Li, N, Sheng, H, Wang, P, Jia, Y, Yang, Z and Jin, Z

Modeling Categorized Truck Arrivals at Ports: Big Data for Traffic Prediction

Modeling Categorized Truck Arrivals at Ports: Big Data for Traffic Prediction

Na Li, Haotian Sheng, Pingyao Wang, Yulin Jia, Zaili Yang* and Zhihong Jin

[1]

Abstract—**Accurate truck arrival prediction is complex but critical for container terminals. A deep learning model combining Gated Recurrent Unit (GRU) and Fully Connected Neural Network (FCNN), is proposed to predict daily truck arrivals using fusion technology. The model can efficiently analyze sequence and cross-section data sets. The new feature in the new model lies in that it, for the first time, incorporates the new parameters influencing traffic volumes such as the vessel-related information, arrival weekdays, and weather conditions into the long-time series of truck arrivals. Furthermore, truck arrivals are predicted in three groups based on their movement purposes: pick-up, delivery, and dual. it also contributes to the literature in a sense that the performance of the model is tested using real big data from a world-leading container port in Southern China. The results generate insightful managerial implications for guiding port traffic management in a generic manner. It reveals the relation of export container arrivals with the Container Yard (CY) closing time of a specific vessel. It is demonstrated the proposed model outperforms the currently available methods with an improved accuracy rate of prediction by 23.44% (dual), 32.09% (pick-up), and 26.99% (delivery), respectively. As a result, the model can better reflect reality compared to the existing ones in the literature. It is also evident that the 3-categorized prediction model can significantly help increase prediction accuracy in comparison with the 2-categorized methods used in practice.**

Index Terms—**Container terminal; Truck arrival; Prediction; Gated Recurrent Unit- Fully Connected Neural Network; Big data.**

I. Introduction

THE truck arrivals at terminals fluctuate significantly on various days. As the size of the largest container ships went up continuously in recent years, container handling becomes more concentrated in certain major ports [1]. Gains from the economies of scale resulting from the deployment of larger vessels do not necessarily benefit ports and inland transport service providers [2]. On the landside thousands of containers that need to be picked up and/or delivered often cause traffic management difficulties for container terminals. For example, the daily truck arrival number at a container terminal in Sothern China reached 19,000 on average and 23,000 at the max in 2019.

Within the context of container truck arrivals, previous studies in the literature reveal various methods from a control perspective such as an appointment system [3-5], time-varying tolls [6], and time windows [7]. They, aiming to smooth the peak arrivals to match the capacity of the equipment, are all base on a certain distribution of arrival rates and use operational research methods to realize mathematical modeling and optimization.

Different from controlling truck arrivals, an alternative way is to forecast them in accuracy. An effective truck arrival prediction system represents an important decision basis for the yard operation plan at a terminal. Furthermore, it would improve the overall operational efficiency of a terminal and reduce truck turnover time, queueing, and exhaust emissions for better air quality. Drayage companies want to estimate the turn time based on the number of truck arrivals. They might choose a period with fewer truck arrivals to save time at a terminal. With the new opportunities emerging from the current wave of digitalization, truck arrival prediction also needs to be revisited by taking a data-driven perspective [8].

In the past decade, due to the disruptions such as climate change risks (e.g. flooding and storms) and COVID-19, the maritime industry is encountering more and more port congestion, serious delay of pick-ups, and shortage of empty containers, etc. One of the main contributors is the ineffective use of trucks at container terminals. The disruptions such as COVID-19 have made the prediction of yard workloads even worse. Resources at terminals are either idling or overloaded. Terminal operators are eager for an accurate forecast of truck arrivals in order to deploy the equipment and labor at the terminal rationally.

Accurate truck flow prediction will benefit drayage dispatch decisions and help the terminal operators schedule the equipment and labor rationally [9]. Hence, the external trucks' efficiency could be improved and the impact of disruptions such as COVID-19 could be reduced. New methods become crucial and beneficial not only for addressing the drayage but also for mitigating its further impact on the supply chains.

This study was partially supported by the National Natural Science Foundation of China (Grant Nos. 72171032, 71702019, 72172023), Humanities and Social Sciences Fund of Chinese Ministry of Education (21YJA630048, 21YJAZH070).

Na Li, Haotian Sheng, Pingyao Wang, Yulin Jia and Zhihong Jin are with the College of Transportation Engineering, Dalian Maritime University, Dalian, Liaoning, China (e-mail: lina@dlmu.edu.cn; shenghaotian@dlmu.edu.cn ; 328474897@qq.com ; 18334752256@163.com ; jinzhihong@dlmu.edu.cn). Zaili Yang is from Liverpool Logistics, Offshore and Marine Research Institute, Liverpool John Moores University, Liverpool, UK (email: Z.Yang@ljmu.ac.uk)

In recent years, deep learning techniques (e.g. neural networks) have been widely used in various fields. Rumelhart, et al. [10] propose a multi-layer perceptron based on backpropagation, also known as the "BP neural network". Lecun and Bottou [11] develop a Convolutional Neural Network (CNN) to break through the limitations of the accuracy of image recognition. Hinton, et al. [12] construct the Deep Belief Nets (DBN), which form the basis for deep learning discipline identification. Yang and Mehmed [13] use two different dynamic Fully Connected Neural Network (FCNN) models to improve the accuracy of forecasting shipping freight rates. Current existing prediction methods are inadequate to deal with emerging challenges of disruptive events of low likelihood and high consequence (e.g., COVID).

Variants of Recurrent Neural Network (RNN) named Long Short-Term Memory (LSTM) [14] and Gated Recurrent Unit (GRU) networks were often employed for sequential data with relatively large time steps. GRU models have fewer parameters and less complicated structures compared to LSTM models, and it performs equally well as LSTM models. GRU may be the preferred method in short-term predictions since it requires less time for model training [15-17]. The FCNN model is good at prediction with cross-sectional data. Therefore, the GRU models deal with sequential truck arrival data, while the FCNN models cope with non-operational data. The merged GRU-FCNN models could perform well with both the sequential data and cross-section data.

The novelty of this study includes 1) the precise prediction of truck arrivals towards container terminal using the proposed GRU-FCNN model for big data applications; 2) the merging of characteristics of drayage service into the prediction models, including three categories of service types (i.e. pickup, delivery, and dual), weekdays and weather, which is proven to be able to significantly improve prediction accuracy; and 3) involvement of container yard closing time to help predict the outbound arrivals, which shows the relationship between vessel and trucks arrivals.

The following paper is organized as follows. The literature review is presented in Section II. Section III describes the problem statement. Section IV outlines the new GRU-FCNN model. Section V presents the descriptive statistics of the data used and the parameter calibration. Section VI analyzes the results, detailing the one-step and multi-step prediction, sensitivity analysis, advantages of prediction based on three categories, and prediction of truck arrivals from vessel information. Finally, Section VII concludes the paper.

II. Literature Review

Deep learning techniques are widely used in the prediction of traffic flows and passenger demand. For road traffic flow, Li, et al. [18] propose a gradient boosting procedure in combination with hierarchical reconciliation for short-term forecasting of traffic flow. Lv, et al. [19] apply a deep architecture model using autoencoders as building blocks to represent traffic flow features for its prediction. Zheng, et al. [20] propose a deep learning-based model which uses hybrid and multiple-layer architectures to automatically extract inherent features of traffic flow data. Wang, et al. [21] propose a novel hierarchical traffic flow prediction protocol based on a spatial-temporal graph convolutional network (ST-GCN), to achieve a more accurate traffic flow prediction. Arguedas, et al. [22] propose an unsupervised system that has a two-layer network, to monitor maritime vessel flows. For the prediction of passenger flow demand, Yang, et al. [23] develop an improved spatiotemporal long short-term memory model (Sp-LSTM) to forecast short-term outbound passenger volume at urban rail stations. Taxi passenger demand has also aroused lots of attention. Zhang, et al. [24] propose an end-to-end multi-task learning temporal convolutional neural network (MTL-TCNN) to predict the short-term passenger demand at a multi-zone level. Cheng, et al. [25] employ the vector autoregression (VAR) model and the CNN-LSTM hybrid neural network model to predict a short-term traffic flow. Truck arrival volume prediction at a container terminal is inherently different from the traffic flow of the urban highway or railway. Trucks towards a container terminal are affected by many external factors, including arrival and departure of vessels [26], service types, weekdays, as well as weather conditions.

The prediction of truck arrivals towards a container terminal has attracted lots of attention from both academics and practice. The existing daily truck arrival prediction research can be classified into two categories: parameter and non-parametric models. Parameter models require manual input of parameters. Yang, et al. [26] obtain the parameter value of the BETA distribution by the least square method, which improves the prediction accuracy of the truck traffic volume towards a container terminal. Guo, et al. [27] establish a simulation model based on the random time-varying characteristics of a container port area and discuss the varying rules of truck arrivals.

The non-parametric models have the potential to learn a nonlinear model structure and are not necessary to rely on empirical parameters. Al-Deek, et al. [28] and Al-Deek and Haitham [29] compare a neural network model and multiple regression, using the traffic flow and the week factor in Port of Miami. The forecasting of the arrivals of import and export container trucks in the Port of Miami was improved. Al-Deek and Haitham [30] use time series models to successfully predict the arrival of the container trucks and pick-up trucks in the port. Pradeep, et al. [31] compare the performance between Back Propagation Neural Network and Recurrent Neural Network (RNN) in the Florida port and prove that RNN has shown better robustness. Xie and Huynh [32] establish GPs (Gaussian processes) and ε-SVMs (ε-support vector machines, based on a Sigmoid kernel function to predict the daily arrival volume of pick-up and delivery trucks at Bayport and Barbours Cut container terminals at the Port of Houston. Zhang, et al. [33] use a neural network to predict the traffic flow according to the schedule of the liner seasonality. Kourounioti, et al. [34] use a Fully Connected Neural Network(FCNN) with the auxiliary factor (container and yard information) to improve the fitting performance for the distribution of container dwell time at the port and analyze the influence of each information. Gao, et al. [35] use LSTM to predict the daily volumes of containers that will enter their investigated storage yard. Nadi, et al. [36]

develop a short-term prediction model for outbound truck flows around major container seaports. In the paper, they use scheduled pick-ups to predict the truck flows. The related literature is summarized in Table I.

Although showing some attractiveness, previous studies still reveal various practical challenges in their applications, including

1) The condition of weather as an important variable influencing truck arrivals is often overlooked. The weather influence on yard operation is not the same as normal road traffic. In bad weather, e.g., fog or storm, the yard crane will not be able to load/discharge containers due to low visibility and stability. Contrarily, heavy rain and temperature will generate less impact on truck service at the yard.
2) In practice, dual transactions involving both pickups and deliveries are increasingly used by terminal operators and drayage companies for cost-saving. It therefore should be taken into account in the truck arrival prediction for improving its accuracy and effectiveness for yard planning.
3) Big data analysis based on deep learning techniques has attracted numerous attempts in city traffic or railway demand studies. With trajectory data accumulated in container terminals, a deep learning model could be proposed to capture the inherent correlations between vessel and drayage but also occupy a high training efficiency.
4) The big difference between city traffic prediction and truck forecasting arriving at the terminal is the influence of vessel schedules. Trucks arrive to deliver containers before the Container Yard (CY) closing time of a specific vessel, and trucks come to pick up containers after the vessel has been unloaded. Mining the relation between vessel schedule and truck arrivals with a non-parametric method remains unexplored fully.

As the pioneering work to address the above challenges, this paper has made contributions as follows: (1) this paper establishes a GRU-FCNN model based on the trajectory datasets of truck arrivals, which for the first time takes into account weekday and weather data; (2) based on the prevailing practice in container terminals, this paper innovatively predicts truck arrivals in the three categories of pickup, delivery and dual; (3) it indicates the influence of a specific vessel's yard closing datetime on the arrival of outbound container trucks.

