers, followed by the subsequent loading of these containers into the yard by the YC. Conversely, the Gate Out procedure involves the YC removing containers from the yard for them to be transported by the ET.

When establishing the operational plan for a CT, it is crucial to carefully evaluate the precedence relationships of

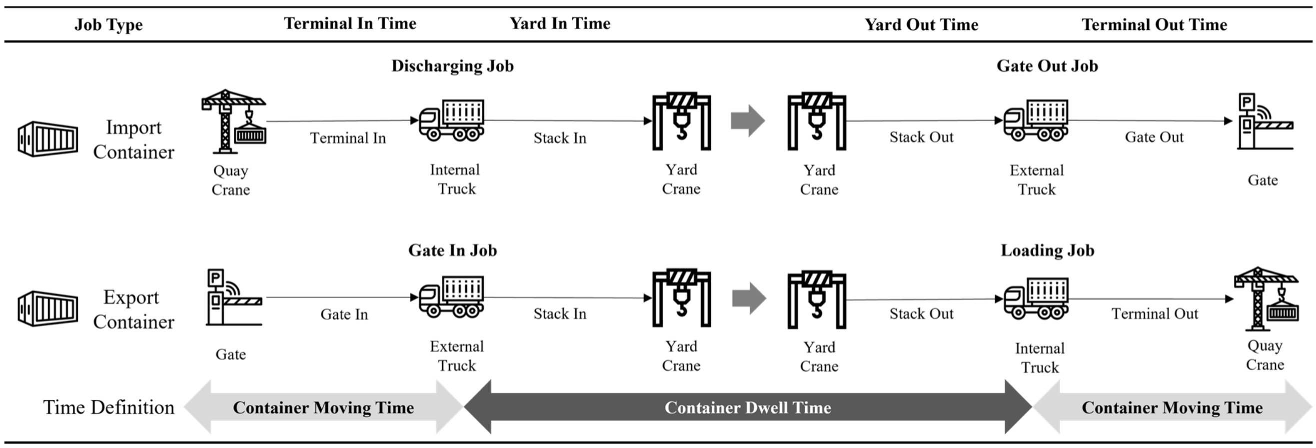



**FIGURE 2.** **The process of jobs in container terminal (CT).**

activities. This is because new tasks inside the terminal cannot begin until the previous operations by the equipment have been completed. Nevertheless, because of the intrinsic intricacy involved in thoroughly examining the many forms of activities and the hierarchical linkages among equipment at CTs, several research endeavors have often split the matters to be tackled into subsystems [15]. Mathematical models have been proposed to handle the Quay Crane Scheduling Problem (QCSP), which pertains to the discharging and loading activities of vessels at container ports. These models include considerations for the operating efficiency of QCs as well as the safety of the vessels involved.

However, it is important to note that there are limits when the operating state of IT is not considered concurrently [16]. On the other hand, research examining the operational strategies of QCs and ITs in conjunction has presented operational planning approaches that include both the time required for QCs to perform operations and the time taken by ITs to travel. Nevertheless, it is important to note that these approaches are predicated on the assumption of continuous assistance from YCs in order to facilitate handling operations [17].

Although previous research on CT operational planning approaches have shown impressive results, a notable restriction remains in their lack of complete consideration for the operating state of all equipment. Hence, to thoroughly evaluate the operational strategy of a CT, several studies have been undertaken to include the operating conditions of QCs, ITs and YCs [18], [19], [20], [21]. These studies have presented approaches for integrating QC, IT and YC factors into the development of operational plans for CTs. The approaches suggested in these studies fix the average values for operation times of each CHE. Nevertheless, as elucidated in Section I, the imposition of preset values for the operating durations of each equipment entails certain constraints, resulting in notable discrepancies between the expected values and the real operational results of CTs. Hence, in order to accurately predict the operational outcomes of a CT, it is essential to enhance the predictive capabilities by using data-driven approach.

### *B. PREDICTION OF CONTAINER TERMINAL OPERATION TIME*

The prediction problem at CT is a matter of great importance, and several studies have been conducted to address this issue. However, a significant portion of these study primarily concentrate on topics such as container throughput or the forecasting of vessel arrival timings. There is a scarcity of study about the prediction of operating times in CT [4], [39]. In the context of studies pertaining to the prediction of operation times in CTs, related studies have predominantly presented approaches aimed at predicting the completion time of operation in CT focusing on the turnaround time of vessels and QCs. The earliest approach suggested for predicting the turnaround time of vessels included the use of a regression model. This model incorporated many input variables, such as the quantities of QCs and YCs, along with the volumes of discharging and loading [22]. This study examined the comprehensive workload and equipment deployment at the CT. This research examined the comprehensive workload and equipment deployment at the CT, but its scope was limited to aggregated elements, thereby restricting the provision of information.

Subsequently, an approach using the CatBoost was presented to predict the turnaround time of vessels [23]. Cargo type was included as a variable in this study to predict the makespan of vessels. Another study considered factors influencing QC operations and proposed an approach for predicting QC's handling time [9]. According to previous study, QC's handling time prediction considered factors such as the status of container discharging or loading, QC's hatch movement, changes in operation types, worker changes, the presence of twin operations and Full/Empty factors. Based on the findings of the experiment, it was elucidated that employing the Multi-Layer Perceptron (MLP) yields higher precision in predicting QC's handling time in comparison to employing a constant value. In another study that utilized MLP for QC's makespan prediction, an approach for predicting the productivity rate of QCs was presented [7].

**TABLE 1.** Related works for simulation applications for container terminal (CT).

| Author | Simulation Method | Assumption for job handling time | Research target |
|---|---|---|---|
| Cai et al., 2023 [8] | MIP | Average | Minimize makespan |
| Zeng et al., 2015 [26] | MIP | Average | Minimize makespan |
| Bowei et al., 2021 [28] | MIP | Average | Minimize makespan |
| Hsu et al., 2022 [21] | Heuristic | Average | Minimize makespan |
| Zeng and Yang 2009 [24] | MIP | Probability Distribution | Minimize makespan |
| Homayouni et al., 2014 [25] | MIP & GA | Constant | Minimize delays of cranes |
| Sha et al., 2021 [13] | DES | Constant | Determine the equipment utilization rates |
| Cahyono et al., 2022 [18] | DES | Average | Simultaneous allocation |
| Stojaković et al., 2021 [27] | DES | Constant | Determine the optimal number of IT |

The limitations of this study lie in their focus on a macroscopic operational perspective of the CT, which restricts their ability to offer precise predictions of operational outcomes derived from operational plans. In addition, the predicts for QC's and vessel's turnaround time can undergo substantial variations depending on the deployment plans and strategies employed by YTs and YCs. Therefore, in order to obtain precise predictions, it is important to utilize an integrated methodology that combines prediction models and simulation methodologies.

### C. SIMULATION FOR CONTAINER TERMINAL OPERATION

As mentioned in section I, the task of generating and evaluating operational plans in a CT becomes challenging when considering the operational states of QC, IT and YC. As a result, various approaches utilizing simulation techniques have been suggested for the purpose of integrated scheduling. Table 1 presents a comprehensive overview of the different approaches employed in the simulation of port operation plans in CTs. These approaches encompass Mixed Integer Programming (MIP), Genetic Algorithm (GA), and DES. Studies utilizing MIP for port operation plans generally aim to explore plans that minimize the makespan [8], [24], [26], [28]. In a similar vein, studies utilizing GA and heuristic approaches also center their attention on the objective of minimizing makespan and mitigating operational delays [21], [25]. Studies conducted on DES encompasses various areas of topics, such as the determination of optimal equipment placement ratios, quantities, and deployment strategies [13], [18], [27].

The assumptions pertaining to the job handling time (event process time) of simulations typically entail the utilization of average values derived from historical data or fixed parameters. Moreover, it is common to make use of probability distributions to model and replicate stochastic scenarios. Nevertheless, as previously stated, the task of accounting for the inherent variability in CT operations becomes challenging when assuming constant job handling time for equipment. Hence, it is imperative to develop simulation techniques that demonstrate resilience in the face of variability. This study focuses on the issue of predicting real operation times in a CT using simulation techniques, particularly in the presence of uncertain operational conditions. In order to address this issue, an approach called as PDES is proposed. This approach considers the operational attributes of QCs, ITs, and YCs, as well as the characteristics of the data collected in the TOS.

### D. DISCRETE EVENT SIMULATION

The utilization of DES in conjunction with System Dynamics (SD) is a commonly employed technique to support Decision Support Systems (DSS) [29]. The DES utilizes event occurrences as the fundamental mechanism for state transitions, thereby introducing a representative approximation of real-world situations by employing irregular discrete time intervals. DES, which is widely acknowledged for its effectiveness in analyzing complex flows that are crucial in simulations, is particularly skilled in assessing potential scenarios at critical decision points [30]. The advantages of DES contribute to its wide range of applications in various fields, including logistics, healthcare, disaster management, and manufacturing [29], [30], [31], [32].

The flow diagram of DES is illustrated in Figure 3 [33]. In order to commence the DES, it is crucial to initially generate the states and event lists that are relevant to the tasks and equipment involved in the simulation. When an event occurs, the simulation time is modified to match the timestamp of the event, and subsequently, the event is executed. Following this, the simulation advances by modifying its state, removing, or creating events based on the results of the executed events. State changes are indicative of fluctuations in operational conditions that align with specific events, whereas the addition or removal of events transpires when the current events exert an impact on subsequent events. For example, if an event is appended to a queue, or a circumstance emerges where a machine becomes incapable of processing a task, supplementary events may be generated, or pre-existing ones may be removed. Following the conclusion of the simulation for all events, the simulation is finalized, and a comprehensive summary of the outcomes is subsequently generated.

The effectiveness of DES is highly dependent on the accurate specification of parameters, such as task initiation and processing times, which underscores the criticality of parameter configuration. In the initialization phase of the DES, the necessary event list is typically generated by either utilizing pre-defined event sequences or employing a random number generator. Nevertheless, within the realm of CTs, a notable attribute is the concurrent happening of various events originating from multiple sources, such as vessels, QCs, ITs and YCs. In order to address this issue, it becomes crucial to effectively manage events that are occurring simultaneously. Effectively managing concurrent events requires careful consideration of task relationships, which are frequently encountered in CT operations. Efficiently handling simultaneous events becomes important in situations when the timing and expected results of future occurrences are greatly affected by the predicted outcomes of previous events. Moreover, when it comes to the implementation of events if the duration of events adheres to an average or distributed pattern can result in significant discrepancies between the actual outcomes and the errors in prediction. Therefore, it is essential to conduct an in-depth examination of event processing time in order to achieve accurate results in simulations.

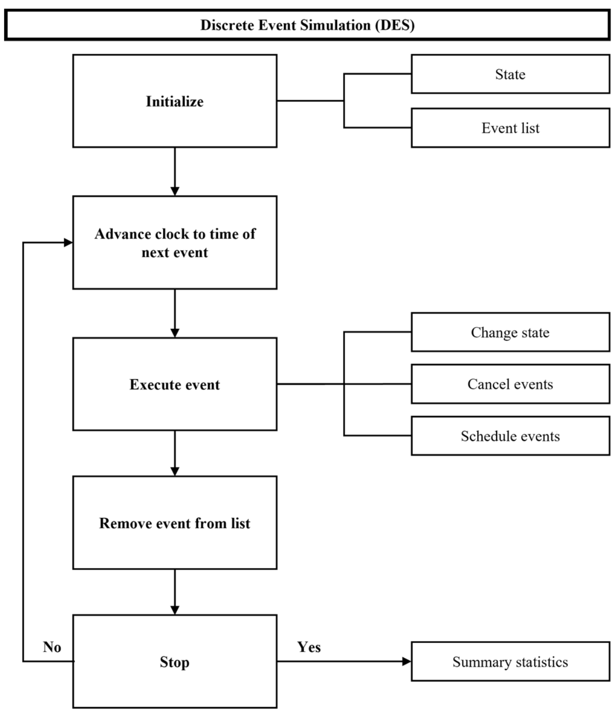



**FIGURE 3. The flow diagram of discrete event simulation (DES).**

In order to address these challenges, it is essential to employ a Data-driven Simulation (DDS) approach, which integrates simulation techniques with data-driven methodologies. DDS has been widely utilized across diverse domains, encompassing manufacturing, healthcare, and transportation, among others. ML algorithms have been implemented in decision support methods within the manufacturing industry [39]. An approach has been made in the field of healthcare to utilize ML algorithms for the purpose of learning from the outcomes of DES and making predictions [40]. A study has introduced an approach to improve simulation performance in the transportation sector by incorporating destination prediction [41].

