Prediction of truck turnaround time based on machine learning approach

A case study at Port of Rotterdam

TIL5060: TIL Thesis

Haoyang Chen

Delft University of Technology

Prediction of truck turnaround time based on machine learning approach

A case study at Port of Rotterdam

by

Haoyang Chen

Chair supervisor:	M. Saeednia
Supervisor:	F. Schulte
Project Duration:	Apr, 2024 - Feb, 2025
Faculty:	Faculty of Civil Engineering and Geosciences, Delft

TUDelft

Preface

As this thesis is submitted, my two-year journey at TU Delft is also drawing to a close. Over two years ago, I left my hometown for the first time, setting foot on this land with both anticipation and trepidation, embarking on an unknown journey. Looking back, this journey has not only significantly enhanced my academic capabilities but also allowed me to explore the world outwardly while reflecting inwardly. In the days ahead, I hope to continue pursuing my passion for the transportation, logistics, and maritime, whether in academia or industry, and to strive toward becoming a better version of myself.

I would like to extend my heartfelt gratitude to everyone who has supported and helped me throughout my master's journey and the completion of this thesis. Without your guidance and encouragement, I would not have made it this far on my own. First and foremost, I want to thank my supervisors, Dr. Mahnam Saeednia and Dr. Frederik Schulte, for their unwavering support and invaluable academic insights. I am equally grateful for the positive research environment you fostered, and your assistance that enables me to improve the quality of my thesis. I would also like to thank the professionals from the Port of Rotterdam for their generous contributions and insights.

Furthermore, I am deeply thankful to my family, classmates, and friends for accompanying me through both the highs and lows. Your companionship, support, and encouragement have been my greatest source of motivation. During every joyous or challenging moment I experienced in the Netherlands, I am glad to have had you by my side.

Although I do not yet know where the future will lead, I hope to carry forward the knowledge and skills I have acquired during my graduation project and academic journey, and make meaningful contributions to the logistics and maritime industries.

Haoyang Chen
Delft, February 2025

Summary

Container terminals are vital hubs in global trade, facilitating the seamless transfer of goods between maritime and inland transportation networks. Truck turnaround time serves as a critical performance metric for container terminals, provides direct feedback on port congestion and efficiency. This study focuses on developing a predictive framework for truck turnaround time (TTT) at the Port of Rotterdam by integrating multi-source data and employing advanced machine learning techniques.

Previous studies on truck turnaround time prediction have largely relied on limited datasets and methodologies, such as utilizing historical truck arrival flows or terminal operation logs and using statistical methods. This research employs diverse datasets, including Bluetooth detection records, container arrival information, and environmental condition. By combining these data sources, a harmonized dataset was constructed to represent the complexities of port operations. A stacked Long Short-Term Memory (LSTM) network was employed as the predictive model, utilizing its ability to capture temporal dependencies and nonlinear interactions between variables. This approach allows for more comprehensive and accurate TTT predictions compared to conventional methods.

To process the noisy and incomplete Bluetooth data, a robust trip identification pipeline was developed. The pipeline employed spatial clustering, temporal filtering, and dual verification to accurately identify container truck trips, achieving an accuracy exceeding 90%. Using this processed data, the stacked LSTM model demonstrated superior predictive performance, effectively capturing periodic trends and long-term dependencies. Benchmarking results showed that the stacked LSTM outperformed traditional methods, including Random Forest and XGBoost. Sensitivity analysis highlighted the critical role of truck arrival flows and wind conditions in determining truck turnaround time variability.

In summary, this study provides a novel and scalable framework for TTT prediction, integrating multi-source data and advanced modeling techniques to address key limitations of existing approaches. The findings offer actionable insights for optimizing terminal operations and reducing congestion. Future research could focus on expanding data sources, enhancing model interpretability, and validating the framework across diverse port environments to ensure broader applicability.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Maritime transport has long been the backbone of the global transportation network. With containerized shipping becoming the standard model for exchanging goods in maritime transportation worldwide, the increased volume of container transportation leads to a surge in container port activities. However, the growth has not come without its challenges. One of the most pressing issues facing container ports is the congestion of container trucks, which has become a bottleneck in the logistics chain, impeding the flow of commerce and affecting the overall efficiency of port operations. The queuing up idle trucks at terminal gates have caused further congestion upstream and also extra emissions, costs and delays [44, 49].

The issue of gate congestion at container terminals is indeed a significant challenge that has been extensively documented. Namboothiri and Erera [37] noted the decrease of drayage operation efficiency caused by gate congestion. When container trucks arrive during peak hours, this generally results in longer total turnaround times within the port. Du et al. [20] defined the turnaround time of container trucks in a container terminal from different perspectives. For an external truck itself, the turnaround time refers to the time between entering the port entrance and departure from the exit gate. Fig 1.1 shows how truck turnaround time in a container terminal is calculated. However, due to the different interest of stakeholders, there exist various resources of data regarding the operation time of trucks in container terminals. For example, terminal operators mostly concern the operation time in yard, and truck operators would start to calculate turnaround time when the truck enters the special road connecting the container terminal.

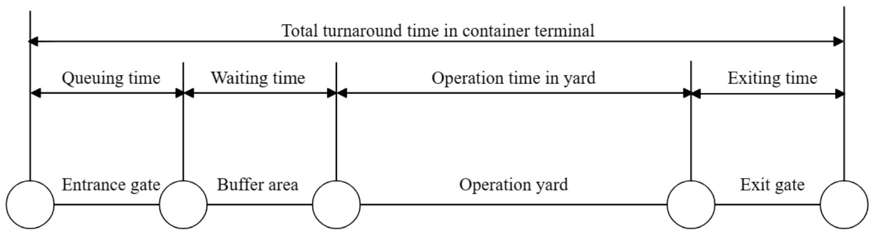


Figure 1.1: Truck turnaround time in a container terminal

For ports, unexpected truck congestion and extended truck turnaround times would prevent the yard from making full use of operational capacity to load and unload containers [52]. For trucks, both operators and drivers do not expect to be affected by the extended turnaround time which may lead to increased idle time and decreased fleet utilization. Therefore, predicting and optimizing truck turnaround time at container terminals has become a focus of academics and industry.

Many papers analyze the factors affecting turnaround time and propose various suggestions to predict and reduce turnaround time. In terms of turnaround time prediction, several methods such as factor analysis, system dynamic (SD), machine learning and so on are used to predict truck turnaround time. Du et al. [21] introduced SD method to build a simulation model of truck operation system in container terminals, and considered the impact of multiple state variables to predict turnaround time. Van der Spoel et al. [50] proposed a prediction model using both regression and classification methods, and established a benchmark to evaluate the prediction results. Sidarta [46] adopted multiple machine learning methods to predict truck cycle time in earthworks. In reducing turnaround time, research on truck appointment system (TAS) dominates the mainstream. Torkjazi et al. [48], Namboothiri and Erera [37] and Zhang et al. [60] employed various methods to build and optimize the TAS to reduce the turnaround time of external trucks at the gate and yard. In addition, research represented by Chen et al. [15] considers using time-varying tolls to control arrival traffic at the terminal gate, but there are few practical applications to prove its effectiveness.

However, few studies have considered in-port operations when predicting and optimizing container truck turnaround times. Most of the TAS models developed only concern quote assignment of time intervals for container trucks, assuming that the basic operating time of container trucks within the terminal is consistent across time periods. In contrast to this, other factors including the arrival of containers and weather conditions may also affect the turnaround time of trucks within the terminal. Similarly, in terms of turnaround time prediction, there is few literature that take container arrival information into account. In terms of data, the studies above basically considered container truck arrival data from the entry gate, and did not explore other data sources.

1.2 Research objectives

Considering the limitations and research problems proposed above, the main objective and contribution of this research is to develop a prediction model of truck turnaround time based on historical truck arrivals and departures identified from the data sets, considering container arrival information and other factors like weather or traffic condition around the port.

In the meantime, several sub-objectives are proposed to support the main objective as follows:

- To identify the truck tours from Bluetooth detections around the port.
- To identify the influencing factors in truck turnaround time considering data availability.
- To analyze the relationship between environmental variables and truck turnaround time.
- To develop and compare different predictive models and methodologies for truck turnaround time.
- To create an automated prediction tool that can provide real-time insights for port authorities, terminal operators, and truck drivers.

For the second sub-objective, the important factors are determined by data availability and sensitivity analysis, such as traffic information and wind condition. The influencing factors will be investigated through literature review. The factors will be used to determine data input, variables, assumptions, and limitations of the model.

1.3 Research questions and scope

This research attempts to find ways to predict a truck's turnaround time at terminals based on container information, truck arrival information and external factors, as this will enable both terminal operators and truck operators to coordinate and enhance their management efficiency. The predicted turnaround time would be a supplement to the truck appointment system and portbase information platform in Port of Rotterdam, and make truck operators aware of expected container truck operating status. To achieve above goals, the following research questions are proposed.

How can a holistic model be built to predict truck turnaround time at terminals based on the data available?

The main research question is jointly answered by the following sub-questions:

- How to identify container truck trips from the Bluetooth detection records around the port area?
- Which factors are crucial in predicting truck turnaround time and also available in the research scope?
- How can container information and Bluetooth data sets be processed into a finalized dataset as the input to the model?
- What kind of approach can achieve higher prediction accuracy based on a fixed level or comparing results?
- How can the prediction results be verified and validated?

The project focuses on developing a truck turnaround time prediction model that provides real-time predictions for both port authorities, terminal operators, and truck drivers. The research scope is limited due to the availability of datasets. Specifically, this study focuses on the main container terminals of the Port of Rotterdam in 2017, including the 5 container terminals located in Maasvlakte I and II.

1.4 Research framework

The research framework describes the approach and methodology for developing a predictive model for the truck turnaround time in container terminals. This section will detail the research design, data collection methods, model development processes, and validation techniques.

The research is structured to systematically address the key objectives and questions outlined in the previous sections. The design includes a combination of quantitative data analysis, model development, and validation stages to ensure robust and reliable outcomes.

Firstly, a literature review is conducted in Chapter 2 to summarize similar research and provide a comprehensive overview of existing studies related to the topic. The main aspects are defined and investigated within the scope of the research, and research gaps are identified to establish the need for this study.

Then, Chapter 3 establishes the research scope by defining the physical and operational boundaries of the container terminal and analyzing port productivity indicators to identify factors influencing truck turnaround time and potential areas for operational improvement.

Chapter 4 outlines the methodology, detailing the techniques used in this study, including data collection and preprocessing, vehicle trip identification via Bluetooth detection, development of predictive models, introduction of the LSTM model, formalization of model inputs, and processes for model training, evaluation, and post-processing.

Chapter 5 data collection and processing for relevant activities are a critical component of this research. The primary data sources include historical truck arrival and departure data derived from Bluetooth sensors around the port area, which could be used to calculate the turnaround times and traffic flow of trucks. Container arrival information and weather conditions are also included, which impact the operational situation and scheduling at the terminals. Data will be provided by the Port of Rotterdam, also collected from publicly available databases from KNMI. The collected data will undergo thorough preprocessing to ensure accuracy and completeness. Key steps include:

- Data cleaning: Removing duplicates and correcting inconsistencies in the dataset.
- Feature extraction: Determine truck tours and other types of vehicles based on the Macaddress and behavior of vehicles. Identifying and extracting relevant features that influence truck turnaround time, such as traffic flow, average speed, congestion index, and weather conditions.
- Time alignment: Handle time series variables by decomposing trends, seasonality, and residual components and converting them into appropriate time intervals. Ensuring that all data points are time-aligned to facilitate accurate analysis and model training.
- Data analysis: To explore data features, descriptive and exploratory statistical analysis will be employed to understand the data distribution, identify patterns, and determine correlations between different variables.

Chapter 6 involves the design and testing of different LSTM model architectures to obtain better performance using the same data set input. To ensure the reliability of the predictive models, rigorous validation and testing will be conducted. After the prediction model is developed, a benchmark would be implemented to evaluate the prediction accuracy or prediction results of the model, compared with other approaches or a given level. This would be part of the evaluation and validation process.

Finally, chapters 7 and 8 provide a discussion of the work together with the conclusions, recommendations, limitation and ideas for future research.

Chapter 2

Literature Review

As mentioned previously, the studies on truck turnaround time mainly focus on prediction and optimization. This literature review aims to examine existing research on Bluetooth trip identification and truck turnaround time prediction and optimization. Then key limitations and challenges are discussed to delineate the research gaps that this study seeks to address.

2.1 Research on truck turnaround time

Truck appointment system is a two-dimensional decision-making system that relies on both space and time, which is used to optimize truck turnaround time and operation cost. Murty et al. [36] were the first research team to explore the arrival time schedule of external trucks, and they developed a TAS system used by Hong Kong International Terminals. With the development of research in this area, some studies began to consider the impact of truck operating hours and terminal appointment time windows on truck operating costs when designing TAS. Research results from Zhao and Goodchild [61] show that the application of TAS can reduce container truck congestion at the terminal and the arrival information of trucks can reduce rehandles at the yard. Zhang et al. [59] developed a truck appointment model using a Baskett–Chandy–Muntz–Palacios (BCMP) queuing network to describe truck activities at the gate or yard. Numerical results shows that the network could help reduce turnaround time. Phan and Kim [40] developed a coordinated solution between hauling companies and TAS to avoid the negative impact of TAS on trucking company operations. They take into account the truck's work schedule to reduce the overall truck turnaround time. Schulte et al. [43] developed a graph-based mathematical model based on the m-TSPTW. Their model optimizes travel costs and emissions for tasks that can be performed by multiple trucks by collaboratively merging tasks, and also alleviates terminal congestion. Torkjazi et al. [48] considered truck tour when designing TAS, and formulated and solved the TAS problem as a mixed integer nonlinear problem (MINLP). However, even when they introduced a hinterland network for truck tour, they didn't take into account possible traffic congestion within the network.

Another approach for managing truck arrivals and reduce truck turnaround time is called vessel-dependent time windows (VDTWs). Yang et al. [56] found that the distribution of truck arrivals with outbound containers could be described using a Beta distribution within a time window based on vessel-calling schedule. Chen and Yang [13] developed a heuristic algorithm to find optimal time window for external trucks to reduce the total cost of gate congestion. Chen and Jiang [12] proposed an approach to manage truck arrival time window based on truck-vessel service relationship. This point of view considers the corresponding relationship between trucks and vessels, and the operating conditions in the port have been taken into consideration, but the impact of trucks being affected by congestion at gates and external roads has not been considered.

There are various directions of prediction regarding terminal operation and truck turnaround time with a focus on predicting methods and influencing factors. Traditional turnaround time prediction systems used in terminals are merely based on the historical truck flows and empirical experience from employees. However, numerous factors can make such predictions inaccurate: port operations, weather conditions, political concern, etc.

In the field of port operations, more research focuses on time-related indicators of vessels. Numerous studies aim to predict estimated vessel arrival time (ETA), vessel turnaround time (VTT) and departure time. Chu et al. [16], Yu et al. [58] both developed evaluation and prediction methods for estimated arrival time of vessels. Chu et al. [17], Filom et al. [22] both predict vessel turnaround time and departure time based on static ship data. Chu et al. [17] developed an XGBoost regression model to enhance the precision of VTT predictions. Abreu et al. [1] proposed a decision tree model to predict vessel turnaround time considering cargo and port information. These time-related indicators focus on similar but different influencing factors than truck turnaround time.

Other studies pay attention to road arrival or turn around time prediction in different scenarios. Sjoerd van der Spoel and van Hillegersberg [47] investigated factors affecting truck arrival times at distribution centers. Du et al. [20] used SD method to predict truck operation time in terminals, but only considering input truck flow and yard capacity. Wang et al. [52] took container information, weather condition and traffic condition into account and used a combination of data mining methods to predict container arrival time at terminals. In addition, Table 2.1 discussed possible influencing factors in time prediction covered in different studies. The listed influencing factors provide possible options to be dealt with as the input of prediction model.

Although many studies have made progress in truck turnaround time prediction and optimization, there is a research gap that needs to be addressed. While previous studies have made significant contributions to understanding and managing truck turnaround time, there is a lack of comprehensive models that consider the holistic operation of container terminals, including in-port operations and external factors. Specifically, few studies have incorporated container arrival information, weather conditions, and traffic conditions into truck turnaround time prediction models. Additionally, existing TAS models often assume consistent truck operating times within the terminal across different time periods, neglecting the potential impact of various factors on in-port operations.

2.2 Trip identification based on Bluetooth data

Bluetooth technology has increasingly been adopted in traffic management due to its ability to collect real-time, cost-effective data that is useful for various traffic analysis tasks. Bluetooth sensors detect devices by capturing Media Access Control (MAC) addresses, which can provide time-stamped data as vehicles pass through specific checkpoints. These data points can generate travel time estimates, congestion metrics, and vehicle flow analyses across monitored areas[5].

However, the data quality obtained through Bluetooth sensors has been widely discussed. A major limitation highlighted across studies is data sparsity and noise, particularly under mixed traffic conditions. Araghi et al. [5] noted that large detection zones can introduce multiple detection events, leading to potential ambiguity in vehicle positioning. Margreiter et al. [32] pointed out that not all vehicles carry Bluetooth-enabled devices. The limited detection rate reduces the availability of Bluetooth data, and may also introduce a sampling bias where the detected data may not fully represent the entire traffic flow.

To address these challenges, researchers have proposed various methodologies to improve the accuracy and utility of Bluetooth data for trip identification, path estimation, and vehicle type recognition. Araghi et al. [5] applied a classification method based on the device type and signal strength to differenciate travel modes. Araghi et al. [6] employed a classification method to distinguish between motorized vehicles and bicycles based on estimated mode-specific travel times. Bachmann et al. [7] explored data fusion between Bluetooth traffic monitoring system and loop detectors to improve freeway speed

Table 2.1: Influencing factors involved in each paper

Reference	Year	Method	Influencing factors						
			Weather	Traffic Condition	Time period	Cumulative Flow	Cargo info	Vehicle/vessel info	Port congestion
Wu et al. [55]	2004	SVR	√	√		√			
Hollander and Liu [27]	2008	Microsimulation	√			√			
Khosravi et al. [28]	2011	Genetic algorithm	√		√	√			
Lederman and Wynter [31]	2011	Data expansion	√	√					
Sjoerd van der Spoel [47]	2017	Data mining	√			√			
Abualhaol et al. [2]	2018	LSTM						√	
Balster et al. [9]	2020	ML	√		√	√	√		
Pruyn et al. [41]	2020	MCA				√			
Antamis et al. [4]	2021	ML regression			√		√		
Wang et al. [52]	2023	Data mining	√		√	√	√		
Du et al. [20]	2022	System dynamics				√			√
Wenhao Peng and Wu [53]	2023	LSTM						√	√
Chu et al. [16]	2023	Random forest						√	
Chu et al. [17]	2024	XGBoost			√			√	
This study		LSTM	√	√		√		√	

estimation. Garrido-Valenzuela et al. [23] presents a Bayesian inference-based methodology to address the issue of missed Bluetooth detections in route choice modeling. Sharma et al. [45] designed a method for clustering based on trip duration time to identify vehicle types, and proposed a bi-objective optimization model to estimate route choices for truck drivers based on sparse Bluetooth data and loop-detector data. Crawford et al. [18] captured repeated trip behaviors from Bluetooth data to class road users and measured spatial similarity using Sequence Alignment.

In summary, these studies collectively highlight the advancements and methodologies developed to enhance the use of Bluetooth technology in traffic analysis, especially in the context of trip identification and vehicle type recognition. These studies support the research by providing tools to improve Bluetooth data quality and identify truck trips, and hence help to predict truck turnaround time more reliably and realistically.

2.3 Port operation and congestion estimation

Huge port congestion is always affecting the whole global supply chain. In major ports that deal with a huge volume of vessels, containers, and bulk cargo every day, port congestion can happen in various places such as terminal gates, container yards and berths. Once port congestion happens, significant time loss and inconvenience could be observed at terminals, gates and/or the hinterland[34]. Port congestion estimates can help terminal operators and other stakeholders properly allocate existing resources and assess the need for infrastructure upgrades in the long term. Therefore, it is crucial to make precise predictions of port congestion for port operator participants.

Some of the existing studies measuring port congestion rely on the data from industry or port authorities. Leachman and Jula [29] built a queuing theory model to estimate the container flow times among loading drays and on-dock rail cars. Yeo et al. [57] explored the port congestion mechanism of Busan Port using flow estimation and sea port traffic simulation. Chen et al. [14] discovered that port performance indicators, including berth utilization and container turnaround time, are found to be directly related to the number of container vessel arrivals and the number of containers handled at each vessel. The studies above mostly uses exclusive data that is heterogenous across ports, thus it is often only applicable to the case under study and does not have universal applicability.

Other recent studies begin to employ machine-learning based technologies and publicly available data sets such as AIS data to forecast port congestion [35]. AbuAlhaol et al. [3] first introduced port congestion measuring using Automatic Identification System (AIS) data. Based on the versatility of AIS data, Peng et al. [39] used the LSTM model to evaluate and predict the congestion levels of different ports around the world. Later, they developed a density-based spatial clustering algorithm to automatically identify vessels and calculate port turnaround times[8].

Overall, existing studies on port congestion prediction are partly based on localized heterogeneous datasets, while others assess port congestion using publicly available AIS data to reflect information on vessel arrivals and vessel density. However, these methods mostly consider congestion related to quay cranes and yard cranes, primarily using data from the vessel side. In contrast, this study needs to comprehensively consider container arrival information at the terminal and truck arrival information at the port gate, evaluating port congestion from both aspects and using this as input for predictions.

2.4 Conclusion

This literature review discussed the existing body of research on truck turnaround time and port congestion. However, there still remains significant research gaps between the studies and this project, particularly in the integration of various influencing factors and the holistic consideration of both in-port and external conditions.

Firstly, while truck appointment systems have been extensively studied and applied to equalize truck traffic and reduce turnaround time, most of the studies fail to account for the dynamic nature of port operation. The data used was also limited to truck arrival data from terminal gates, and failed to expand new data sources. Secondly, the prediction models for truck and vessel turnaround times often focus on isolated factors, such as historical traffic flows, or vessel-specific data like ETA and VTT. Such a narrow focus makes these studies in many cases only able to obtain information from the changing characteristics of a single data, making it difficult to explore the regularity behind the changes. Furthermore, the research on port congestion primarily concentrates on vessel-related operations, utilizing AIS data or other data sources to forecast congestion at the quay and yard levels. Few studies in this area have considered integrated operations from both vessel and landside. Lastly, regarding Bluetooth trip identification, although existing research has made progress in road trip and vehicle type identification, there has been no research on identifying vehicle entry and exit trips based on Bluetooth data in a certain area, which could propose more challenge in such situations.