TABLE I
LITERATURE REVIEW SUMMARY

No.	Author (Year)	Datasets			Service types				Methodologies		
		D1	D2	D3	S1	S2	S3	S4	M1	M2	M3
1	Al-Deek et al. (2000)		√	√	√	√				√	
2	Al-Deek & Haitham (2001)		√	√	√	√				√	
3	Al-Deek & Haitham (2002)	√		√	√	√				√	
4	Pradeep et al. (2005)	√		√	√	√				√	
5	Yang et al. (2010)		√					√	√		
6	Xie & Huynh (2010)	√			√	√			√		
7	Zhang et al. (2014)	√						√		√	
8	Kourounioti et al. (2016)	√		√	√				√		
9	Guo et al. (2017)		√		√						√
10	Gao et al. (2019)		√					√		√	
11	Nadi, et al. (2021)		√			√				√	
	This paper	√	√	√	√	√	√			√	

Note: Dataset：D1-liner schedules, D2- port traffic flow dataset, D3- Auxiliary Information；
Classifications of result: S1- pick-up truck, S2- delivery truck, S3- dual truck, S4- Unclassified；
Methodology: M1- parametric models, M2- nonparametric models, M3-others.

III. AN EMPIRICAL INVESTIGATION ON THE NON-LINEAR CHARACTERISTICS

The data of truck arrival flows are collected from the Terminal Operation System (TOS) at a container terminal in Southern China. The data are indexed by container Identity (ID). For each container, the serving time point including the gate-in time, the confirmation time of loading or unloading, gate-out time, as well as truck ID, etc. are recorded in the system. We collected two sets of data. The first set covers 12 months from September 1, 2018, to August 31, 2019. The other set covers six consecutive months (186 days) from January 1 to June 30, 2020, which was during the COVID pandemic period. Besides, the second set contains the corresponding vessel code, berthing time, container yard closing time, etc.

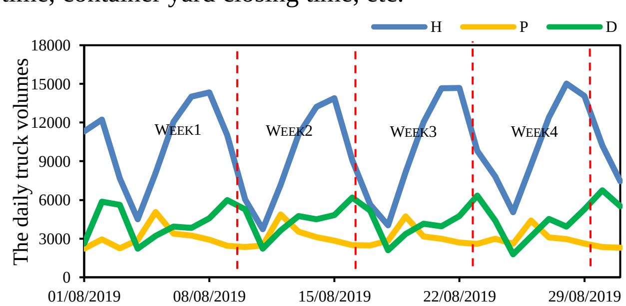


Fig.1. The weekly fluctuation of daily truck arrivals

The real data in Fig.1 shows a significant difference among the three categories of truck arrivals. Here we symbolize three services Pick-up task (P), Delivery task (D), and Dual tasks (H). The dual tasks, dominating more than half of truck arrivals, are often substituted by outbound or inbound tasks respectively in the literature. Usually, the outbound tasks need to match other inbound tasks to have dual transactions. The inbound and outbound ones are not quite independent of each other. If more dual transactions are deployed, the number of total truck arrivals would be reduced. There are some transformations among the three categories. Therefore, it makes sense to have three categories of tasks. Figure 2 illustrates the three service types, including 1) picking up an inbound container at a terminal; 2) delivering an outbound container to a terminal; 3) dual transactions that a truck delivers a container and then picks up another one in a single trip. From Fig.3, we can find the distributions of three categories of truck arrivals. In most cases, the number of dual trucks is around 5,000 to 10,000 per day. In contrast, the number of pick-ups is around 2,000 to 3,000, while the number of deliveries is between 4,000 and 5,000. It reveals the fact that the terminal is mainly an export port. Further, some import empty containers would not be picked up but stored in the yard under some special contracts with shipping lines.

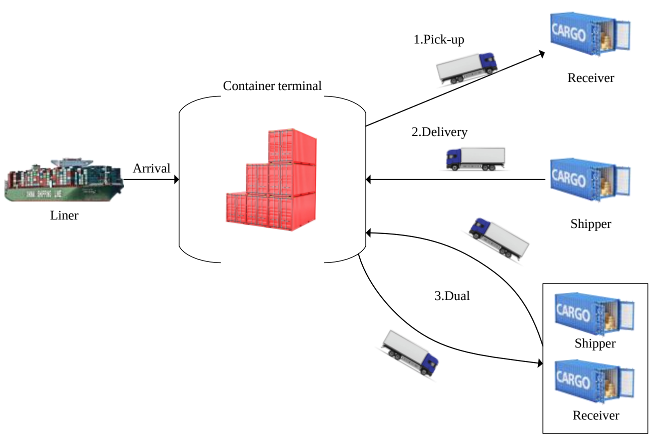


Fig.2. Three types of truck arrivals

Table II shows that in the dual process, 99.01% of the first operations (deliveries) are full (loaded) containers, and 84.00% of the second operations (pick-ups) are empty containers. In single pick-up tasks, the empty container accounts for 96.38%.

TABLE II
THE RELATIVE OCCUPANCY OF TWO ASSIGNMENTS OF DUAL TRUCKS

Pick-ups / Deliveries	Empty import container	Full import container	SUM
Empty export container	74.10%	25.90%	0.99%
Full export container	84.09%	15.91%	99.01%
SUM	84.00%	16.00%	100%

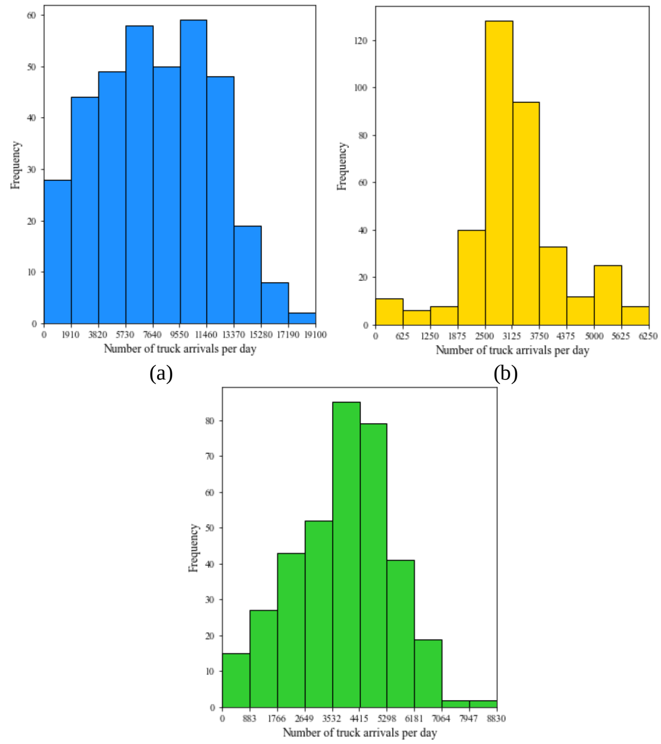


Fig.3. Daily arrival distribution of trucks of (a) dual; (b) pick-up; (c) delivery

The big difference between city traffic and trucks arriving at the terminal is the influence of vessel schedules. Trucks arrive before the Container Yard (CY) closing time to catch up with the schedule of vessels. Inbound containers are picked up after they are unloaded from the vessel. From the 1645 voyages in the first half-year of 2020, outbound truck arrivals of the first 100 ones are illustrated in Fig.4. Each subgraph represents the outbound truck arrivals during the time window for receiving containers. The red dotted line in each subgraph is the CY closing date. In most cases, there would be a peak number some day before the deadline. Export container arrivals are related to historical arrivals over the past few periods. At the same time, on each weekday there is a similar pattern of truck arrivals. It is consistent with the weekly schedules of vessels.

However, for inbound containers, the relation with vessel berth time is not significant as illustrated in the figure in Appendix B. It is owing to the feature of the export terminal in Southern China. Some inbound empty containers are not going to be picked up, but just for transshipment or storage in the terminal. It would be contrary if the terminal is an import one.

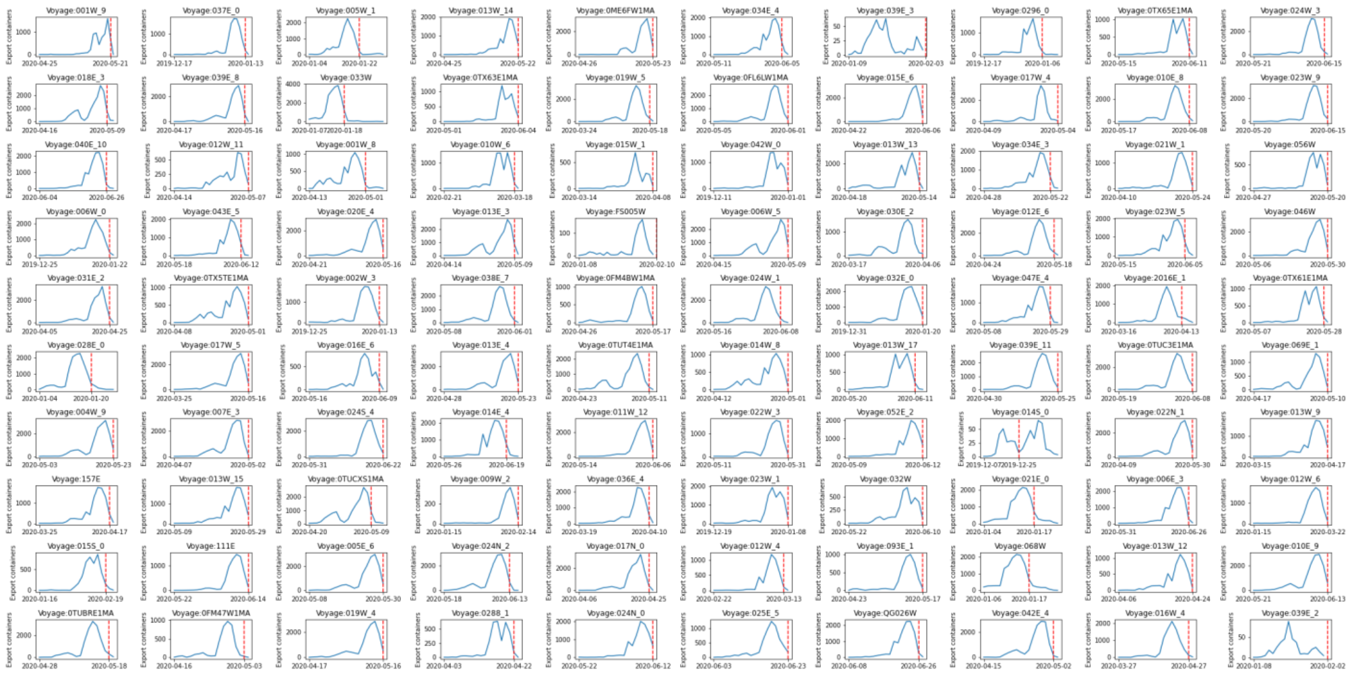


Fig.4. Outbound truck arrivals of each voyage
Note: Each subgraph is a ship's export arrivals sequence. the red dashed line is CY closing time. The subplot title is the voyage number_arrival order.

Fig.5 describes the correlation between meteorology and the average arrival of dual trucks. When it is sunny or overcast, the influence of weather is the least. However, if it involves a rainstorm or shower, truckers tend to change their schedule to a different day. It is interesting to find that a thundershower has less influence than a shower. Because it often has a thundershower in Southern China. Several minutes after the thunder, it will turn sunny again. On a thundershower day, it often means a long sunny period for terminal operations. Although it is not determinant, the effect of weather conditions on truck arrivals is clearly seen in Fig.5.