An additional approach aimed at improving simulation performance involves the application of ML algorithms to predict outcomes by analyzing event execution within simulations. The study proposed the use of ML algorithms to predict truck turnaround times in the context of truck dispatch, which resulted in enhanced simulation accuracy [36]. Another study aimed to incorporate predictive models for machine temperature and the corresponding time required to achieve that temperature into the simulation methodology, with a specific focus on simulating synthetic rubber recipes. The study aimed to incorporate predictive models for machine temperature and the corresponding time required to achieve that temperature into the simulation methodology. It showcased enhanced predictive capabilities by integrating predictive models with the simulation approach, surpassing the predictive performance achieved solely using Random Forest (RF) algorithms [37]. The proposed studies have put several data-driven approaches to improve the performance of simulations. While these studies have proposed data-driven approaches to enhance simulation performance, a limitation arises from the assumption or fixation of the sequence of event occurrences.

The requirement for predefined scenarios in simulation necessitates a constraint on the execution of simulations. In summary, the precise estimation of event process time holds significant importance in the effective utilization of DES for generating authentic outcomes. This study proposes an approach to improve the predictive capabilities of basic DES by incorporating real-world data.

Moreover, in the process of simulation, it is imperative to consider assumptions pertaining to the sequential order of job occurrence and equipment allocation. It is known that proposing cooperative control-based task assignments and real-time adjustment of processing time based on event progress as strategies can improve the efficiency and accuracy of multi-agent systems [42], [43]. In the realm of CT operations, cooperative-control based task assignments are necessary in CTs to improve the precision of simulations.

This study introduces a mechanism to replicate real operational strategies by exchanging the work status of internal trucks. Therefore, it is important to develop an approach that imitates authentic CHE allocation strategies that accurately reflect the simulation states. This study introduces a method to replicate real operational strategies by exchanging the work status of ITs.

## III. PROBLEM DEFINITION AND DATA DESCRIPTION

### A. PROBLEM DEFINITION

As mentioned in section II, to utilize the simulation technique for predicting operation times at CT, it is necessary to consider the event process time and equipment allocation strategies. The primary objective of this study is to provide a comprehensive understanding of a specific issue pertaining to the improvement of predictive accuracy using simulation.

- Problem 1: Selection of ML algorithms for accurately predicting CHE operation times
- Problem 2: The integration of predictive information-based simulation methods and ML algorithms to enhance the prediction performance of operation times
- Problem 3: Improvement of operation time prediction performance through emulation of actual CHE allocation strategies at a CT

The primary objective of this study is to ascertain a ML algorithm that can effectively predict the attributes associated with these operation times. The primary issue examined in this research pertains to the investigation of a ML algorithm's efficacy in accurately predicting the job handling time of QC. The difficulty in predicting the job handling time of operations at CTs arises from the significant variability observed even among tasks of similar nature, compounded by the categorical nature of the variables that are accessible for analysis. In order to address this problem, we suggest a framework that combines BN for selecting important variables and ML predictive algorithm for predicting QC operation times.

The second problem pertains to the anticipate of container operation times by means of incorporating a pre-trained ML algorithm and simulation techniques that rely on predictive data. Predictive information-based simulation is employed as a strategy to address the constraints associated with DES, which requires predetermined event sequences and durations, as outlined in Section II. When tasks are performed simultaneously at different locations and demonstrate substantial variability in terms of time, relying on predetermined task sequences for predictions is insufficient. As a result, this study presents a novel approach for arranging sequences by utilizing predictive information to anticipate potential events prior to performing the simulation.

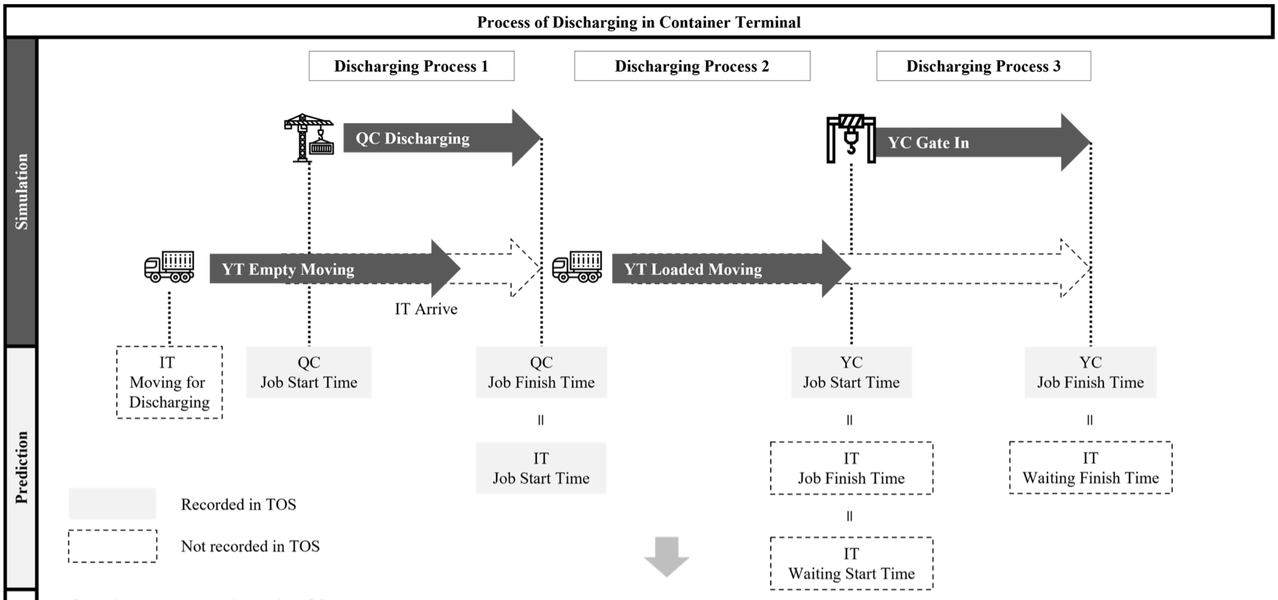
The final problem revolves around the prediction of operation times by simulating CHE allocation strategies in CTs. The berth planning for vessel and the operational planning for QCs are often pre-established due to the intrinsic characteristics of CT operations. The allocation of IT and YC demonstrates a dynamic characteristic, which is susceptible to modifications depending on the advancement of operations. according on previous study, a Model Predictive Allocation (MPA) has been introduced to determine the CHE allocation aiming to minimize makespan [18]. While the objective of reducing makespan in CTs aligns with the goal of minimizing waiting time for CHE assignment, it is important to acknowledge that practical scenarios in these terminals may not always adhere to such practices in CHE allocation. As a result, the implementation of a strategy focused on reducing makespan presents a challenge whereby the projected operating time, obtained from simulation, may be underestimated in relation to the real operation times, leading to a decline in the precision of predictions. In order to address this issue, this study presents a sampling model to the allocation strategy of IT. The sampling model is an approach used to allocate IT resources to the job in QC based on the observed job allocation pattern of IT at the CT. The allocation approach is implemented by a simulation process that accurately replicates the actual job distribution pattern seen in real IT activities at the CT.

### B. DATA DESCRIPTION

In this section, the data used for predicting the operation times in CT and data recorded process are introduced. The data utilized in this study pertains to information pertaining to the historical operation of containers. Figure 4 depicts the procedure employed for recording data during the discharging operation. According to Figure 4, the process of discharging can be observed to occur in a series of three distinct sequential processes. The initial phase of the discharging operation entails the QC descending containers from the vessel. In the given context, prior to the commencement of QC's operation, the IT initiates the empty moving operation, guaranteeing its arrival prior to the conclusion of QC's operation. The exact commencement time of the IT's empty moving is not documented, and the initiation of the unloading operation by QC is recorded as the start time. The recorded finish time corresponds to the point at which the QC concludes the operation. The termination time of IT's empty moving is equivalent to QC's completion time and is not documented separately.

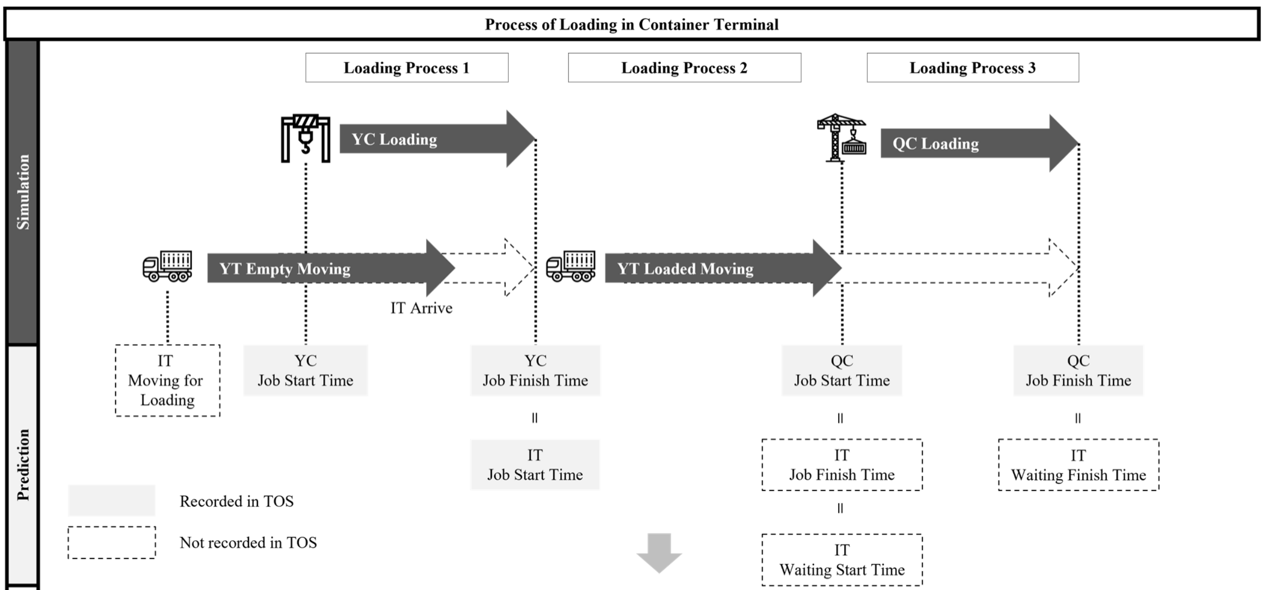
The second phase of the discharging operation entails the IT receiving containers from the QC and subsequently engaging in the transportation of loaded containers towards the YC. The initial process's completion time is documented as the commencement time for the subsequent process, while the assigned start time for the YC is recorded as the conclusion time of the second process. Within this context, the designated initiation time for the YC is established as the conclusion time of the preceding operation. The difficulty in collecting data arises when the YC begins the task immediately after completing the previous operation. However, if there is a delay between the completion of the previous task and the start of the YC's operation, resulting in a prolonged idle time, accurately estimating the precise start time of the YC's operation becomes challenging.

The inherent nature of TOS data presents challenges in accurately determining the durations of movement and




| Vessel ID | Hatch No | Container ID | Equipment Type | Equipment No | Job Type | Start Time | Finish Time | Yard Block |
|---|---|---|---|---|---|---|---|---|
| Vessel 1 | Hatch 1 | Container 1 | QC | QC 1 | QC Discharging | 2020-03-01 04:02:00 (QC Job Start Time) | 2020-03-01 04:05:30 (QC Job Finish Time) | |
| Vessel 1 | | Container 1 | YT | YT 1 | YT Loaded Moving | 2020-03-01 04:05:30 (IT Job Start Time) | 2020-03-01 04:08:10 (IT Job Finish Time) | Block 1 |
| Vessel 1 | | Container 1 | YC | YC 1 | YC Gate In | 2020-03-01 04:08:10 (YC Job Start Time) | 2020-03-01 04:10:20 (YC Job Finish Time) | Block 1 |

**FIGURE 4.** Data recording process for discharging job in TOS.




| Vessel ID | Hatch No | Container ID | Equipment Type | Equipment No | Job Type | Start Time | Finish Time | Yard Block |
|---|---|---|---|---|---|---|---|---|
| Vessel 1 | Block 2 | Container 1 | QC | YC 2 | YC Gate Out | 2020-03-01 05:02:00 (YC Job Start Time) | 2020-03-01 05:05:30 (YC Job Finish Time) | |
| Vessel 1 | Block 2 | Container 1 | YT | YT 2 | YT Loaded Moving | 2020-03-01 05:05:30 (IT Job Start Time) | 2020-03-01 05:08:10 (IT Job Finish Time) | |
| Vessel 1 | | Container 1 | YC | QC 2 | QC Loading | 2020-03-01 05:08:10 (QC Job Start Time) | 2020-03-01 05:10:20 (QC Job Finish Time) | Hatch 2 |