To bridge this research gap, the proposed research aims to develop a prediction model for truck turnaround time that incorporates historical truck arrivals and departures, along with factors such as container arrival information, weather conditions, and traffic conditions around the port. The truck arrivals and departures would be determined from the Bluetooth detection records through trip identification and vehicle type identification methods. By considering a wide range of influencing factors, the proposed model seeks to provide a more accurate and comprehensive prediction of truck turnaround time at container terminals. Specifically, the research will leverage data from both the container side to describe terminal operations and the input volume of trucks, as well as external data sources such as weather and Bluetooth data to describe traffic conditions around the port.

Chapter 3

Problem identification

In this chapter, the demarcation of the container transport chain will be discussed to identify the research boundaries and identifying key performance indicators of truck turnaround time within the whole transport chain. This chapter will also discuss and analyze the relevant information and operational processes of the specific research scope involved in this project - the Port of Rotterdam.

3.1 Demarcation of the container turnaround process

3.1.1 Terminal structure description

A container terminal operates as an open material flow system where containers are processed through several stages, each involving specific operations and equipment. The terminal can be divided into three primary areas: quayside, yard, and landside operation zones. Each of these areas plays a distinct role in the handling of containers, and their performance directly impacts the overall efficiency of container movement.

In the inbound flow, containers arrive at the terminal via ships, where they are unloaded using quay cranes. These cranes transfer containers to Automated Guided Vehicles (AGVs), terminal trucks, or straddle carriers, which then move the containers to the yard. Efficient quayside operations ensure that container unloading is synchronized with yard availability, minimizing ship turnaround time (STAT) and preventing congestion at the quay.

The yard serves as an intermediary zone between quayside and landside operations. Inbound containers are stacked in the yard based on their priority for retrieval and further movement. Yard cranes manage the storage and retrieval of containers. The performance of the yard operations is influenced by the container dwell time (CDT), which measures how long a container stays in the yard, and yard occupancy rates, which indicate the level of congestion. Truck turnaround times are affected by yard operations and the resulting container re-handling. Reducing delays in retrieving containers from landside operations can effectively reduce truck turnaround times.

On the landside, containers are either received from or dispatched to trucks and trains through the terminal gates. The efficiency of this stage is measured by gate waiting time, which reflects the time trucks spend waiting for processing at the gate, and truck turnaround time, which measures the time from when a truck enters the terminal to when it exits after collecting or delivering a container. Effective landside operations depend on streamlined inspection, processing, and efficient truck scheduling systems such as Truck Appointment System. Trucks arrive at the terminal's in-gate where the data of the containers have to be checked and filed into the computer system or actualized in case of pre-advice. Trucks then drive to transition points where the containers are loaded or unloaded by internal equipment. Large container terminals serve some thousand trucks a day. Transition points are located either at the stack crane or inside the yard in case of straddle carrier operation. The arrival time of the trucks at the transition points cannot be precisely foreseen, i.e., transport jobs for the internal equipment cannot be released until the truck arrives at the transition point. Because of the permanently changing traffic volume, the turnaround time of trucks would fluctuate in various scenarios. The terminal structure is displayed as figure 3.1.

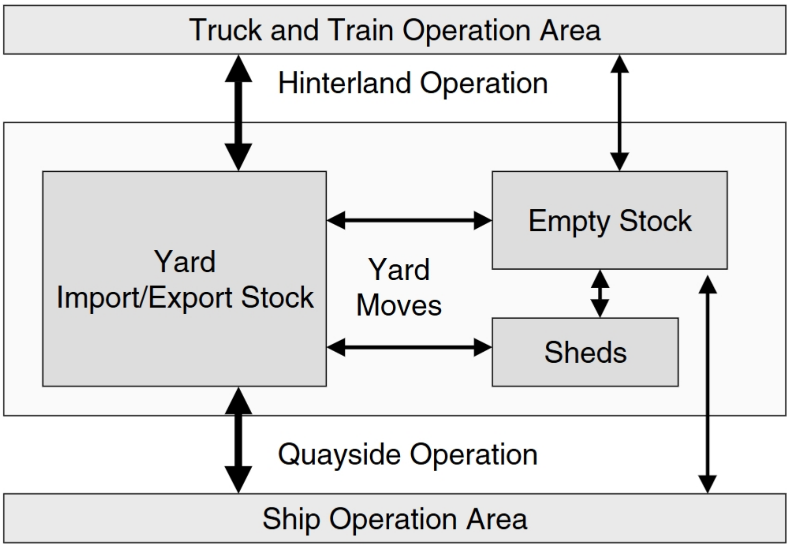


Figure 3.1: Container Terminal Structure[51]

3.1.2 General factors and KPI's within the process

The performance of a container terminal and its ability to manage truck turnaround time is influenced by several key factors. These factors are monitored through specific KPIs that measure operational efficiency at each stage of the terminal process.

- Quayside Operation KPIs. Ship turnaround time (STAT) reflects the time a vessel spends at the port, from arrival to departure. STAT is generally considered a crucial factor to characterize port congestion and port productivity. Gross crane productivity measures the number of containers moved per hour, which reflects the terminal operation capacity and has a strong relationship with container dwell time.
- Yard operation KPIs. Container dwell time indicates the average time containers spend in the yard. Longer container dwell time increases yard congestion and delays container retrieval, negatively affecting truck turnaround. Yard occupancy also reflects how the yard space utilized. Higher occupancy rates can slow down container handling operations, increasing both yard rehandling and truck awaiting times.
- Landside operation KPIs. Truck turnaround time refers to the total time a truck spends within the terminal from entry to exit. Gate waiting time measures the time trucks spend waiting to be processed at the terminal gates. Long gate waiting time can indicate inefficiencies in gate operations or high traffic volumes.
- External factors. Beyond terminal operation indicators, other factors may also influence truck turnaround time and port operations. Traffic conditions and congestion on roads may cause delays or synchronized arrivals of trucks, which may pose a challenge to the ability to pick up and deliver containers. Adverse weather can disrupt both quayside and yard operations, increasing ship turnaround times and delaying container handling. Severe wind conditions may even cause

quayside operations to cease or terminal shut down, which will also reflect on ship and truck turnaround time.

3.2 Case study background in the Port of Rotterdam

The Port of Rotterdam is Europe's largest seaport and one of the most crucial logistics hubs in the world. It plays a vital role in facilitating global trade, primarily serving as a gateway for goods entering and exiting the European continent. In 2023, the port handled approximately 13.4 million TEU (Twenty-foot Equivalent Unit) of container throughput, maintaining its leading position in Europe and around the world[42].

The Maasvlakte, a man-made extension of the Port of Rotterdam, carries the vast majority of container operations in the Port of Rotterdam today. Its development, divided into two phases-Maasvlakte I and Maasvlakte II-has played a crucial role in enhancing Rotterdam's status as Europe's leading container port.

The Maasvlakte area is home to several container terminals, including ECT Euromax, ECT Delta, APM Rotterdam, APM Maasvlakte II and Rotterdam World Gateway (RWG). Each terminal has different handling capacity and level of automation. Figure 3.2 shows the location and the distribution of these terminals within the Maasvlakte area. All these terminals are connected to the Dutch and European highway networks through direct access to major motorways such as A15. Additionally, each major terminal in Maasvlakte has dedicated rail terminals for container transport, but the proportion of containers transported by rail varies among the terminals. Barges also play a crucial role for container transport from Maasvlakte to hinterland areas. The development of multimodal transport in ports is conducive to coping with more complex transportation service requirements and reducing port carbon emissions, but it also places higher requirements on port operation levels. Maasvlakte's five container terminals and three hinterland transport modes reflect the complexity of port operations.

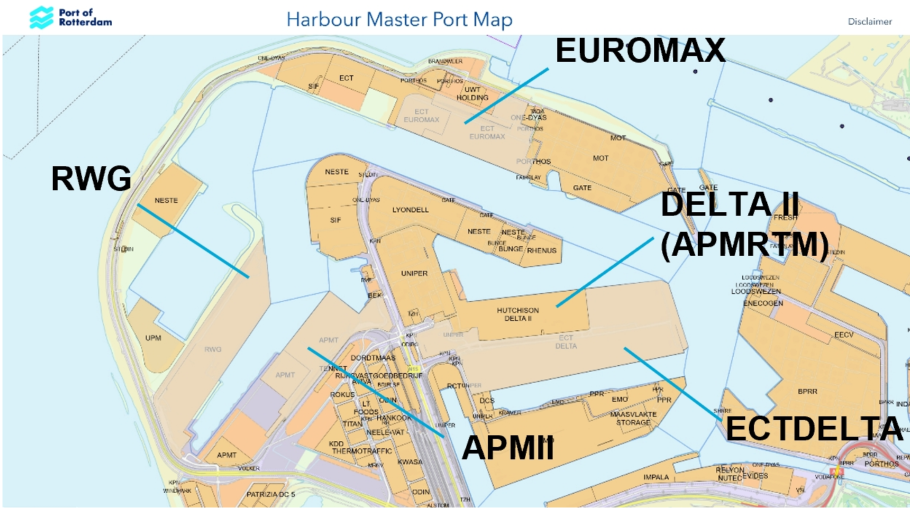


Figure 3.2: Maasvlakte map

As terminals like RWG and APM Maasvlakte II are equiped with advanced automation systems, they may share shorter turnaround times. The gate operation efficiency also differs in different terminals. Gate congestion frequently occurs during peak arrival hours, which introduce delays to the subsequent process. Similarly, the operational efficiency of the inland side of the terminal directly affects the turnaround time. Therefore, these influencing factors are strongly related to the terminals.

Hinterland connectivity is another critical determinant. The Maasvlakte area's connection to the A15 motorway provides essential access for trucks; however, road traffic congestion often disrupts synchronized arrivals at terminal gates, resulting in uneven demand. Multimodal transport options, including rail and barge, also influence turnaround time. Terminals with higher reliance on rail and barge can reduce pressure on road infrastructure.

Environmental and external factors also introduce additional complexity. Weather conditions, particularly severe wind conditions, frequently disrupt quayside crane operations, thus delaying container operations. Driven by irregular schedules of deep-sea vessels may create operational peaks and troughs on sea-side operations. It not only indicates the busyness of port operations, and also displays the peak load of trucks. Moreover, temporal patterns in daily and weekly truck flow show consistent fluctuations, with the highest variability observed during the start of the week and business hours.

This section provides a framework for selecting data features for the predictive model, ensuring the integration of key operational, environmental, and connectivity factors. It bridges theoretical models with practical realities, offering actionable insights for terminal operators and policymakers to enhance efficiency. By capturing the interplay between terminal operations, hinterland connectivity, and external influences, this section underlines the importance of a holistic approach to modeling and optimizing truck turnaround times in a modern port logistics system.

Chapter 4

Methodology

This chapter aims to define and describe the research trajectory followed in this study, as well as the potential research methods that may be employed. In order to reduce the complexity of the problem, the entire prediction problem is divided into several independent but connected tasks.

4.1 Data Collection and Analysis

This section aims at providing a comprehensive overview of the data collection and analysis process. This process is divided into three main parts: Dataset description, data pre-processing and data analysis.

4.1.1 Dataset Collection and Description

This section provides a detailed description of the datasets utilized in the research for predicting truck turnaround time at sea terminals. The research collects and employs several datasets to gain more understanding and contribute to predicting truck turnaround time at sea terminals. The datasets involve container arrival dataset, vehicle Bluetooth detection dataset, and additional datasets. The datasets are introduced as below.

- Container arrival dataset. The dataset is provided by Port of Rotterdam and contains all arriving container records at the five container terminals in the Maasvlakte area in 2017. It reveals several key data fields of the arrived containers. Table 4.1 shows the key data fields of this data set.

Table 4.1: Container arrival data field description

Key Fields	Field meaning
DEEPSEA_ARRIVAL_TIME	Container ship arrival time
EQUIPMENT_DISCHARGE_TIME	Container unloading time
CONTAINER_TYPE_TYPE_CODE	Container type
TERMINAL	Container unloading terminal
HINT_EQP_ETA_VISIT	Estimated hinterland transit time
VESSEL_NAME	Container ship name

- Vehicle Bluetooth detection dataset. Bluetooth stations log the timestamp and identifier of each vehicle with an active Bluetooth sensor as it passes. This identifier is recorded as a media access control (MAC) address. Travel time between two Bluetooth stations can be calculated by comparing these timestamps. In this study, the Bluetooth data are sourced from the Port of Rotterdam's Bluetooth service, but it do not provide information about vehicle type. This Bluetooth data set is measured in days and contains all Bluetooth detection information on the roads surrounding the Rotterdam port area for 365 days in 2017. The daily data set contains approximately 4 million detection records. Each sensor records the passage of vehicles, the MAC Address of the vehicle, along with the time stamp and the latitude and longitude of the sensor. This dataset is instrumental in understanding vehicular movement patterns, traffic density, and congestion levels around the port terminals. Vehicle turnaround time is also calculated based on the timestamp in this dataset. Table 4.2 shows the key data fields of this data set.

Table 4.2: Vehicle Bluetooth data description

Key Fields	Field meaning
LAT	Latitude of the sensor
LONG	Longitude of the sensor
LOCATION_ID	Sensor ID
MACADDRESS	Unique ID of vehicles
PASSAGETIME	Vehicle passing time

- Wind condition dataset. Provided by the Port of Rotterdam Authority, this data set describes the wind speed conditions every ten minutes around the Maasvlakte area in 2017. The data fields include factors such as wind level, maximum wind gust speed and wind direction.

The data spans from January 1, 2017, to December 31, 2017. Since these datasets are collected independently and are only linked by their time stamps, pre-processing is required to transform them into a format suitable for model input.

4.1.2 Data Pre-processing

The following steps outline the pre-processing procedure:

1. Identification of Influencing Factors: The first step involves identifying the key factors that will serve as inputs to the prediction model. These include variables such as traffic flow, terminal operation intensity, and weather conditions. This step is comprehensively defined by the data sets available for this study as well as container transport chain demarcation conducted previously.
2. Processing of Bluetooth Data: The next step is to map the sensors in the Bluetooth dataset to visualize their locations and determine which sensors are relevant to the port and surrounding areas. Once the relevant sensors are identified, the data is filtered to retain only those records associated with these sensors. The filtered data is then used to estimate traffic flow in the road network and calculate the estimated operation time of vehicles within the port area. To remove noise from the dataset, duplicate and irrelevant data are discarded, and records showing vehicle operation times outside the typical range of 1 to 4 hours are excluded.
3. Incorporation of Additional Data Sources: Weather data and port operation data are collected from external sources such as KNMI and the Port of Rotterdam. These datasets are aligned with the Bluetooth and container arrival data based on time stamps, so as to ensure that all data points are synchronized.
4. Employ descriptive and exploratory data analysis on all these datasets. Using histograms, bar or box charts to visualize the distribution and observe basic statistics such as means, medians, and standard deviations. Based on the trends observed, the relationships within one variable and between different variables would be explored to identify which factors are most strongly associated with truck turnaround times.
5. Necessary new features are extracted from the existing data through feature engineering. For instance, from the container arrival dataset, we might derive features such as the number of container arrivals per hour or the average loading/unloading time. These new variables help capture the operational dynamics more accurately and improve the prediction model's performance.

4.2 Prediction Model Development

The prediction model development phase is the core component of this research. Given the sequential nature of the data involved, we elaborate on the choice of employing time series-based prediction methods for forecasting truck turnaround times at sea terminals. Additionally, we discuss the reason behind selecting Long Short-Term Memory (LSTM) networks as the primary modeling technique.

4.2.1 Introduction to LSTM Model

The choice of LSTM is driven by the need to model complex sequential relationships between variables. As a type of recurrent neural network (RNN), LSTM was firstly introduced by Hochreiter and Schmidhuber, as a solution to the vanishing gradient problem commonly encountered in traditional RNNs[26].

The primary advantage of LSTMs lies in their unique architecture designed to balance complexity and performance. The introduced memory cells, input gates, output gates, and forget gates allow LSTMs to selectively retain or discard information over long sequences[24]. The core of a LSTM model consists of one or more LSTM layers. Each LSTM layer is configured with a specific number of memory units (cells) that control the flow of information through the network, enabling the model to maintain a memory of past events and use it to make accurate predictions. The inner structure of one cell is displayed in figure 4.1. In a typical LSTM neuron, it includes n cells where n equals to the time length of the long-term memory. The information from the preceding cell will be transmitted to the next cell. With the help of gate units, the necessary information can be retained for a long time.

As shown in fig 4.1, at time step t, the cell involves three inputs, three gate units and two outputs. Two of the inputs comes from the unit of previous time step, namely the cell state C_{t-1} and hidden state h_{t-1}. Another input is the current system variable vector x_t. The outputs of the cell are the updated cell state C_t and the hidden state h_t, which are passed to the next time step.

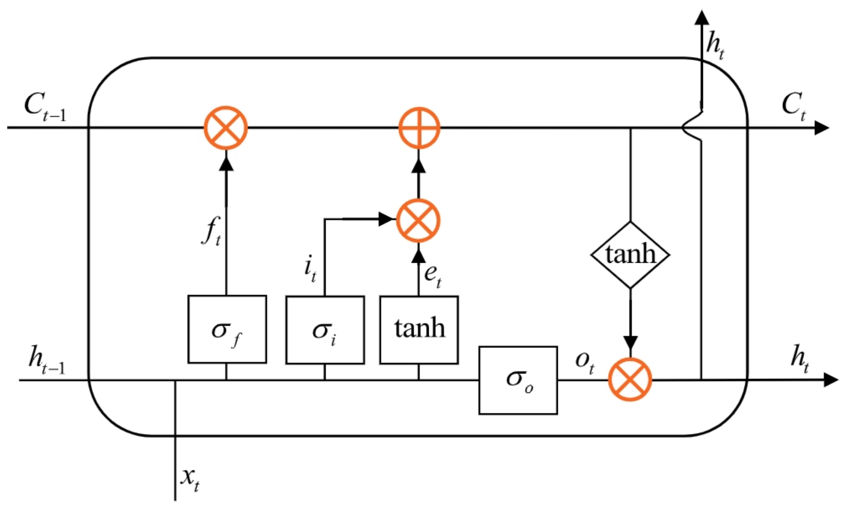


Figure 4.1: Typical repeating module in an LSTM[19]

Three gates are structured within a unit. The forget gate determines which information from the previous cell state C_{t-1} should be discarded. It is represented mathematically as:

$$f_t = \sigma_f(W_f \cdot [h_{t-1}, x_t] + b_f), \tag{4.1}$$

where σ_f is the sigmoid activation function, W_f is the weight matrix, b_f is the bias vector, and $[h_{t-1}, x_t]$ represents the concatenation of the hidden state and input vector. The forget gate outputs a value $f_t \in [0, 1]$, where values close to 1 retain information and values close to 0 discard information.

The input gate controls which new information should be added to the cell state. This process involves two steps:

1. A candidate cell state $\tilde{C}_t$ is generated using the hyperbolic tangent function:

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c). \tag{4.2}$$

2. The sigmoid activation function determines how much of this candidate state to incorporate:

$$i_t = \sigma_i(W_i \cdot [h_{t-1}, x_t] + b_i). \tag{4.3}$$

The updated cell state is computed by combining the outputs of the forget and input gates:

$$C_t = f_t \otimes C_{t-1} \oplus i_t \otimes e_t, \tag{4.4}$$

where $\otimes$ denotes element-wise multiplication, $\oplus$ denotes direct sum operation.

The output gate determines what part of the cell state C_t should be output as the hidden state h_t and convert to next cell:

$$o_t = \sigma_o(W_o \cdot [h_{t-1}, x_t] + b_o). \tag{4.5}$$

The hidden state is calculated as:

$$h_t = o_t \otimes \tanh(C_t). \tag{4.6}$$

Finally, the weights and biases will be updated through back propagation for minimizing the loss function. In this study, the loss function is determined by the prediction of the state, which could be written as:

$$\text{Loss}(\hat{x}_{T+1} - x_{T+1}) = \text{Loss}(h_T - x_{T+1}) \tag{4.7}$$

According to Cui and so on, a multi-layer architecture can capture inner layers and latent variables, which can effectively enhance the power of neural networks[19]. In this study, different structures of LSTM models would be adopted to predict truck turnaround time.

Truck turnaround time prediction inherently involves analyzing sequential data, where past observations play a crucial role in forecasting future outcomes. Time series analysis provides a framework to capture temporal dependencies and patterns within the data. Among various time series modeling approaches, LSTM networks offer distinct advantages that align well with the characteristics of the dataset and the complexity of the prediction task. Moreover, LSTM networks are well-suited for handling variable-length input sequences, making them adaptable to the diverse temporal resolutions present in the dataset. This flexibility is essential for accommodating the irregularities and fluctuations inherent in real-world transportation data, where the timing and frequency of events may vary widely.

4.2.2 Model Development

In this research, a stacked-LSTM model is developed to predict the turnaround time. Existing studies have proved that a more complicated LSTM architecture is more probably able to build the representations of sequential data [30]. The stacked-LSTM model features a multi-layer architecture, allowing it to process sequential data hierarchically, capture both short-term and long-term dependencies, and build richer representations. In a stacked multi-layer LSTM architecture, the hidden layer's output serves as the input to the next layer in the sequence. This architecture allows for embedding multiple LSTM components to capture richer hidden features in the data.

The stacked LSTM model in this study employs a multi-layered architecture designed to effectively process and analyze complex sequential data, as illustrated in fig 4.2. The model integrates multiple LSTM components to capture both short-term and long-term temporal dependencies while extracting rich features from two primary data sources: container-related information and road traffic data. These two data streams are first processed independently through separate LSTM layers, allowing the model to focus on the unique characteristics and temporal patterns of each source.

The output from these initial LSTM layers is then combined and passed to a shared LSTM layer, where the model integrates the extracted features to understand the interdependencies between container operations and road traffic conditions. This consolidated information is further processed through a dense layer, which reduces dimensionality and refines the feature representation, before being fed into the output layer to generate predictions of truck turnaround time. By isolating and then integrating data from these two sources, the architecture ensures a comprehensive representation of the factors influencing turnaround time. The hierarchical design of stacked LSTM layers enhances its ability to capture intricate temporal relationships and dependencies, enabling more accurate and robust predictions. Furthermore, this design is scalable and can accommodate additional data sources or layers, offering flexibility for future applications in container terminal operations.

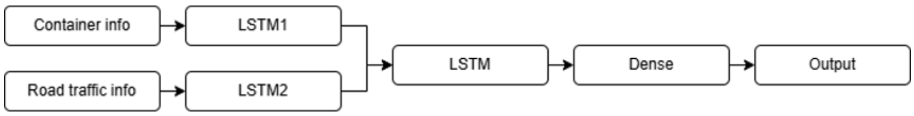


Figure 4.2: Architecture of stacked-LSTM model

The input variables of the model and related data-preprocessing expressions are formulated below.