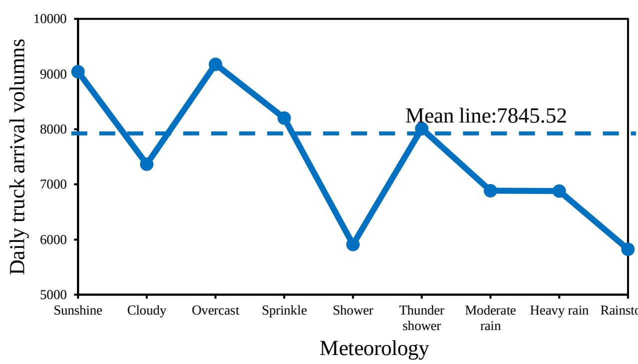


Fig.5. Illustrated correlation between meteorology and average arrival of dual trucks
Note: the mean line represents the average daily arrivals of dual trucks

IV. METHODOLOGY

The GRU model can efficiently deal with series data. Contrarily the FCNN is broadly applied in the prediction of cross-section data. For the advantages of the GRU model with sequential data, as well as that of the FCNN model with cross-sectional data, we design and train a GRU-FCNN model to predict the three categories of truck arrivals, considering the vessel information, as well as weekdays and weather.

A. The formulation and data preprocessing

In this paper, the prediction value of the s type of truck arrivals at t day is $\hat{y}_{s,t}$. For the prediction of truck arrivals for pick-ups at $t+Q$ days ahead, it is denoted by $\hat{y}_{P,t+Q}$ (Q>1). Notations for trucks with Delivery task (D) and Dual tasks (H) are similarly denoted as $\hat{y}_{D,t+Q}$ and $\hat{y}_{H,t+Q}$. The predicted number of truck arrivals with export containers on day t is $\hat{D}_t^{'}$. The following notations are defined as input features.

1) *Time series of daily truck volumes:*

$Y_{s,t}$ a vector representing the ground truth of the s type of truck arrivals at t day, $t=\{\mathrm{k},...,N\}$, $s\in\{P,D,H\}$, N is the number of sample size.

The time series of daily arrivals of the Pick-up task (P), Delivery task (D), and Dual tasks (H), are denoted by $Y_{P,t}$, $Y_{D,t}$, $Y_{H,t}$.

2) *Weather-related parameters:*

S_t^1 a vector representing the first weather condition of t day， $t\in\{\mathrm{k},...,N\}$

S_t^2 a vector representing the second weather condition at t day， $t\in\{\mathrm{k},...,N\}$

All the 9 weather conditions in Fig.5 are the value of S_t^1 or S_t^2. For example, it is "cloudy to light rain" during the t day. Then S_t^1 is cloudy and S_t^2 is light rain.

V_t^1 a vector representing the maximum temperature at t day, $t \in \{k,...,N\}$

V_t^2 a vector representing the minimum temperature at t day, $t \in \{k,...,N\}$

For example, the temperature is 15℃~20℃, then the V_t^1 is 20, and V_t^2 is 15.

3) Weekday parameters:

W_t a vector representing the weekday at t day, $t \in \{k,...,N\}$

4) Vessel-related variables:

$A_{c,t}$ a vector representing the observed value of outbound truck arrivals on the t day with closing time c, $A_{c,t} = [\alpha_{c,t-1},...,\alpha_{c,t-k}]$, $c \in \{1,...,C\}$, $t \in \{c-15,...,c+2\}$, k is the sequence length;

W_c the weekday of the cut-off period c;

$L_{c,t}$ length of time from the t day till the closing time c;

$\beta_{c,t}$ the ratio of truck arrivals with outbound containers on the t day with closing time c;

Question 1: The learning function $f(\cdot): \mathbb{R}^{(N-k)\times F} \rightarrow \mathbb{R}^{N-k}$ maps input data features to the arrivals of the next time interval, where F is the dimension of the input features. Formula (1) is the prediction problem.

$$[Y_{s,t}, S_t^1, S_t^2, V_t^1, V_t^2, W_t] \xrightarrow{f} \hat{y}_{s,t} \qquad (1)$$

Before training the model, pre-processed data techniques are used to model the defined six factors. It includes the following tasks:

i) Independent Factors are converted into one row and m-column array with only one bit being 1 and the rest being 0 using the One-Hot encode method. m is the number of different values in the feature set. Take weekdays for an example. W_t has seven variable forms from "Monday" to "Sunday". Using the One-Hot encoding, "Monday" is transformed into a vector $[1,0,0,0,0,0,0]$, "Tuesday" $[0,1,0,0,0,0,0]$, etc.

ii) Factors with obvious size relationships (e.g. S_t^1 and S_t^2) are converted into an integer using the method of label encoding.

iii) To improve the prediction accuracy and convergence speed of the model normalization is used. All factors are transformed between [0,1].

In the proposed model GRU-FCNN, the FCNN1 component is fed with non-operational factors which contain 11 dimensions including W_t (seven), S_t^1 (one), S_t^2 (one), V_t^1 (one), V_t^2 (one), and then the GRU component is fed with the truck arrival records $Y_{s,t-1}$ (one) but K sequence length.

After transforming every K consecutive truck arrival record as groups, the data are split into two data sets: (a) The first 70% of the data is used for model training, which includes $(365-K)\cdot 70\%$ days of H type, $(365-K)\cdot 70\%$ days of P type, and $(365-K)\cdot 70\%$ days of D type; (b) the remaining 30% of the data is used for model testing, which includes $(365-K)\cdot 30\%$ days of H type, $(365-K)\cdot 30\%$ days of P type and $(365-K)\cdot 30\%$ days of D type [32].

Question 2: Where F is the dimension of the input features, $g(\cdot): \mathbb{R}^{M\times F} \rightarrow \mathbb{R}^{M}$ is the learning function that describes the input data features to export container arrivals in the next time interval of CY closing time.

$$[A_{c,t}, W_c, L_{c,t}] \xrightarrow{g} \hat{\alpha}_{c,t} \qquad (2)$$

Specifically, the raw data are categorized according to the remaining time till the CY closing date $t = c-15,...,c+2$. A small portion of the export containers can get later than the CY closing date in less than 2 days due to uncertain factors. The containers with the same CY closing date are treated as the same category. The arrival number $\alpha_{c,t}$ is summed in the same sub-figure. Totally there are 182 sub-figures for the 182 CY closing date in Appendix A, which includes 3348 samples.

Inbound containers can only be picked up after the vessel's berthing time. Basically, they do not have any relationship with the CY closing date. However, the data shows that the relation with berthing time is not significant either as in Appendix B. It is owing to the fact that the terminal in Southern China is an export port, and most of the inbound containers are empty. Under some special contracts between shipping lines and the terminal, some empty containers can be stored in the yard for future use. It is normal for a container to stay in the yard for half a month. As a result, the trend of pick-up arrivals is not as obvious as delivery arrivals.

In practice, the total number of export containers to be loaded on the ship is known. We can predict the ratio of truck arrivals $\beta_{c,t}$ instead of $\alpha_{c,t}$, in which $\beta_{c,t} = \alpha_{c,t} / E_c$. E_c is the total number of export containers with the CY closing time c. The learning function can be transformed into equation (3) and the export container arrivals volume $\hat{D}_t^{'}$ on day t is calculated by equation (4).

$$[B_{c,t}, W_c, L_{c,t}] \xrightarrow{g} \hat{\beta}_{c,t} \qquad (3)$$

$$\hat{D}_t^{'} = \sum_c \hat{\beta}_{c,t} * E_c \qquad (4)$$

To reduce the impact of outliers on the training model, we delete some $\alpha_{c,t}$ with a small total number E_c. As illustrated in Fig 6, in which the blue points are original ones and the orange points are those filtered. If E_c >2000, then $\alpha_{c,t}$ is kept; otherwise, the data is deleted. Furthermore, some deleted data are relating to the season of the Chinese Spring Festival, with a trough in export container volume for 18 days before and after the holiday. Employees of the factories, trucking companies, and the terminal were on vacation during the holiday season. Because lack of previous data with both vessel and holiday information to train the model, $\alpha_{c,t}$ on those days are deleted to reduce the caused bias. The other cases include the dates 2020-01-02, 2020-06-29, and 2020-06-30, which have missing data. Fig.6 shows a comparison of data before and after screening.

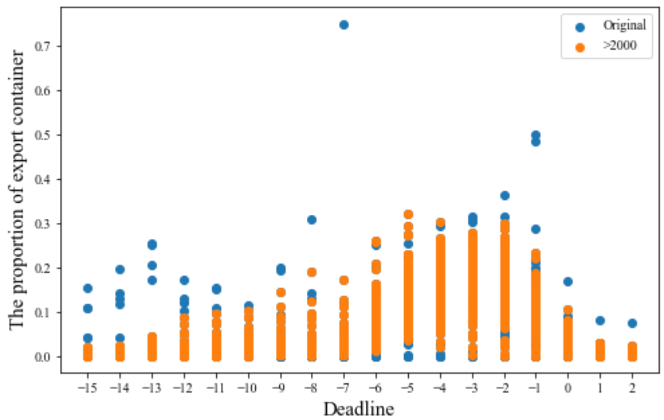


Fig.6.Outbound arrivals rate before and after screening

Data preprocessing techniques are applied to the input features. W_c and $L_{c,t}$ are converted using the One-Hot encode technology. Because they are discrete integers that have no direct impact on the total number of export containers, e.g., "The volume of export containers on Tuesday must be smaller than Wednesday's" is incorrect. Except for outliers, data1 (2790 records) is divided into a training set (1953 records) and a test set (837 records) in a 70:30 ratio to ensure data integrity.

Varying sequence length is adopted to improve the fitting effect of sequence models, including RNN, LSTM, and GRU-FCNN. It is noted that k is the sequence length in the input data $A_{c,t} = [\alpha_{c,t-1},\ldots,\alpha_{c,t-k}]$. However, the number of possible values t is restricted. The traditional fixed sequence length input will have the following two effects if k is set to a large value: a) The amount of data available will be reduced, which will have a significant impact on the training effect and lead to underfitting. b) The predictable range is limited to the arrival volume of export containers on previous days. For example, if $k=3$, the arrival volume of export containers must be collected for at least 3 days to meet the demand. However, the ratios before the 3 days, e.g., $\hat{\beta}_{c,c-15}, \hat{\beta}_{c,c-14}, \hat{\beta}_{c,c-13}, \ldots$ will not be contained. We abandon fixed sequence lengths in favor of informative sequences of varying lengths, inspired by the widespread success of variable-length deep learning models in the field of semantic recognition. The sequence model currently only accepts inputs of equal length, and the missing points in the export container arrival sequence must be filled in. The arrivals with export containers before the deadline gradually approach 0, as shown in Fig. 4. Therefore, we fill in the missing value of the unequal length export container sequence $B_{c,t} = [\beta_{c,t-k}, \beta_{c,t-k+1},\ldots,\beta_{c,t-1}]$ with the value of 0, as shown in Fig.7.