**FIGURE 5.** Data recording process for loading job in TOS.

**TABLE 2.** Features of container movement history in TOS.

| Notation | Feature | Description |
|---|---|---|
| $k$ | Job sequence of all QCs | The sequence of event $k$, $K = \{1,2,\ldots,n(K)\}$ |
| $v$ | Vessel ID | The id of Vessel. $V = \{1,2,\ldots,n(V)\}$ |
| $q$ | QC ID | The id of QC, $Q = \{1,2,\ldots,n(Q)\|v\}$ |
| $i$ | IT ID | The id of IT, $I = \{1,2,\ldots,n(I)\}$ |
| $c$ | YC ID | The id of YC, $C = \{1,2,\ldots,n(C)\}$ |
| $j$ | Job sequence of QC ($q$) | The sequence of jobs QC ($q$), $J = \{1,2,\ldots,n(J)\|q\}$ |
| $x^{cid}(k)$ | Container ID | The container id of event $k$ |
| $x^{vid}(k)$ | Vessel ID | The vessel id of event $k$ |
| $x^{berth}(k)$ | Berth location | The berth location of event $k$ |
| $x_v^h(k)$ | Hatch number | The number of hatches of event $k$ of vessel $v$ |
| $x^{block}(k)$ | Block location | The location of block of event $k$ |
| $x^{job}(k)$ | Job type | Unloading or loading job (Discharging: 0, Loading: 1) |
| $x^{tw}(k)$ | Single/Twin | Single job or twin job (if single = 1, else 2) |
| $x_q^{wid}(k)$ | QC worker ID | The worker id of QC ($q$) of event $k$ |
| $y_q^s(k)$ | QC job start time | The job start time of QC ($q$) of event $k$ |
| $y_t^s(k)$ | IT job start time | The job start time of IT ($i$) of event $k$ |
| $y_c^s(k)$ | YC job start time | The job start time of YC ($c$) of event $k$ |
| $y_q^f(k)$ | QC job finish time | The job finish time of QC ($q$) of event $k$ |
| $y_i^f(k)$ | IT job finish time | The job finish time of IT ($i$) of event $k$ |
| $y_c^f(k)$ | YC job finish time | The job finish time of YC ($c$) of event $k$ |
| $y_q(k)$ | QC container handling time | The job handling time of QC ($q$) of event $k$ |
| $y_i(k)$ | IT container moving time | The job handling time of IT ($i$) of event $k$ |
| $y_c(k)$ | YC container handling time | The job handling time of YC ($c$) of event $k$ |

operational activities for both IT and YC, thereby complicating the predictive phase. This study involves the utilization of TOS data to perform simulation, prediction, and result output. The simulation, situated at the uppermost part of Figure 4, is accountable for the regular updating of the operational status of the equipment, as illustrated. The prediction segment, situated centrally, fulfills the function of offering anticipatory data regarding the initiation and completion times of operations, which in turn facilitate the updating of equipment states within the simulation. Finally, the results derived through the integration of simulation and prediction are stored. Figure 5 provides a detailed illustration of the data recording process during the loading operation. In contrast to the discharging operation, the loading operation involves the IT commencing the movement of empty items prior to the initiation of the YC's operation. The initial time at which the first process commences is documented as the operation start time of YC, while the time at which it concludes is recorded as the finish time. Within the given context, akin to the process of discharging, the initiation time of the YC is documented as the termination time of its antecedent operation.

The second procedure, which bears resemblance to the discharging operation, involves documenting the initiation time of the loaded movement by the IT. Additionally, the start and finish times of the loading operation's second and third processes are recorded when the QC begins and completes the loading operation. Like the discharging process, the loading process is also divided into simulation, prediction, and result stages. The recording of container operation history data by the TOS is comprehensive, yet not exhaustive, thereby presenting difficulties in the application of analytical methodologies. The analysis of TOS data reveals that a key factor in assessing the duration of operations at the CT is the identification of commencement and conclusion times, which are contingent upon the interrelationships among different types of equipment. The scheduling of the operation is dependent upon the synchronization of every individual piece of equipment. For example, if the QC commences its operation promptly, the conclusion of said operation is contingent upon the timely arrival of the IT. In contrast, if the IT arrives ahead of schedule yet encounters delays in the processing of the operation by the QC or YC, the IT is unable to successfully carry out the task and is compelled to wait.

Table 2 provides a concise overview of the container movement history that has been collected within the TOS. According to the table 2, the sequence of jobs of CT is defined as $k$. The numerical representations of vessel, QC, IT and YC are denoted as $v$, $q$, $i$ and $c$ respectively. The container movement history consists of features representing the id of container ($x^{cid}$ $(k)$), the id of vessel, the berth location of vessel ($x^{berth}$ $(k)$), the number of hatch ($x_v^h$ $(k)$) and the location of block ($x^{block}(k)$). Job type ($x^{job}(k)$) indicates whether the job $k$ is discharging or loading. Single/Twin ($x^{tw}(k)$) indicates

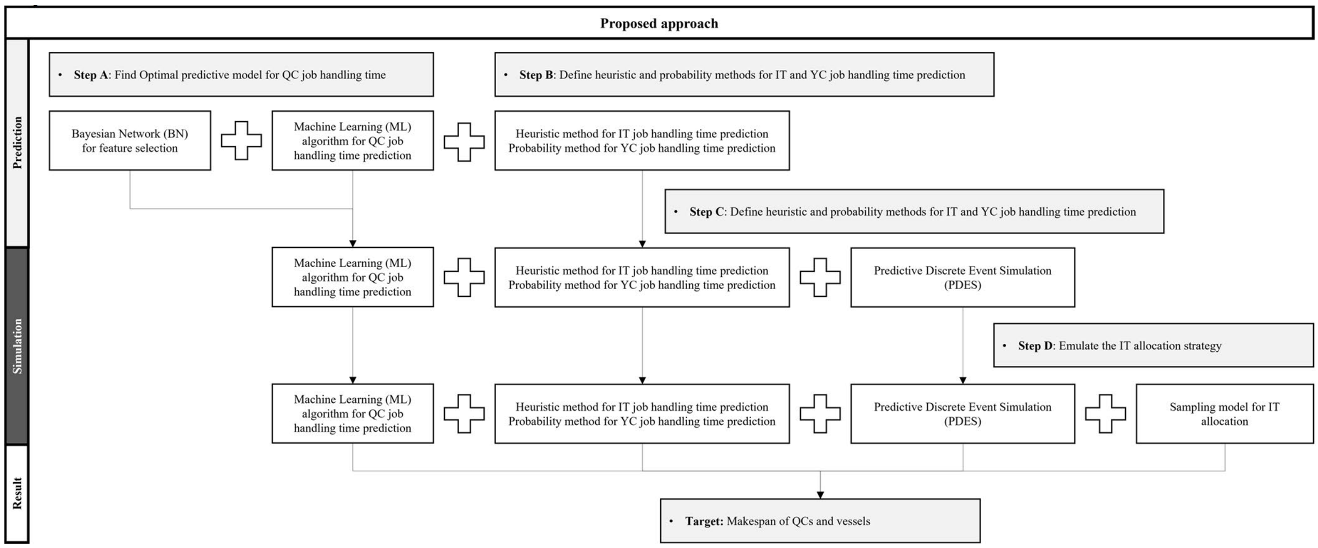



**FIGURE 6. Proposed approach for Container Terminal (CT) operation time prediction.**

whether the job $k$ is single container job or twin container job. The worker id of QC is defined as $x_q^{wid}(k)$.

The subsequent features pertain to the particulars of the CHE operation times. The job start time of QC, IT and YC are denoted as $y_q^s(k)$, $y_i^s(k)$ and $y_c^s(k)$ respectively, while $y_q^f(k)$, $y_q^f(k)$ and $y_c^f(k)$ represent the job finish time. Finally, the job handling time of QC, IT and YC are defined as $y_q(k)$, $y_i(k)$ and $y_t(k)$, respectively. The $k-$th job's finish time is calculated by adding the job handling time to the $k-$th job's start time, in compliance with the operational procedures at the CT. When considering the QC, for example, the calculation proceeds as follows.

$$y_q^s(k) \leftarrow y_q^f(k-1) \quad (1)$$

$$y_q^f(k) \leftarrow y_q^s(k) + y_q(k) \quad (2)$$

Equation (1) represents the procedure by wherein the finish time of the $k-1$th job for the QC, denoted as $y_q^f(k-1)$, is updated to the start time of the $k-$th job ($y_q^s(k)$). Equation (2) represents the process of calculating the finish time of the $k-$th job by adding the $k-$th job handling time ($y_q(k)$) to the start time of the $k-$th job. Similar to QC, for IT and YC, the previous job's task time is also the current job's start time. After that, the update is completed by adding the necessary job handling time to it. In the following section, a discussion of the utilization of these principles for the purpose of running simulations and predicting operation times will be presented.

## IV. PROPOSED APPROACH

This paper introduces an approach for predicting the operation times of CT by utilizing a ML algorithm in conjunction with simulation techniques. In order to predict the operation times of CT using container movement history collected from TOS, three issues need to be addressed. Firstly, the selection of a prediction model for QC job handling time is imperative. The second issue involves the enhancement of predictive performance by combining PDES with trained predictive ML algorithm. Thirdly, this study focuses on emulating the CHE allocation strategy implemented by a real CT. This is done to address the problem of reduced predictive accuracy when applying CHE allocation strategies that minimize makespan.

The approach utilized in this study to predict the operation times of a CT, is depicted in Figure 6, which showcases the PDES methodology. According to Step A, an approach is introduced for predicting the job handling time of QC, which are important in CT. The prediction of job handling time for QC entails the application of BN for the identification of relevant features, as well as the utilization of ML algorithms for the training of predictive models. In the next step, heuristic method and probability method are utilized for IT and YC job handling prediction. In the step C, the models that were acquired in Step A and B are employed to construct a simulation model based on predictions. In the final step, the prediction is executed by simulating the IT allocation strategy of a real CT through the utilization of a sampling model.

### A. ML ALGORITHM FOR QC JOB HANDLING TIME PREDICTION

Previous study has identified several factors that have an impact on the job handling time of QCs. These factors encompass the nature of the operation, specifically whether it involves discharging or loading, as well as the movement of QC hatches. Additionally, variations in job types, changes in worker, the presence of twin operations, and consideration related to Full/Empty status have also been found to influence QC job handling time [9].

Additionally, it has been noted that the proficiency of both QC spec and workers also impacts QC operations [18], [34]. This study aims to identify the pertinent elements that have a

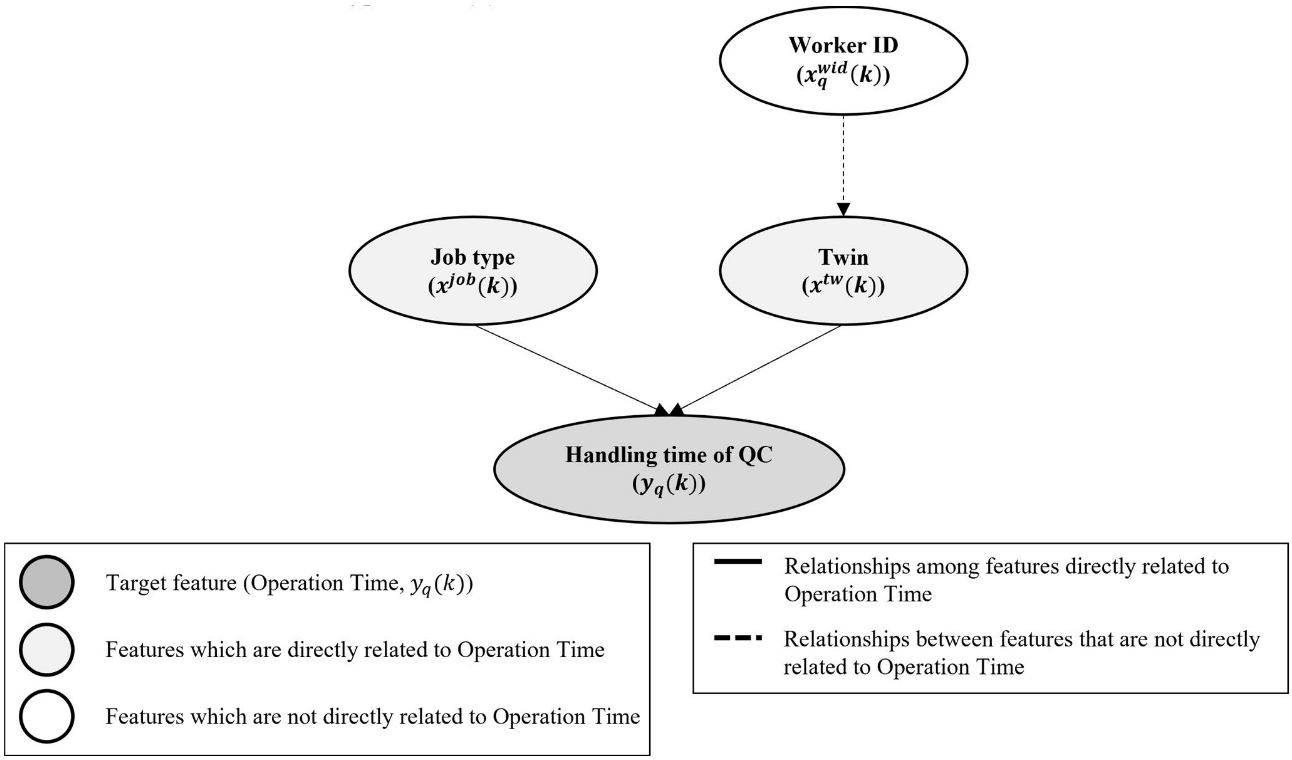



**FIGURE 7.** Directed acyclic graph (DAG) explored by Bayesian Network (BN).

direct impact on job handling time in QC by analyzing data from the TOS. The prediction of job handling time in QC is thereafter performed by means of training ML algorithms. BN is employed for feature selection, and BN utilizing Hill Climbing (HC) has demonstrated commendable performance in feature selection. It is known to effectively address prediction problems when combined with ML algorithms such as Artificial Neural Network (ANN) and SVM [35].