In this research, the prediction value of truck turnaround time at time step t is expressed as $\hat{y}_t$, $t \in \{1, \ldots, N\}$, where N is the number of time series sample size. For the predicted truck turnaround time at m time periods ahead, it is denoted by $\hat{y}_{t+m}$. Additionally, the following notations are defined as model inputs.

- **Truck Arrival Flow (λ_t):** The number of container trucks arriving at the port during time step t. This represents the truck arrival flow.
- **Lagged Average Turnaround Time (T_{t-i}):** The average truck turnaround time at $t-i$ time steps, as the model aims at predicting in advance.
- **Wind Components (w_x and w_y):** The two-dimensional components of wind speed, w_x and w_y, representing the wind's horizontal and vertical force directions at time step t.
- **Container Arrival Volumes at Terminals ($A_{i,t}$):** The number of containers arriving at terminal i during time step t. For example:
  - $A_{\text{APMII},t}$: Container arrivals at APMII terminal.
  - $A_{\text{APMRTM},t}$: Container arrivals at APMRTM terminal.
  - $A_{\text{ECTDELTA},t}$: Container arrivals at ECT DELTA terminal.

In addition to the above features obtained from the data sets, periodicity parameters are also introduced. $T_{d,t}$ is a vector representing the weekday of time step t, and $T_{h,t}$ represents the hour of day of time step t. The temporal features are transformed using sine and cosine functions. This transformation generates periodic variables and avoids discontinuities introduced by the cyclical nature of time.

Day of the Week ($T_{d,t}$) Assuming a week consists of 7 days (from Monday to Sunday), the day $d \in \{0, 1, \ldots, 6\}$ at time step t is transformed into periodic variables as follows:

$$T_{d,t} = \begin{bmatrix} \sin\left(\frac{2\pi \cdot d}{7}\right) \\ \cos\left(\frac{2\pi \cdot d}{7}\right) \end{bmatrix}. \tag{4.8}$$

Hour of the Day ($T_{h,t}$) Assuming a day consists of 24 hours (from 0 to 23), the hour $h \in \{0, 1, \ldots, 23\}$ at time step t is transformed into periodic variables as follows:

$$T_{h,t} = \begin{bmatrix} \sin\left(\frac{2\pi \cdot h}{24}\right) \\ \cos\left(\frac{2\pi \cdot h}{24}\right) \end{bmatrix}. \tag{4.9}$$

Before training the model, some data pre-processing techniques are implemented in order to improve data quality and prediction accuracy. Specifically, two key steps are performed:

1. Outlier Removal: A significant portion of errors in container truck turnaround time data arises due to low sample sizes during certain time periods. Outlier predictions are handled through quantile clipping, where predictions falling outside a certain percentile range are adjusted or removed to improve the model's general accuracy. To address this, we first identify and clean these low-sample-size outlier periods, ensuring that the resulting dataset better represents the actual trends and patterns.

2. Masking Invalid Time Steps: After removing outliers, the cleaned data is combined with the original missing time steps to create a comprehensive mask. This mask is represented mathematically as follows:

Let $\mathbf{X} \in \mathbb{R}^{T \times F}$ denote the input time series data, where T is the total number of time steps and F is the number of features. Define a binary mask $\mathbf{M} \in \{0, 1\}^{T \times 1}$ such that:

$$M_t = \begin{cases} 1, & \text{if time step } t \text{ is valid} \\ 0, & \text{if time step } t \text{ is invalid (missing or removed as an outlier)} \end{cases} \tag{4.10}$$

During training, the mask $\mathbf{M}$ is applied to the input sequence $\mathbf{X}$ such that the model only processes valid time steps. This is achieved using a masking layer in the architecture:

$$\tilde{\mathbf{X}}_t = \mathbf{M}_t \cdot \mathbf{X}_t \tag{4.11}$$

where $\tilde{\mathbf{X}}_t$ is the effective input at time step t, and invalid time steps ($M_t = 0$) are ignored.

3. Time-series forecasting models like LSTM can produce predictions that may fluctuate due to minor noise in the data, especially when dealing with complex operational environments like port terminals. In an LSTM-based prediction problem, the presence of missing or null values in the input time series poses a significant challenge, as the model cannot process null values during training. Simply imputing missing values with predefined values, such as zeroes, the mean of historical observations, or the last observed value, introduces bias into the model inputs. This bias can lead to inaccurate parameter estimation during the training process, ultimately affecting the model's predictive performance[11]. To address this, data quality improving techniques are applied to reduce these fluctuations and produce more stable predictions.

After resampling the data set into a data form characterized by time steps and preprocessing it, the finalized data are split into training data and test data. The first 80% of the data are used for model training and the remaining are used for test and validating. Different sequence lengths are applied to strike a balance between forecast accuracy, data availability and practical significance.

To standardize the input data, a min-max scaling technique is applied in order to prevent features with larger magnitudes from dominating the learning process and to accelerate model convergence during training. For a given feature x_t at time step t, the scaling is performed as follows:

$$x_t^{\text{scaled}} = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \tag{4.12}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the feature x_t in the dataset. The same scale is applied to both train and test set.

4.2.3 Hyperparameter Selection

The stacked-LSTM model has many hyperparameters that significantly affect its performance. The parameters can be determined through testing and evaluation to achieve optimal performance on domain-specific problems. These hyperparameters can be categorized into two groups: model architecture hyperparameters and train process hyperparameters.

- **Model Architecture Hyperparameters**
  - Number of LSTM layers:Controls the depth of the network. Deeper networks may capture more complex patterns but also increase the risk of overfitting or vanishing gradients.
  - Number of neurons per layer: There are multiple neurons within each layer of LSTM. It determines the model's capacity to learn features. Large numbers of neurons can describe one input unit in more perspectives, allowing for capturing more patterns but require more computational resources.
  - Dropout rate: Dropout randomly selects a subset of neurons to be deactivated under preset probability conditions. The set of dropout rate encourages the activate neurons to adapt more independently and less cohesively to the data, reducing the risk of overfitting[25].
  - Input time steps: It Defines the historical time steps used for prediction. Longer sequences capture more long-term dependencies but may introduce noise.
  - Activate function: Standard `tanh` and `sigmoid` functions are used as default activate functions, as they are effective in handling non-linear temporal patterns in LSTMs.
  - Loss function and optimizer: A loss function measures the difference between the predicted output of a model and the actual target values, while an optimizer updates the model's parameters to minimize the loss function. Different loss functions and optimizers behave differentially in various scenarios and problems.
- **Training Process Hyperparameters**
  - Batch size: It refers to the number of training samples processed before the model updates its parameters during one iteration. A small batch size allows the model to update more frequently, leading to faster but noisier convergence, which can help escape local minima but may cause instability. A large batch size results in smoother and more stable updates but requires more memory and may converge more slowly.
  - Learning rate: It controls how much the model needs to change in response to the estimated error for each time when the model's weights are updated. A high learning rate may cause the model to diverge, while a learning rate that is too low can result in slow convergence.
  - Epochs: An epoch could be described as an entire training cycle of the model. The number of epochs refers to the number of iterations that updating the model's parameters. Theoretically, more epochs can increase the fitting level and improve the model accuracy, but there is generally an upper limit on accuracy in the actual training process. Hence, early stopping is applied if the validation loss does not improve within 10 consecutive epochs.

In the model construction of this study, some parameters were selected as the benchmark that are generally considered to be better solutions by academic circles, while the optimal values of other parameters were obtained through repeated trials. The test process and results will be explained later.

4.2.4 Model Evaluation and Validation

Model Evaluation

To evaluate the prediction performance of the developed LSTM model, commonly used regression evaluation metrics are employed. These metrics provide quantitative assessments of the model's accuracy and consistency in predicting truck turnaround time. In this study, the primary evaluation metrics include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). These metrics are chosen for their effectiveness in capturing different aspects of model accuracy.

- **Root Mean Square Error (RMSE)**: RMSE measures the square root of the average squared differences between predicted and actual values. It is particularly sensitive to large prediction errors, making it suitable for identifying models with significant outliers.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{y}t - yt)^2} \tag{4.13}$$

- **Mean Absolute Error (MAE)**: MAE computes the average absolute differences between predicted and actual values. It provides an intuitive measure of average error without giving extra weight to large deviations.

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}|\hat{y}_t - y_t| \tag{4.14}$$

- **Coefficient of Determination (R^2)**: R^2 quantifies the proportion of variance in the actual data explained by the model. It ranges from 0 to 1, with values closer to 1 indicating better model fit.

$$R^2 = 1 - \frac{\sum_{t=1}^{N}(\hat{y}_t - y_t)^2}{\sum_{t=1}^{N}(y_t - \bar{y})^2} \tag{4.15}$$

  where $\bar{y}$ is the mean of actual values.

These metrics are applied to evaluate different model architectures, hyperparameter settings, and comparative baseline methods such as linear regression, random forest, and simple LSTM models. The evaluation results provide a basis for selecting the best-performing model for prediction.

Model Validation

To ensure the reliability and generalizability of the developed model, a k-fold cross-validation strategy is employed [10]. Cross-validation is a robust validation method that divides the dataset into k mutually exclusive subsets (folds) of approximately equal size. The model is trained and tested k times, each time using one fold as the test set and the remaining $k - 1$ folds as the training set.

The process for k-fold cross-validation can be described as follows:

1. The dataset is randomly partitioned into k subsets of equal size.
2. For each iteration, one subset is used as the test set, and the remaining $k - 1$ subsets are combined to form the training set.
3. The model is trained on the training set and evaluated on the test set using the chosen evaluation metrics (RMSE, MAE, R^2).
4. This process is repeated k times, with each subset serving as the test set exactly once.
5. The final evaluation metrics are computed as the average of the metrics across all k iterations.

For this study, a 5-fold cross-validation is applied to balance computational efficiency and statistical reliability. This method ensures that the model's performance is not overly dependent on any specific train-test split and minimizes the risk of overfitting.

Chapter 5

Data Analysis

This chapter aims to explain the sources, characteristics, and data processing methods of the data sets used in this study, and to conduct statistical descriptive analysis and exploratory data analysis on data sets such as truck turnaround time.

5.1 Data Pre-processing

The obtained data sets contain a large number of missing and misaligned values, and their original format is not directly suitable for analyzing parameters related to port truck turnaround time. Therefore, a lot of preprocessing work is required on these data sets.

5.1.1 Bluetooth-based Road Traffic Data Processing

Since 2010, the Port of Rotterdam has begun to deploy Bluetooth detectors on the port area and surrounding roads to monitor the road traffic conditions in the port area. Specifically, Bluetooth data around the port area is used to determine the driving time of vehicles on the road, or to extract movement data and driving time of specific freight flows from the measurements.

Bluetooth is a communication technique based on electromagnetic waves. It is used to establish connection between devices in order to send data back and forth. In order to distinguish which device is being communicated with, each Bluetooth has a unique identity which is called Media Access Control (MAC) address. By placing Bluetooth detection devices along the road, vehicles with Bluetooth enabled could be detected. The disadvantage of this technology is that it cannot match Bluetooth devices with vehicles. Bluetooth devices are not completely popular in vehicles, and the ability to detect vehicles equipped with on-board Bluetooth devices depends on the combination of on-board devices and detection antennas[54]. When dealing with Bluetooth devices in road traffic, a critical issue is the detection rate, or penetration. This indicator represents the ratio of vehicles that can be detected by Bluetooth detectors to the total traffic amount, and should be as high as possible so that the collected data can be considered reliable. Margreiter et al. [33] studied the Bluetooth detection rates from 2010 to 2016 and found that the detection rate was of about 25%.

According to a 2011 pilot study by the Port of Rotterdam and MAPtm, Bluetooth technology is generally effective for traffic monitoring. Research shows that travel time measured using Bluetooth technology has similar trends compared to traditional license plate recognition (LPR) systems, and shows higher accuracy on road sections with heavy traffic volumes. However, since the penetration rate of Bluetooth signals is approximately between 40% and 50%, especially when traffic density is low, the coverage and reliability of Bluetooth data will decrease [54].

In addition, the study found that at specific locations, such as truck terminal entrances, the Bluetooth system had a significantly higher detection rate for heavy vehicles than light vehicles, reaching approximately 75 percent. This high penetration rate shows that Bluetooth technology has high applicability in areas with large freight vehicle traffic, especially in areas around container terminals.

5.1.2 Identification of Vehicle Trips and Categories

In our data, some Bluetooth sensors, mapped to specific locations, provide insights into vehicle routes, travel times, and the turnaround times associated with entering and exiting port terminals. A pair of Bluetooth sensors along the A15 route to Port of Rotterdam area are found to cluster travel time observations and thus identify the corresponding vehicle category. Other sensors, deployed within the Maasvlakte area, can track the sequence of vehicle passes.

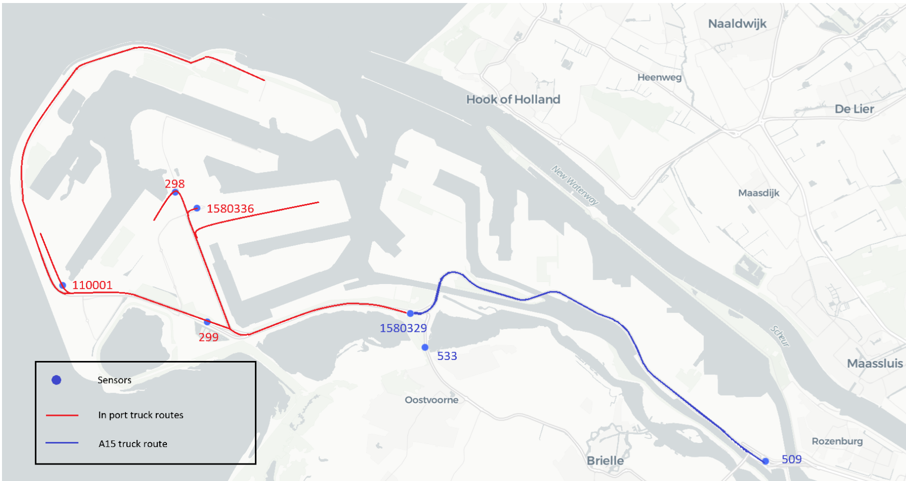


Figure 5.1: Utilized sensor locations and truck routes

Fig 5.1 displays the distribution of selected sensors and corresponding vehicle routes. Before identifying truck tours from the Bluetooth data, we need to extract vehicle categories from the dataset. The flow chart presented in fig 5.2 illustrates the steps for extracting truck-specific data.

For a given time period and specific Bluetooth sensors, we eliminate outliers from Bluetooth observations based on the driving behavior of different types of vehicles. Then, the filtered Bluetooth observations are clustered into two representative groups to analyze the driving behavior of these vehicles. The selected A15 route is about 10 km long. Therefore, after calculating the travel time, we delineated a reasonable travel time range based on the speed limit of Dutch highways. Observations exceeding this travel time range are excluded to avoid the impact of missing observations or longer travel times caused by congestion.

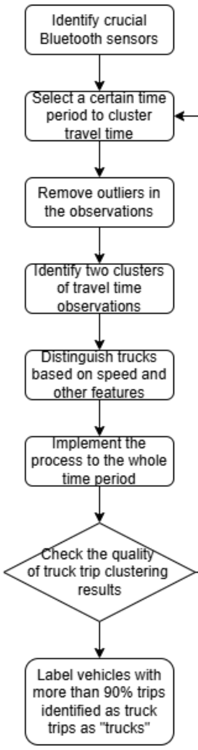


Figure 5.2: Identifying trucks from Bluetooth observations

After clustering vehicle observations, we categorized the driving behaviors based on the speed limit for trucks on Dutch highways. Vehicles meeting these conditions were then labeled as "likely to be trucks". Selecting January 2017 as the time period, with the timestamp of entry into the route as the horizontal axis and the travel time as the vertical axis, the observed distribution of data points is shown in Fig 5.3.

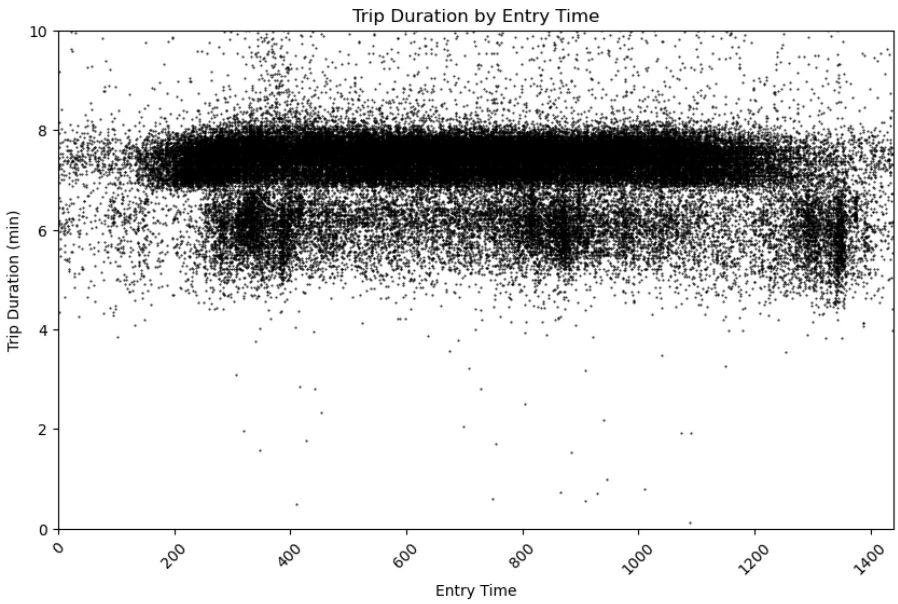


Figure 5.3: Trip duration time observations

The majority of the observations appear to cluster around a narrow range of trip durations between about 7 to 8 minutes. For the chosen route, the presence of a clear boundary between the clusters centered around 6 and 7 minutes suggests distinct driving behaviors or vehicle types. Fig 5.4 presents the clustered results of Bluetooth observations after excluding outliers from the data. The slower group of vehicles is marked red. As the regulatory speed limit for trucks and some other types of vehicles in the Netherlands is 80 km/h, these observations generally meet the speed limit requirements. This can also explain the cause of the cluster boundary.

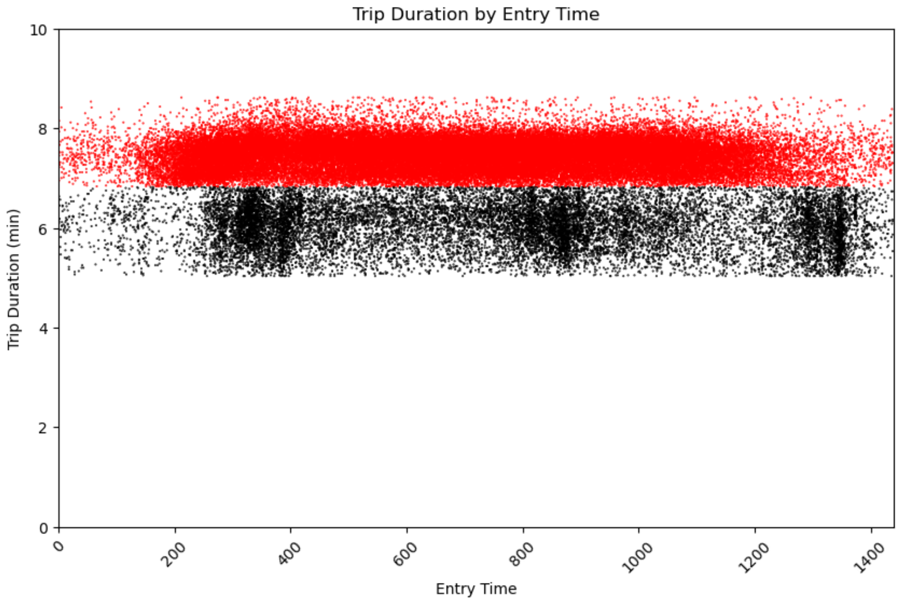


Figure 5.4: Clustered trip duration time observations

The identified clusters of slower vehicles account for approximately 75% of all vehicles. Considering the speed limits of different types of vehicles, the penetration rate of Bluetooth devices, and other factors, these vehicles cannot be completely labeled as trucks, pending subsequent verification. Therefore, these data points are labeled "likely to be trucks". For each MAC address, if more than 90% of its trips are marked as such, it is initially classified as a truck.

5.1.3 Identification of Truck Trips within Port Area

As the Maasvlakte area hosts five major container terminals, understanding the movement of container trucks within this area is essential and crucial. This section describes the methodology for identifying truck trips and extracting travel routes, destinations, and turnaround times of truck trips. As displayed in fig 5.1, the Maasvlakte area's sensors are positioned at entry points, terminal access roads, and within key road segments, which provides opportunities to track container trucks' movement across the five terminals.

A vehicle trip within the Maasvlakte area can be defined as the journey a truck undertakes from its entry into the area, through terminals, and until its exit. Each terminal, within its unique entry route, represents a potential destination within a tour. By investigating the traffic network within the Maasvlakte area through Google Maps and geographic information, specific routes to each container terminal are delineated with corresponding Bluetooth sensors placed along the routes. Sensor ID 1580329 is located at the entry to the Maasvlakte area. As a result, a pair of records of a vehicle being detected twice within a specific time period by sensor 1580329 is considered as the entry and exit time stamp. Other sensors reveal the routes taken by vehicles after entering the port area. Based on Google Maps and satellite images, the locations of these sensors can be identified along key paths to various terminals. Table 5.1 displays the correspondence between each sensor and the respective terminal.

Building upon the historical analysis of vehicle trip times and the identification of valid trips, the

Table 5.1: Sensor and Terminal Mapping

Sensor	Terminal
298	APMII
299	RWG
110001	ECT_EUROMAX
1580336	APMRTM
If only 1580329 but no other sensors	ECTDELTA

process for defining vehicle tours within the Maasvlakte area relies on detecting consistent entry and exit behavior for each vehicle.Considering the possibility of missing Bluetooth detection records and their impact on subsequent data, an effective trip detection step needs to be added. A single trip is considered valid only if it lasts more than 20 minutes and occurs within the same day, or if it spans multiple days but is within four hours. Among the identified trips, based on information provided by sources such as the Port of Rotterdam, the turnaround time for truck trips is generally less than four hours. Therefore, trips lasting less than four hours are labeled as "likely to be a truck".