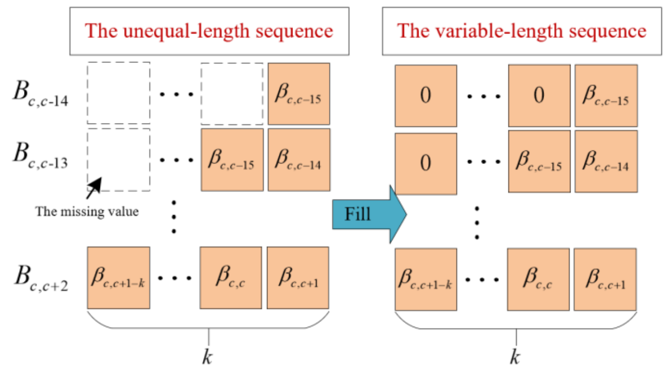


Fig.7.Filling process of unequal length sequences

B. *The GRU-FCNN Model*

We propose a novel hybrid deep learning method fusing GRU and FCNN components. It consists of two FCNN components and one GRU component. The inputs of FCNN components are cross-sectional data sets, including non-operational factors (weather-related variables and weekday variables) at $t+1$ day. The inputs of the GRU component are sequences, w.r.t., the consecutive truck arrival records at k pre-steps. For Question 2, there is no information exchanged between $B_{c,t}$, W_c, and $L_{c,t}$. It is inefficient to put them all into the sequence model. Therefore, cross-sectional data W_c and $L_{c,t}$ are fed into the FCNN component 1; and time series data $B_{c,t} = [\beta_{c,t-k}, \beta_{c,t-k+1},\ldots,\beta_{c,t-1}]$ are fed into the GRU component.

The FCNN component is constructed from the multi-layer network, each of which has the same number of neutrals and uses the ReLU activation function. FCNN is the most basic neural network model. Each layer is fully connected, and the data flow passes through the input layer and the hidden layer before reaching the output layer[37]. The GRU component contains multi-layer GRU units to increase the accuracy of learning. The GRU model is similar to the LSTM model and is often proposed to solve problems like long-term memory and backpropagation gradients. The reset and update gates in the internal structure of the former improved learning efficiency significantly, and the computational complexity is lower than that of the latter. The GRU can achieve the same accuracy as the LSTM with the same number of layers and neurons, and the calculation time will be shorter[16]. The hidden state h_t of the GRU at time step t is obtained based on the current input x_t and the information h_{t-1} retained from the previous step, assuming that the sequence $X = (x_{t-k}, x_{t-k+1},\ldots,x_t)$ is sequentially passed into the GRU. As a result, the GRU can remember the previous element's effect on the next element. Formulas (5)-(8) are used to express the above process:

$$r_t = \sigma(W_{ir}x_t + b_{ir} + W_{hr}h_{t-1} + b_{hr}) \tag{5}$$

$$z_t = \sigma(W_{iz}x_t + b_{iz} + W_{hz}h_{t-1} + b_{hz}) \tag{6}$$

$$n_t = \tau(W_{in}x_t + b_{in} + r_t \odot (W_{hn}h_{t-1} + b_{hn})) \tag{7}$$

$$h_t = (1 - z_t) \odot n_t + z_t \odot h_{t-1} \tag{8}$$

Where $\odot$ is the Hadamard product of two matrices, and $\sigma(x)$ is a GRU unit activation function that converts the variable between 0 and 1, indicating how much information is remembered. $\tau(\mathrm{x})$ can reduce the occurrence of gradient disappearance and gradient explosion. In the model proposed in this paper, the FCNN uses the activation function ReLU as formula (11) to map the nonlinear relationship between historical data and non-operational factors.

$$\sigma(x) = \mathrm{sigmoid}(x) = \frac{1}{1+e^{-x}}, x \in (-\infty, \infty) \tag{9}$$

$$\tau(\mathrm{x}) = \mathrm{tahn}(x) = \frac{1-e^{-2x}}{1+e^{-2x}}, x \in (-\infty, \infty) \tag{10}$$

$$\mathrm{ReLU}(x) = \begin{cases} x, x \geq 0 \\ 0, x < 0 \end{cases}, x \in (-\infty, \infty) \tag{11}$$

Fig.8 illustrates the overall architecture of the proposed model. It is proposed to include the FCNN1 component to capture NOF; and the GRU component in the model to reflect the influence of the truck's historical data. The former captures the influence of NOF, and the latter learns the long-term and short-term influences hidden in the sequence. To describe the correlation between NOF and historical truck arrivals, we use a fusion technique to connect the output tensors of FCNN component 1 and the GRU component. Then we pass them into FCNN component 2. It means that the input dimension of the FCNN component 2 is $\mathrm{N_{FCNN1}}$+$\mathrm{N_{LSTM}}$. The final output of FCNN component 2 predicts the number of truck arrivals. The loss function, which compares the predicted and observed values to calculate the loss, measures the model's prediction accuracy. A backpropagation algorithm is then used to update the model's weights and biases.

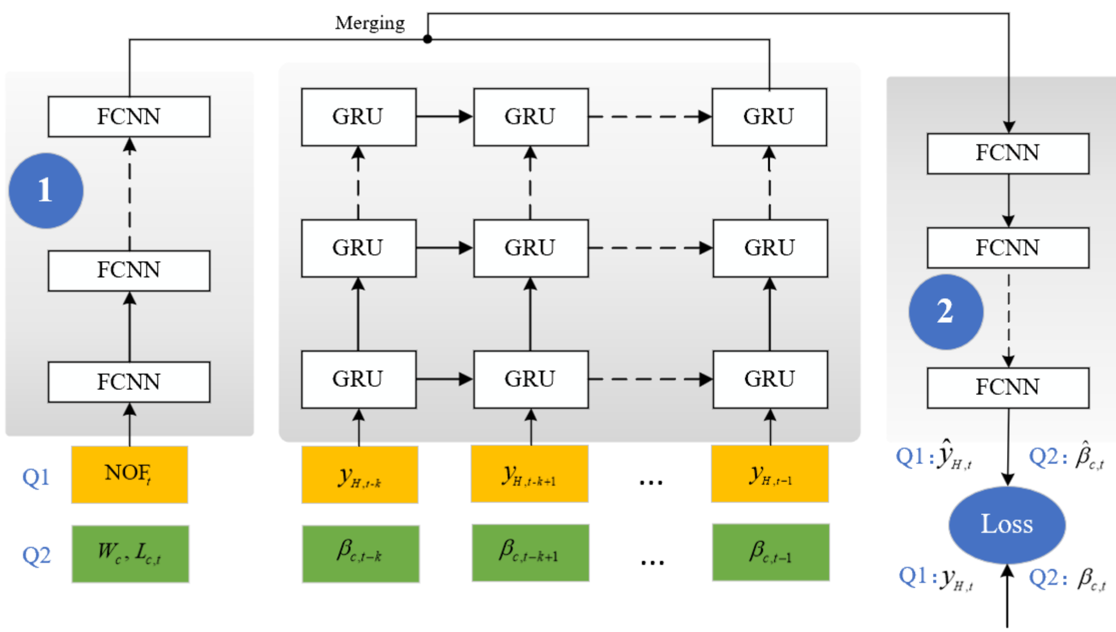


Fig.8. The structure of the GRU-FCNN model

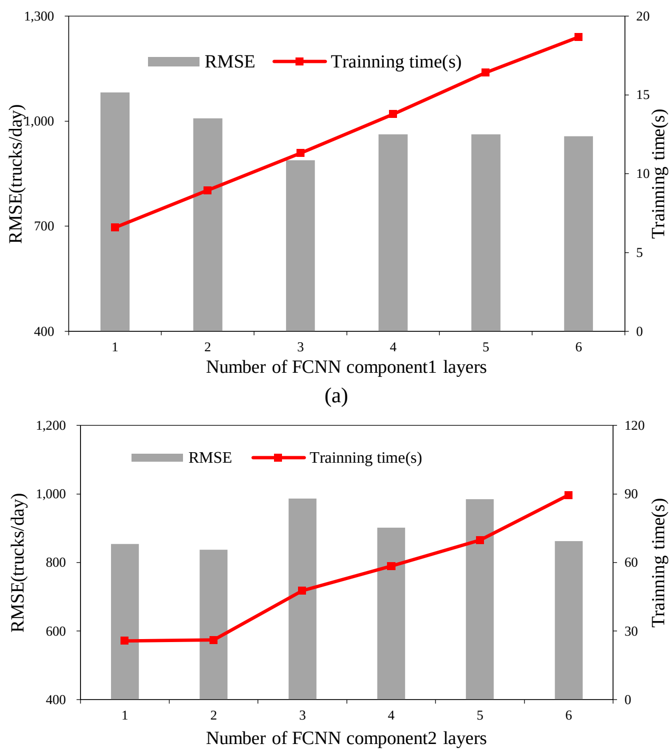
V. Model Training

The GRU-FCNN model has many variables and model parameters, which can be determined through calibration and evaluation to achieve the best performance for a specific domain problem. In this section, the parameters of GRU-FCNN are proved through experiments. For the sake of brevity, this section only takes the dual truck arrival forecast model as an illustrative example. Similar steps can be repeated for the predictions of pick-up and delivery tasks.

Before the model is built, the number of hidden layers and the number of neurons in each layer should be determined first which depends on each other. Generally, we first manually input the number of neurons in each layer and then find the corresponding optimal number of hidden layers [38, 39]. Considering the number of non-operational factors (Eleven) and historical data of truck arrivals (One), 64 units are manually set in each layer at FCNN component1 and 16 units are manually set in each GRU layer, and the output layer has the same neurons as the hidden layer. The number of FCNN2 input layer units is the sum of FCNN1 and GRU outputs, which is 80.

Each hidden layer has been assigned the same number of units as the input layer.

To discover the best-hidden layers in FCNN1 and FCNN2, we firstly train the model of hidden layers 1. Then we gradually increase the hidden layers until the error is no longer reduced. As shown in Fig.9, the error gradually decreases as the number of hidden layers increases. As it achieves 3, the loss doesn't decrease any longer but needs more training time. The performance of the GRU component is affected by the number of hidden layers and k sequence length. Then we gradually increase the number of hidden layers and k sequence length until the loss error stops decreasing. Fig.10 shows the effect the number of hidden layers, as well as the value k, has on RMSE (root mean square error). The best parameters of other types of trucks are also summarized in Table III.



(a)

(b)

Fig. 9. Model performance (RMSE) vs. the number of hidden layers at (a)FCNN component1 and (b) FCNN component2.

The performance of the neural network model is also affected by the application of technology. To improve the efficiency of training, Adam (Adaptive moment estimation) [40] is used. Adam is more efficient than the gradient descent algorithm. ReduceLROnPlateau technology allows setting a larger learning rate manually to help the parameters fall quickly in the high-dimensional space and save time. If the loss value does not change within 10 steps, the learning rate will be reduced to 10% of the original [41]. L2 regularization[42] and drop-out technology[43] are commonly used methods to prevent overfitting in deep l

earning.

To evaluate the prediction accuracy, RMSE and MAPE are used as in Equations (12-13). Both of them are the commonly used cost functions in the study of prediction, which can accurately show the current fitting level.

$$\mathrm{RMSE}=\sqrt{\left(\sum_{t=1}^{N}(y_{s,t}-\hat{y}_{s,t})^2\right)/N} \tag{12}$$

$$\mathrm{MAPE}=\frac{1}{N}\sum_{t=1}^{N}\frac{|\hat{y}_{s,t}-y_{s,t}|}{y_{s,t}}*100\% \tag{13}$$

Among them, N is the amount of data, $y_{s,t}$ is the observed number of truck arrivals, $\hat{y}_{s,t}$ is the predicted one, $s\in\{P,D,H\}$.