Figure 7 is the Directed acyclic graph (DAG) that explored by BN using HC. The lines indicate the relationships between each feature. The dark-grey circle represents the target feature, job handling time of QC ($y_q(k)$), while the light-grey circle represents the features that directly affect job handling time of QC. The white circle represents the feature that indirectly affect job handling time. The dotted line indicates the relationship between feature that indirectly affect job handling time of QC, while the solid lines represent the relationships that indirectly affect job handling time. According to the DAG, the features that directly consists of two features: Job type ($x^{job}(k)$) and Twin ($x^{tw}(k)$). The job type, representing whether a discharging or loading job, significantly influences the handling time of QCs due to differences in the container handling process. Furthermore, the handling time of QCs is directly affected by twin jobs, which involves the simultaneous transportation of two containers.

$$\hat{y_q}(k) = f_q(x^{job}(k), x^{tw}(k)) \tag{3}$$

Equation (3) represents $f_q$(ML algorithm) trained to predict the handling time of QCs using features deemed to directly influence QC handling time.

### B. HEURISTIC AND PROBABILITY METHODS FOR IT AND YC JOB HANDLING TIME

In order to predict for CT operation times, it is imperative to consider the handling time of the designated IT ($i$) and YC ($c$). The handling time of IT and YC are defined as follows:

$$\hat{y_i}(k) = \begin{cases} \frac{1}{ve} dijkstra\left(x^{berth}(k), x^{block}(k)\right), \\ \quad if\ x^{job}(k) = 0 \\ \frac{1}{ve} dijkstra\left(x^{block}(k), x^{berth}(k)\right), \\ \quad if\ x^{job}(k) = 1 \end{cases} \tag{4}$$

The job handling time of IT ($\hat{y_i}(k)$) defined in equation (4), represents the calculation method for the travel time of IT considering on the location of the $k-$th container operations. The calculation of travel time for IT involves dividing the estimated distance covered by IT, as determined by the layout of the CT, by the velocity ($ve$). The determination of distance is accomplished through the utilization of the Dijkstra algorithm, a computational procedure that systematically explores and identifies the shortest path.

The layout of BPT, South Korea is described in Figure 8, which provides information on the positions of berths and blocks. Additionally, the defined directions in which IT can

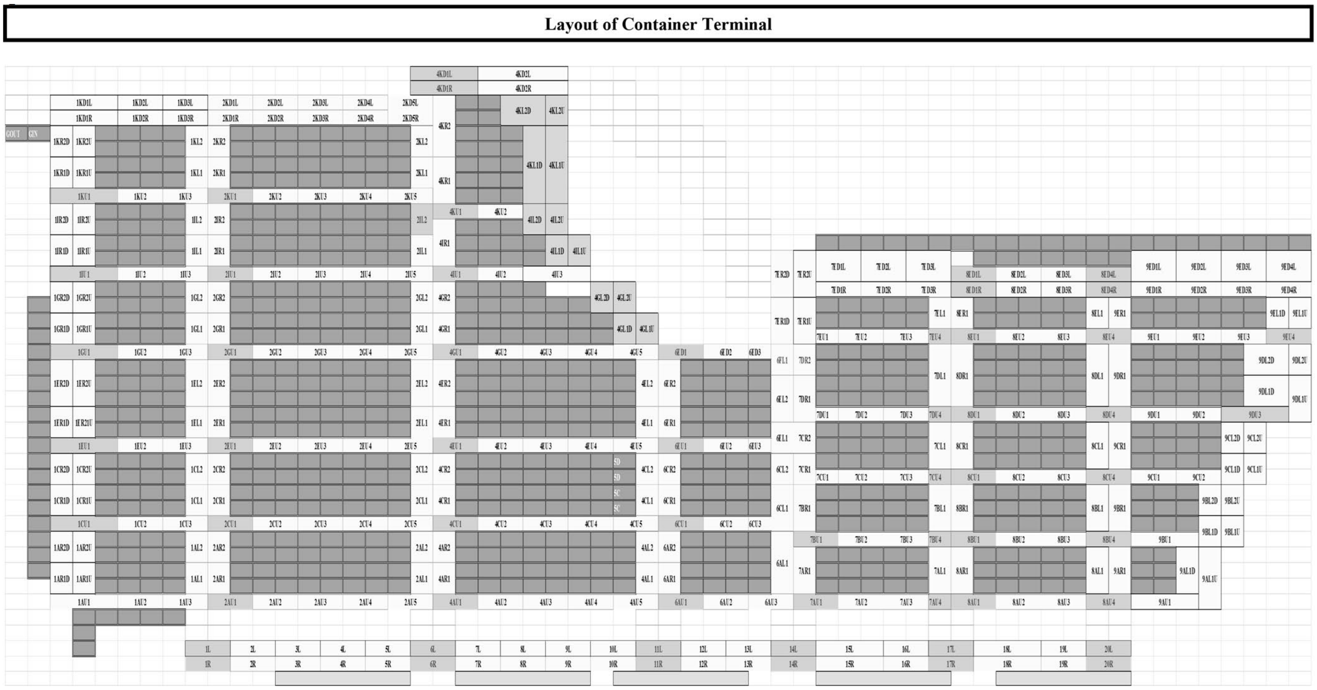



**FIGURE 8.** The layout of the Busan Port Terminal (BPT), South Korea.

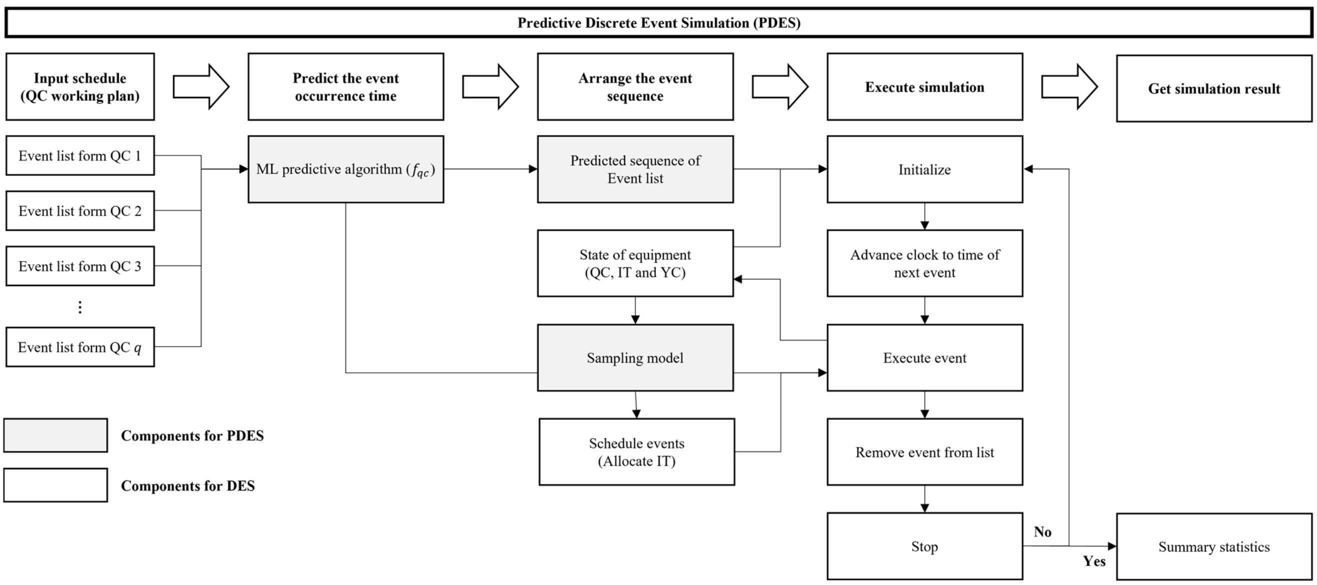



**FIGURE 9.** Flowchart of predictive discrete event simulation (PDES).

move for each node are indicated. According to equation (4), if the $k-$th job is a discharging, the travel distance is calculated as $dijkstra\left(x^{berth}(k), x^{block}(k)\right)$, representing the distance from the berth to the yard. Conversely, if the $k-$th job is a loading, the travel distance is calculated as $dijkstra\left(x^{block}(k), x^{berth}(k)\right)$.

As mentioned in section III, accurately documenting the start and finish times of YC activities poses difficulties within the data acquisition procedures of the TOS. Therefore, the handling time of YC is represented using a probability distribution in the model. The probability distribution for YC job handling time is assumed to follow a Gamma distribution. The justification for utilizing the Gamma distribution stems from the fundamental property that job handling time cannot be negative. Therefore, it is necessary to employ a distribution that exclusively encompasses positive values.

$$\hat{y_c}(k)\ Gamma(\alpha, \beta) \tag{5}$$

The job handling time of YCs ($\hat{y_c}(k)$) is represented by equation (5), which assumes a Gamma distribution. The application of Maximum Likelihood Estimation (MLE) is utilized to estimate the job handling time of YC, which is obtained from data collected by the TOS. The MLE of the estimated Gamma distribution is defined with scale $\alpha$ and shape $\beta$.

### C. PREDICTIVE DISCRETE EVENT SIMULATION

In this section, a PDES approach utilizing the ML predictive algorithm and a sampling model for IT allocation is presented. Figure 9 describes the procedural of the PDES. The initial step entails entering the operational plan into the simulation. The operational plan is determined by the sequential container operations scheduled for the QCs involved in the primary operations, which are detailed in table 2. The second step utilizes a ML predictive algorithm to predict the start and end times of individual container operations assigned to each QC. It is assumed in this step that the ITs and YCs provide uninterrupted support for all operations of the QC. In the third step, the sequence of events is sorted according to the predicted start and end times of container operations that have been allocated to QCs in the second step. This process establishes the order in which events happen at different locations within the CT and produces the necessary list of events for the simulation. The simulation is carried out in the fourth step, adhering to the operational principles of DES outlined in Figure 3. Upon the execution of each event, the states of the QCs, ITs, and YCs are promptly updated. During the process of updating states, the start and finish time of QCs, ITs, and YCs are computed. These updated states are then utilized as input for the sampling model, which determines the allocation of ITs. Executed events are excluded from the list. After the simulation concludes, the final step entails summarizing the results.

#### 1) PREDICT THE SEQUENCE OF EVENT

In order to carry out the simulation, it is necessary to establish an event list. The event list is generated by utilizing ML predictive algorithm to predict the start and finish times of QC operations. This approach is employed to effectively arrange the simultaneous occurrences of QC jobs. Since the assignment of ITs has not taken place in this phase, it is presumed that all ITs efficiently facilitate QC operations without any interruptions during event list generation. Assuming that the jobs assigned to QC ($q$) are denoted as $J_q = \{1, 2, \ldots, n(J_q)\}$, the anticipated start and finish times for $J_q$ are predicted by referring to the equations (1) - (3). Following the completion of computations for all jobs across all QCs, an event list is generated based on the anticipated job start times. The event list is defined as $K = \{1, 2, \ldots, n(K)\}$.

#### 2) INITIALIZATION STAGE

After the event list is generated, the CHE states are initialized as follows:

$$y_q^s(k), y_q^f(k) \leftarrow y_q^s(1) \tag{6}$$
$$y_s^i(k), y_f^i(k) \leftarrow y_q^s(1) \tag{7}$$
$$y_c^s(k), y_c^f(k) \leftarrow y_q^s(1) \tag{8}$$

In the initialization stage for the simulation, as defined in equations (6) – (8), at the initial job sequence ($k = 1$), all start and finish times for each QC ($q$), IT ($i$), and YC ($c$) are initialized to the initial job start time ($y_q^s(1)$). After this initialization, the simulation advances according to the given event list $K$.