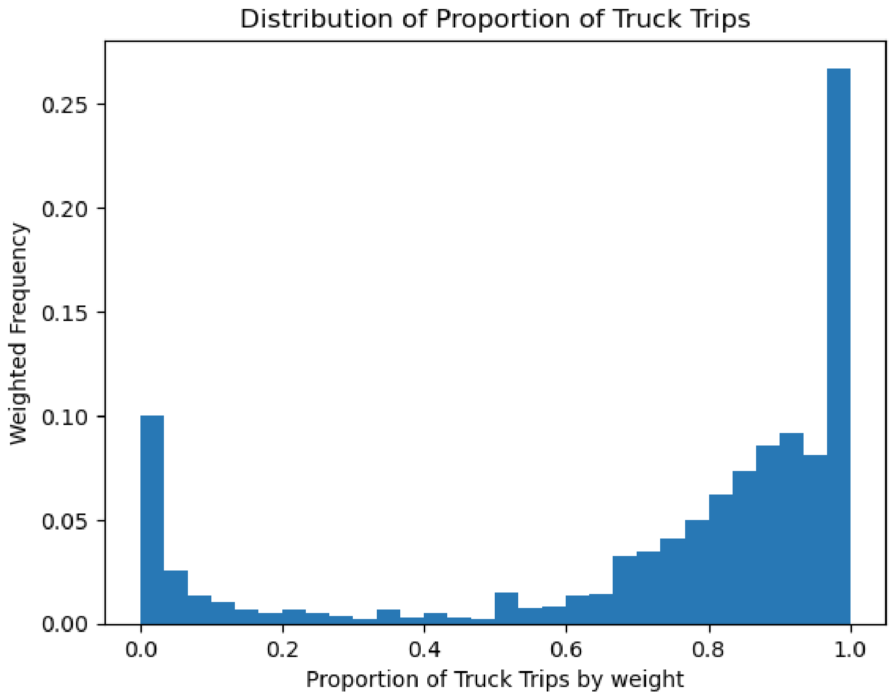


Figure 5.5: Weighted Frequency Distribution of Truck Trip Proportion by Vehicle

After the trip detection and classification process described earlier, each vehicle's trips are divided into two categories: a certain proportion is labeled as "truck" trips, while the remaining trips are labeled as "others." This plot shows the weighted frequency distribution of the proportion of trips labeled as "truck" for each vehicle. Each vehicle's weight in the distribution is determined by its total number of trips, giving more weight to vehicles with a higher trip count. Fig 5.5 shows the weighted frequency distribution of truck trip proportion. As the frequency of truck trips gradually increases after the 70% quantile value, we initially label vehicles with over 70% truck trips as "trucks".

According to Wijbenga et al. [54], the penetration rate of Bluetooth vehicles in port areas is no more than 75%. Since there are other facilities in the port area besides container terminals, and the same road speed limits apply to other types of vehicles other than container trucks, we are not able to capture precise road traffic flows, container truck flows and turnaround times. As a result, in order to improve the authenticity of the turnaround time prediction input, we use double verification methods. Only vehicles that are identified as trucks in both identification methods will be labeled as trucks in the final processed data set. Finally, based on the vehicles identified to be trucks according to their trip duration time on the A15 route, 85% of the vehicles were further confirmed to be trucks. These identified trucks carries 64% of the trips within the port area.

5.1.4 Container Arrival Flow Processing

The process of container arrival flow analysis act as the main step for capturing and analyzing port operation conditions. This section describes the procedure to prepare and process the container arrival data to ensure the consistency, accuracy, and temporal alignment across datasets.

To begin, the raw container arrival data is filtered to address any inconsistencies in format and alignment. Due to the nature of data collection at multiple points across the port, initial observations show a variety of time zones, formats, and missing entries. We first convert all timestamps to a standard time zone (CET) and ensure that each record includes consistent date and time formats. Entries with missing values on key fields are also removed for data integrity in subsequent analysis. Among all 1,299,859 container unloading records, only a few dozen contain missing values on key fields, so removing them would have minimal impact on final results.

Following the cleaning and filtering process, the left container arrivals are resampled into fixed time intervals of each hour in order to generate a normalized time series. The time series is measured in hours and is divided according to the terminal where the container arrives and the corresponding hinterland transportation mode.

Fig 5.6 presents the results of aggregated container flow and visualizes the arrival of vessels at various terminals over a one-month period, with each data point representing a vessel arrival. The horizontal axis denotes the time of arrival, while the vertical axis specifies the terminal where each vessel arrived. The size of each scatter point corresponds to the container count associated with the respective vessel, with larger circles indicating vessels carrying more containers. By visualizing the aggregated container flow, we can observe the obvious operational differences of different terminals and pave the way for model input.

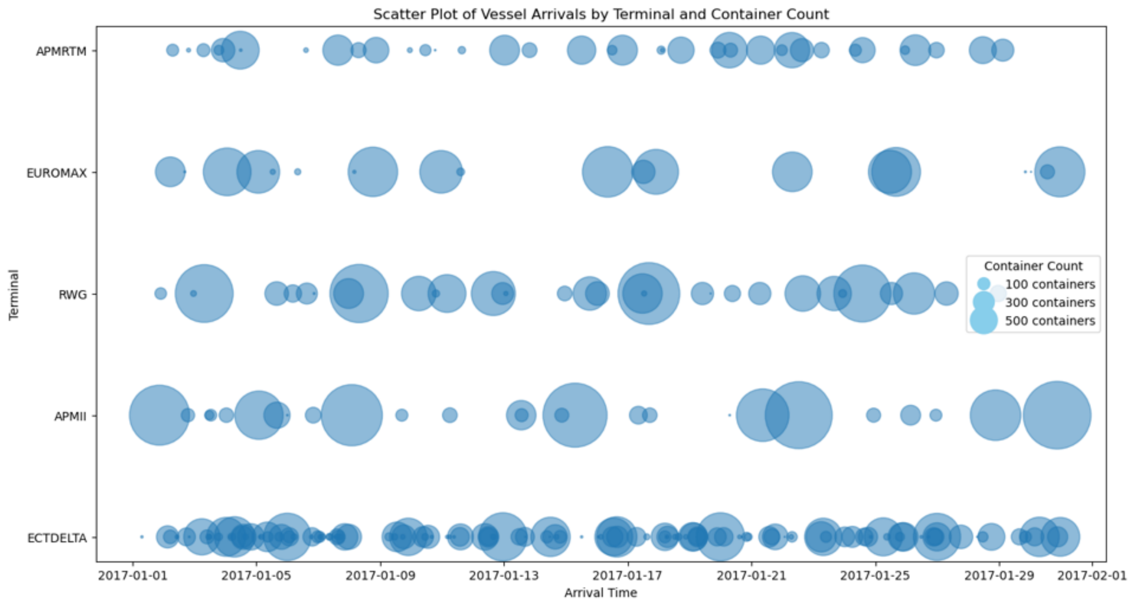


Figure 5.6: Vessel Arrivals by Terminal and Container Volume

5.1.5 Wind Data Processing

To utilize this dataset effectively in analyzing the impact of environmental conditions on port operations, it is necessary to inspect the dataset for misalignments and missing values. Due to the sensor downtime which is common in high-frequency environmental data, there are many missing values in the three key fields of wind speed, instantaneous gust and wind direction, and the absence of these fields will also lead to misalignment of the data set.

For the 10 minutes time interval in the time series, a moving average imputation technique is applied. Specifically, the mean wind speed and direction over a 60-minute window centered on each missing entry is calculated. For larger data gaps, where a moving average may not provide reliable estimates, we use forward-filling based on the last available valid value. This method is only applied to gaps that do not exceed one hour, as extending beyond this time frame would risk introducing unrealistic continuity in dynamic weather conditions.

While the wind speed along with wind direction (in degree) is not an ideal model input, convert wind conditions to wind vectors would helps to explain the model. In the wind direction, 360° and 0° should approach each other and wrap smoothly. If the wind isn't blowing, the direction doesn't matter. As a result, Wind speed and direction are converted into u (east-west) and v (north-south) vector components using the following formulas:

$$u = -\text{wind speed} \times \sin(\text{wind direction})$$
$$v = -\text{wind speed} \times \cos(\text{wind direction})$$

In these equations, wind direction is represented in degrees, with 0° indicating wind blowing from the north, 90° from the east, and so forth in a clockwise direction. By converting wind conditions into u and v components, we enable straightforward integration into models that require vector inputs and facilitate the calculation of resultant wind effects in specific directions relevant to port activities.

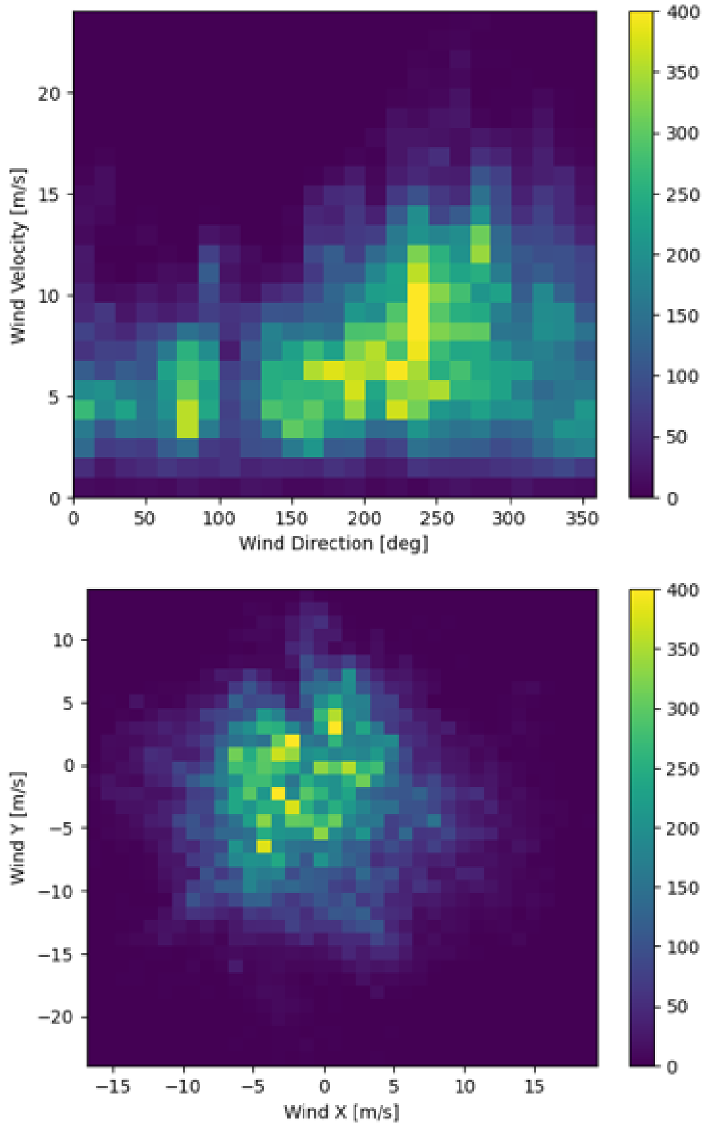


Figure 5.7: Comparison of Wind format

Fig 5.7 shows the results before and after wind vectorization processing. The wind vectors are primarily concentrated near the center of the plot, while points farther from the center represents higher wind speeds. This vectorization approach provides a more intuitive view of wind distribution across different directions, helping to reveal potential impacts of wind from various orientations. After vectorization, wind data shows a clearer directional distribution, and the format is better suited for integration into further modeling. This enables models to more effectively capture changes in wind speed and directions, and their potential impacts on port operations.

5.2 Truck Behavior Analysis

In this section, we conduct an in-depth analysis of the processed data to uncover patterns and insights relevant to port operations and truck turnaround times. Identifying and analyzing the distribution characteristics of the data can help us make more targeted adjustments to the model input. In addition, the data periodicity revealed by distribution characteristics can help us better capture underlying temporal patterns, especially when dealing with indicators with obvious periodic fluctuations (such as turnaround time or traffic fluctuations). This type of periodic information helps the model perform better in Effectively capture cyclical changes in data when forecasting.

Fig 5.8 illustrate the distribution of turnaround times, one representing all vehicles and the other specifically focusing on identified trucks. The x-axis stands for turnaround time length, while the y-axis coordinate is the frequency of occurrence. The distribution of turnaround times for all vehicles in the first figure displays a bimodal pattern. The first peak occurs between 1 and 2 hours, and the turnaround time for most vehicle trips are distributed around this peak. The secondary peak, around 8-10 hours, likely represents non-container vehicles involved in different types of operations within the port, especially vehicles that regularly enter and exit the port area during working hours. In contrast, the distribution of turnaround time for identified trucks (second figure) shows a unimodal and positively skewed distribution. Similar to the first peak in the first figure but with a lower frequency, the majority of turnaround times are clustering between 1 and 2 hours. This concentration suggests that, under normal operating conditions, most container trucks can complete their tasks within this time period. The observed distribution features and differences demonstrate the effectiveness of the proposed truck identification method.

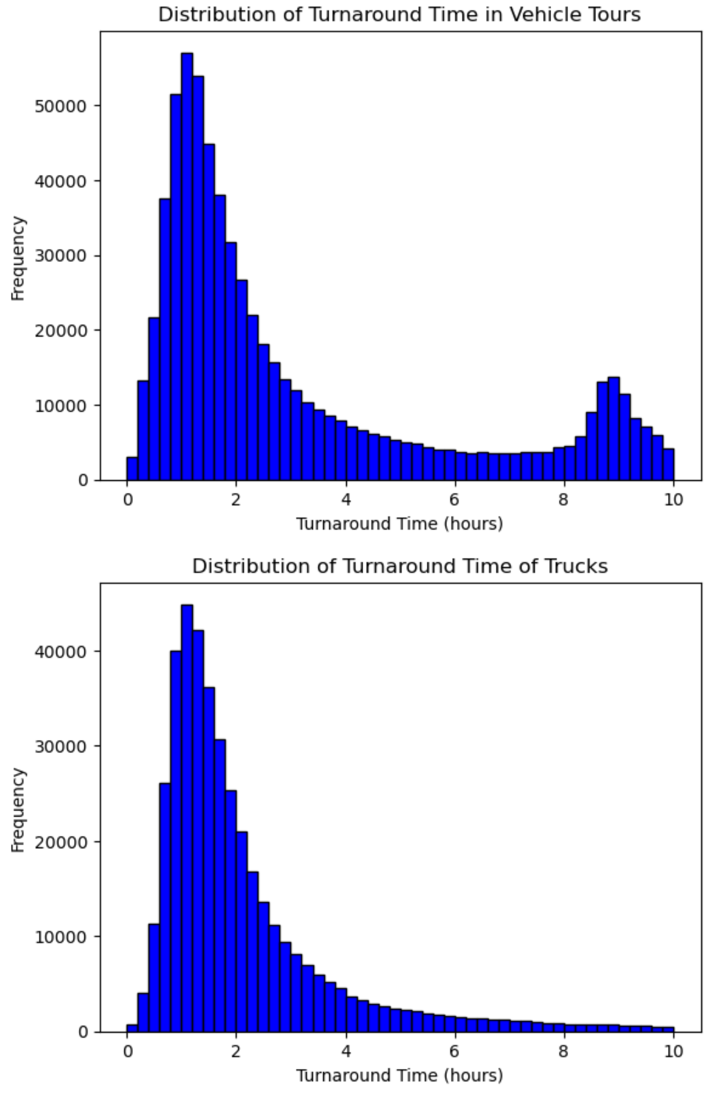


Figure 5.8: Comparison of Turnaround Time Distributions

Fig 5.10 displays the monthly distribution of truck flow into various terminals. Fig 5.9 shows the hourly truck flow difference in a day. The results indicate that different terminals exhibit distinct traffic patterns over time. Among the five terminals, ECT DELTA has the highest traffic volume, with the most noticeable fluctuations. Truck flow at each terminal shows clear daily periodicity, while remaining generally stable throughout the year. Only ECT DELTA and APMRTM experience significant demand changes in October, which may reflect infrastructure maintenance at the port.

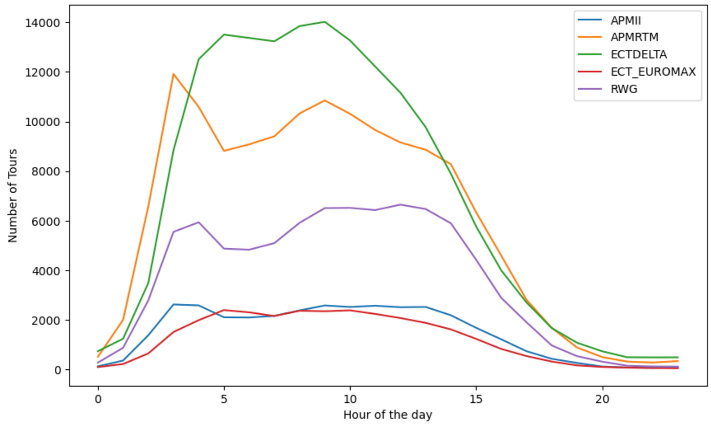


Figure 5.9: Input Truck Flow into Different Terminals by Hour

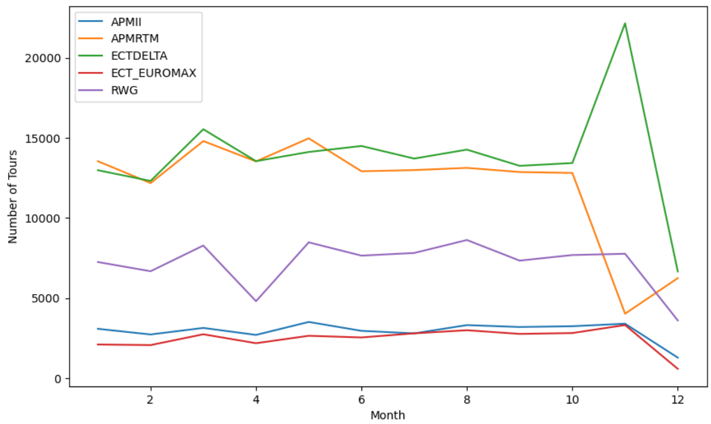


Figure 5.10: Input Truck Flow into Different Terminals Each Month

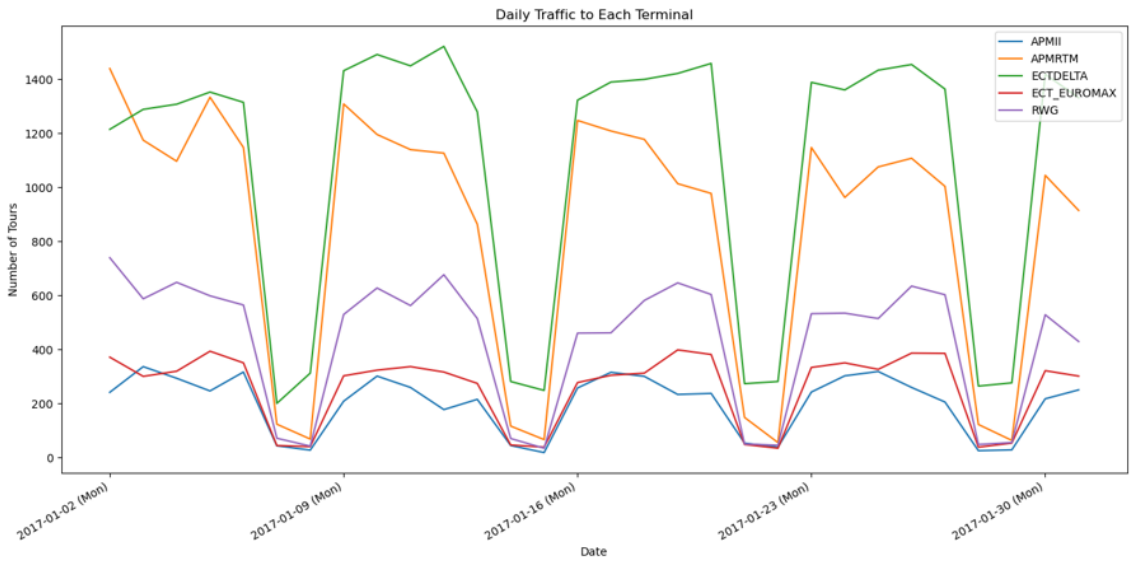


Figure 5.11: Daily Truck Flow to Each Terminal

Fig 5.11 displays daily truck traffic to each terminal over a month. There is a clear recurring weekly pattern, where traffic peaks at the beginning of each week (around Monday or Tuesday) and then gradually declines towards the weekend, reaching a minimum typically on Sundays. Evidence from the figures displayed above proves the apparent periodicity in port operations and truck traffic. Therefore, we also perform a periodic analysis of truck turnaround times.

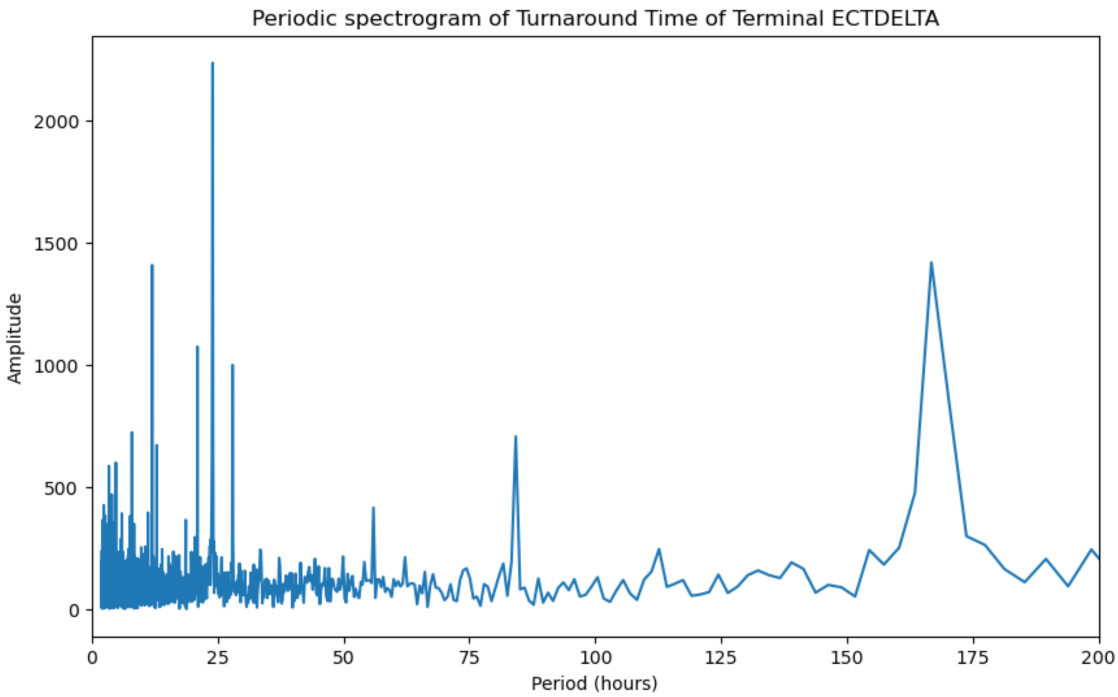


Figure 5.12: Frequency Domain of Truck Turnaround Time in Terminal ECTDELTA

To further investigate the periodicity in truck turnaround times, a frequency domain analysis was conducted using the Fourier Transform. The Fourier Transform is a mathematical technique that decomposes a time series into its constituent frequencies, allowing us to observe periodic components within the data. By converting time-domain data (turnaround times over time) into the frequency domain, we can identify dominant cycles that may reflect operational rhythms or systematic delays in the port's workflow.