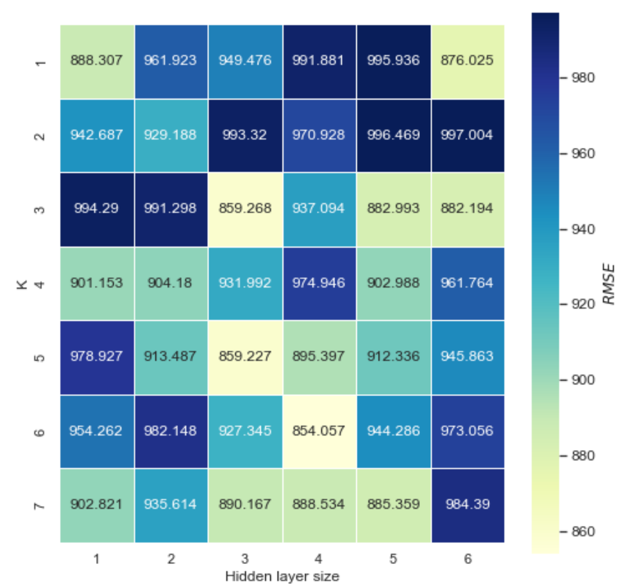


Fig. 10. The heatmap of RMSE of combinations of k and hidden layer size

TABLE III

STRUCTURE AND PARAMETERS OF GRU-FCNN

	Dual truck	Pick-up truck	Delivery truck
Model architecture	FCNN component 1: 3 hidden layers, each with 64 neurons; FCNN component 2: 2 hidden layers, each with 80 units. LSTM component: 4 hidden layers, each with 16 units and 6 sequences;	FCNN component 1: 3 hidden layers, each with 64 neurons; FCNN component 2: 1 hidden layer, each with 80 units. LSTM component: 3 hidden layers, each with 16 units and 7 sequences;	FCNN component 1: 4 hidden layers, each with 64 neurons; FCNN component 2: 2 hidden layers, each with 80 units. LSTM component: 3 hidden layers, each with 16 units and 7 sequences;
Optimizer		Adam	
Initial learning rate	0.001	0.001	0.001

Techniques to avoid over-fitting		ReduceLROnPlateau; Ridge Regression; Dropout.	
Regularization coefficient	0.00001	0.00005	0.00003
Epochs	500	600	600

VI. EXPERIMENTS AND PERFORMANCE ANALYSIS

A. Benchmarks

To evaluate the performance of the proposed model, six commonly used models for truck arrival prediction are compared, namely Linear Process model (LP), Support Vector Regression (SVR), Random Forest (RF), Fully Connected Neural Network (FCNN), Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM). The basic principles and parameters of these models are as follows.

1) Linear Process model (LP) has become one of the most popular models in the field of machine learning and is often used for prediction and classification problems. The basic principle is to assume that the data obey the multivariate linear function distribution, and then use the least square method or gradient descent method to find the parameters [44]. The commonly used kernel functions in linear regression are Poly (Polynomial) and RBF (Radial Basis Function). The model requires the target value and the feature to have a linear relation and the RBF kernel function can ensure the linear regression model fits the data that follows a Gaussian distribution. Because the distribution of the dual trucks (Fig.3(a)) is not a Gaussian process, we use the polynomial linear process instead. The model is implemented using the Scikit-learn package [45].

2) Support Vector Regression (SVR) is a powerful supervised machine learning model that can realize classification and regression. It is a linear classifier with the largest interval defined in the feature space [46]. This model has been used to predict the arrival of trucks [32]. Parameters can affect the performance of the SVR model, so the Scikit-learn package is used [45], and it can optimize the kernel function, penalty factor, loss function, and regularization. The commonly used kernel functions in SVR are linear, polynomial, sigmoid, and RBF. The penalty factor can be selected from 0.01, 0.1, 0.5, 1, 2, 5, 10, 100. The loss function can be either a hinge loss function or a square hinge loss function. More details are documented in Swami and Jain [45]. On the prediction of dual truck arrivals, the kernel of the SVR was a polynomial kernel function, the penalty function coefficient was 2 and the degree was 3. However, the RBF kernel function is used in the prediction of pick-up and delivery.

3) Random Forest (RF) is a simple and powerful machine learning technique that uses ensemble learning to merge several decision trees. From an intuitive standpoint, if each decision tree conducts an individual regression operation (assuming that it is a regression problem), then N trees will have N results for an input sample. The random forest incorporates all of the regression voting results, and the category with the most votes is assigned as the final output. This model is widely used in various fields, including the problem of truck flow prediction[47-49]. The model is established using the Scikit-learn package [45]. The amount of decision trees on the forecast dual trucks is manually set at 70.

4) Fully Connected Neural Network(FCNN) is a type of fully connected front-end neural network with multiple layers, multiple neurons, and adjacent hidden layers. In the model training, the data passes from the input layer through the hidden layer to the output layer, and the gradient descent algorithm and backpropagation algorithm optimize the weight and bias [37]. This model is widely used in various fields, including the problem of truck arrival prediction [29-31, 33, 34]. The FCNN model can be established by using PyTorch which can optimize the hidden layer, the number of neurons, and the activation function [41]. Finally, the FCNN used in this study was optimized with three hidden layers, each of which had 35 neurons. The activation function was chosen from identity, logistic, tanh, and relu. Finally, we used the relu activation function.

5) Recurrent Neural Network (RNN) is a type of recurrent neural network that takes sequence data as input, recursively in the evolution direction of the sequence, and all recurrent units are connected in a chain. RNN has short-term memory capabilities. Its neurons cannot only receive information from other neurons but also receive their information, forming a network structure with loops. At the same time, the network will memorize the previous information and apply it to the calculation for the current output, which means, the nodes between the hidden layers are no longer unconnected but connected, and the input of the hidden layer includes not only the output of the input layer but also the output of the last hidden layer. At present, RNN has been widely used in natural language processing [50]. The RNN model can be established by using PyTorch, and the optimization of the hidden layer, the number of neurons, and the activation function can be realized [41]. The RNN used in this study was optimized with three hidden layers, each of which had 64 neurons, and with the 3 sequences.

6) Long Short-Term Memory (LSTM) is a sort of recurrent neural network that is used to handle the problem of vanishing and exploding gradients. The most important part is the memory cell that performs information addition and removal operations through the gate structure (including

the ‘forget’ gate, input gate, and output gate). This model is widely used to process long-length datasets, including traffic flow[18], train delay[51] and stock price movement[52], etc. Using PyTorch, the LSTM model can be created, and the hidden layer, number of neurons, and activation function can all be optimized[41]. The LSTM used in this study was optimized with 5 hidden layers, each of which had 64 neurons, and with the 3 sequences for dual truck arrival.

B. Model Performance on One-step Prediction

First of all, we make single-step predictions for dual trucks, delivery trucks, and pick-up trucks. The results of the GRU-FCNN model and baseline models are shown in Table IV. The training time is also recorded and shown in the last column of Table IV. Based on the test data-sets from May 19, 2019, to August 31, 2019, it presents an average improvement(Gap1) of 23.44%, 32.09%, and 26.99%, compared to the mean values of other models, concerning dual, pick-up, and delivery trucks respectively. GRU-FCNN shows a distinctive advantage over other models in prediction accuracy. Moreover, the gap in the estimation can also be seen in Fig.11, where the prediction data by GRU-FCNN is the closest to the real data. The inaccuracy during the period from July 31, 2019, to August 3, 2019 (within the red square in the dual of Fig.11(a)) may be due to human factors.

$$\text{Gap1}=\frac{\text{RMSE(Baseline model)-RMSE(GRU-FCNN)}}{\text{RMSE(GRU-FCNN)}}\times 100\% \tag{14}$$

$$\text{Gap2}=\frac{\text{MAPE(Baseline model)-MAPE(GRU-FCNN)}}{\text{MAPE(GRU-FCNN)}}\times 100\% \tag{15}$$

TABLE IV
THE OVERALL PERFORMANCE OF MODELS ON TEST-DATASET

Truck type	**Model**	**RMSE**	**MAPE (%)**	**Gap1(%)**	**Gap2(%)**	**Time Cost(s)**
Dual	**GRU-FCNN**	**837.38**	**11.75**	-	-	**26.115**
	Linear process(poly)	1162.34	17.47	27.96	32.74	0.761
	SVR(poly)	1123.92	18.25	25.49	35.62	0.250
	RF	1227.21	16.11	31.77	27.06	0.065
	FCNN	1133.74	16.26	26.14	27.74	51.72
	RNN	1138.20	18.49	26.43	36.45	5.63
	LSTM	862.28	13.19	2.89	10.92	140.98
Pick-up	**GRU-FCNN**	**228.33**	**8.53**	-	-	**20.411**
	Linear process(RBF)	285.91	10.82	20.14	21.16	0.23
	SVR(RBF)	407.34	14.37	43.95	40.64	0.21
	RF	452.08	15.65	49.49	45.50	0.063
	FCNN	331.54	11.73	31.13	27.28	24.6
	RNN	328.01	11.62	30.39	26.59	23.00
	LSTM	276.47	9.84	17.41	13.31	152.59
Delivery	**GRU-FCNN**	**365.10**	**10.23**	-	-	23.00
	Linear process(RBF)	555.22	15.80	34.24	35.25	0.23
	SVR(RBF)	563.57	16.59	35.22	38.34	0.20
	RF	566.39	15.25	35.54	32.92	0.046
	FCNN	439.35	13.00	16.90	21.31	15.40
	RNN	459.46	12.77	20.54	19.89	6.70
	LSTM	453.64	12.95	19.52	21.00	138.22

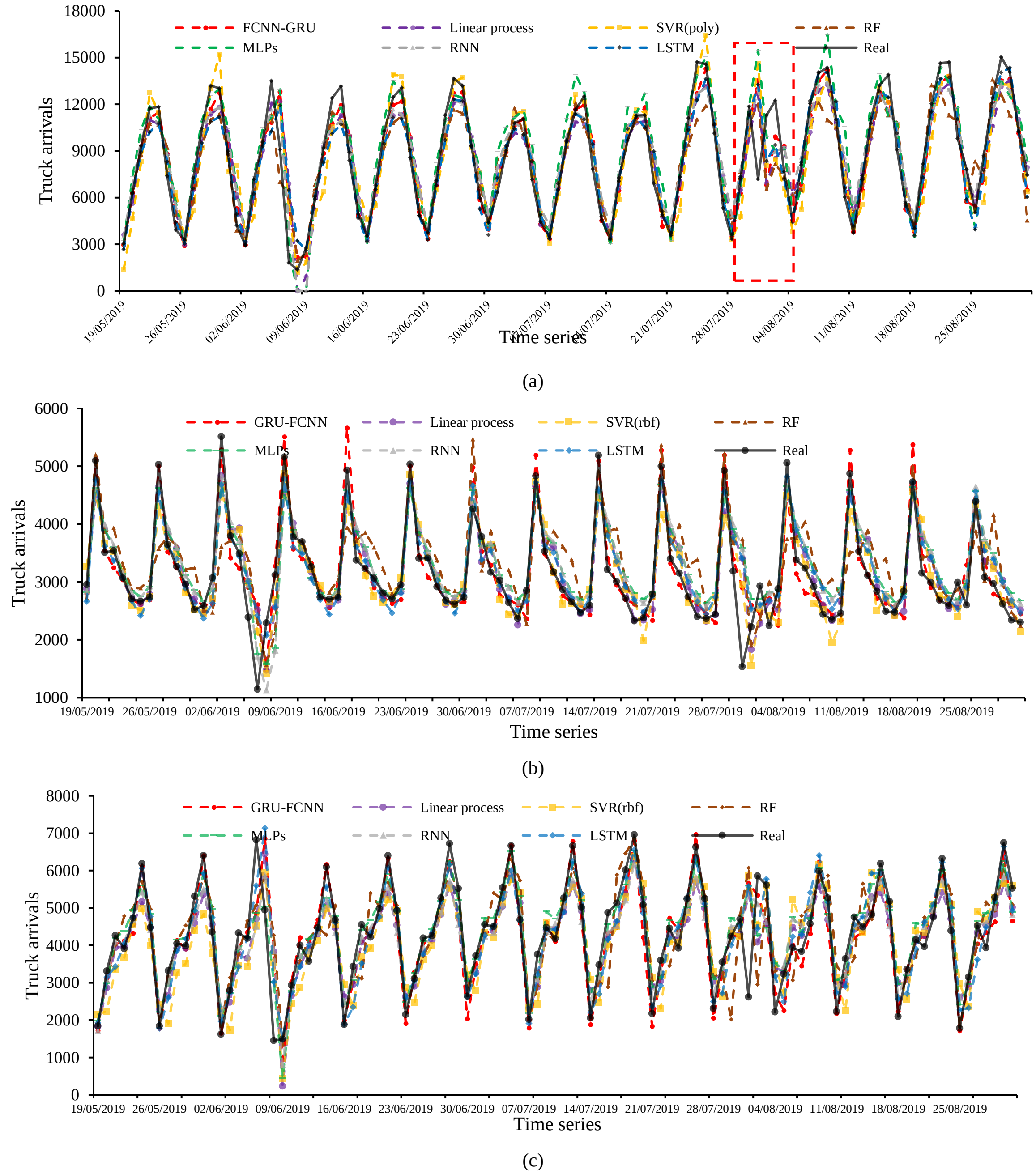
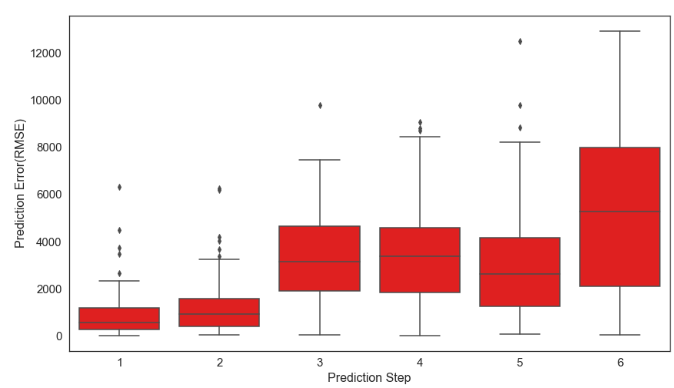