#### 3) STATE UPDATING RULE FOR SIMULATION

Throughout the simulation, the process of updating the state of CHE is differentiated based on discharging and loading jobs. The state update for discharging jobs adheres to the sequence of QC, IT and YC as depicted in Figure 4.

$$y_q^s(k) \leftarrow max(y_q^f(k-1), y_i^f(k-1)) \tag{9}$$
$$y_q^f(k), y_i^s(k) \leftarrow y_q^f(k) + \hat{y}_q(k) \tag{10}$$
$$y_c^s(k) \leftarrow max(y_c^f(k-1), y_s^i(k) + \hat{y}_i(k)) \tag{11}$$
$$y_i^f(k), y_c^f(k) \leftarrow y_c^s(k) + (\hat{y}_c(k) \times x^{tw}(k)) \tag{12}$$

Equation (9) updates the start time of the $k-$th job for QC ($q$). It is updated to the larger of the finish times of $k-$ 1th job for QC ($q$) and IT ($i$). This equation considers the precedence relationship in the CT, indicating that the $k-$th job can only start after the preceding job has been completed. Equation (10) calculates the finish time of the $k-$th job for QC ($q$), utilizing the ML predictive algorithm defined in equation (3). The start time of the $k-$th job for IT ($i$) is updated to match the finish time of the QC ($q$). Equation (11) updates the start time of the job for YC ($c$), selecting the larger value between the arrival time of IT ($y_s^i(k) + \hat{y}_i(k)$) and the finish time of the preceding job for YC ($c$). If the $k-$th job is twin job, the finish time is increased by the product of the number of handling ($x^{tw}(k)$) and the expected job handling time of YC referring to equation (5).

Conversely, if the $k-$th job is loading job, the state of the CHE is updated according to equations (13) – (16). The sequencing of state updates for loading jobs is reversed compared to discharging jobs, with YC, IT, and QC being updated in that order, as illustrated in Figure 5.

$$y_c^s(k) \leftarrow max(y_c^f(k-1), y_i^f(k-1)) \tag{13}$$
$$y_c^f(k), y_i^s(k) \leftarrow y_c^f(k) + \left(\hat{y}_c(k) \times x^{tw}(k)\right) \tag{14}$$
$$y_q^s(k) \leftarrow max\left(y_q^f(k-1), y_s^i(k) + \hat{y}_i(k)\right) \tag{15}$$
$$y_q^f(k), y_i^f(k) \leftarrow y_q^s(k) + \hat{y}_q(k) \tag{16}$$

Equation (13) updates the start time of the $k-$th job for YC ($c$). It is updated to the larger of the finish times of $k-1$th job for YC ($c$) and IT ($i$). Equation (14) calculates the finish time of the $k-$th job for YC ($c$), and the start time of the $k-$th job for IT ($i$) is updated to match the finish time of the YC ($c$). Similarly, in the case of twin jobs, the update is performed by multiplying the job handling time by the number of handling occurrences. Equation (15) updates the start time of the job for QC ($q$), selecting the larger value between the arrival time of IT and the finish time of the preceding job for QC ($q$).

Finally, the finish time of QC ($q$) is updated by adding the job start time to the job handling time of the QC.

#### 4) SUMMARY STATISTICS

Upon the completion of the simulation, the makespan ($t$) for QC and vessel is computed to derive the predicted results for operational time.

$$y_q^t = y_q^f(n(K)) - y_q^s(1) \tag{17}$$

$$y_v^t = max\, y_{q\in v}^f(n(K)) - min\, y_{q\in v}^s(1) \tag{18}$$

Equation (17) is utilized to calculate the QC's makespan ($y_q^t$). The makespan of QC ($q$) is defined as the interval of finish time of final job ($y_q^f(n(K))$) and the start time of the initial job ($y_q^s(1)$). Likewise, the formula for calculating the vessel 's makespan time is equation (18). It is determined by the interval between the final job finish time for QCs involved in operations with vessel ($v$) and the start time for QCs involved in vessel.

### D. SAMPLING MODEL FOR CHE (IT) ALLOCATION

As mentioned in section III, the operational sequence of the QCs is predetermined, but the deployment plans for ITs and YCs demonstrate flexibility depending on the changing operational conditions. In order to address these specific attributes, this study presents a sampling model for the allocation of ITs. The allocation of ITs utilizing simulation model is executed through the following 3 steps.

- Step 1: Approximately the probability distribution for the hourly frequency of job assignments for ITs based on the container movement history
- Step 2: Estimate the expected frequency of job assignments per hour for the intended ITs
- Step 3: Assign jobs to ITs during the simulation process involves considering the expected frequency of job assignments per hour for ITs

The first step involves deriving the deployment pattern of IT from the container movement history. The deployment pattern of IT is defined by the frequency of job assignments per hour and follows a multinomial distribution. In second step, the derived probability distribution is employed to estimate the frequency of job assignments per hour for intended ITs. ITs which have expected job frequencies, face limitations that prevent them from being assigned jobs that exceed their projected frequency. The ascertained probability distribution denoting the pattern of IT ($i$) is denoted as $p_i^n$, with the total number of intended ITs designated as $n(I)$. The expected frequency of job assignments per hour for IT ($i$) is computed by the product of $n(I)$ and $p_i^n$. For example, considering a deployment of $n(I)$ ITs, $n(I) \times p_i^1$ ITs are assigned jobs only once per hour, and no more. Similarly, the number of ITs that can be assigned a maximum of two jobs per hour is calculated as $n(I) \times p_i^2$. The sampling model employed in this study prevents the allocation of all available equipment using a minimization strategy, thereby enhancing predictive accuracy.

If equipment allocation adheres to the minimize approach for all resources, a predicament occurs whereby the projected operation times are reduced in comparison to the actual operating times. Previous studies have demonstrated that employing a minimization strategy for CHE allocation poses challenges in accurately predicting operation times compared to actual operational outcomes. The approach proposed in this study is specifically designed to overcome these limitations. During the simulation, the selection of an IT for the $k-$th job is determined at the $k - 1$th state by randomly choosing from ITs that have a job input frequency per hour below a predefined expected frequency, and whose finish time for the previous job is one of the earliest among eligible ITs.

Algorithm 1 addresses the algorithm of the proposed PDES in this study. In the first line, the QC plan ($J_q$) is invoked. Lines 2 to 6, as elucidated in Section IV-A.1, involve retrieving the scheduled task lists of QCs, predicting their expected job start time and finish time. Subsequently creating the event list $K$ by sorting them in ascending order based on the commencement sequence. The line 7 signifies the initialization of the operational states of QCs, ITs, and YCs prior to commencing the simulation. Lines 8 to 21 entail the simulation process for the $k-$th job, during which the operational states of CHE are updated. The update occurs in accordance with the processes of discharging and loading jobs. Upon the conclusion of the simulation, the makespan of the QC and vessel is computed, and the corresponding values are returned.

### E. ANALYZE THE COMPLEXITY OF ALGORITHM

This study aims to examine real-world circumstances, so an examination of algorithm complexity has been carried out. The algorithmic complexity is analyzed by iteratively changing the quantity of event lists ($K$) in the input data of the proposed approach to measure its computational complexity. Based on the analysis results, the computational complexity shows a linear relationship which leads to the definition as Big O as $O(|K|)$.

## V. EXPERIMENTS

In this section, the experimental results of comparative studies conducted to validate the predictive performance of the approach proposed in this study are discussed.

### A. DESCRIPTION OF DATA USED IN EXPERIMENTS AND TYPE OF EXPERIMENTS

The data used in the experiments were collected from January 2020 to February 2020 at BPT. The container movement history data obtained from BPT included information on more than 180 thousand containers movement. The data used for the experiments are identified in table 3. According to table 3, to train the ML algorithm, container movement history from January 2020 was used. The period for predictive performance was set after February 2020. BPT operates a total of 15 QCs, 78 ITs, and 58 YCs. The assessment data

**Algorithm 1** Predictive Discrete Event Simulation for predicting operational times in Container Terminal (CT)

**Set**QC plans $J_q$# $J_q = \{1, 2, \ldots, n(J_q)\}$
**for each** QC $q \in Q$
**for each** $j \in J_q$
$y_q^s(j) \leftarrow y_q^f(j-1)$
$y_q^f(j) \leftarrow y_q^s(j) + \hat{y}_q(j)$
Generate the event list $K$ by sorting $y_q^s(j)$# $K = \{1, 2, \ldots, n(K)\}$
$y_q^s(k), y_q^f(k) \leftarrow y_q^s(1); y_s^i(k), y_f^i(k) \leftarrow y_q^s(1); y_c^s(k), y_c^f(k) \leftarrow y_q^s(1)$# Initialize state
**while** $k \leq n(K)$ **do**
select IT $(i)$ for job $k$ according to sampling model
**if** $x^{job}(k) = 0$# Discharging job
$y_q^s(k) \leftarrow max(y_q^f(k-1), y_i^f(k-1))$
$y_q^f(k), y_i^s(k) \leftarrow y_q^s(k) + \hat{y}_q(k)$
$y_c^s(k) \leftarrow max(y_c^f(k-1), y_i^s(k) + \hat{y}_i(k))$
$y_i^f(k), y_c^f(k) \leftarrow y_c^s(k) + (\hat{y}_c(k) \times x^{tw}(k))$
**else** # Loading job
$y_c^s(k) \leftarrow max(y_c^f(k-1), y_i^f(k-1))$
$y_c^f(k), y_i^s(k) \leftarrow y_c^s(k) + \left(\hat{y}_c(k) \times x^{tw}(k)\right)$
$y_q^s(k) \leftarrow max\left(y_q^f(k-1), y_i^s(k) + \hat{y}_i(k)\right)$
$y_q^f(k), y_i^f(k) \leftarrow y_q^s(k) + \hat{y}_q(k)$
$k \leftarrow k + 1$
**end while**
$y_q^t = y_q^f(n(K)) - y_q^s(1); y_v^t = max\, y_{q\in v}^f(n(K)) - min\, y_{q\in v}^s(1)$# Get result
**return** $y_q^t, y_v^t$

**TABLE 3.** Data used for experiment.

| Train/Test | Dates | Total number of CHE | | | | |
|---|---|---|---|---|---|---|
| | | QC | | IT | | YC |
| Train | 2020-01-01 ~ 2020-01-31 | 15 | | 78 | | 58 |
| Test | 2020-02-01 ~ 2020-02-14 | | | | | |
| Test Case | Count | Number of scheduled jobs, vessel, and CHE input's range | | | | |
| | | Job | Vessel | QC | IT | YC |
| Case 1 | 9 | 596 ~ 1040 | 3 ~ 6 | 7 ~ 12 | 37 ~ 66 | 30 ~ 46 |
| Case 2 | 18 | 524 ~ 1478 | 7 ~ 9 | 10 ~ 13 | 50 ~ 68 | 34 ~ 46 |
| Case 3 | 3 | 1195 ~ 1335 | 10 | 12 ~ 13 | 61 ~ 70 | 39 ~ 47 |

**TABLE 4.** Descriptions of experiments.

| Experiments | Data | Input CHE plan | Model | Benchmark |
|---|---|---|---|---|
| Exp 1 | Train | Historical plan | ML, ML + PDES | |
| Exp 2 | Test | Historical plan | ML + PDES | ML, DL |
| Exp 3 | Test | Dynamic plan | ML + PDES + Sampling model | DES + MPA |

for the experiment was divided into three discrete cases. Each case is characterized by a 12-hour operational plan, which is differentiated by the quantity of vessels scheduled for operations within that specific period. Regarding case 1, the planned number of vessels for a 12-hour period varies from a minimum of 3 to a maximum of 6. In Case 2, the range is from 7 to 9 vessels, and in Case 3, it includes 10 vessels in the operational plan. Furthermore, the planed number of CHE has been specified for each case. Case 1 involves the deployment of 7 to 12 QCs, 37 to 66 ITs, and 30 to 46 YCs. Due to their larger operational scale, Case 2 and 3 allocate a greater amount of CHE in the operational plans compared to Case 1. Finally, then, three experiments for CT operation times prediction were performed.