In this frequency domain analysis, shown in Fig 5.12, the x-axis represents the period in hours, while the y-axis shows the amplitude of each periodic component. Peaks in amplitude indicate strong periodic components in the turnaround time data. The two peaks appear at 23.97 and 166.76 hours, indicating that truck turnaround time has strong periodic characteristics on daily and weekly basis. This periodic analysis not only confirms the presence of the previous figures, but also revealed that time signals can be transformed into "periods of a day" or "periods of a week" for interpretation to improve prediction accuracy.

Chapter 6

Model Training and Results

Building on the processed and analyzed datasets from previous sections, we aim to utilize the temporal features and operational patterns identified to develop a holistic forecasting model. Based on the methodology and model architecture defined in the previous sections, this chapter elaborates on the training process of the proposed LSTM model and presents the prediction results. To evaluate the performance of the model's predictions, results obtained from different forecasting methods are compared. On this basis, a sensitivity analysis of the input factors is discussed.

6.1 Model Preparation and Training

The developed LSTM model has many variables and model parameters. The training performance of the neural network model is influenced by both the input data and the network structure. In this section, we describe the methods used to improve the quality of input data and determine the optimal parameters of the LSTM network through experiments.

6.1.1 Data Cleaning and Imputation

The selection of valid input data plays a crucial role in enhancing model performance, particularly in the context of time-series forecasting using LSTM. In this study, outliers in the truck turnaround time data were identified and removed based on domain-specific knowledge obtained through expert interviews with professionals from the Port of Rotterdam.

Given the limitations of Bluetooth-based trip identification methods and data quality constraints, certain low-frequency periods with turnaround times outside the predefined range were considered invalid. The acceptable range for truck turnaround time was determined to be between 1 and 2.5 hours. Statistical analysis revealed that more than 95% of the time intervals with average turnaround times outside this range contained only a single recorded truck trip. According to expert insights, such occurrences are considered operationally implausible due to the nature of port operations. Consequently, these low-frequency intervals were labeled as invalid and excluded from the dataset.

The original dataset spanned from January 2, 2017, to December 15, 2017, comprising a total of 8,338 time intervals. Each time period is one hour apart. Among these, 627 intervals had no recorded truck trips, resulting in undefined turnaround times. Additionally, 428 intervals were identified as invalid based on the above-mentioned criteria, leaving 7,283 valid intervals for further analysis.

Given the sequential nature of LSTM-based models and the dependencies inherent in time-series data, removing invalid intervals directly would disrupt the temporal continuity of the input features. Therefore, after filtering out invalid intervals, the remaining missing values are filled using interpolation techniques. This step is essential to maintain data integrity and ensure that the model received a consistent input sequence, which is crucial for improving prediction performance.

To address missing values after outlier removal, various interpolation methods were evaluated to determine their impact on model performance. The effects of these interpolation methods were compared based on the most basic LSTM neural network. The methods considered include mean imputation, linear interpolation, K-Nearest Neighbors (KNN) interpolation, and Multiple Imputation by Chained Equations (MICE).

- Mean Imputation: Replaces missing values with the mean or a specific quantile value of the observed data.
- Linear Interpolation: Estimates missing values by linearly connecting the preceding and succeeding observations.
- K-Nearest Neighbors (KNN) Interpolation: Predicts missing values by averaging or weighting the values of the nearest neighbors based on a distance metric, capturing patterns from similar data points.
- Multiple Imputation by Chained Equations (MICE): Generates multiple imputations for missing data using iterative regressions, accounting for variability and potential relationships among variables.

For fixed value imputation, two scenarios were tested: filling missing values with 0 values and the 5th percentile of the truck turnaround time distribution. The evaluation was conducted using three performance metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2-score). Table 6.1 summarizes the interpolation experiment results, highlighting the performance differences among the tested methods.

Table 6.1: Performance Comparison of Interpolation Methods

Interpolation Method	**RMSE**	**MAE**	**R^2-score**
Fixed Value (0 value)	0.50	0.33	0.24
5 percentile imputation	0.30	0.21	0.17
Linear Interpolation	0.36	0.25	0.08
KNN Imputer	0.35	0.24	0.05
MICE	0.37	0.22	0.11

Considering the performance of different interpolation methods, mean imputation is selected as the final imputation approach to enhance the quality of the dataset. Additionally, as illustrated in Section 4.2.2, the masking technique is employed in the subsequent prediction process to further mitigate the influence of missing and anomalous values on the prediction results.

6.1.2 Hyperparameter tuning

After completing the dataset imputation, the proposed stacked-LSTM model was developed. The model implementation in this study was carried out in the following development environment: in terms of hardware and operating system, an Intel Ultra 7 155H processor, Nvidia RTX 4060 Laptop GPU, and 32 GB RAM running on Windows 11 were utilized. For software implementation, the model was built using the TensorFlow machine learning engine and the Keras deep learning library.

Before constructing the model, it is essential to determine the model parameters. Some of the general parameters in the model are set based on recommended optimal parameter configurations, as shown in Table 6.2. All the RNN-based models are trained by minimizing the mean square error using the Adam optimization method. The early stopping mechanism is used to avoid over-fitting.

Table 6.2: Common settings of each layer.

Layers	Common settings
LSTM layer	Activation = Sigmoid
Dense layer	Activation = Linear
Training process	Optimizer = 'Adam'
	Learning rate = 0.001
	Loss = MSE
Dropout layer	Dropout rate = 0.2

In this study, hyperparameter tuning was performed to determine the optimal configuration for the proposed stacked-LSTM model. The hyperparameters considered for experimentation included time steps, neuron units in each layer, number of epochs, and batch size. The candidate values for each hyperparameter are as follows:

- ***time_steps***: 24, 72, 168
- ***units***: 32, 64, 128
- ***epochs***: 25, 50, 100
- ***batch_size***: 16, 32, 64, 128

For the epochs parameter, while longer training typically improves model performance, it comes with a significant increase in training time. However, due to the implementation of the early stopping mechanism, most experiments exhibited convergence within 50 epochs. As a result, 50 epochs were chosen as the default value for the training process.

The other three hyperparameters (time_steps, neuron units, and batch size) were evaluated using a grid search approach. Specifically, experiments were conducted by varying the combination of these parameters, and their effects on the model's performance were assessed. The mean absolute error (MAE) was selected as the evaluation metric for hyperparameter optimization. Figure 6.1 presents the 3D grid search results, highlighting the best hyperparameter set results.

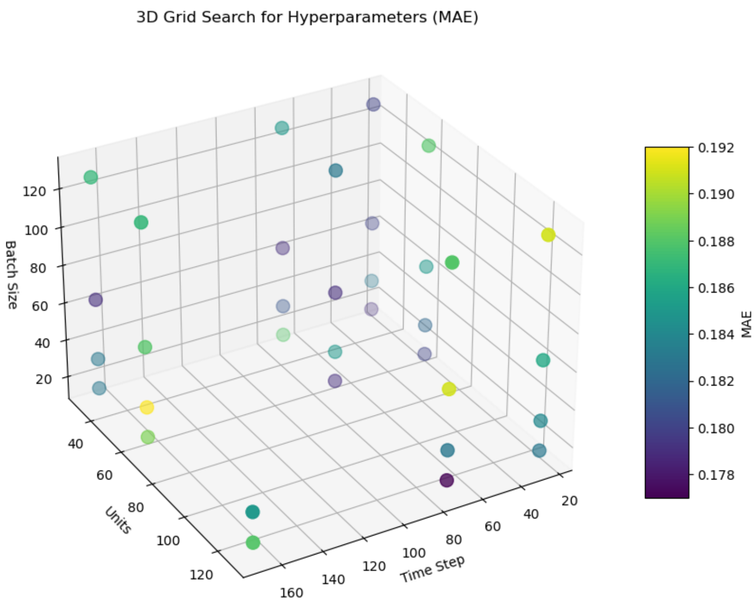


Figure 6.1: Hyperparameter tuning results for time_steps, neuron units, and batch size, evaluated using MAE.

Through hyperparameter experiments, this study can evaluate the performance of the LSTM model and find the optimal parameter combination. For the hyperparameters of the LSTM model, it is necessary to find a balance between its running time and performance. For instance, increasing the number of neuron units generally improves the model's capacity to capture complex patterns in the data. However, an excessive number of units significantly increases training time and may lead to overfitting. Similarly, the choice of batch size affects both training dynamics and prediction accuracy. A smaller batch size allows for finer gradient updates, often resulting in lower prediction errors but at the cost of longer training times and greater gradient oscillations. Conversely, larger batch sizes stabilize gradient updates and reduce training time but are more prone to getting trapped in local minima.

As shown in Fig 6.1, through experimental evaluation, it was observed that a time step of 72 hours, representing a prediction window of 3 days, achieved a favorable balance between predictive accuracy and computational requirements. Regarding batch size, smaller batch sizes consistently yielded better MAE values, suggesting that finer gradient updates were advantageous for this dataset despite the associated increase in training time. From the figure, it can be observed that the combination of *time_steps* = 72, *neuron units* = 128, and *batch size* = 16 achieves the best performance with the lowest MAE. Based on the numerical experimental results and empirical insights, further adjustments to the model's hyperparameters were made. Under the finalized model architecture, the following settings were selected for subsequent training: the number of hidden layer units was set to 128, the number of epochs to 50, the activation function to Sigmoid, the dropout rate to 0.2, the time step to 72 hours, and the batch size to 16.

6.1.3 Imputation method validation

To validate the effectiveness of the proposed data cleaning pipeline, we conducted a controlled experiment comparing model performance on raw unprocessed data versus the cleaned dataset. The same Stacked LSTM architecture and hyperparameters were applied to isolate data quality impacts.

Table 6.3: Performance Comparison Between Raw and Cleaned Datasets

Dataset	R^2	RMSE	MAE
Raw Data	0.209	0.338	0.234
Improved Data	0.418	0.249	0.173

The results indicate a significant enhancement in model performance after applying the data cleaning and imputation methods. The R^2-score, which measures the proportion of variance explained by the model, increased from 0.209 to 0.418, signifying a considerable improvement in predictive accuracy. Additionally, both RMSE and MAE, which quantify prediction errors, saw notable reductions, with RMSE decreasing from 0.338 to 0.249 and MAE from 0.234 to 0.173. These findings confirm that the removal of implausible data points, coupled with appropriate imputation strategies, leads to a more robust and reliable model.

Further analysis of the residual distribution of model predictions on both datasets revealed that predictions on the cleaned dataset exhibited a tighter error distribution with fewer extreme deviations. This suggests that the imputed dataset not only improves accuracy but also enhances the model's generalizability.

In sum, the data cleaning and imputation approach significantly contributes to the model's performance. By filtering out operationally implausible records and applying suitable imputation techniques, the model is better equipped to learn meaningful patterns, ultimately leading to more accurate and reliable truck turnaround time predictions.

6.2 Prediction Results and Performance Analysis

6.2.1 Model Performance

Under the selected hyperparameter combination, the prediction results of the model are shown in the figure 6.2.

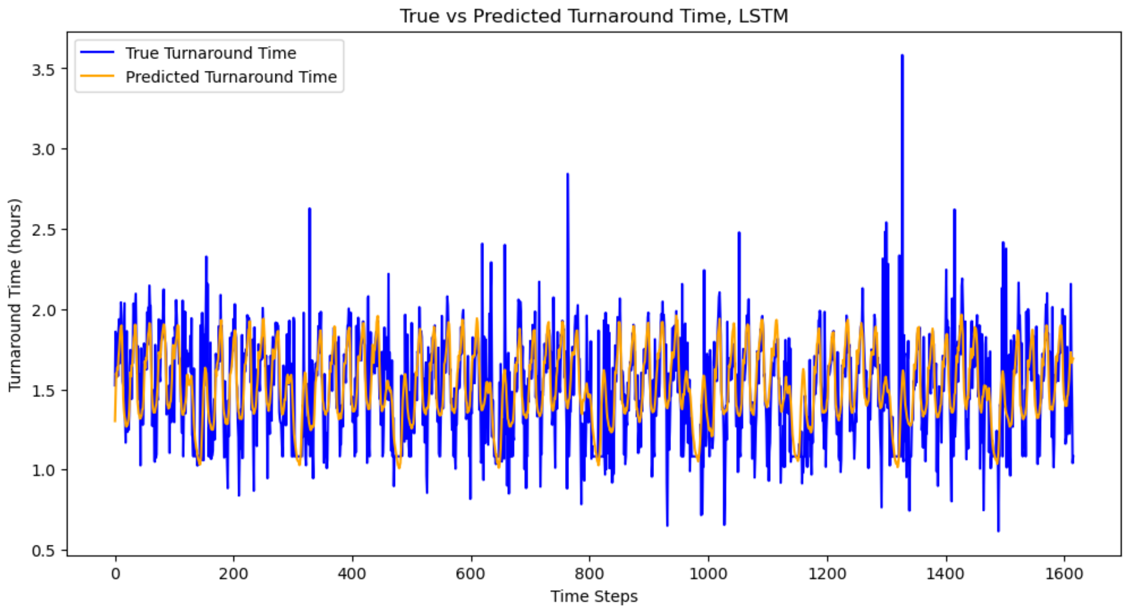


Figure 6.2: Prediction Results of the proposed stacked-LSTM model

The blue line in the figure represents the actual truck turnaround time, and the orange line represents the predicted turnaround time. The predicted values generally follow the fluctuation trend of the real values well. A clear pattern of fluctuations can be observed from the prediction results. Generally, the proposed LSTM model well captured the cyclical changes, and in most cases the predicted values align well with the true values at the peaks and troughs in the time series.

Specifically, some local peaks, such as those around time steps 600 and 1400 in the validation set, were not accurately predicted. Abnormal fluctuations in peaks are often caused by short-term external events such as sudden terminal congestion, equipment failure, which may not be fully reflected in the obtained data and selected input features. The LSTM model captures long-term dependencies through memory units, which has a certain smoothing effect on noise and short-term extreme changes in the input sequence. When faced with severe fluctuations, models tend to predict values close to the overall trend rather than capturing these changes according to the set of the loss function. Additionally, the model fails to capture the deviation between some of the predicted troughs and true values.

Although the prediction errors at some time steps are large, the overall error distribution is relatively uniform, and there is no significant systematic overestimation or underestimation of the model's predicted values. This shows that the model has good generalization ability in the overall range and does not cause obvious bias problems.

6.2.2 Benchmarking

To evaluate the performance of the proposed model, several kinds of models for truck turnaround time prediction are compared with the prediction results. This study selected linear regression models, XGBoost models, random forest (RF) models, single-layer LSTM models, and multi-layer LSTM models based only on turnaround time for comparison. The rationale and selected parameters of these models are presented below.

- Linear Regression: Linear regression is a simple and interpretable baseline model that assumes a linear relationship between input features and the target variable. This model was included to establish a reference point for performance comparison. The parameters of the linear regression model were left at their default settings, as the simplicity of the model requires minimal tuning. The model is implemented using the Sklearn package in python[38], with the kernel function RBF (Radial Basis Function) selected.
- XGBoost Algorithm: XGBoost is a highly efficient gradient boosting algorithm that can handle non-linear relationships between input features and the target variable. It is widely used for predictive tasks due to its flexibility and robustness. In this study, key hyperparameters such as the number of trees, learning rate, and maximum depth were optimized using grid search. The final model settings included a learning rate of 0.1, a maximum depth of 6, and 100 estimators.
- Random Forest: Random forest is an ensemble learning method that builds multiple decision trees and aggregates their outputs to improve prediction accuracy and reduce overfitting. Its capability to model complex interactions between features makes it a strong benchmark. In this study, the number of trees was set to 100, and the maximum depth was adjusted to 10 to balance accuracy and computational efficiency. The model is implemented using the Sklearn package, and the amount of estimators is set at 100, max depth at 10.
- Single-layer LSTM: A basic LSTM structure is implemented to compare the results. The network contains one LSTM layer with 64 hidden units. The configuration settings are the same as the stacked model. The model is implemented using the TensorFlow and Keras libraries in python.
- Multi-layer LSTM: Multi-layer long short-term memory network (Multi-layer LSTM) can capture patterns at different time scales by stacking multiple LSTM layers. The network consists of 3 LSTM layers with no pre-designed structure. The configuration settings are the same as the single-layer model.

Under the same dataset and sample division, we compared the performance of the proposed stacked LSTM model with other models to evaluate their advantages and disadvantages in predicting truck turnaround time within the context of this study. The performance of the models was measured using RMSE, MAE, and R-square metrics. The prediction comparison results are shown in the figure 6.3 and figure 6.4.

Compared to the basic LSTM models and the linear regression model, the proposed Stacked LSTM model demonstrates superior performance on the test dataset in terms of RMSE, MAE, and R-square. When compared with XGBoost and Random Forest, the overall prediction accuracy is similar. As shown in Figure 6.3, the Linear Regression model has the shortest training time, while the three LSTM models require relatively longer training times.

Analyzing the prediction curves of each model in more detail, it can be observed from Figure 6.4 that due to the introduction of the mask mechanism in the dataset, where masked data were filled with fixed values, the tree-based models (Random Forest and XGBoost) performed particularly well in predicting the invalid data periods. This is likely because tree-based methods excel at identifying patterns in such fixed-value inputs. RF has the advantages of fast training speed, good generalization performance, and strong resistance to over-fitting. During the training process, RF will randomly select a part of the features to build each decision tree to ensure the randomness of the features and thereby reduce the variance of the model. XGBoost utilizes an additive model that sequentially builds decision trees, where each new tree is trained to correct the residual errors of the previous ones. In this case, both decision tree models have similar performance, with Random Forest performing slightly better.

However, in terms of overall prediction accuracy, these models show a slight disadvantage in predicting valid data compared to the Stacked LSTM model, which is better at capturing temporal dependencies and generating smoother predictions. This indicates a trade-off: while tree-based models are robust in handling invalid data segments due to the mask, their performance in predicting valid data points is slightly weaker, making the Stacked LSTM model more suitable for datasets with strong temporal patterns and a focus on valid data predictions.

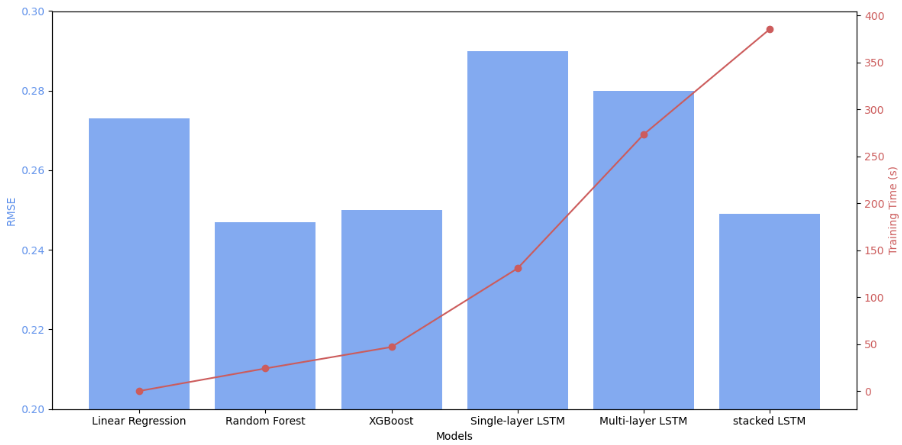
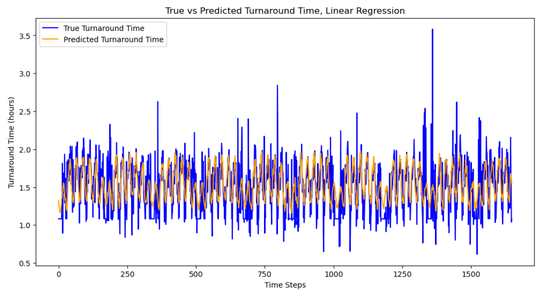
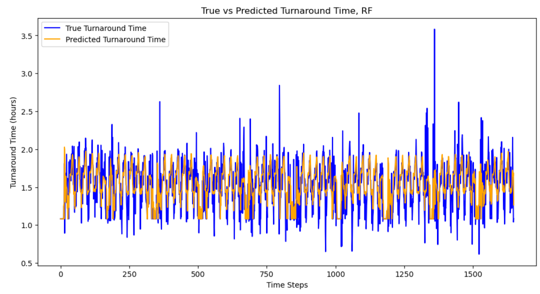


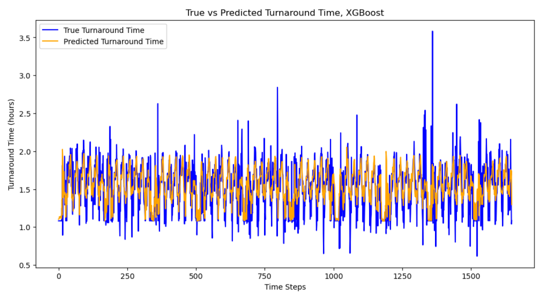
Figure 6.3: Result comparison of different models



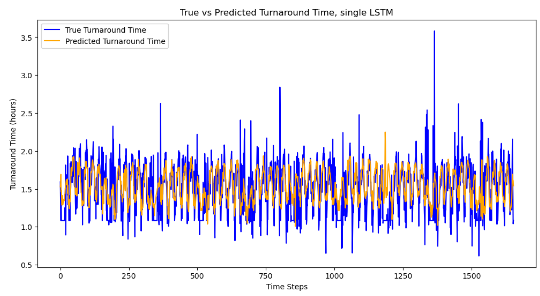
(a) Linear Regression



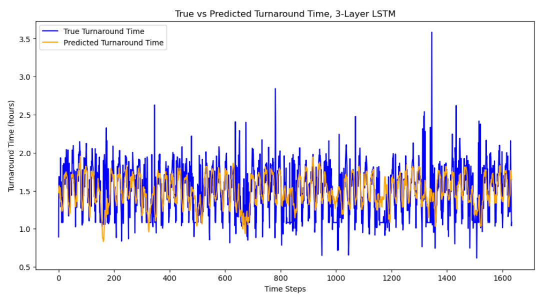
(b) Random Forest



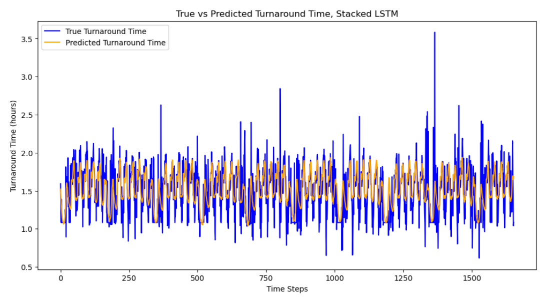
(c) XGBoost



(d) Single-layer LSTM



(e) Multi-layer LSTM



(f) Stacked LSTM

Figure 6.4: Comparison of results for different models

Based on the Table 6.4, the proposed Stacked LSTM model demonstrates significant advantages in the given prediction scenario. It achieves an RMSE of 0.249 and an MAE of 0.173, which are comparable

Table 6.4: Comparison of Model Performance Metrics

Model	RMSE	MAE	R-square	(RMSE) GAP%
Stacked LSTM	0.249	0.173	0.418	-
Linear Regression	0.273	0.191	0.367	9.64%
Random Forest	0.247	0.168	0.420	-0.80%
XGBoost	0.250	0.172	0.414	0.40%
Single LSTM	0.290	0.213	0.290	16.47%
Multilayer LSTM	0.280	0.202	0.320	12.45%

to the best-performing Random Forest model (RMSE = 0.247, MAE = 0.168) and slightly better than XGBoost (RMSE = 0.250, MAE = 0.172). Furthermore, the Stacked LSTM model shows a notable improvement in R-square (0.418), indicating its superior ability to explain the variance in the data compared to other LSTM-based models and Linear Regression.