Fig.11. Truck arrivals prediction of each model on test datasets. (a) Dual. (b) Pick-up. (c) Delivery.

C. *Model Performance on Q-step Prediction*

The models are also used to make Q-step predictions (Q>=2). The model performance is investigated using three categories of the dual(H), pick-up(P), and delivery(D) trucks respectively in a container terminal in Southern China.

Take the variance of RMSE of dual trucks (H) as an example. As illustrated in box plots in Fig.12, the median of RMSE gradually increases with the lengthening of prediction steps. At the same time, the fluctuation of prediction is enlarged. It is consistent with the common applications of the GRU-FCNN model for short-term prediction [23, 53]. It is explained that the past information becomes less valuable to facilitate the prediction in a further future, which causes the reduction of the prediction accuracy [54].



(a)

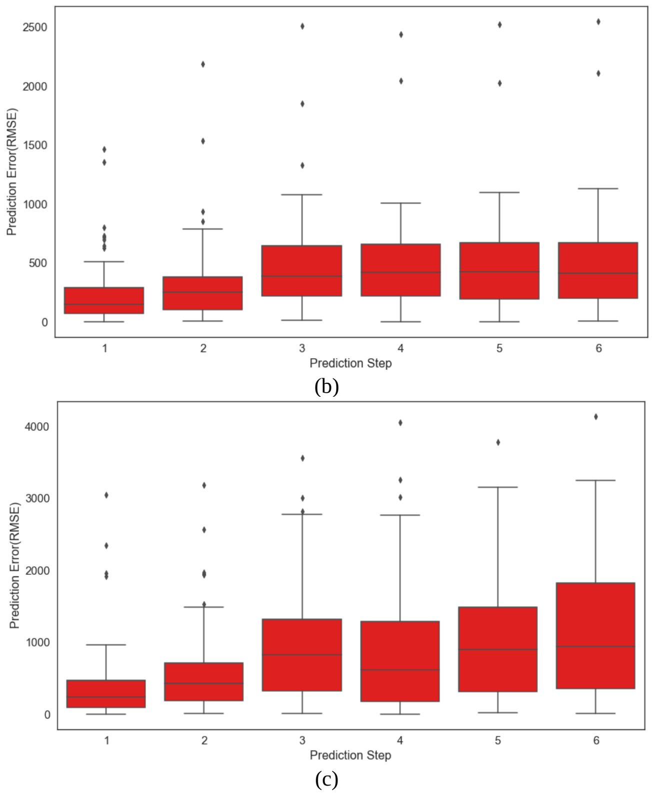


Fig.12. The multi-step prediction for GRU-FCNN. (a) Dual. (b) Pick-up. (c) Delivery.

D. *Sensitivity Analysis of Input Factors*

To determine the influence of different factors, the input is decomposed and analyzed. In detail, in this section, we investigate the extent to which the meteorological factors (temperature and weather conditions), and weekly data affect the accuracy of prediction. The following four types of experiments with different input factors are therefore set up. Their results are compared with the control group.

Test 1: GRU-FCNN is fed with the data sets without "weekday" (W_t) information ;

Test 2: GRU-FCNN is fed with the data sets without “weather-related” ($S_t^1, S_t^2, V_t^1, V_t^2$) information;

Test 3: GRU-FCNN is fed with daily volumes of truck arrivals only. The FCNN component1 is removed and all the non-operational factors (NOF) $(S_t^1, S_t^2, V_t^1, V_t^2, W_t)$ are not included.

Control group: GRU-FCNN is fed with all the input data sets.

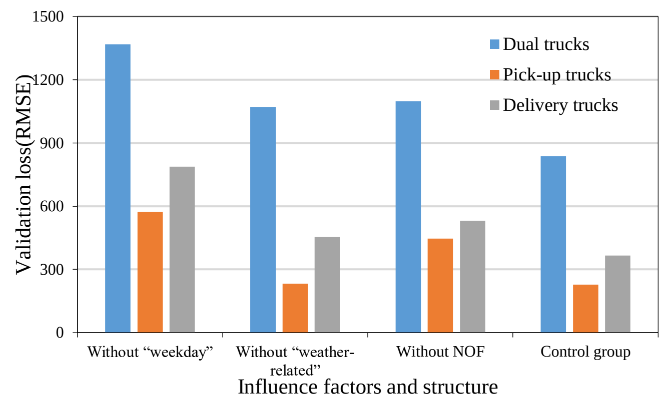


Fig.13. The comparison results with different input data and structure

Fig.13 shows the RMSE of the experiments with dual, pick-up, and delivery trucks as an illustrative example. Firstly, the difference between Test 1 (without “weekday”) and the control group is the most significant, which is 531.40(H), 345.75(P), and 422.52(D). It demonstrates the importance of the weekday factor in the prediction. The weekday factor reflects the arrival pattern of liner vessels. Experimental results prove the close relationship between truck arrivals and vessel schedules. Secondly, the results of Test 2 (without “weather-related”) show that the weather-related factors contribute to the improved accuracy of prediction. It means truck arrivals are affected by weather conditions. There was no extreme condition during the period from May to August 2019 in Southern China. If there was extreme weather, like hurricanes or typhoons, the influence would be more serious. Lastly, a comparison between Test 3 (without all the NOF) and the control group once more proves the importance of weekday and weather-related information, as well as the advantages of the proposed method. In tests 1 and 3, it can be found that the prediction results with the weather feature only (in Test 1) are even worse than those obtained without any NOF. However, the week feature (in Test 2) enhances prediction accuracy compared with Test 3. Interestingly, the control group performs best with the least divergence from the real value, since the interaction between the weather and weekday features enhances prediction accuracy further.

E. *Advantages of Prediction with Three Categories*

In the literature [28-30, 32], the truck is usually classified into two categories: pick-up trucks and delivery trucks. However, from a practical view, dual transactions are used in a growing manner and could not be overlooked anymore [55]. To show the advantages of the three categories, data is preprocessed with only two categories. Because each truck delivers a container and then picks up another one in dual transactions. The number of pick-ups (P) adds by the number of the dual (H). So does the number of deliveries (D). It is noted that P_t' and D_t' represent the number of transactions of pickup and delivery truck arrivals respectively in this context. $\hat{P}_t'$ and $\hat{D}_t'$ represent the estimated number of transactions. Equations (16) and (17) are the calculation of mean values of errors. Using one-step prediction, the comparison results are shown in Table V. It is remarkable to find that the errors of three-category of the truck reduce by at least 38.49% across all six models. In Fig.14, the illustration results of GRU-FCNN demonstrate that the three-category datasets have lower prediction errors on almost days.

$$\text{Errors of two-category} = \frac{1}{N}\sum_{t=1}^{N}\left|(P_t' + D_t') - (\hat{P}_t' + \hat{D}_t')\right| \tag{16}$$

$$\text{Errors of three-category} = \frac{1}{N}\sum_{t=1}^{N}\left|(P_t + D_t + H_t) - (\hat{P}_t + \hat{D}_t + \hat{H}_t)\right| \tag{17}$$

where P_t, D_t, H_t are ground truth of trucks with a pick-up, a delivery, or dual tasks.

TABLE V

THE ERRORS FOR TWO OR THREE CATEGORIES DATA-SETS

Models	two-category(a)	three-category(b)	|(a)-(b)|/(a)(%)
GRU-FCNN	**1880.77**	**997.44**	**46.97**
LSTM	2057.85	1083.04	47.37
Linear process	2360.01	1189.16	49.61
SVR	3038.84	1295.29	57.37
RF	2626.26	1449.80	38.49
FCNN	2460.52	1253.24	49.07
RNN	5691.30	1018.11	82.11

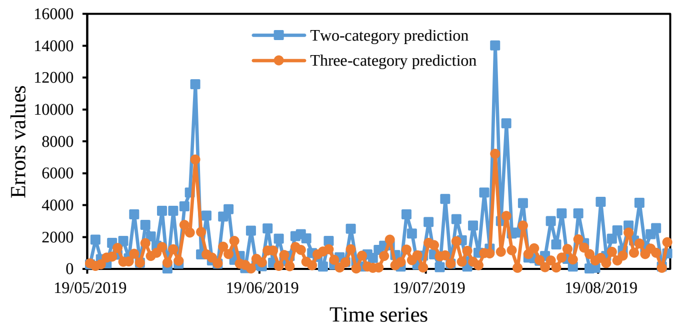


Fig.14. The advantages of predictions based on three categories of GRU-FCNN results

F. The prediction with vessel information

Following Question 2, experiments are undertaken to reveal the effects of vessel information, e.g., the CY closing date of each vessel. The prediction errors of different models are shown in Table VI. The prediction without vessel information has been conducted by using the same data set based on the method in Section VI. It demonstrates that vessel information would significantly help in accuracy improvement. Except for the model of SVR, the other six models have improved the performance variously with vessel information. Among the models, the GRU-FCNN has the best results, which reflects its advantages in dealing with both time series and cross-sectional data. The RMSEs of $\hat{\beta}_{c,t}$ and $\hat{D}_t^{'}$ vary proportionally, which proves the reasonability of the experiment with vessel information.

According to the experimental results, the daily arriving number of outbound trucks is illustrated in Fig. 15. We can find the prediction curve of GRU-FCNN in red almost overlap with that of the ground truth.

In Fig.16, for each CY closing date, the prediction curves of different models are shown. Although they have various features on different days, the GRU-FCNN does best in most cases, owing to the vessel information and weekday as important input factors.