- Exp 1: Experimentation on the optimal ML algorithm for predicting the job operational time
- Exp 2: Experimentation on comparing the predictive performance of operational time between non-utilization and utilization of simulation
- Exp 3: Experimentation on comparing the predictive performance of operational time between DES and PDES

Table 4 describes the details of experiments. Exp 1 presents the results of exploring the optimal ML algorithm for accurately predicting the job handling time of QCs in the simulation. The train data is utilized for experiment. ML and ML + PDES are applied for exploring the optimal ML algorithm for prediction. In Exp 1, the validation of the performance of the ML + PDES involves utilizing the recorded actual job assignments for QC, IT, and YC from the TOS. In Exp 2, comparative experiments assessed the predictive performance of the proposed approach for operational time in comparison with various ML algorithms using test data. The main goal of the comparative experiment in Exp 2 is to present proof that a combined approach utilizing simulation and ML algorithm is superior in predicting operational time in CTs. Similarly, in Exp 2, the actual job assignments recorded in the TOS are directly utilized for prediction. Exp 3, like Exp 2, carries out a performance comparison experiment to predict the operation times of the CT. The objective of Exp 3 is to benchmark the DES and MPA approach proposed in previous study, specifically focusing on CHE allocation in CTs [18]. Through this comparison, the aim is to substantiate the predictive performance of the PDES, and the sampling model proposed in this study for CHE allocation.

The error in this study were evaluated using the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE). The metrics are defined by the prediction errors for the job handling time of QCs, as well as the prediction errors for the makespan of both QCs and vessels.

$$MAE_1 = \frac{1}{n(K)} \sum_{k=1}^{n(K)} |y_q(k) - \hat{y_q}(k)| \tag{19}$$

$$MAPE_1 = \frac{1}{n(K)} \sum_{k=1}^{n(K)} |\frac{y_q(k) - \hat{y_q}(k)}{y_q(k)}| \tag{20}$$

$$RMSE_1 = \sqrt{\frac{1}{n(K)} \sum_{k=1}^{n(K)} (y_q(k) - \hat{y_q}(k))^2} \tag{21}$$

The first metric is the numerical error for QC job handling time. In equations (19), (20) and (21), $MAE_1$, $MAPE_1$ and $RMSE_1$ represent the prediction error, where the actual job handling time is $y_q(k)$. The unit of $MAE_1$ is second. Similarly, the second and third errors for the makespan of both QCs and vessels are described in equations (22) – (27).

$$MAE_2 = \frac{1}{n(Q)} \sum_{q=1}^{n(Q)} |y_q^t - \hat{y_q^t}| \tag{22}$$

$$MAPE_2 = \frac{1}{n(Q)} \sum_{q=1}^{n(Q)} |\frac{y_q^t - \hat{y_q^t}}{y_q^t}| \tag{23}$$

**TABLE 5. Parameter for experiment.**

| Notation | Description | Value |
|---|---|---|
| $ve$ | The velocity of IT | 20Km/h |
| $\alpha$ | The scale value of gamma distribution for YC job handling time | 8.2 |
| $\beta$ | The shape value of gamma distribution for YC job handling time | 16.1 |

$$RMSE_2 = \sqrt{\frac{1}{n(Q)} \sum_{q=1}^{n(Q)} (y_q^t - \hat{y_q^t})^2} \tag{24}$$

$$MAE_3 = \frac{1}{n(V)} \sum_{v=1}^{n(V)} |y_v^t - \hat{y_v^t}| \tag{25}$$

$$MAPE_3 = \frac{1}{n(V)} \sum_{v=1}^{n(V)} |\frac{y_v^t - \hat{y_v^t}}{y_v^t}| \tag{26}$$

$$RMSE_3 = \sqrt{\frac{1}{n(V)} \sum_{v=1}^{n(V)} (y_v^t - \hat{y_v^t})^2} \tag{27}$$

Equations (22) - (24) represent the prediction errors for the QC makespan, while equations (25) - (27) denote the prediction errors for the vessel makespan. The unit of $MAE_2$ and $MAE_3$ is minute.The actual makespan of QC and vessel are denoted as $y_q^t$ and $y_v^t$, respectively.

Table 5 describes the parameters for experiment. The velocity of IT ($ve$) is set to 20Km/h. The MLE of gamma distribution for YC job handling time is set to 8.2 and 16.1, respectively. This study utilized the R 4.3.2 software to conduct comparative experiments, aiming to seamlessly integrate several ML predictive algorithms and simulation models.

### B. EXP 1: EXPERIMENTAL RESULTS FOR EXPLORING OPTIMAL ML ALGORITHMS FOR QC JOB HANDLING TIME PREDICTION

In order to obtain a precise ML algorithm for QC job handling time prediction, experiments are conducted to evaluate the

**TABLE 6. Experimental results for optimal ML algorithm for QC job handling time prediction.**

| ML algorithm | $MAE_1$ | $MAPE_1$ | $RMSE_1$ |
|---|---|---|---|
| Constant | 130.6 | 0.80 | 204.35 |
| Gamma | 166.5 | 1.09 | 248.56 |
| LR | 130.5 | 0.80 | **203.45** |
| DT | 130.9 | 0.80 | 204.62 |
| RF | 131.2 | 0.81 | 204.57 |
| **SVM** | **114.0** | **0.47** | 215.1 |
| XGB | 133.8 | 0.83 | 207.39 |
| ANN | 134.02 | 0.85 | 207.49 |
| LR + BN | 131.1 | 0.81 | 204.34 |
| DT + BN | 130.9 | 0.80 | 204.60 |
| RF + BN | 131.2 | 0.81 | 204.56 |
| **SVM + BN** | **115.3** | **0.48** | 215.5 |
| XGB + BN | 131.1 | 0.81 | **204.33** |
| ANN + BN | 134.02 | 0.85 | 207.48 |

**TABLE 7.** Experimental results for optimal ML + PDES for makespan prediction.

| ML + PDES | Makespan of QC ($y_q^t$) | | | Makespan of vessel ($y_v^t$) | | |
|---|---|---|---|---|---|---|
| | $MAE_2$ | $MAPE_2$ | $RMSE_2$ | $MAE_3$ | $MAPE_3$ | $RMSE_3$ |
| Constant | 73.77 | 0.66 | 86.11 | 78.08 | 0.53 | 90.23 |
| Gamma | 90.21 | 0.83 | 104.57 | 97.29 | 0.68 | 111.07 |
| LR | 74.99 | 0.67 | 87.49 | 79.2 | 0.54 | 91.55 |
| DT | 78.61 | 0.69 | 91.45 | 84.06 | 0.55 | 96.23 |
| RF | 76.34 | 0.68 | 88.95 | 81.19 | 0.55 | 93.46 |
| **SVM** | **28.44** | **0.14** | **33.92** | **28.89** | **0.11** | **35** |
| XGB | 74.99 | 0.67 | 87.49 | 79.2 | 0.54 | 91.55 |
| ANN | 84.53 | 0.74 | 97.99 | 90.73 | 0.6 | 103.34 |

predictive performance of ML algorithms. Additionally, the experiment results are organized based on the application of feature selection derived through BN. The experiment utilizes various ML algorithms, including Linear Regression (LR), Decision Tree (DT), RF), SVM, Extreme Gradient Boosting (XGB), and ANN. Moreover, the constant value and gamma distribution, employed in previous studies, are incorporated.

Table 6 presents the results of QC job handling time predictions when employing ML algorithms. According to the table, it is observed that there is not a significant difference in prediction performance when feature selection is applied and when it is not. As previously explained, the job handling time of QC exhibits a substantial variance due to the cascade effect, and the features utilized ($x^{job}(k)$, $x^{tw}(k)$) are categorical in nature.

According to table 6, it indicates that the prediction accuracy of methods other than SVM is relatively poor. Excluding SVM, these methods exhibit an approximately 130 in $MAE_1$, a 0.80 in $MAPE_1$, and a 200 or higher in $RMSE_1$, reflecting suboptimal predictive performance. The prediction outcomes are significantly negative, especially in cases assuming a gamma distribution. This shows that estimating job handling time based on probability distributions is insufficient. In the case of SVM, while the $RMSE_1$ is relatively higher compared to other methods, the $MAE_1$ and $MAPE_1$ are lower at 115 and 0.48, respectively. This suggests that when dealing with significant variance and categorical features, the predictive capability through kernel transformation is more suitable.

The prediction outcomes for the makespan of QC and vessels utilizing ML + PDES are available in Table 7. According to table 7, SVM + BN + PDES exhibited the best predictive performance for makespan of QC, with a $MAE_2$ of 28.44 and a $MAPE_2$ of 0.14 and $RMSE_2$ of 33.92. Similarly, SVM + BN + PDES exhibits the most favorable predictive performance for vessel makespan with a $MAE_3$ of 28.89 and a $MAPE_3$ of 0.11 and $RMSE_3$ of 35. Therefore, it can be concluded that the SVM is the most suitable for predicting the makespan of QC and vessel. The terms of ML algorithms refer to a predictive model that incorporates feature selection by BN.

### C. EXP 2: COMPARATIVE EXPERIMENTS USING SIMULATION MODEL

The prediction results for QC and vessel makespan can be found in table 8. In Exp 2, an extended set of comparative experiments was conducted involving additional ML algorithm. Bayesian Additive Regression Trees (BART) and CatBoost were incorporated into the experiment. Table 8 presents the calculated prediction error for each case. It can be observed that in all cases, the SVM + PDES model had the smallest MAE, MAPE, and RMSE and the highest r-squared value ($R^2$). The predictive errors for the makespan of the vessel show similar values for MAE and coefficient of $R^2$, while MAPE and RMSE exhibit relatively lower values.

However, ML algorithm with non-simulation models generally performed poorly, reflecting the difficulty of predicting the makespan of QC and vessel using ML algorithms. As mentioned in section II, in order to predict the operation times in CT, it is necessary to consider the detail job information such as bottleneck situation during operation. In the application of simulation for predictive purposes, the predictive outcomes may be influenced by the states of CHE, thereby allowing the manifestation of bottleneck situations in the prediction results. It is difficult to achieve precise learning solely by using information such as container job counts and equipment deployment quantities.

### D. EXP 3: COMPARATIVE EXPERIMENTS UTILIZING CHE (IT) ALLOCATION STRATEGIES

Exp 3 presents the comparative experimental results of applying DES and PDES for prediction. It also discusses the outcomes of using MPA for IT allocation and the application of a sampling model. Table 9 presents the summarized results of the comparative performance in predicting the makespan for both QC and vessel. It is evident that the predictive performance of the DES + MPA method, corresponding to the benchmark, is notably inferior. The $MAE_2$ is 54 to 73, $MAPE_2$ is 0.32 to 0.40, and $RMSE_2$ is 76 to 100. The $R^2$ is around 0.80. The reason for this difference can be ascribed to the fact that the updating of CHE statuses through DES is triggered by events. When updating the status of the QC for

**TABLE 8.** Experimental results for QC and vessel makespan.

| Model | Case | Makespan of QC ($y_q^t$) | | | | Makespan of vessel ($y_v^t$) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $MAE_2$ | $MAPE_2$ | $RMSE_2$ | $R^2$ | $MAE_3$ | $MAPE_3$ | $RMSE_3$ | $R^2$ |
| **SVM + PDES** | **Case 1** | **24.25** | **0.12** | **24.25** | **0.98** | **23.44** | **0.08** | **23.44** | **0.98** |
| | **Case 2** | **24.12** | **0.21** | **24.12** | **0.98** | **23.61** | **0.13** | **23.61** | **0.98** |
| | **Case 3** | **20.15** | **0.12** | **20.15** | **0.98** | **23.37** | **0.1** | **23.37** | **0.97** |
| LR | Case 1 | 126.46 | 0.51 | 154.62 | 0.8 | 154.56 | 0.43 | 187.35 | 0.8 |
| | Case 2 | 122.75 | 0.87 | 149.4 | 0.75 | 137.28 | 1.51 | 159.42 | 0.64 |
| | Case 3 | 119.91 | 0.44 | 135.96 | 0.77 | 123.62 | 0.28 | 148.43 | 0.55 |
| DT | Case 1 | 143.21 | 0.59 | 182.52 | 0.66 | 194.26 | 0.47 | 231.64 | 0.62 |
| | Case 2 | 133.62 | 1.04 | 167.77 | 0.59 | 164.2 | 1.45 | 198.77 | 0.61 |
| | Case 3 | 114.12 | 0.47 | 141.81 | 0.68 | 156.3 | 0.35 | 212.59 | 0.67 |
| RF | Case 1 | 127.74 | 0.48 | 162.06 | 0.71 | 155.12 | 0.46 | 183.5 | 0.69 |
| | Case 2 | 125.96 | 0.51 | 171.73 | 0.62 | 134.25 | 1.16 | 167.26 | 0.65 |
| | Case 3 | 99.53 | 0.39 | 118.8 | 0.74 | 120.76 | 0.29 | 150.39 | 0.74 |
| BART | Case 1 | 134.69 | 0.58 | 134.69 | 0.75 | 150.96 | 0.37 | 150.96 | 0.72 |
| | Case 2 | 118.84 | 0.47 | 118.84 | 0.72 | 136.67 | 1.39 | 136.67 | 0.64 |
| | Case 3 | 113.13 | 0.4 | 113.13 | 0.65 | 134.48 | 0.29 | 134.48 | 0.65 |
| SVM | Case 1 | 102.58 | 0.38 | 127.91 | 0.79 | 113.66 | 0.29 | 139.78 | 0.75 |
| | Case 2 | 90.96 | 0.57 | 113.52 | 0.73 | 96.13 | 0.86 | 116.75 | 0.72 |
| | Case 3 | 83.44 | 0.35 | 102.27 | 0.77 | 74 | 0.15 | 96.61 | 0.71 |
| XGB | Case 1 | 146.02 | 0.55 | 184.08 | 0.58 | 187.93 | 0.46 | 218.82 | 0.61 |
| | Case 2 | 161.77 | 0.58 | 219.34 | 0.47 | 188.58 | 0.9 | 258.65 | 0.54 |
| | Case 3 | 108.85 | 0.32 | 141.95 | 0.65 | 63 | 0.13 | 90.44 | 0.92 |
| CatBoost | Case 1 | 172.34 | 1.04 | 210.68 | 0.69 | 213.34 | 0.64 | 245.84 | 0.52 |
| | Case 2 | 145.44 | 1.26 | 181.46 | 0.58 | 174.56 | 2.29 | 216.91 | 0.45 |
| | Case 3 | 145.82 | 0.70 | 184.09 | 0.43 | 180.02 | 0.372 | 252.54 | 0.25 |
| ANN | Case 1 | 145.81 | 0.56 | 190.70 | 0.60 | 204.34 | 0.62 | 270.55 | 0.41 |
| | Case 2 | 119.63 | 0.84 | 153.55 | 0.60 | 160.21 | 1.11 | 200.01 | 0.42 |
| | Case 3 | 97.05 | 0.59 | 121.84 | 0.73 | 112.55 | 0.24 | 157.69 | 0.34 |