While Random Forest slightly outperforms the Stacked LSTM in RMSE (-0.8% GAP), it is important to note that Random Forest's superior handling of masked data might have contributed disproportionately to its performance. In contrast, the Stacked LSTM model excels in capturing temporal dependencies, making it better suited for valid data predictions in this context. Additionally, the Stacked LSTM surpasses both Single-layer LSTM and Multilayer LSTM models by a significant margin in all metrics, highlighting the effectiveness of stacking layers in leveraging sequential patterns.

6.2.3 Prediction Results by Terminal

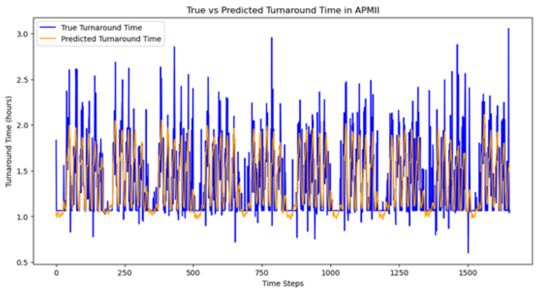
In this section, we analyze the prediction results for individual terminals, where the same input features were used across all terminals. However, due to the sparsity of the dataset for certain time periods at specific terminals, a significant proportion of the data corresponds to invalid periods, during which the truck count is zero. These invalid periods were represented as fixed turnaround time in the dataset due to the interpolation method implemented during preprocessing. To mitigate the influence of such invalid periods on prediction accuracy, a masking mechanism was implemented.

The masking method is implemented through the prediction procedure and the invalid periods were excluded during the computation of evaluation metrics. The results presented below in Table 6.5 include key metrics such as RMSE, MAE, and R^2, which were computed after removing the masked invalid periods. Furthermore, the true vs. predicted turnaround time plots for each terminal provide a visual comparison, highlighting the model's performance in predicting the valid periods.

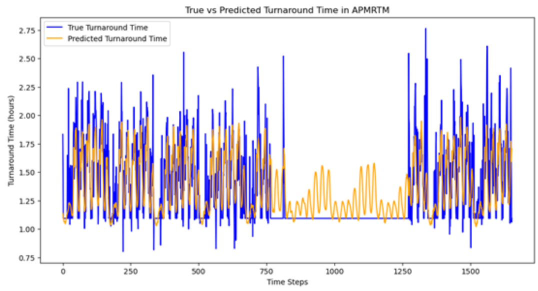
Table 6.5: Prediction Performance Metrics by Terminal

Terminal	RMSE	MAE	R-square
APMII	0.323	0.238	0.369
APMRTM	0.252	0.184	0.451
ECTDELTA	0.254	0.196	0.375
ECT_EUROMAX	0.320	0.232	0.410
RWG	0.269	0.203	0.466

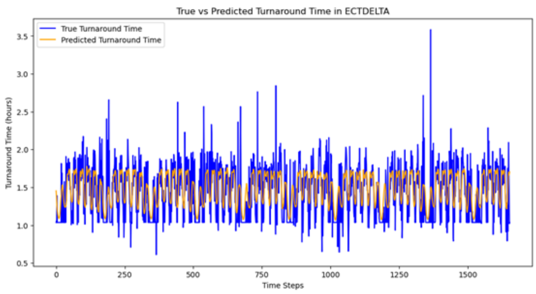
The prediction results differ among terminals, reflecting the varying data quality and different characteristics. From the evaluation metrics, it is evident that APMRTM and RWG exhibit the best predictive performance among the terminals. APMRTM achieves an RMSE of 0.252 and an R-suqare of 0.451, while RWG records an RMSE of 0.269 and the highest R-square value of 0.466. In contrast, EUROMAX shows moderate results, with an RMSE of 0.320 and an R-square of 0.410, while APMII and ECTDELTA demonstrate relatively poorer performance, with RMSE values of 0.323 and 0.254 and R-square values of 0.369 and 0.375, respectively.



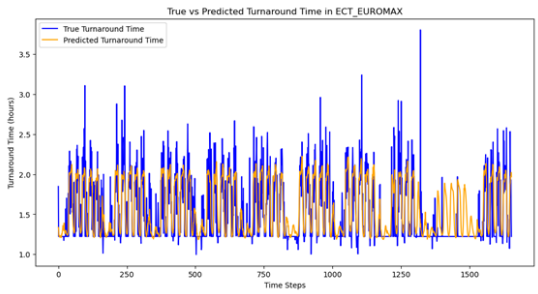
(a) APMII



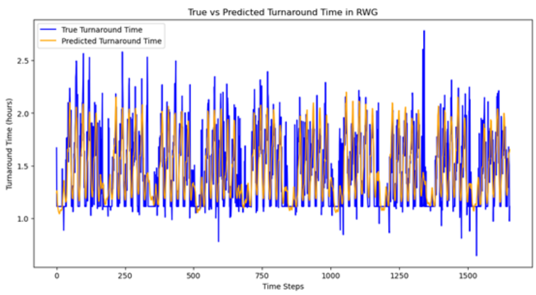
(b) APMRTM



(c) ECTDELTA



(d) ECT_EUROMAX



(e) RWG

Figure 6.5: True vs Predicted Turnaround Time for Different Terminals

From a visual perspective, the comparison plots further substantiate these findings. Among the terminals, RWG demonstrates the best fit, with the predicted curve closely following the true values throughout most valid data periods. The prediction results also benefit from having the most extensive valid data periods, enabling the model to learn consistent temporal patterns effectively.

Apart from the strong metrics APMRTM has reached, a prolonged period of data missing creates a gap in the data where no predictions are required. In the valid periods, the predictions in APMRTM captures the overall trend, except for occasional sharp peaks.

For ECTDELTA and APMII, both terminals show clear signs of overfitting to the periodicity of the data. This is evident from the predicted curves, where the model replicates periodic patterns excessively, even when they deviate from the true values. The evaluation metrics reflect this limitation, with lower R-square values and relatively higher RMSE and MAE scores compared to better-performing terminals. The overfitting to cyclical trends indicates that the model struggles to generalize well for terminals with more sparse or irregular data distributions.

From an overall perspective, the developed prediction model behaves varingly among different terminals due to their unique operational characteristics and data quality. While the masking mechanism successfully eliminates the impact of invalid periods, the model's performance is still contingent on the quality and quantity of valid data. Additionally, the limited ability of the selected input features may not be able to fully describe the operational characteristics of the terminals. Due to constraints in data availability, the input features may not capture critical aspects of terminal operations. This limitation likely affects the model's ability to identify and predict sudden peaks in turnaround times. Therefore, when applying prediction models to different terminals, it is still necessary to adjust the model according to the characteristics of the terminal itself to achieve better prediction results.

6.3 Factor Sensitivity Analysis

In order to explore the impact of different input features on the average turnaround time prediction accuracy, this section decomposes the input factors of the model. These input factors include historical average turnaround time, wind speed, container arrivals at each terminal, and truck arrivals. To systematically evaluate the contribution of each feature, the experiments were designed to start with a baseline model containing all input features and gradually eliminate individual features or specific combinations to observe the resulting changes in prediction performance. This elimination method enables the identification of critical features whose removal significantly impacts the model's accuracy.

The experiments were conducted using the Stacked LSTM model with consistent architecture, training and testing datasets, and training parameters across all experiments. The design includes tests where each individual feature is removed (e.g., historical average turnaround time, wind speed, container arrivals, or truck arrivals) to analyze its standalone impact on the predictions. In addition, specific feature combinations were excluded to evaluate the influence of the combined effect of internal and external factors, also to observe the inner connection of certain features. For example, historical turnaround time and truck arrivals were removed together to assess the effect of excluding traffic-related dynamics, while wind conditions and container arrivals were excluded to test the significance of terminal-side operational factors. The experimental design for this sensitivity analysis is shown in Table 6.6.

Table 6.6: Experimental Design for Sensitivity Analysis Based on Input Features

Experiment Number	TTT	Wind	Container Count	Truck Arrival	Periodicity
1	✓	✓	✓	✓	✓
2	-	✓	✓	✓	✓
3	✓	-	✓	✓	✓
4	✓	✓	-	✓	✓
5	✓	✓	✓	-	✓
6	✓	✓	✓	✓	-
7	✓	-	-	✓	✓
8	✓	✓	-	-	✓
9	-	✓	✓	-	✓

The experimental results of sensitivity analysis are shown in the figure 6.6. In the single-feature analysis, removing any feature negatively impacted prediction accuracy. Among them, current wind speed (environmental factor) and truck arrivals had the most significant influence, with their removal causing a noticeable increase in RMSE. From an operational perspective, adverse wind conditions can render terminal operations infeasible or significantly inefficient, leading to simultaneous shutdowns on the landside and quayside. The absence of wind condition information makes it difficult for the model to capture unexpected shutdowns. This parameter is crucial for all terminals in the Maasvlakte area. Similarly, truck arrivals have a clear real-world correlation with average truck turnaround times. A higher queue of trucks inside the port directly impacts the turnaround time of subsequent trucks.

In contrast, container arrival information and periodicity did not significantly affect prediction accuracy. While previous analyses have shown that port turnaround times exhibit notable daily or weekly periodicity, the model appears capable of capturing these temporal patterns from the time-series data

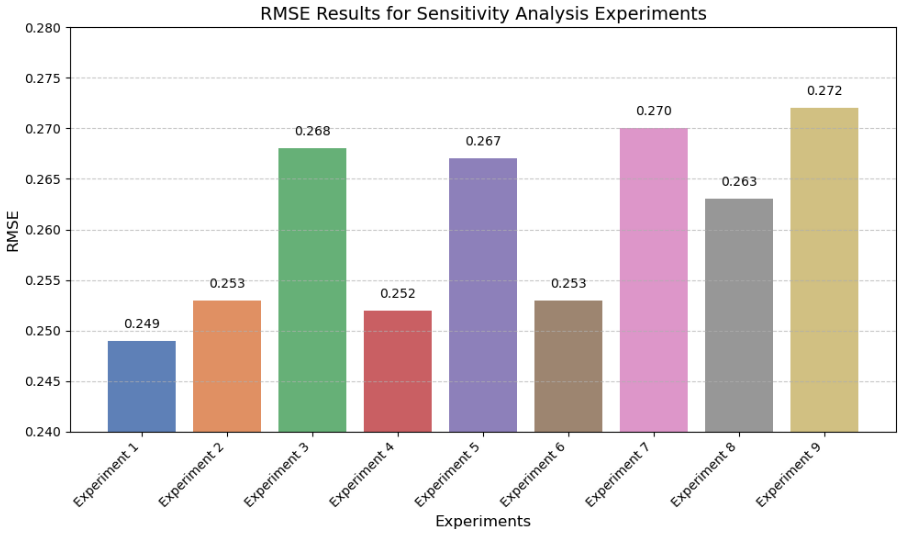


Figure 6.6: Result of Sensitivity Experiments

itself. In the double-feature combination experiments, the removal of any of the three designed feature combinations also led to a significant increase in RMSE. The results of Experiments 2 and 9 highlight the substantial contribution of truck arrivals to improving prediction accuracy. Meanwhile, Experiment 7 performed slightly worse than Experiment 3, suggesting that container arrival counts, as a workload indicator, can complement other features to enhance the model's performance. The experimental results demonstrate that removing any single feature or feature combination results in predictions that are less accurate than those obtained by using all features together.

6.4 Model Validation

To evaluate the robustness and generalizability of the proposed Stacked LSTM model, a 5-fold cross-validation approach was adopted. This method divides the dataset into five equal-sized subsets, or "folds." For each iteration, one fold is used as the validation set while the remaining four folds are used for training. The process is repeated five times, with each fold serving as the validation set once. The final performance metrics are computed as the average across all folds, providing a more reliable estimate of the model's predictive capabilities compared to a single train-test split.

The results of the cross-validation experiments are summarized below. Each fold was trained and evaluated independently, with key metrics such as Root Mean Square Error (RMSE) and R-square recorded for every fold. The aggregated results from the 5-fold cross-validation are as follows:

Table 6.7: Cross-Validation Results for Each Fold

Fold	RMSE	R^2
Fold 1	0.256	0.395
Fold 2	0.249	0.405
Fold 3	0.259	0.388
Fold 4	0.247	0.412
Fold 5	0.252	0.384

The aggregated results from the 5-fold cross-validation are as follows:

- Mean RMSE: 0.254
- Mean R^2: 0.397

The cross-validation results indicate that the proposed Stacked LSTM model achieves a mean RMSE of 0.254 and a mean R^2 of 0.397, demonstrating reliable performance in predicting truck turnaround times. The low RMSE reflects the model's accuracy in minimizing prediction errors, while the R^2 value shows that the model captures a significant portion of the variance in the data. These results confirm the model's robustness and generalizability across different data splits, providing a strong basis for its application in real-world scenarios.

6.5 Reflection from the Prediction Results

This chapter focuses on the training of the proposed model architecture and compares the results obtained using different numerical imputation strategies and parameter combinations. Based on the optimal numerical imputation strategy and parameter configuration, the model was trained to obtain prediction results and conduct performance analysis. Furthermore, a factor sensitivity analysis was performed on the input features of the model, providing insights into parameter importance. Finally, k-fold cross-validation was applied to the training results, demonstrating the model's relatively reliable performance in predicting truck turnaround times.

Several findings emerged from this chapter.

- All tested models struggled to accurately predict the sharp peaks in the dataset. This limitation is largely due to the quality and scope of the original dataset, which may lack the detailed information needed to explain these extreme variations. Improving data quality and adding more informative features could help address this issue in the future.
- In terms of the prediction results of each terminal, the model shows different performance, showing the different operational characteristics and operational differences of the terminal. It also shows that in actual operations and practical application of the model, it is necessary to adjust accordingly and make decisions based on the particularity of each case.
- Sensitivity analysis and k-fold cross-validation confirmed the model's effectiveness and reliability. The sensitivity analysis provided helpful insights into the relationship between input features and truck turnaround times. However, due to the limitations of neural networks, the model's ability to explain these relationships in detail is somewhat restricted.

Chapter 7

Discussion and Insights

This chapter first provides a systematic answer to the research questions of this article, and then discusses the problems in the research process and the insights found on the basis of answering the research questions.

7.1 Answers to the research questions

The research questions and sub-questions defined in this article are as follows:

How can a holistic model be built to predict truck turnaround time at terminals based on the data available?

- How to identify container truck trips from the Bluetooth detection records around the port area?
- Which factors are crucial in predicting truck turnaround time and also available in the research scope?
- How can container information and Bluetooth data sets be processed into a finalized dataset as the input to the model?
- What kind of approach can achieve higher prediction accuracy based on a fixed level or comparing results?
- How can the prediction results be verified and validated?

The sub-questions are answered as follows.

How to identify container truck trips from the Bluetooth detection records around the port area?

To address this question, the Bluetooth detection records around the port area are collected to identify vehicle trajectories. The original Bluetooth data set obtained is not for the special purpose of this project, so it contains a large amount of invalid and noisy data and is in the format of m file. In order to convert the detections to vehicle trips, a data preprocessing pipeline was developed, including filtering for unique MAC addresses, spatial clustering to associate detections with entry and exit points, and temporal analysis to calculate trip durations. Rules were established based on trajectory patterns typical of container trucks. The obtained vehicle trips are combined with GIS information to reveal the behavioral characteristics of different types of vehicles. Therefore, vehicle categories are double verified in terms of travel time and turnaround time.

The method successfully identifies a significant proportion of container truck tours. More than 90% of the vehicles identified as container trucks passed the dual verification of travel time and turnaround time, demonstrating the accuracy of the identification method.

Which factors are crucial in predicting truck turnaround time and also available in the research scope?

To address this question, an analysis of the demarcation of the container turnaround process is conducted. Starting from the problem identification, the study scopes influencing factors into operational, container-specific, and environmental categories. Operational factors include truck arrival flow, terminal congestion levels, and handling type. Container-specific information involves arriving vessels and container handling requirements. Environmental factors include weather conditions influencing the operation of the terminals.

Based on the data we can obtain, preliminary descriptive analysis reveals the data features and their correlation with truck turnaround time. Then, factor refinement aims to implement feature engineering and create refined factors that are more representative of the raw data. Features selected as having a strong correlation with truck turnaround time include truck arrivals, container arrivals, wind speed, and time periodicity.

How can container information and Bluetooth datasets be processed into a finalized dataset as the input to the model?

The processing included time alignment of records, handling missing data using interpolation techniques, and feature engineering to derive relevant metrics (e.g. waiting times, entry-to-exit durations). After integrating data from all sources into one dataset, the resampling results at different scales are tested. Considering the practical significance and the large number of missing values in smaller time windows, one hour is selected as the final time step for prediction. Normalization and one-hot encoding were applied for numerical and categorical variables, respectively.

What kind of approach can achieve higher prediction accuracy based on a fixed level or comparing results?

This study selected stacked LSTM as the model for expected development and compared it with other traditional or machine learning prediction methods during the model training process. All the models have been parameter-tuned to achieve the best prediction results. During the training process, some other mechanisms are also tested to improve the quality of training and the accuracy of prediction results, some of which are adopted and some are not. The mechanisms that have not been adopted include the attention mechanism, while the methods that have been adopted include the masking mechanism, etc. Hyperparameter sensitivity experiments are performed to obtain the optimal parameter set.

The developed stacked-LSTM model demonstrated the best prediction effects, displaying its ability to capture temporal dependencies in the data. Random Forests and XGBoost also perform well, but they are less effective when considering the artificially introduced fixed value interpolation. This undermines their credibility in accurately capturing true changes in turnaround times over changing time domains.

How can the prediction results be verified and validated?

K-cross validation test is implemented on the developed model, and metrics such as root mean square error (RMSE) and R-squared value are used to evaluate prediction accuracy. The internal validation on the data set proves the model's performance. In the context of this study, there is no condition to implement external validation by comparing predictions with actual terminal operations.

7.2 Challenges and Discussions

After addressing all these sub-questions, the framework and results of this research are clearly listed. However, it is essential to discuss the challenges faced during the study.

7.2.1 Bluetooth Data Quality and Availability

One major challenge is the inconsistency and noise in the Bluetooth detection data. The original data source was used to investigate the traffic flow information around the port area, and the data quality occasionally limited the precision of truck trip identification. Due to the nature of the Bluetooth detection technique, that detectors communicate with all Bluetooth-enabled devices within their range, the dataset contains numerous duplicate detection records within small time intervals. Some of these intervals may be as long as 50 to 100 seconds or more. These duplicates are difficult to filter out entirely, as the fine-grained distinctions required to confirm their redundancy could inadvertently lead to the removal of valid data.

Furthermore, Bluetooth detection has its inherent limitations. A non-negligible proportion of vehicles are either not equipped with Bluetooth devices, or might not be detected by the detectors sometimes due to signal occlusions or technical issues. This could lead to incomplete trip records, or erroneous representations of traffic flow. A significant number of detected vehicle passages could not be paired to form complete trips due to missing corresponding entry or exit records. Also, in some scenarios, the detected trips failed to accurately represent the actual flow of vehicles into and out of the port. The resulting non-systematic errors are unpredictable and will also lead to insufficient information fed back to the model, which may affect the prediction accuracy.

In related research, the above-mentioned Bluetooth data set disadvantages are difficult to be effectively avoided. Especially in terms of trip identification, since this study identifies the turnaround time of container trucks entering and leaving the port area, it is difficult to identify trips simply through travel time. These issues underline the need for additional data validation mechanisms and complementary data sources. Data collected from GPS-based fleet management system or automated license plate recognition (ALPR) could help mitigate the limitation of Bluetooth detection. Bluetooth data may be useful in mapping traffic around the port, but it falls short in terms of data quality.

7.2.2 Data Comprehensiveness and Feature Representativeness

The comprehensiveness of information and the representativeness of extracted features contribute to the predictive model for truck turnaround time. Multi-source data integration enhances the ability to capture the complexity of port operations, theoretically enabling better predictions compared to models relying solely on historical traffic flow or TTT data. However, the extent of information available and the model's ability to extract meaningful features from these data sources determine the achievable accuracy and applicability of the predictions.

In this study, multi-sourced data is integrated to build a holistic dataset. This approach enables the extraction of representative features which significantly contribute to improving the model's prediction accuracy. However, there is more information that cannot be reflected in the limited data we have. For example:

- Driver behavior and decision logic in port area: The operational processes of container truck drivers within the port are more complex than simply entering, picking up cargo, and exiting. Some drivers may spend extended periods, from tens of minutes to several hours, in rest areas, while others might deliver a container at one terminal and then pick up another at a different terminal. In this study, the available data does not capture such route selection choices or waiting strategies, which constrains the model's ability to represent the nuanced micro-level operational dynamics accurately.
- Equipment capacity and container handling dynamics: Terminal operation dynamics are difficult to obtain, while standard operating procedures for container handling within the port and their influence on container turnaround times are absent from the dataset. It would be better to have data linking container arrivals with hinterland transport documents available for this study.