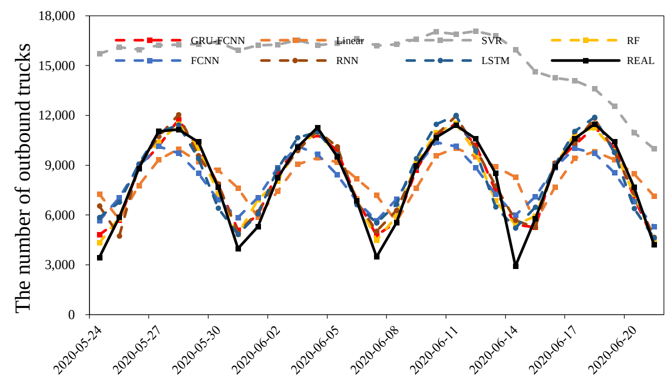


Fig. 15 The prediction contrasts of outbound containers

TABLE Ⅵ

THE CONTRASTS OF THE CURACY OF VARIOUS MODELS

			Linear	SVR	RF	FCNN	RNN	LSTM	GRU-FCNN
With vessel information	$\hat{\beta}_{c,t}$	RMSE	0.0153	0.0698	0.0107	0.0148	0.0115	0.0130	0.0091
		MAPE	90.61%	4079.30%	51.65%	254.94%	192.23%	253.90%	183.70%
	$\hat{D}_t^{'}$	RMSE	1256.74	6058.22	495.78	1014.90	590.03	760.90	461.44
		MAPE	29.84%	3509.25%	19.56%	273.73%	24.28%	20.36%	15.67%
Without vessel information	$\hat{D}_t^{'}$	RMSE	1262.16	2666.64	1306.04	1175.79	1125.76	892.12	807.99
		MAPE	74.82%	53.21%	60.43%	81.21%	77.49%	224.40%	95.59%

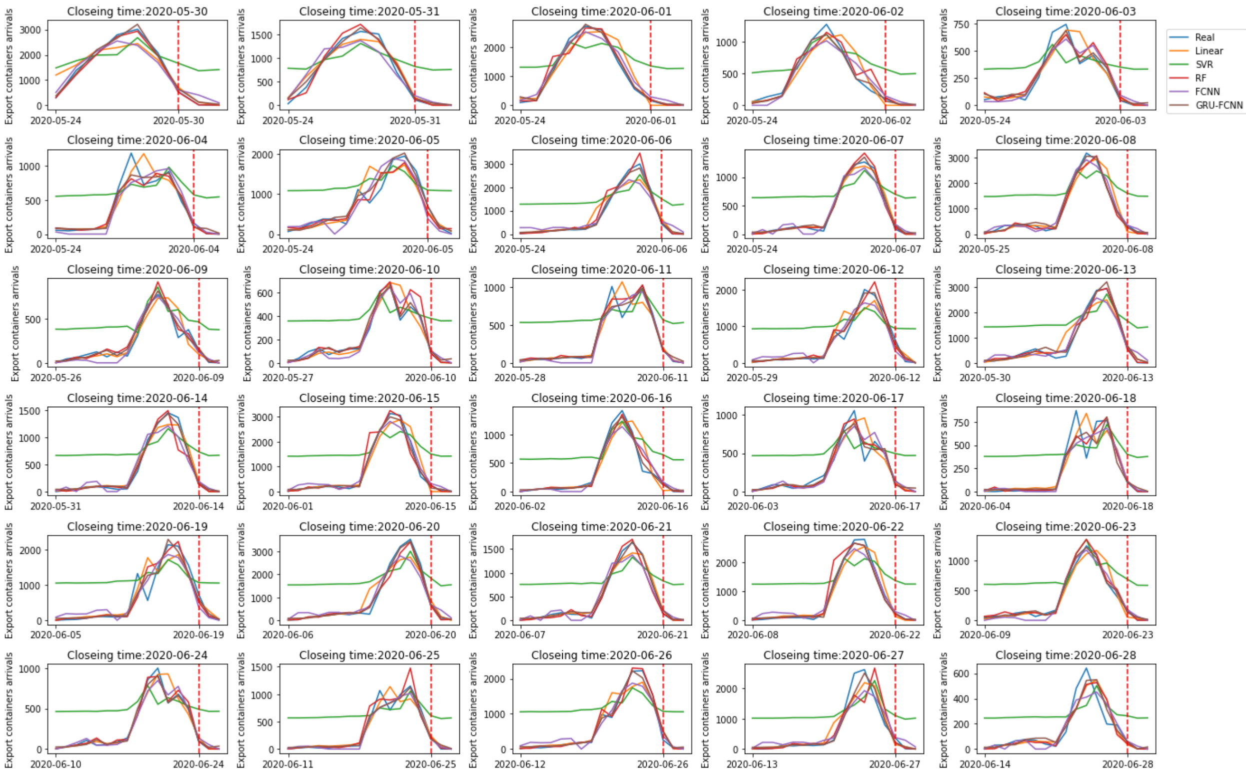


Fig.16. The prediction of outbound trucks with a specific CY closing date

VII. CONCLUSIONS

Container terminals serve as important transfer hubs for international trades and goods flows. An accurate prediction of daily truck arrivals would benefit the yard plan as well as equipment allocations to solve the associated bottleneck problems. This paper proposes a new GRU-FCNN model that enables the terminals to predict the accurate number of pick-up, delivery, and dual trucks coming in the future days. The model illustrates the close relationship between the non-operational factors (weekday and weather), CY closing time, and the daily truck arrivals. It outperforms the other widely-used models by 23.44%, 32.09%, and 26.99%, concerning dual, pick-up, and delivery trucks respectively. The comparison between the GRU-FCNN model and the six commonly used truck arrival prediction models in the existing literature discloses the new model's advantages in predictive performance, generalizability, and computational requirement. With the 3 categories of trucks (i.e. P, D, and H), as well as vessel information, the prediction accuracy is significantly improved within each model. The main findings and the associated contributions from these experiments are summarized as follows.

1) It is beneficial to consider the influence of weather-related and weekday data to improve the prediction accuracy of daily volumes of truck arrivals in a container terminal.
2) The proposed model GRU-FCNN considerably outperforms other widely used truck arrival prediction models.
3) The prediction with three categories of trucks is demonstrated to be more accurate than the one with two categories.
4) Adding vessel information as input data, e.g., the CY closing date of a vessel enhances the accuracy of export container arrival predictions greatly.

Multi-stakeholders including terminal managers, hauler companies, and shipping lines would therefore benefit from the improved effectiveness of our truck arrival prediction model for reasons beyond the generally expected basic planning rationale.

Terminal operators could undertake a better allocation of yard equipment and infrastructure according to the number of dual, pick-up, and delivery trucks. Yard cranes or front-end loaders could work more efficiently. The yard plan on where and how to store the containers could better perform by following the number of pick-ups and deliveries. Idling emission is more concerned compared to the pollution of trucks at other operations. Based on our findings, idling time could be cut down through a better allocation of equipment and more efficient loading/discharging operations. Container supply and demand could strike a balance by an accurate prediction of truck arrivals.

Drayage companies could also save time and personal cost with the rational matching of truck supply and demand. As a result, terminals could better schedule their infrastructure and equipment, and congestion around the terminal could be released and hence emissions from trucks.

Finally, our work is to be further developed to conduct the hourly prediction of truck arrivals or during any specified period of the day upon the research need and significance. The target is to establish a truck arrival prediction framework as a reference for the stakeholders (e.g. terminal operator) for better traffic management involving truck arrivals at a container terminal.

Appendix A

Fig.A. Outbound truck arrivals rate towards each closing time
Note: Each subgraph is an export arrival sequence at a closing time. The red dashed line is CY closing time.

Appendix B

Fig.B. Inbound truck arrivals towards each closing time
Note: Each subgraph is an export arrival sequence at a berthing time

REFERENCES

[1] UNCTAD, "Review of Maritime Transport." New York: United Nations Conference on Trade and Development, 2020.

[2] H. E. Haralambides, "Gigantism in container shipping, ports and global logistics: a time-lapse into the future," *Marit Econ Logist,* vol. 21, no. 1, pp. 1-60, 2019/03/01 2019.

[3] M.-H. Phan and K. H. Kim, "Collaborative truck scheduling and appointments for trucking companies and container terminals," *Transportation Research Part B: Methodological*, vol. 86, pp. 37-50, 2016.

[4] N. Huynh, "Reducing Truck Turn Times at Marine Terminals with Appointment Scheduling," *Transp Res Rec,* vol. 2100, no. 1, pp. 47-57, 2009/01/01 2009.

[5] N. Li, G. Chen, M. Ng, W. K. Talley, and Z. Jin, "Optimized appointment scheduling for export container deliveries at marine terminals," *Marit. Policy Manag.*, vol. 47, no. 4, pp. 456-478, 2020.

[6] X. Chen, X. Zhou, and G. F. List, "Using time-varying tolls to optimize truck arrivals at ports," *Transportation Research Part E: Logistics and Transportation Review,* vol. 47, no. 6, pp. 965-982, 2011.

[7] G. Chen, K. Govindan, and Z. Yang, "Managing truck arrivals with time windows to alleviate gate congestion at container terminals," *International Journal of Production Economics,* vol. 141, no. 1, pp. 179-188, 2013.

[8] L. Heilig, R. Stahlbock, and S. Voß, "From Digitalization to Data-Driven Decision Making in Container Terminals," in *Handbook of Terminal Planning*, J. W. Böse Ed. Cham: Springer International Publishing, 2020, pp. 125-154.

[9] R. Chen, Q. Meng, and P. Jia, "Container port drayage operations and management: Past and future," *Transportation Research Part E: Logistics and Transportation Review,* vol. 159, p. 102633, 2022/03/01/ 2022.

[10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature,* vol. 323, no. 6088, pp. 533-536, 1986.

[11] Y. Lecun and L. Bottou, "Gradient-based learning applied to document recognition," *P IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[12] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," *Neural Comput,* vol. 18, no. 7, pp. 1527-1554, 2006/07/01 2006.

[13] Z. Yang and E. E. Mehmed, "Artificial neural networks in freight rate forecasting," *Maritime Economics & Logistics,* vol. 21, no. 3, pp. 390-414, 2019.

[14] X. Li *et al.*, "Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation," *Environ Pollut,* vol. 231, pp. 997-1004, Dec 2017.

[15] S. Gao *et al.*, "Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation," *J Hydrol (Amst),* vol. 589, Oct 2020, Art no. 125188.

[16] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling," 12/11 2014.

[17] D. Zhang, G. Lindholm, and H. Ratnaweera, "Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring," *J Hydrol (Amst),* vol. 556, pp. 409-418, Jan 2018.

[18] Z. Li, Z. Zheng, and S. Washington, "Short-Term Traffic Flow Forecasting: A Component-Wise Gradient Boosting Approach With Hierarchical Reconciliation," *IEEE T Intell Transpt,* vol. PP, pp. 1-13, 10/25 2019.

[19] Y. S. Lv, Y. J. Duan, W. W. Kang, Z. X. Li, and F. Y. Wang, "Traffic Flow Prediction With Big Data: A Deep Learning Approach," (in English), *IEEE Trans. Intell. Transp. Syst.*, Article vol. 16, no. 2, pp. 865-873, Apr 2015.

[20] H. F. Zheng, F. Lin, X. X. Feng, and Y. J. Chen, "A Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction," *IEEE Trans. Intell. Transp. Syst.,* vol. 22, no. 11, pp. 6910-6920, Nov 2021.

[21] H. Q. Wang, R. Q. Zhang, X. Cheng, and L. Q. Yang, "Hierarchical Traffic Flow Prediction Based on Spatial-Temporal Graph Convolutional Network," *IEEE Trans. Intell. Transp. Syst.*, to be published. doi: 10.1109/tits.2022.3148105.