**TABLE 9.** Experimental results for QC and vessel makespan utilizing IT allocation strategy.

| Model | Case | Makespan of QC ($y_q^t$) | | | | Makespan of vessel ($y_v^t$) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $MAE_2$ | $MAPE_2$ | $RMSE_2$ | $R^2$ | $MAE_3$ | $MAPE_3$ | $RMSE_3$ | $R^2$ |
| DES + MPA | Case 1 | 64.98 | 0.32 | 83.88 | 0.82 | 63.61 | 0.20 | 82.42 | 0.84 |
| | Case 2 | 73.14 | 0.40 | 108.7 | 0.80 | 69.36 | 0.30 | 103.34 | 0.84 |
| | Case 3 | 54.98 | 0.34 | 76.19 | 0.77 | 56.42 | 0.23 | 80.11 | 0.75 |
| PDES + MPA | Case 1 | 37.12 | 0.16 | 46.69 | 0.97 | 38.58 | 0..13 | 48.44 | 0.97 |
| | Case 2 | 35.13 | 0.16 | 47.10 | 0.97 | 33.23 | 0.12 | 44.95 | 0.98 |
| | Case 3 | 32.75 | 0.16 | 42.09 | 0.97 | 33.54 | 0.12 | 43.64 | 0.97 |
| **PDES + Sampling model** | **Case 1** | **19.44** | **0.10** | **27.13** | **0.98** | **21.40** | **0.08** | **29.14** | **0.98** |
| | **Case 2** | **24.97** | **0.15** | **34.74** | **0.97** | **24.03** | **0.09** | **34.85** | **0.97** |
| | **Case 3** | **15.08** | **0.08** | **22.99** | **0.98** | **16.43** | **0.06** | **24.41** | **0.98** |

the handling job related to the $k$-$th$event, the update considers both the tasks of the IT and the YC. Consequently, this creates a problem in simulation-based prediction where the duration of the task is inaccurately estimated to be longer than it is. Furthermore, it is common in DES to assume average or constant values for CHE operation times. However, this practice also demonstrates results that are unsuitable for prediction purposes.

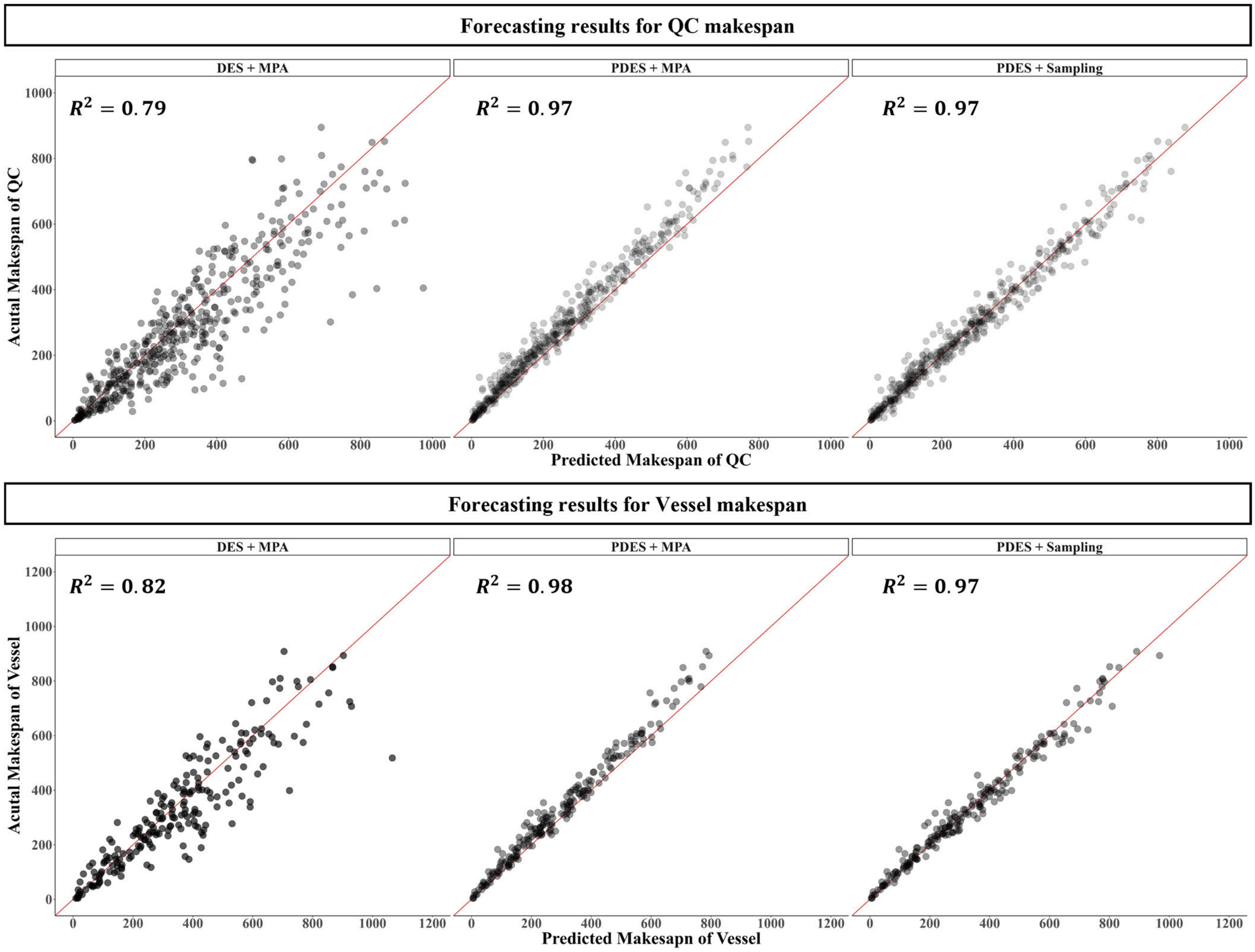



**FIGURE 10.** Comparison between actual and predicted makespan.

After analyzing the results using PDES + MPA for prediction, it is clear that the predictive performance has roughly doubled in comparison to DES + MPA. The MAE varies from 30 to 39, the MAPE ranges from 0.12 to 0.16, and the RMSE is reported between 42 and 48. Significantly, the $R^2$ has risen to 0.97. Nevertheless, the MPA may have constraints in accurately forecasting the durations of real tasks, despite its objective of minimizing makespan. Thus, in this study, the Sampling Model for IT allocation to specifically tackle this potential limitation. Upon examining the experimental results, it is evident that the ultimately proposed PDES + Sampling model proposed in this study exhibits the most favorable performance. There is a significant decrease in all measurements, such as MAE, MAPE, and RMSE, when compared to PDES + MPA. Consequently, in order to predict the operation times at a CT, it is more appropriate to employ a prediction technique that replicates actual data and strategies.

Figure 10 represents the comparison between actual and predicted makespan. Each point represents the actual and predicted values for each test scenario, and the dashed line, a reference line of $y = x$ shape, compares the predicted and actual values. Referring to the figure, the proposed approach (PDES + Sampling model) yielded a cluster of points around the reference line. PDES + MAP also showed points cluster around the reference line, but the predicted values are smaller than actual values. The DES + MPA method reveals a wide dispersion of points around the reference line, indicating a considerable spread in performance.

## VI. CONCLUSION

This study presents a method that employs PDES to precisely assess the operational strategies of CTs. Due to the intricate nature of CTs, where multiple pieces of equipment simultaneously perform handling operations, the creation and assessment of operational plans using simulation methods are considered essential. However, numerous studies that utilize simulation techniques in CTs frequently assume of average or constant task durations. This limitation hampers their capacity to accommodate variations in task durations caused by cascade effects in real terminal operations, leading to discrepancies between simulated and actual operational outcomes. In order to address this constraint, this study presents a methodology that effectively forecasts real terminal operations utilizing PDES by leveraging data obtained from TOS.

BN is utilized to identify the specific features that directly impact the job handling time of QC. Empirical studies are carried out to ascertain the ML algorithm that exhibits the most accurate predictive capabilities. The utilization of BN-based feature selection enables the direct identification of factors that have a significant impact on QC operations using data, distinguishing it from other similar studies. In addition, the study presents the use of SVM for prediction in order to handle job variability caused by cascade effects. It shows that SVM achieves more accurate predictive performance compared to traditional simulation methods. In addition, this study suggests the use of PDES as a solution to the limitations of simulation approaches that rely on DES for predicting operation times in CTs. PDES conducts simulations based on predictive information.

Within CTs, the allocation of jobs for QCs is predetermined, while assignments for ITs are not pre-established. Previous study endeavors focused on reducing job handling time by utilizing simulation and employing a strategy that minimizes equipment idle time. Nevertheless, this approach is frequently unfeasible in real CTs. In order to tackle this inequality, this study suggests a sampling-based IT assignment strategy that replicates the real IT assignment strategy in CTs, thereby improving predictive performance.

The proposed method for predicting operational times of container terminals in this study could be enhanced as follows. By predicting fluctuations in operation times according to timing or frequency, it may be feasible to better anticipate uncertain conditions, thus improving prediction accuracy. Furthermore, by including predicted fluctuations into DES, it is expected that a predictive model capable of more precisely representing different cascade effects might be established.

Future research for this study entails developing a precise forecasting model for QC planning and determining the ideal quantity of ITs using a simulation method that can accurately predict the makespan of CT. When formulating operational plans for QC, it is crucial to minimize the duration required to finish jobs. This can be accomplished by carefully deciding where to stack import and export containers, considering significant variations in lead time. The predictive simulation model presented in this study has the benefit of precisely forecasting the duration of QC processes. When integrated with optimization models for QC planning, it enables the formulation of more accurate and comprehensive plans. Moreover, in CT operations, it is important to determine the appropriate number of ITs in advance. This will improve the practical applicability of the operational plan in real CT scenarios.

## REFERENCES

[1] *Container Shipping—Statistics & Facts*, Statistia Search Dept., New York, NY, USA, May 2022.

[2] *Review of Maritime Transport 2022*, UNCTAD, Geneva, Switzerland, 2022.

[3] A. K. Lange, A. Schwientek, and C. Jahn, ‘‘Reducing truck congestion at ports—Classification and trends,’’ in *Proc. Hamburg Int. Conf. Logistics (HICL)*, vol. 24, 2017, pp. 37–58.

[4] K. Park, S. Sim, and H. Bae, ‘‘Vessel estimated time of arrival prediction system based on a path-finding algorithm,’’ *Maritime Transp. Res.*, vol. 2, 2021, Art. no. 100012.

[5] H. Lin, W. Zeng, J. Luo, and G. Nan, ‘‘An analysis of port congestion alleviation strategy based on system dynamics,’’ *Ocean Coastal Manage.*, vol. 229, Oct. 2022, Art. no. 106336, doi: 10.1016/j.ocecoaman.2022.106336.

[6] P. Bolat, G. Kayisoglu, E. Gunes, F. Kizilay, and S. Ozsogut, ‘‘Weighting key factors for port congestion by AHP method,’’ *J. ETA Maritime Sci.*, vol. 8, pp. 252–273, Feb. 2020, doi: 10.5505/jems.2020.64426.