The findings of this study underscore the potential of multi-source data integration in enhancing predictive capabilities for TTT. By leveraging diverse data sources, the study demonstrated the feasibility of extracting representative features that capture the temporal and operational dynamics of port logistics. The limitations also point to opportunities for further advancements. Future research could focus on the directions listed below:

- Expanding data sources to enhance the granularity of information
- Advanced feature engineering techniques to improve the representativeness of extracted features
- Addressing missing data through imputation frameworks or supplementary systems
- Expanding the research scope to include interactions between container operations and hinterland logistics

7.3 Prediction Results and Model Validation

While the model shows potential in predicting TTT, the achieved prediction accuracy did not fully meet initial expectations. Some of the reasons are discussed above, and this section talks about the model. Although advanced machine learning techniques such as stacked LSTM captured temporal patterns effectively, inherent variability in TTT due to unmeasured factors limited overall performance.

Compared with other methods, the stacked LSTM model demonstrates better prediction accuracy, but the characteristics of the neural network model pose challenges in terms of interpretability. The model's black-box nature made it difficult to quantify the influence of individual features. While techniques such as feature importance rankings and sensitivity analyses were explored, the dynamic interactions between features and the temporal dependencies modeled by the LSTM limited the precision of these assessments.

The validation process faced significant constraints due to the absence of external validation opportunities using real-world operational data. The lack of live testing under actual port conditions limited confidence in its practical applicability.

The developed model presents both strengths and limitations in terms of its practical application and generalizability. A key advantage is the creation of a generalizable framework for integrating multi-source data to predict TTT. However, the model's adaptability comes with the need for significant customization at each port to ensure optimal performance. Variability in operational practices, data availability, and infrastructure across terminals poses hurdles to broader adoption. Future enhancements could focus on reducing the customization effort required for deployment while maintaining the model's robustness and flexibility.

7.4 Further Development Directions

The proposed truck turnaround time prediction model offers possible insights for optimizing port operations. By expanding the model's scope to address systemic challenges and leveraging real-time data integration, the Port of Rotterdam can achieve higher efficiency, sustainability, and stakeholder satisfaction.

- the currently PortAlert app used by truck drivers and terminal operators is an ideal platform for deploying truck turnaround time predictions. By embedding the model's outputs into PortAlert, drivers could receive real-time forecasts with a certain confidence intervals (e.g. 1.9 to 2.3 hours at 10:00 AM to 11:00 AM. Terminal operators, meanwhile, could use a backend dashboard to monitor predicted congestion and adjust operations preemptively such as opening overflow lanes in terminals. Over time, user feedback from PortAlert could refine the model's accuracy, creating a closed-loop system where predictions and operations co-evolve. The proposed model is more accurate and responds to changes faster than existing experience-based inferences in the APP, which could benefit truck drivers, fleet managers and port operators to better allocate their time and resources.
- A system dynamic approach can transform the prediction model from a tool to a component of a broader decision-support system. SD models capture interdependencies between port subsystems—truck arrivals, terminal operations, and hinterland logistics—enabling the evaluation of long-term policy impacts. Embedding TTT predictions in an SD framework would help quantify

how dynamic truck appointment pricing influences congestion, emissions, and equipment utilization. Incorporating environmental and external factors, such as wind conditions affecting quay operations or vessel schedule disruptions, allows the model to forecast ripple effects across the logistics chain, supporting proactive mitigation strategies. Additionally, The current model predicts TTT at individual terminals, but congestion often stems from imbalanced truck distribution across terminals. A cross-terminal collaboration framework, powered by real-time TTT predictions, could dynamically allocate trucks to underutilized terminals, smoothing workload distribution. This strategy requires integrating terminal-specific data—such as yard occupancy, crane availability, and hinterland connectivity—into the prediction model. Additionally, collaboration could extend to shared resource pools, such as mobilizing straddle carriers or personnel between terminals during peak periods.

Chapter 8

Conclusions and Recommendations

Container terminals serve as crucial hubs in international maritime logistics, acting as transshipment points for goods transported globally. Container trucks are vital for connecting terminals with inland logistics, and accurately predicting truck turnaround time can significantly enhance terminal operations. This study proposes a machine learning-based framework for turnaround time prediction, utilizing multi-source data integration and advanced neural network models. By improving turnaround time prediction accuracy, this research contributes to optimizing terminal operations, reducing congestion, and supporting better decision-making for logistics companies, terminal operators, and port authorities.

The research of this article mainly focuses on Bluetooth trip identification, prediction model construction and performance comparison. First, by filtering and integrating Bluetooth detection data, an efficient itinerary identification method was developed, achieving an identification accuracy of more than 90% under double verification. In the construction of the prediction model, this article designed and implemented a stacked long short-term memory (stacked LSTM) network. This model uses multi-source data to capture the temporal dependencies in port operations and improves the prediction accuracy through the optimization of hyperparameters. In the sensitivity analysis, this paper found that container truck arrival volume and wind speed are the variables that have the greatest impact on turnaround time, further verifying the advantages of multi-source data integration. In the sensitivity analysis, this paper found that container truck arrival volume and wind speed are the variables that have the greatest impact on turnaround time, further verifying the advantages of multi-source data integration.

Compared with traditional forecasting research, this paper not only relies on historical traffic flow and terminal operation data, but also introduces environmental variables and container dynamic information, thereby comprehensively capturing the complexity of terminal operations. At the same time, the use of stacked LSTM models breaks through the limitations of traditional regression models and can more effectively model nonlinear characteristics and time series patterns of data. In addition, this article innovatively introduces a masking mechanism to deal with the problem of missing data, which significantly improves the robustness of the model in scenarios with incomplete data quality.

The research in this article has important implications for managers. The prediction results can not only guide logistics companies to adjust the arrival time of container trucks to avoid congestion during peak periods, but also provide a decision-making basis for terminal managers to optimize resource allocation. Compared with the previous method of forcibly allocating container truck arrival times through quotas, this paper's method is more flexible. It guides container truck companies to spontaneously adjust scheduling strategies by publishing forecast information, effectively balancing terminal operation efficiency and flexibility. In addition, the results of the sensitivity analysis emphasize the importance of traffic and environmental variables in prediction, and managers should make full use of these data to improve port operation efficiency.

Future research directions include expanding the dataset to introduce more turnaround time-related features, such as equipment operating status and intermodal connection information; further optimizing the model to improve prediction accuracy and enhance interpretability; and studying logistics companies through behavioral logic, and deeply exploring the impact of predictive information on decision-making. At the same time, verifying and promoting the model in this article in other ports and exploring its universality and scalability will provide new ideas and methods for smart logistics management in global ports.

References

Abreu, L. R., Maciel, I. S., Alves, J. S., Braga, L. C., and Pontes, H. L. (2023). A decision tree model for the prediction of the stay time of ships in brazilian ports. *Engineering Applications of Artificial Intelligence*, 117:105634.

Abualhaol, I., Falcon, R., Abielmona, R., and Petriu, E. (2018). Data-driven vessel service time forecasting using long short-term memory recurrent neural networks. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 2580–2590.

AbuAlhaol, I., Falcon, R., Abielmona, R., and Petriu, E. (2018). Mining port congestion indicators from big ais data. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.

Antamis, T., Medentzidis, C.-R., Skoumperdis, M., Vafeiadis, T., Nizamis, A., Ioannidis, D., and Tzovaras, D. (2021). Ai-supported forecasting of intermodal freight transportation delivery time. In *2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*, pages 1–6.

Araghi, B. N., Christensen, L. T., Krishnan, R., Olesen, J. H., and Lahrmann, H. (2013). Use of low-level sensor data to improve the accuracy of bluetooth-based travel time estimation. *Transportation Research Record Journal of the Transportation Research Board*, 2338(1):29–34.

Araghi, B. N., Krishnan, R., and Lahrmann, H. (2015). Mode-specific travel time estimation using bluetooth technology. *Journal of Intelligent Transportation Systems*, 20(3):219–228.

Bachmann, C., Roorda, M. J., Abdulhai, B., and Moshiri, B. (2012). Fusing a bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation. *Journal of Intelligent Transportation Systems*, 17(2):152–164.

Bai, X., Jia, H., and Xu, M. (2024). Identifying port congestion and evaluating its impact on maritime logistics. *Maritime Policy & Management*, 51(3):345–362.

Balster, A., Hansen, O. E., Friedrich, H., and Ludwig, A. (2020). An ETA prediction model for intermodal transport networks based on machine learning. *Business and Information Systems Engineering*, 62(5):403–416.

Berrar, D. (2019). Cross-validation. In Ranganathan, S., Gribskov, M., Nakai, K., and Schönbach, C., editors, *Encyclopedia of Bioinformatics and Computational Biology*, pages 542–545. Academic Press, Oxford.

Che, Z., Purushotham, S., Cho, K., Sontag, D., and Liu, Y. (2018). Recurrent neural networks for multivariate time series with missing values. *Scientific Reports*, 8(1).

Chen, G. and Jiang, L. (2016). Managing customer arrivals with time windows: a case of truck arrivals at a congested container terminal. *Annals of Operations Research*, 244(2):349–365.

Chen, G. and Yang, Z. (2010). Optimizing time windows for managing export container arrivals at Chinese container terminals. *Maritime Economics and Logistics*, 12(1):111–126.

Chen, L., Zhang, D., Ma, X., Wang, L., Li, S., Wu, Z., and Pan, G. (2015). Container port performance measurement and comparison leveraging ship gps traces and maritime open data. *IEEE Transactions on Intelligent Transportation Systems*, 17(5):1227–1242.

Chen, X., Zhou, X., and List, G. F. (2011). Using time-varying tolls to optimize truck arrivals at ports. *Transportation Research Part E: Logistics and Transportation Review*, 47(6):965–982.

Chu, Z., Yan, R., and Wang, S. (2023). Evaluation and prediction of punctuality of vessel arrival at port: a case study of Hong Kong. *Maritime Policy and Management*, pages 1–29.

Chu, Z., Yan, R., and Wang, S. (2024). Vessel turnaround time prediction: A machine learning approach. *Ocean and Coastal Management*, 249:107021.

Crawford, F., Watling, D., and Connors, R. (2018). Identifying road user classes based on repeated trip behaviour using bluetooth data. *Transportation Research Part a Policy and Practice*, 113:55–74.

Cui, Z., Ke, R., Pu, Z., and Wang, Y. (2020). Stacked bidirectional and unidirectional lstm recurrent neural network for forecasting network-wide traffic state with missing values. *Transportation Research Part C: Emerging Technologies*, 118:102674.

Du, Y., Zhao, Y., and Gao, D. (2022a). Predictive simulation of external truck operation time in a container terminal based on traffic big data. *Journal of Shanghai Jiaotong University (Science)*.

Du, Y., Zhao, Y., and Gao, D. (2022b). Predictive simulation of external truck operation time in a container terminal based on traffic big data. *Journal of Shanghai Jiaotong University (Science)*.

Filom, S., Amiri, A. M., and Razavi, S. (2022). Applications of machine learning methods in port operations – a systematic literature review. *Transportation Research Part E: Logistics and Transportation Review*, 161:102722.

Garrido-Valenzuela, F., Raveau, S., and Herrera, J. C. (2022). Bayesian route choice inference to address missed bluetooth detections. *IEEE Transactions on Intelligent Transportation Systems*, 23(3):1865–1874.

Gers, F. A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: Continual prediction with lstm. *Neural Computation*, 12(10):2451–2471.

Hinton, G. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Hollander, Y. and Liu, R. (2008). Estimation of the distribution of travel times by repeated simulation. *Transportation Research Part C: Emerging Technologies*, 16(2):212–231.

Khosravi, A., Mazloumi, E., Nahavandi, S., Creighton, D., and Van Lint, J. (2011). A genetic algorithm-based method for improving quality of travel time prediction intervals. *Transportation Research Part C: Emerging Technologies*, 19(6):1364–1376.

Leachman, R. C. and Jula, P. (2011). Congestion analysis of waterborne, containerized imports from asia to the united states. *Transportation Research Part E: Logistics and Transportation Review*, 47(6):992–1004.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

Lederman, R. and Wynter, L. (2011). Real-time traffic estimation using data expansion. *Transportation Research Part B: Methodological*, 45(7):1062–1079.

Margreiter, M., Busch, F., and Carstensen, C. (2019a). The evolution of bluetooth detection rates. In *Transportation Research Board*.

Margreiter, M., Busch, F., and Carstensen, C. (2019b). The evolution of bluetooth detection rates. In *Transportation Research Board*.

Meersman, H., Van de Voorde, E., and Vanelslander, T. (2012). Port congestion and implications to maritime logistics. In *Maritime Logistics*, pages 49–68. Emerald Group Publishing Limited.

Munim, Z. H., Dushenko, M., Jimenez, V. J., Shakil, M. H., and Imset, M. (2020). Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. *Maritime Policy & Management*, 47(5):577–597.

Murty, K. G., Wan, Y., Liu, J., Tseng, M. M., Leung, E., Lai, K.-K., and Chiu, H. W. C. (2005). Hongkong International terminals gains elastic capacity using a Data-Intensive Decision-Support system. *Interfaces*, 35(1):61–75.

Namboothiri, R. and Erera, A. L. (2008). Planning local container drayage operations given a port access appointment system. *Transportation Research Part E: Logistics and Transportation Review*, 44(2):185–202. Selected Papers from the National Urban Freight Conference.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830.

Peng, W., Bai, X., Yang, D., Yuen, K. F., and Wu, J. (2023). A deep learning approach for port congestion estimation and prediction. *Maritime Policy & Management*, 50(7):835–860.

Phan, M.-H. and Kim, K. H. (2016). Collaborative truck scheduling and appointments for trucking companies and container terminals. *Transportation Research Part B: Methodological*, 86:37–50.

Pruyn, J., Kana, A., and Groeneveld, W. (2020). Chapter 4 - analysis of port waiting time due to congestion by applying markov chain analysis. In Vanelslander, T. and Sys, C., editors, *Maritime Supply Chains*, pages 69–94. Elsevier.

Rotterdam, H. (2024). Havenbedrijf rotterdam, 2023 jaarverslag.

Schulte, F., Lalla-Ruiz, E., González-Ramírez, R. G., and Voß, S. (2017). Reducing port-related empty truck emissions: A mathematical approach for truck appointments with collaboration. *Transportation Research Part E: Logistics and Transportation Review*, 105:195–212.

Sharif, O., Huynh, N., and Vidal, J. M. (2011). Application of el farol model for managing marine terminal gate congestion. *Research in Transportation Economics*, 32(1):81–89. Container Transport Operations.

Sharma, S., Van Lint, H., Tavasszy, L., and Snelder, M. (2022). Estimating route choice characteristics of truck drivers from sparse automated vehicle identification data through data fusion and bi-objective optimization. *Transportation Research Record Journal of the Transportation Research Board*, 2676(12):280–292.

Sidarta, M. (2021). Predicting truck cycle time in earthworks using a machine learning approach. Master's thesis, TU Delft, http://resolver.tudelft.nl/uuid:0f3c5ea9-c2b5-447d-b282-d6d25843f97d.

Sjoerd van der Spoel, C. A. and van Hillegersberg, J. (2017). Predictive analytics for truck arrival time estimation: a field study at a european distribution centre. *International Journal of Production Research*, 55(17):5062–5078.

Torkjazi, M., Huynh, N., and Shiri, S. (2018). Truck appointment systems considering impact to drayage truck tours. *Transportation Research Part E: Logistics and Transportation Review*, 116:208–228.

van Asperen, E., Borgman, B., and Dekker, R. (2013). Evaluating impact of truck announcements on container stacking efficiency. *Flexible Services and Manufacturing Journal*, 25(4):543–556.

Van der Spoel, S., Amrit, C., and Hillegersberg, J. (2016). A benchmark for predicting turnaround time for trucks at a container terminal.

Vo, S., Stahlbock, R., and Steenken, D. (2004). Container terminal operation and operations research - a classification and literature review. *OR Spectrum*, 26(1):3–49.

Wang, R., Li, J., and Bai, R. (2023). Prediction and analysis of container terminal logistics arrival time based on simulation interactive modeling: A case study of ningbo port. *Mathematics*, 11(15).

Wenhao Peng, Xiwen Bai, D. Y. K. F. Y. and Wu, J. (2023). A deep learning approach for port congestion estimation and prediction. *Maritime Policy & Management*, 50(7):835–860.

Wijbenga, A., Bac, J., and Boerma, S. (2011). Een nieuwe manier van verkeerswaarneming: Bluetooth. *B50*.

Wu, C.-H., Ho, J.-M., and Lee, D.-T. (2004). Travel-time prediction with support vector regression. *IEEE transactions on intelligent transportation systems*, 5(4):276–281.

Yang, Z., Chen, G., and Moodie, D. R. (2010). Modeling road traffic demand of container consolidation in a chinese port terminal. *Journal of Transportation Engineering*, 136(10):881–886.

Yeo, G.-T., Roe, M., and Soak, S.-M. (2007). Evaluation of the marine traffic congestion of north harbor in busan port. *Journal of waterway, port, coastal, and ocean engineering*, 133(2):87–93.

Yu, J., Tang, G., Song, X., Yu, X., Qi, Y., Li, D., and Zhang, Y. (2018). Ship arrival prediction and its value on daily container terminal operation. *Ocean Engineering*, 157:73–86.

Zhang, X., Zeng, Q., and Chen, W. (2013). Optimization model for truck appointment in container terminals. *Procedia - Social and Behavioral Sciences*, 96:1938–1947. Intelligent and Integrated Sustainable Multimodal Transportation Systems Proceedings from the 13th COTA International Conference of Transportation Professionals (CICTP2013).

Zhang, X., Zeng, Q., and Yang, Z. (2018). Optimization of truck appointments in container terminals. *Maritime Economics and Logistics*, 21(1):125–145.

Zhao, W. and Goodchild, A. V. (2010). The impact of truck arrival information on container terminal rehandling. *Transportation Research Part E: Logistics and Transportation Review*, 46(3):327–343.

Appendix A

Scientific Paper

Predicting Truck Turnaround Time Using Machine Learning: A Case Study at the Port of Rotterdam

Haoyang Chen[1]

[1]Delft University of Technology, Postbus 5, 2600 AA, Delft, The Netherlands

February 10, 2025

Abstract

Container terminals play a critical role in global trade, with truck turnaround time (TTT) serving as a key performance indicator for port efficiency. This study develops a predictive framework for TTT at the Port of Rotterdam by integrating multi-source data, including Bluetooth sensor records, container arrivals, and environmental conditions. A stacked Long Short-Term Memory (LSTM) model is proposed, leveraging its ability to capture temporal dependencies and complex nonlinear interactions. To address noisy data, a robust trip identification pipeline was implemented, achieving over 90% accuracy in identifying container truck trips. The stacked LSTM model outperformed traditional methods, such as Random Forest and XGBoost, in predictive accuracy. Sensitivity analysis revealed the critical influence of truck arrivals and wind conditions on TTT variability. This framework provides actionable insights for terminal operators, truck drivers, and policymakers to optimize terminal operations and reduce congestion.

Keywords

Truck Turnaround Time; Port Operations; Bluetooth Data; Machine Learning; LSTM

1 Introduction

Maritime transport has long been the backbone of the global transportation network. With containerized shipping becoming the standard model for exchanging goods in maritime transportation worldwide, the increased volume of container transportation leads to a surge in container port activities. However, the growth has not come without its challenges. One of the most pressing issues facing container ports is the congestion of container trucks, which has become a bottleneck in the logistics chain, impeding the flow of commerce and affecting the overall efficiency of port operations. The queuing up idle trucks at terminal gates have caused further congestion upstream and also extra emissions, costs and delays [1], [2].

The issue of gate congestion at container terminals is indeed a significant challenge that has been extensively documented. Rajeev and Alan [3] noted the decrease of drayage operation efficiency caused by gate congestion. When container trucks arrive during peak hours, this generally results in longer total turnaround times within the port. Du, Zhao and Gao [4] defined the turnaround time of container trucks in a container terminal from different perspectives. For an external truck itself, the turnaround time refers to the time between entering the port entrance and departure from the exit gate. The total truck turnaround time in a container terminal comprises four sequential components: queuing time at the entrance gate, waiting time in the buffer area, operation time in the yard, and exiting time at the exit gate. However, due to the different interest of stakeholders, there exist various resources of data regarding the operation time of trucks in container terminals. For example, terminal operators mostly concern the operation time in yard, and truck operators would start to calculate turnaround time when the truck enters the special road connecting the container terminal.

For ports, unexpected truck congestion and extended truck turnaround times would prevent the yard from making full use of operational capacity to load and unload containers [5]. For trucks, both operators and drivers do not expect to be affected by the extended turnaround time which may lead to increased idle time and decreased fleet utilization. Therefore, predicting and optimizing truck turnaround time at container terminals has become a focus of academics and industry.

Many papers analyze the factors affecting turnaround time and propose various suggestions to predict and reduce turnaround time. In terms of turnaround time prediction, several methods such as factor analysis, system dynamic (SD), machine learning and so on are used to predict truck turnaround time. Du, Zhao and Gao [4] introduced SD method to build a simulation model of truck operation system in container terminals, and considered the impact of multiple state variables to predict turnaround time. Sjoerd, Amrit and Hillegersberg [6] proposed a prediction model using both regression and classification methods, and established a benchmark to evaluate the prediction results. Sidarta [7] adopted multiple machine learning methods to predict truck cycle time in earthworks. In reducing turnaround time, research on truck appointment system (TAS) dominates the mainstream. Literature [8], [3] and [9] employed various methods to build and optimize the TAS to reduce the turnaround time of external trucks at the gate and yard. In addition, research represented by literature [10] considers using time-varying tolls to control arrival traffic at the terminal gate, but there are few practical applications to prove its effectiveness.

However, few studies have considered in-port operations when predicting and optimizing container truck turnaround times. Most of the TAS models developed only concern quote assignment of time intervals for container trucks, assuming that the basic operating time of container trucks within the terminal is consistent across time periods. In contrast to this, other factors including the arrival of containers and weather conditions may also affect the turnaround time of trucks within the terminal. Similarly, in terms of turnaround time prediction, there is few literature that take container arrival information into account. In terms of data, the studies above basically considered container truck arrival data from the entry gate, and did not explore other data sources.