[22] V. F. Arguedas, G. Pallotta, and M. Vespe, "Maritime Traffic Networks: From Historical Positioning Data to Unsupervised Maritime Traffic Monitoring," (in English), *IEEE T Intell Transpt,* Article vol. 19, no. 3, pp. 722-732, Mar 2018.

[23] X. Yang, Q. Xue, M. Ding, J. Wu, and Z. Gao, "Short-term prediction of passenger volume for urban rail systems: A deep learning approach based on smart-card data," *Int J Pord Econ,* vol. 231, p. 107920, 2021/01/01/ 2021.

[24] K. P. Zhang, Z. J. Liu, and L. Zheng, "Short-Term Prediction of Passenger Demand in Multi-Zone Level: Temporal Convolutional Neural Network With Multi-Task Learning," (in English), *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1480-1490, Apr 2020.

[25] Z. Y. Cheng, J. Lu, H. J. Zhou, Y. B. Zhang, and L. Zhang, "Short-Term Traffic Flow Prediction: An Integrated Method of Econometrics and Hybrid Deep Learning," *IEEE Trans. Intell. Transp. Syst.*, 2021.

[26] Z. Yang, g. Chen, and R. R. Moodie, "Modeling Road Traffic Demand of Container Consolidation in a Chinese Port Terminal," *J Transp Eng,* vol. 136, no. 10, pp. 881-886, 2010.

[27] Guo, Zijian, Fang, Zexing, Wang, and Wenyuan, "Dynamic traffic simulation model for container terminal based on random arrival of vessel," *Port & Waterway Engineering,* vol. 05, no. v.43;No.202, pp. 14-18, 2017.

[28] H. Al-Deek, G. Johnson, A. Mohamed, and A. El-Maghraby, "Truck Trip Generation Models for Seaports with Container and Trailer Operation," *Transport Res Rec,* vol. 1719, pp. 1-9, 2000.

[29] Al-Deek and Haitham, "Which Method Is Better for Developing Freight Planning Models at Seaports—Neural Networks or Multiple Regression?," *Transport Res Rec,* vol. 1763, pp. 90-97, 2001.

[30] Al-Deek and Haitham, "Use of vessel freight data to forecast heavy truck movements at seaports," *Transport Res Rec,* vol. 1804, pp. 217-224, 2002.

[31] S. Pradeep, A. D. Haitham, K. Jack, and A. Georgios, "Evaluation of Two Modeling Methods for Generating Heavy-Truck Trips at an Intermodal Facility by Using Vessel Freight Data," *Transport Res Rec,* vol. 1906, pp. 113-120, 2005.

[32] Y. Xie and N. Huynh, "Kernel-Based Machine Learning Models for Predicting Daily Truck Volume at Seaport Terminals," *J Transp Eng,* vol. 136, no. 12, pp. 1145-1152, Dec 2010.

[33] S. Zhang, M. Peng, H. Du, and Z. Cai, "Discussion on New Kind of T-shaped Interchange Design," *Transportation Science & Technology,* no. 5, pp. 148-151, 2014.

[34] I. Kourounioti, A. Polydoropoulou, and C. Tsiklidis, "Development of Models Predicting Dwell Time of Import Containers in Port Container Terminals – An Artificial Neural Networks Application," in *Transportation Research Procedia*, vol. 14, 2016, pp. 243-252.

[35] Y. P. Gao, D. F. Chang, T. Fang, and Y. Q. Fan, "The Daily Container Volumes Prediction of Storage Yard in Port with Long Short-Term Memory Recurrent Neural Network,", *J Adv Transport,* Article vol. 2019, p. 11, Dec 2019.

[36] A. Nadi, S. Sharma, M. Snelder, T. Bakri, H. van Lint, and L. Tavasszy, "Short-term prediction of outbound truck traffic from the exchange of information in logistics hubs: A case study for the port of Rotterdam," *Transport Res C-Emer,* vol. 127, Jun 2021.

[37] D. Svozil, V. Kvasnicka, and J. í. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometer Intell Lab,* vol. 39, no. 1, pp. 43-62, 1997/11/01/ 1997.

[38] L. Wu, Y. Wang, X. Li, and J. Gao, "Deep Attention-Based Spatially Recursive Networks for Fine-Grained Visual Recognition," *IEEE Trans Cybern,* vol. 49, no. 5, pp. 1791-1802, 2019.

[39] X. Yuan, L. Li, and Y. Wang, "Nonlinear Dynamic Soft Sensor Modeling With Supervised Long Short-Term Memory Network," *IEEE Trans Industr Inform,* vol. PP, pp. 1-1, 02/28 2019.

[40] D. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *ICLR 2015*. Ithaca, New York: arXiv.org, 2015, p. 13.

[41] N. Ketkar, "Introduction to PyTorch," in *Deep Learning with Python: A Hands-on Introduction*, N. Ketkar Ed. Berkeley, CA: Apress, 2017, pp. 195-208.

[42] R. Tibshirani, "Regression shrinkage and selection via the Lasso," in *Journal of the Royal Statistical Society Series B-Methodological*, vol. 58, no. 1): Wiley, 1996, pp. 267-288.

[43] G. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *arXiv preprint,* vol. arXiv, 07/03 2012.

[44] R. Steel and J. H. Torrie, "Principles and procedures of statistics a biometrical approach," in *International Biometric Society*, vol. 37: McGraw-Hill, 1981, pp. 859-860.

[45] A. Swami and R. Jain, "Scikit-learn: Machine Learning in Python," *J Mach Learn Res,* vol. 12, no. 10, pp. 2825-2830, 2013.

[46] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Stat Comput,* vol. 14, no. 3, pp. 199-222, 2004.

[47] D. Xu and Y. J. Shi, "A Combined Model of Random Forest and Multi-layer Perceptron to Forecast Expressway Traffic Flow," in *7th IEEE International Conference on Electronics Information and Emergency Communication (ICEIEC)*, Macau, PEOPLES R CHINA, Jul 21-23 2017, in IEEE International Conference on Electronics Information and

Emergency Communication, 2017, pp. 448-451. Available: <Go to ISI>://WOS:000414279500104. [Online].

[48] S. H. Huang, Z. H. Niu, J. X. Huang, and Ieee, "Road Traffic Congestion Prediction Based on Random Forest and DBSCAN Combined Model," in *5th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, Zhangjiajie, PEOPLES R CHINA, Jun 13-14, 2020, pp. 323-326.

[49] X. Han and Y. J. Shi, "Online Traffic Congestion Prediction Based on Random Forest," in *4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering (ICMMCCE)*, Xian, PEOPLES R CHINA, Dec 12-13 2015, vol. 39, in ACSR-Advances in Comptuer Science Research, 2015, pp. 2683-2689. [Online]. Available: <Go to ISI>://WOS:000373157302141.

[50] A. Graves, A. R. Mohamed, G. Hinton, and Ieee, "Speech recognition with deep recurrent neural networks," in *2013 Ieee International Conference on Acoustics, Speech and Signal Processing*, ICASSP, 2013, pp. 6645-6649.

[51] P. Huang *et al.*, "Modeling train operation as sequences: A study of delay prediction with operation and weather data," *Transp. Res. Pt. e-Logist. Transp. Rev.*, vol. 141, Sep 2020.

[52] S. Chen and L. Ge, "Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction," *Quantitative Finance*, vol. 19, no. 9, pp. 1507-1515, Sep 2019.

[53] X. L. Ma, Z. M. Tao, Y. H. Wang, H. Y. Yu, and Y. P. Wang, "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data," *Transp Res Part C Emerg Technol*, vol. 54, pp. 187-197, May 2015.

[54] K. Zhang, Z. Liu, and L. Zheng, "Short-Term Prediction of Passenger Demand in Multi-Zone Level: Temporal Convolutional Neural Network With Multi-Task Learning," *IEEE T Intell Transpt*, vol. 21, 04/08 2019.

[55] C. Caballini, M. D. Gracia, J. Mar-Ortiz, and S. Sacone, "A combined data mining - optimization approach to manage trucks operations in container terminals with the use of a TAS: Application to an Italian and a Mexican port," *Transp. Res. Pt. e-Logist. Transp. Rev.*, vol. 142, Oct 2020.

Na Li received the B.E. and M.E. degrees in transportation planning and management from Dalian Maritime University, Dalian, China, in 2003 and 2006. She got her Ph.D. degree in logistics engineering and management from Dalian Maritime University, China, in2011.

From 2012 to 2015, she conducted her postdoctoral research at the Dalian University of Technology, China. In 2016, she was visiting Dr. Wayne K. Talley at Old Dominion University, VA, USA. She is an associate professor at the College of Transportation Engineering, Dalian Maritime University. Her research findings have been published in leading journals including TRD. Her research interests include container terminal operation and optimization, truck arrival management, and prediction.

Dr. Li is a member of the International Association of Maritime Economists (IAME).

Haotian Sheng was born in Ningxia province, China in 1997. He received a Bachelor of Engineering degree in logistics engineering from Dalian Maritime University, Dalian, Liaoning, China in 2019. He is currently pursuing a Master's degree in Logistics engineering and management at Dalian Maritime University, Dalian, Liaoning, China. His research interests include traffic flow modeling for container terminals based on the method of machine learning.

Pingyao Wang was born in Xiantao City, Hubei Province on January 1, 1997. He obtained a B.E. degree in logistics engineering from Dalian Maritime University in 2019 and has been studying for an M.E. degree in transportation since 2020.

Since 2015, he has been engaged in the research of logistics and transportation and has won many awards in transportation competitions and a patent in logistics. In 2019, he participated in the summer school organized by Tsinghua University and Munich University of technology to learn the frontier knowledge of transportation. His interests include machine learning in the field of terminal traffic.

Yulin Jia was born in Shuozhou, Shanxi Province, China in 1996. He received a B.M. degree in logistics management from the Taiyuan University of Technology. He is now pursuing an M.E. degree in transportation engineering at Dalian Maritime University. His main research interests are traffic flow modeling.

Zaili Yang received his BEng in Maritime Transportation from Dalian Maritime University, China, in 2001, MSc in International Transport from Cardiff University, UK, in 2003, and Ph.D. in Maritime Safety from LJMU, UK, in 2006.

He is a Professor of Maritime Transport at Liverpool John Moores University (LJMU), UK. His research interests are the analysis and modeling of safety, resilience, and sustainability of transport networks, particularly maritime and logistics systems.

He has published more than 300 refereed papers in risk and supply chain areas, including 3 books and140 SCI-cited journal papers. Prof. Yang is an Associate Editor/Editorial Board Member of 9 transport/maritime journals (e.g. Transport Research Part E).

Zhihong Jin was born in Shenyang, Liaoning, China in 1963. He received his B.E. and M.E. degrees in system engineering from the Northeast University of China, in 1984 and 1987, respectively. He received his Ph.D. degree in system engineering from Nagoya Institute of Technology, Japan, in 2000.

From 1987 to 1989 and from 1990 to 1994, he was a Research Assistant and Lecturer at Northeast University. From 1994 to 1996 and from 2000 to 2003, he was a Visiting Scholar and Assistant Professor at Nagoya Institute of Technology, respectively. Since 2003, he has been a Professor at Dalian Maritime University. He was the dean of Transportation Management College from 2014 to 2020.

His research interests are logistics system planning and management, transportation planning and management technology, and supply chain design and management. He has published 4 books and more than 160 papers and obtained 6 projects from the National Natural Science Foundation of China (NSFC). He also is a member of the Teaching Guiding Committee of the China Ministry of Education, an editorial board member of the International Journal of Shipping and Transport Logistics, and the Editor-in-Chief of the Journal of Dalian Maritime University.