[7] K. Chargui, T. Zouadi, A. El Fallahi, M. Reghioui, and T. Aouam, ‘‘A quay crane productivity predictive model for building accurate quay crane schedules,’’ *Supply Chain Forum, Int. J.*, vol. 22, no. 2, pp. 136–156, Apr. 2021, doi: 10.1080/16258312.2020.1831889.

[8] L. Cai, W. Guo, L. He, and W. Li, ‘‘Port integrated scheduling under uncertain operation time and cascade effects: A complex network structure entropy solution,’’ *Comput. Ind. Eng.*, vol. 182, Aug. 2023, Art. no. 109435, doi: 10.1016/j.cie.2023.109435.

[9] E. Lee, K. Park, D. Kim, H. Bae, and C. Hong, ‘‘Prediction of the quay crane's handling time with external handling factors,’’ *ICIC Exp. Lett., B, Appl.*, vol. 12, pp. 351–358, Jan. 2021, doi: 10.24507/icicelb.12.04.351.

[10] C. Boer and Y. Saanen, ‘‘Testing, tuning and training terminal operating systems. A modern approach,’’ in *Proc. Int. Conf. Logistics Maritime Syst.*, 2012, pp. 1–11.

[11] B. Eduard and W. Ming, ‘‘Discrete event simulation: State of the art,’’ *Int. J. Online Eng.*, vol. 4, pp. 60–63, Jan. 2008, doi: 10.5772/9894.

[12] S. Voß, R. Stahlbock, and D. Steenken, ‘‘Container terminal operation and operations research—A classification and literature review,’’ *OR Spectr.*, vol. 26, no. 1, pp. 3–49, Jan. 2004, doi: 10.1007/s00291-003-0157-z.

[13] M. Sha, T. Notteboom, T. Zhang, X. Zhou, and T. Qin, ‘‘Simulation model to determine ratios between quay, yard and intra-terminal transfer equipment in an integrated container handling system,’’ *J. Int. Logistics Trade*, vol. 19, no. 1, pp. 1–18, Mar. 2021, doi: 10.24006/jilt.2021.19.1.001.

[14] A. H. Gharehgozli, F. G. Vernooij, and N. Zaerpour, ‘‘A simulation study of the performance of twin automated stacking cranes at a seaport container terminal,’’ *Eur. J. Oper. Res.*, vol. 261, no. 1, pp. 108–128, Aug. 2017, doi: 10.1016/j.ejor.2017.01.037.

[15] Z. Zhang, M. Liu, C.-Y. Lee, and J. Wang, ‘‘The quay crane scheduling problem with stability constraints,’’ *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 3, pp. 1399–1412, Jul. 2018, doi: 10.1109/TASE.2018.2795254.

[16] D. Sun, L. Tang, R. Baldacci, and A. Lim, ‘‘An exact algorithm for the unidirectional quay crane scheduling problem with vessel stability,’’ *Eur. J. Oper. Res.*, vol. 291, no. 1, pp. 271–283, May 2021, doi: 10.1016/j.ejor.2020.09.033.

[17] A. Skaf, S. Lamrous, Z. Hammoudan, and M.-A. Manier, ‘‘Integrated quay crane and yard truck scheduling problem at port of tripoli-lebanon,’’ *Comput. Ind. Eng.*, vol. 159, Sep. 2021, Art. no. 107448, doi: 10.1016/j.cie.2021.107448.

[18] R. T. Cahyono, S. P. Kenaka, and B. Jayawardhana, ‘‘Simultaneous allocation and scheduling of quay cranes, yard cranes, and trucks in dynamical integrated container terminal operations,’’ *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8564–8578, Jul. 2022, doi: 10.1109/TITS.2021.3083598.

[19] Y. Zheng, M. Xu, Z. Wang, and Y. Xiao, ‘‘A genetic algorithm for integrated scheduling of container handing systems at container terminals from a low-carbon operations perspective,’’ *Sustainability*, vol. 15, no. 7, p. 6035, Mar. 2023, doi: 10.3390/su15076035.

[20] T. Qin, Y. Du, J. H. Chen, and M. Sha, ‘‘Combining mixed integer programming and constraint programming to solve the integrated scheduling problem of container handling operations of a single vessel,’’ *Eur. J. Oper. Res.*, vol. 285, no. 3, pp. 884–901, Sep. 2020, doi: 10.1016/j.ejor.2020.02.021.

[21] H.-P. Hsu, C.-C. Chou, and C.-N. Wang, ‘‘Heuristic/metaheuristic-based simulation optimization approaches for integrated scheduling of yard crane, yard truck, and quay crane considering import and export containers,’’ *IEEE Access*, vol. 10, pp. 64650–64670, 2022, doi: 10.1109/ACCESS.2022.3180752.

[22] K. Mokhtar and M. Z. Shah, "A regression model for vessel turnaround time," in *Tokyo Academic, Industry & Cultural Integration Tour*. Tokyo, Japan: Shibaura Inst. Technol., 2006, pp. 10–19.
[23] D. Štepec, T. Martincic, F. Klein, D. Vladušic, and J. P. Costa, "Machine learning based system for vessel turnaround time prediction," in *Proc. 21st IEEE Int. Conf. Mobile Data Manage. (MDM)*, Jun. 2020, pp. 258–263.
[24] Q. Zeng and Z. Yang, "Integrating simulation and optimization to schedule loading operations in container terminals," *Comput. Oper. Res.*, vol. 36, no. 6, pp. 1935–1944, Jun. 2009, doi: 10.1016/j.cor.2008.06.010.
[25] S. M. Homayouni, S. H. Tang, and O. Motlagh, "A genetic algorithm for optimization of integrated scheduling of cranes, vehicles, and storage platforms at automated container terminals," *J. Comput. Appl. Math.*, vol. 270, pp. 545–556, Nov. 2014, doi: 10.1016/j.cam.2013.11.021.
[26] Q. Zeng, A. Diabat, and Q. Zhang, "A simulation optimization approach for solving the dual-cycling problem in container terminals," *Maritime Policy Manage.*, vol. 42, no. 8, pp. 806–826, Nov. 2015, doi: 10.1080/03088839.2015.1043362.
[27] M. Stojaković and E. Twrdy, "Determining the optimal number of yard trucks in smaller container terminals," *Eur. Transp. Res. Rev.*, vol. 13, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/s12544-021-00482-6.
[28] B. Xu, D. Jie, J. Li, Y. Yang, F. Wen, and H. Song, "Integrated scheduling optimization of U-shaped automated container terminal under loading and unloading mode," *Comput. Ind. Eng.*, vol. 162, Dec. 2021, Art. no. 107695, doi: 10.1016/j.cie.2021.107695.
[29] A. A. Tako and S. Robinson, "The application of discrete event simulation and system dynamics in the logistics and supply chain context," *Decis. Support Syst.*, vol. 52, no. 4, pp. 802–815, Mar. 2012, doi: 10.1016/j.dss.2011.11.015.
[30] A. Basaglia, E. Spacone, J. W. van de Lindt, and T. D. Kirsch, "A discrete-event simulation model of hospital patient flow following major earthquakes," *Int. J. Disaster Risk Reduction*, vol. 71, Mar. 2022, Art. no. 102825, doi: 10.1016/j.ijdrr.2022.102825.
[31] W. T. de Sousa Junior, J. A. B. Montevechi, R. de Carvalho Miranda, and A. T. Campos, "Discrete simulation-based optimization methods for industrial engineering problems: A systematic literature review," *Comput. Ind. Eng.*, vol. 128, pp. 526–540, Feb. 2019, doi: 10.1016/j.cie.2018.12.073.
[32] Y. Frichi, F. Jawab, L. Aboueljinane, and S. Boutahari, "Development and comparison of two new multi-period queueing reliability models using discrete-event simulation and a simulation–optimization approach," *Comput. Ind. Eng.*, vol. 168, Jun. 2022, Art. no. 108068, doi: 10.1016/j.cie.2022.108068.
[33] M. Laguna and J. Marklund, *Business Process Modeling, Simulation and Design*. Boca Raton, FL, USA: CRC Press, 2018, doi: 10.1201/9781315162119.
[34] J. He, Y. Wang, C. Tan, and H. Yu, "Modeling berth allocation and quay crane assignment considering QC driver cost and operating efficiency," *Adv. Eng. Informat.*, vol. 47, Jan. 2021, Art. no. 101252.
[35] P. Das and K. Chanda, "Bayesian network based modeling of regional rainfall from multiple local meteorological drivers," *J. Hydrol.*, vol. 591, Dec. 2020, Art. no. 125563, doi: 10.1016/j.jhydrol.2020.125563.
[36] S. Park, D. Jung, and Y. Choi, "Prediction of ore production in a limestone underground mine by combining machine learning and discrete event simulation techniques," *Minerals*, vol. 13, no. 6, p. 830, Jun. 2023, doi: 10.3390/min13060830.
[37] K. Park, H. Park, and H. Bae, "A data-driven recipe simulation for synthetic rubber production," *IEEE Access*, vol. 10, pp. 129408–129418, 2022, doi: 10.1109/ACCESS.2022.3228241.
[38] A. K. Schwientek, A.-K. Lange, and C. Jahn, "Effects of terminal size, yard block assignment, and dispatching methods on container terminal performance," in *Proc. Winter Simul. Conf. (WSC)*, Dec. 2020, pp. 1408–1419, doi: 10.1109/WSC48552.2020.9384062.
[39] I. M. Cavalcante, E. M. Frazzon, F. A. Forcellini, and D. Ivanov, "A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing," *Int. J. Inf. Manage.*, vol. 49, pp. 86–97, Dec. 2019, doi: 10.1016/j.ijinfomgt.2019.03.004.
[40] A. Atalan, H. Şahin, and Y. A. Atalan, "Integration of machine learning algorithms and discrete-event simulation for the cost of healthcare resources," *Healthcare*, vol. 10, no. 10, p. 1920, Sep. 2022, doi: 10.3390/healthcare10101920.
[41] S. Shafiei, A.-S. Mihăiţă, H. Nguyen, and C. Cai, "Integrating data-driven and simulation models to predict traffic state affected by road incidents," *Transp. Lett.*, vol. 14, no. 6, pp. 629–639, Jul. 2022, doi: 10.1080/19427867.2021.1916284.
[42] B. Wang, W. Chen, B. Zhang, Y. Zhao, and P. Shi, "Cooperative control-based task assignments for multiagent systems with intermittent communication," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6697–6708, Oct. 2021, doi: 10.1109/TII.2020.3044950.
[43] J. Qi, G. Ran, B. Wang, J. Liu, W. Ma, P. Zhou, and D. Navarro-Alarcon, "Adaptive shape servoing of elastic rods using parameterized regression features and auto-tuning motion controls," *IEEE Robot. Autom. Lett.*, vol. 9, no. 2, pp. 1428–1435, Feb. 2024, doi: 10.1109/lra.2023.3346758.
[44] MAERSK. (2023). *Top 9 Reasons for Shipping Delays*. Accessed: Aug. 28, 2023. [Online]. Available: http://www.maersk.com/logistics-explained/transportation-and-freight/2023/08/28/cargo-delays
[45] S. Margaronis. (2023). *Port of Busan Will Open Its First Automated Terminal in 2023 Using Korean Cranes*. [Online]. Available: http://www.ajot.com/insights/full/ai-port-of-busan-will-open-its-first-automated-terminal-in-2023-using-korean-cranes
[46] Busan Port Authority (BPA), "Current status of smart port progress in Busan port," in *Proc. 31st IAPH World Ports Conf.*, May 2019, pp. 1–8.
[47] (2023). *BPT (Busan Port Terminal)—Action Guide for Delayed Cargo*. [Online]. Available: http://kr.lines.coscoshipping.com/col/col24862/art/2023/art_242f2e05210c4b9ebd02b864cde9529c.html

**KIKUN PARK** received the B.S. degree in applied statistics from Gachon University, Republic of Korea, in 2017. He is currently pursuing the Ph.D. degree with Pusan National University, Busan, South Korea. His research interests include machine learning, scheduling problem, and simulation.

**MINSEOP KIM** received the B.S. degree from Changwon National University, Republic of Korea, in 2020. He is currently pursuing the Ph.D. degree with Pusan National University, Busan, South Korea. His research interests include machine learning and representation learning.

**HYERIM BAE** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in industrial engineering from Seoul National University, Republic of Korea. From 2002 to 2003, he was with Samsung Credit Card Corporation, Seoul, Republic of Korea. Since 2005, he has been a Professor with the Department of Industrial Engineering, Pusan National University, Republic of Korea. He is interested in information system design, cloud computing, business process management systems, process mining, and big data analytics for operational intelligence. His current research activities include analyzing huge volumes of event logs from port logistics and shipbuilding industries using process mining techniques.