This study introduces a novel approach to predicting truck turnaround time at container terminals by integrating multi-source datasets and advanced machine learning techniques. Specifically, the research utilizes historical truck arrival and departure data identified from Bluetooth sensor records, container arrival information, and environmental factors such as weather and traffic conditions. A Long Short-Term Memory (LSTM) model is employed as the core predictive tool, leveraging its capacity to capture temporal dependencies and nonlinear relationships in sequential data. The model development includes preprocessing steps such as data cleaning, feature extraction, and time-series alignment to ensure high data quality and robustness. By incorporating these diverse data sources, the proposed model addresses the limitations of existing studies, which often focus on isolated factors or static terminal operations, providing a more comprehensive and accurate prediction framework.

This research contributes to the existing literature by bridging key gaps. Unlike prior work that primarily emphasizes truck appointment systems (TAS) or uses limited datasets, this study accounts for in-terminal operations and external factors like weather and road congestion. Additionally, the integration of data from container operations and Bluetooth detections introduces a holistic perspective that enhances data integration and operational insights. The automated prediction tool developed in this study has significant practical implications, offering real-time support for terminal operators, truck drivers, and port authorities to optimize operations, reduce congestion, and improve efficiency.

The rest of this study is structured as follows. Chapter 2 provides a comprehensive literature review, summarizing existing research on truck turnaround time prediction and identifying key gaps that this study addresses. Chapter 3 outlines the methodology, detailing the research design, data collection strategies, and the framework for predictive model development. Chapter 4 focuses on data processing and analysis, describing the preprocessing steps, feature engineering, and exploratory analysis conducted to prepare the multi-source datasets for modeling. Chapter 5 presents the model design, training, and evaluation, including the development of the Long Short-Term Memory (LSTM) model, benchmarking against alternative methods, and analyzing the results to validate its predictive performance. Finally, Chapter 6 concludes the study, summarizing the key findings, discussing the practical implications of the developed model, and proposing directions for future research.

2 Literature Review

This literature review aims to examine existing research on Bluetooth trip identification and truck turnaround time prediction and optimization. Then key limitations and challenges are discussed to delineate the research gaps that this study seeks to address.

2.1 Research on truck turnaround time

Truck appointment system is a two-dimensional decision-making system that relies on both space and time, which is used to optimize truck turnaround time and operation cost. Murty et al. [11] were the first research team to explore the arrival time schedule of external trucks, and they developed a TAS system used by Hong Kong International Terminals. With the development of research in this area, some studies began to consider the impact of truck operating hours and terminal appointment time windows on truck operating costs when designing TAS. Research results from Zhao and Goodchild [12] show that the application of TAS can reduce container truck congestion at the terminal and the arrival information of trucks can reduce rehandles at the yard. Zhang, Zeng and Chen [13] developed a truck appointment model using a Baskett–Chandy–Muntz–Palacios (BCMP) queuing network to describe truck activities at the gate or yard. Numerical results shows that the network could help reduce turnaround time. Phan and Kim [14] developed a coordinated solution between hauling companies and TAS to avoid the negative impact of TAS on trucking company operations. They take into account the truck's work schedule to reduce the overall truck turnaround time. Schulte et al. [15] developed a graph-based mathematical model based on the m-TSPTW. Their model optimizes travel costs and emissions for tasks that can be performed by multiple trucks by collaboratively merging tasks, and also alleviates terminal congestion. Torkjazi, Huynh and Shiri [8] considered truck tour when designing TAS, and formulated and solved the TAS problem as a mixed integer nonlinear problem (MINLP). However, even when they introduced a hinterland network for truck tour, they didn't take into account possible traffic congestion within the network.

Another approach for managing truck arrivals and reduce truck turnaround time is called vessel-dependent time windows (VDTWs). Yang, Chen and Moodie [16] found that the distribution of truck arrivals with outbound containers could be described using a Beta distribution within a time window based on vessel-calling schedule. Chen and Yang [17] developed a heuristic algorithm to find optimal time window for external trucks to reduce the total cost of gate congestion. Chen and Jiang [18] proposed an approach to manage truck arrival time window based on truck-vessel service relationship. This point of view considers the corresponding relationship between trucks and vessels, and the operating conditions in the port have been taken into consideration, but the impact of trucks being affected by congestion at gates and external roads has not been considered.

There are various directions of prediction regarding terminal operation and truck turnaround time with a focus on predicting methods and influencing factors. Traditional turnaround time prediction systems used in terminals are merely based on the historical truck flows and empirical experience from employees. However, numerous factors can make such predictions inaccurate: port operations, weather conditions, political concern, etc.

In the field of port operations, more research focuses on time-related indicators of vessels. Numerous studies aim to predict estimated vessel arrival time (ETA), vessel turnaround time (VTT) and departure time. Some ttudies [19], [20] developed evaluation and prediction methods for estimated arrival time of vessels, while others [21], [22] both predict vessel turnaround time and departure time based on static ship data. Chu, Yan and Wang [22] developed an XGBoost regression model to enhance the precision of VTT predictions. Abreu et al. [23] proposed a decision tree model to predict vessel turnaround time considering cargo and port information. These time-related indicators focus on similar but different influencing factors than truck turnaround time.

Other studies pay attention to road arrival or turn around time prediction in different scenarios. Sjoerd [24] investigated factors affecting truck arrival times at distribution centers. Du, Zhao and Gao [4] used SD method to predict truck operation time in terminals, but only considering input truck flow and yard capacity. Wang, Li and Bai [5] took container information, weather condition and traffic condition into account and used a combination of data mining methods to predict container arrival time at terminals. In addition, Table 5 discussed possible influencing factors in time prediction covered in different studies. The listed influencing factors provide possible options to be dealt with as the input of prediction model.

2.2 Trip identification based on Bluetooth data

Bluetooth technology has increasingly been adopted in traffic management due to its ability to collect real-time, cost-effective data that is useful for various traffic analysis tasks. Bluetooth sensors detect devices by capturing Media Access Control (MAC) addresses, which can provide time-stamped data as vehicles pass through specific checkpoints. These data points can generate travel time estimates, congestion metrics, and vehicle flow analyses across monitored areas [25].

However, the data quality obtained through Bluetooth sensors has been widely discussed. A major limitation highlighted across studies is data sparsity and noise, particularly under mixed traffic conditions. Araghi et al. [25] noted that large detection zones can introduce multiple detection events, leading to potential ambiguity in vehicle positioning. Margreiter et al. [26] pointed out that not all vehicles carry Bluetooth-enabled devices. The limited detection rate reduces the availability of Bluetooth data, and may also introduce a sampling bias where the detected data may not fully represent the entire traffic flow.

To address these challenges, researchers have proposed various methodologies to improve the accuracy and utility of Bluetooth data for trip identification, path estimation, and vehicle type recognition. Araghi et al. [25] applied a classification method based on the device type and signal strength to differenciate travel modes. Araghi et al. [27] employed a classification method to distinguish between motorized vehicles and bicycles based on estimated mode-specific travel times. Bachmann et al. [28] explored data fusion between Bluetooth traffic monitoring system and loop detectors to improve freeway speed estimation. Garrido et al. [29] presents a Bayesian inference-based methodology to address the issue of missed Bluetooth detections in route choice modeling. Sharma et al. [30] designed a method for clustering based on trip duration time to identify vehicle types, and proposed a bi-objective optimization model to estimate route choices for truck drivers based on sparse Bluetooth data and loop-detector data. Crawford et al. [31] captured repeated trip behaviors from Bluetooth data to class road users and measured spatial similarity using Sequence Alignment. These studies support the research by providing tools to improve Bluetooth data quality and identify truck trips, and hence help to predict truck turnaround time more reliably and realistically.

This literature review highlights significant gaps in the existing research on truck turnaround time prediction, particularly in the integration of diverse influencing factors and the holistic consideration of both in-port and external conditions. While truck appointment systems (TAS) have been extensively studied to optimize traffic flow and reduce turnaround times, most studies fail to capture the dynamic nature of port operations and rely heavily on limited datasets, such as truck arrival data from terminal gates. Prediction models often focus on isolated factors, such as historical traffic flows or vessel-specific metrics like ETA and VTT, which restrict their ability to identify patterns across multiple variables. Moreover, although Bluetooth data has been widely used for road trip and vehicle type identification, there is a lack of research leveraging this data to identify vehicle entry and exit trips within a specific area, further limiting its application in truck turnaround time modeling. To address these gaps, this study aims to develop a prediction model for truck turnaround time that integrates historical truck arrival and departure data, container arrival information, and external factors such as weather and traffic conditions. By incorporating diverse datasets and leveraging Bluetooth trip identification methods, the proposed model seeks to provide a more comprehensive and accurate framework for predicting truck turnaround times at container terminals.

3 Bluetooth Trip Identification and Empirical Data Analysis

This chapter focuses on the empirical analysis of the datasets utilized in this study and the identification of Bluetooth-based vehicle trips to derive meaningful inputs for the prediction model.

3.1 Data Description

To develop an accurate and robust prediction model for truck turnaround time at container terminals, this study utilizes multiple datasets sourced from the Port of Rotterdam and other reliable sources. The datasets span the entire year of 2017, covering a wide range of temporal and operational variations. The primary datasets are described as follows:

- Container arrival dataset: This dataset contains all arriving container records at the five container terminals in the Maasvlakte area in 2017. It reveals several key data fields including container ship arrival time, container unloading time, container unloading terminal, estimated hinterland transit time and container ship name.
- Vehicle Bluetooth detection dataset. This Bluetooth data set is measured in days and

contains all Bluetooth detection information on the roads surrounding the Rotterdam port area for 365 days in 2017. The daily data set contains approximately 4 million detection records. Each sensor records the passage of vehicles, the MAC Address of the vehicle, along with the time stamp and the latitude and longitude of the sensor.

- Environmental dataset. This data set mainly describes the wind speed conditions every ten minutes around the research area in 2017. The data fields include factors such as wind level, maximum wind gust speed and wind direction.

3.2 Bluetooth-based Trip Identification

In our data, some Bluetooth sensors, mapped to specific locations, provide insights into vehicle routes, travel times, and the turnaround times associated with entering and exiting port terminals. A pair of Bluetooth sensors along the route to the research area are found to cluster travel time observations and thus identify the corresponding vehicle category. Other sensors, deployed within the research area, can track the sequence of vehicle passes. Before identifying truck tours from the Bluetooth data, we need to extract vehicle categories from the dataset. The flow chart presented in fig 1 illustrates the steps for extracting truck-specific data.

For a given time period and specific Bluetooth sensors, outliers are eliminated from Bluetooth observations based on the driving behavior of different types of vehicles. Then, the filtered Bluetooth observations are clustered into two representative groups to analyze the driving behavior of these vehicles. The selected route is about 10 km long. Therefore, after calculating the travel time, we delineated a reasonable travel time range based on the speed limit of Dutch highways. Observations exceeding this travel time range are excluded to avoid the impact of missing observations or longer travel times caused by congestion.

After clustering vehicle observations, we categorized the driving behaviors based on the speed limit for trucks on Dutch highways. Vehicles meeting these conditions were then labeled as "likely to be trucks". Selecting January 2017 as the time period, with the timestamp of entry into the route as the horizontal axis and the travel time as the vertical axis, the observed distribution of data points is shown in Fig 2.

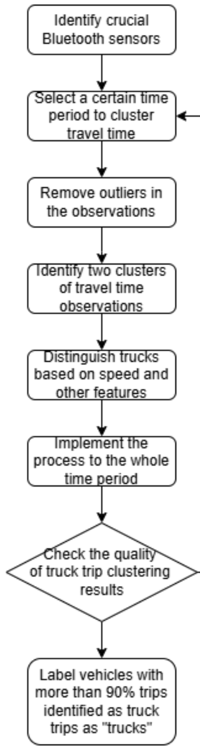


Figure 1: Identifying trucks from Bluetooth observations

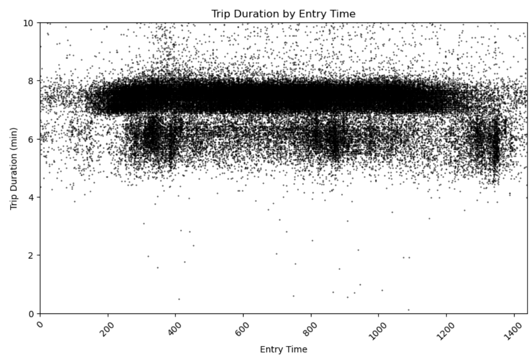


Figure 2: Trip duration time observations

The majority of the observations appear to cluster around a narrow range of trip durations between about 7 to 8 minutes. For the chosen route, the presence of a clear boundary between the clusters centered around 6 and 7 minutes suggests distinct driving behaviors or vehicle types. Fig 3 presents the clustered results of Bluetooth observations after excluding outliers from the data. The slower group of vehicles is marked red. As the regulatory speed limit for trucks and some other types of vehicles in the Netherlands is 80 km/h, these observations generally meet the speed limit requirements. This can also explain the cause of the cluster boundary.

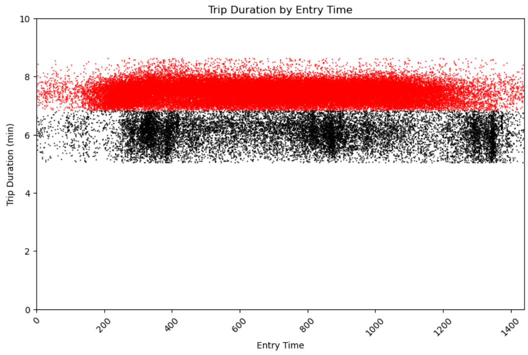


Figure 3: Clustered trip duration time observations

The identified clusters of slower vehicles account for approximately 75% of all vehicles. Considering the speed limits of different types of vehicles, the penetration rate of Bluetooth devices, and other factors, these vehicles cannot be completely labeled as trucks, pending subsequent verification. Therefore, these data points are labeled "likely to be trucks". For each MAC address, if more than 90% of its trips are marked as such, it is initially classified as a truck.

After identifying the vehicle type, it is also necessary to identify the truck trip and extract the driving route, destination and turnaround time of the truck trip. The research area's sensors are positioned at entry points, terminal access roads, and within key road segments, which provides opportunities to track container trucks' movement across the five terminals. A vehicle trip within the research area can be defined as the journey a truck undertakes from its entry into the area, through terminals, and until its exit. Each terminal, within its unique entry route, represents a potential destination within a tour. By investigating the traffic network within the research area through Google Maps and geographic information, specific routes to each container terminal are delineated with corresponding Bluetooth sensors placed along the routes. Table 1 displays the correspondence between each sensor and the respective terminal. Building upon the historical analysis of vehicle

Table 1: Sensor and Terminal Mapping

Sensor	**Terminal**
298	A
299	B
110001	C
1580336	D
If only 1580329	E

trip times and the identification of valid trips, the process for defining vehicle tours within the research area relies on detecting consistent entry and exit behavior for each vehicle.Considering the possibility of missing Bluetooth detection records and their impact on subsequent data, an effective trip detection step needs to be added. A single trip is considered valid only if it lasts more than 20 minutes and occurs within the same day, or if it spans multiple days but is within four hours. Among the identified trips, based on information provided by sources such as the Port of Rotterdam, the turnaround time for truck trips is generally less than four hours. Therefore, trips lasting less than four hours are labeled as "likely to be a truck".

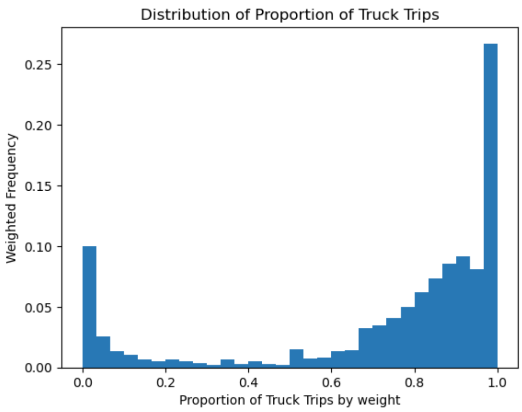


Figure 4: Weighted Frequency Distribution of Truck Trip Proportion by Vehicle

After the trip detection and classification process described earlier, each vehicle's trips are divided into two categories: a certain proportion is labeled as "truck" trips, while the remaining trips are labeled as "others." This plot shows the weighted frequency distribution of the proportion of trips labeled as "truck" for each vehicle. Each vehicle's weight in the distribution is determined by its total number of trips, giving more weight to vehicles with a higher trip count. Fig 4 shows the weighted frequency distribution of truck trip proportion. As the frequency of truck trips gradually increases after the 70% quantile value, we initially label vehicles with over 70% truck trips as "trucks".

According to research [32], the penetration rate of Bluetooth vehicles in port areas is no more than 75%. Since there are other facilities in the port area besides container terminals, and the same road speed limits apply to other types of vehicles other than container trucks, we are not able to capture precise road traffic flows, container truck flows and turnaround times. As a result, in order to improve the authenticity of the turnaround time prediction input, we use double verification methods. Only vehicles that are identified as trucks in both identification methods will be labeled as trucks in the final processed data set. Finally, based on the vehicles identified to be trucks according to their trip duration time on the selected route, 85% of the vehicles were further confirmed to be trucks. These identified trucks carries 64% of the trips within the port area.

3.3 Descriptive Analysis of Input Parameters

In this section, we conduct an in-depth analysis of the processed data to uncover patterns and insights relevant to port operations and truck turnaround times. Identifying and analyzing the distribution characteristics of the data can help us make more targeted adjustments to the model input. In addition, the data periodicity revealed by distribution characteristics can help us better capture underlying temporal patterns, especially when dealing with indicators with obvious periodic fluctuations.

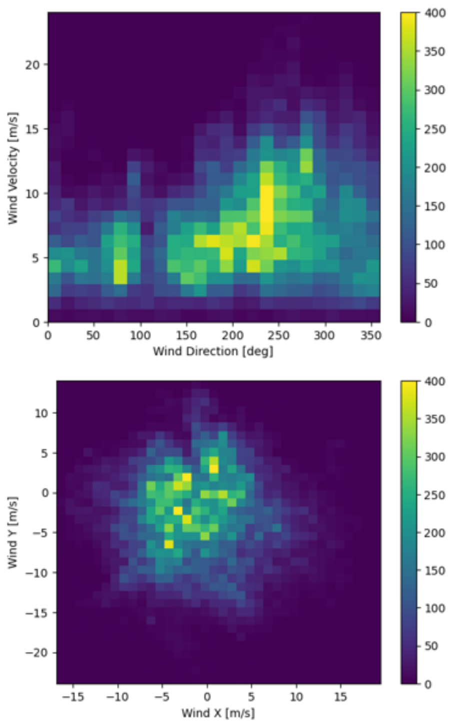


Figure 5: Comparison of Wind format

In terms of environmental wind conditions, while the wind speed along with wind direction (in degree) is not an ideal model input, convert wind conditions to wind vectors would helps to explain the model. In the wind direction, 360° and 0° should approach each other and wrap smoothly. If the wind isn't blowing, the direction doesn't matter. As a result, Wind speed and direction are converted into u (east-west) and v (north-south) vector components. Fig 5 shows the results before and after wind vectorization processing. The wind vectors are primarily concentrated near the center of the plot, while points farther from the center represents higher wind speeds. After vectorization, wind data shows a clearer directional distribution, and the format is better suited for integration into further modeling.

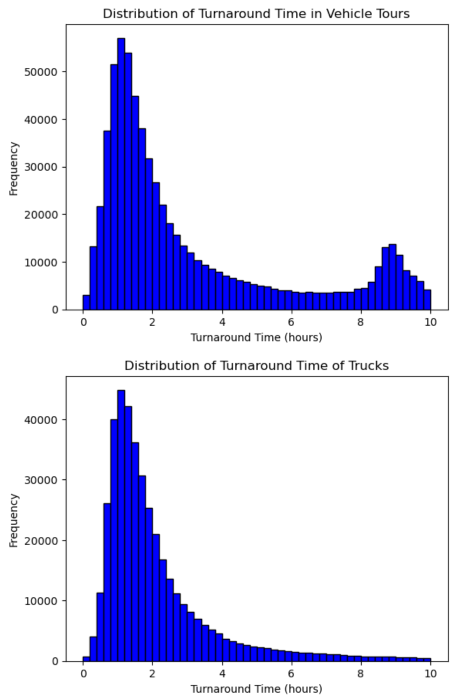


Figure 6: Comparison of Turnaround Time Distributions

Fig 6 illustrate the distribution of turnaround times, one representing all vehicles and the other specifically focusing on identified trucks. The x-axis stands for turnaround time length, while the y-axis coordinate is the frequency of occurrence. The distribution of turnaround times for all vehicles in the first figure displays a bimodal pattern. The first peak occurs between 1 and 2 hours, and the turnaround time for most vehicle trips are distributed around this peak. The secondary peak, around 8-10 hours, likely represents non-container vehicles involved in different types of operations within the port, especially vehicles that regularly enter and exit the port area during working hours. In contrast, the distribution of turnaround time for identified trucks (second figure) shows a unimodal and positively skewed distribution. Similar to the first peak in the first figure but with a lower frequency, the majority of turnaround times are clustering between 1 and 2 hours. This concentration suggests that, under normal operating conditions, most container trucks can complete their tasks within this time period. The observed distribution features and differences demonstrate the effectiveness of the proposed truck identification method.

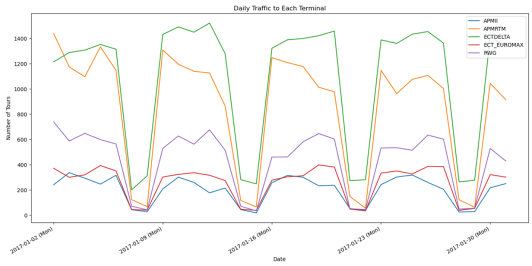


Figure 7: Daily Truck Flow to Each Terminal

Fig 7 displays daily truck traffic to each terminal over a month. There is a clear recurring weekly pattern, where traffic peaks at the beginning of each week (around Monday or Tuesday) and then gradually declines towards the weekend, reaching a minimum typically on Sundays. Evidence from the figures displayed above proves the apparent periodicity in port operations and truck traffic. Therefore, we also perform a periodic analysis of truck turnaround times.

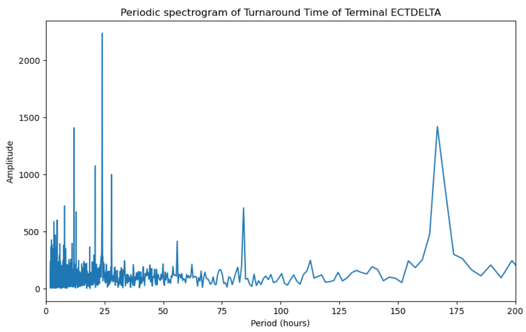


Figure 8: Frequency Domain of Truck Turnaround Time in Terminal ECTDELTA

To further investigate the periodicity in truck turnaround times, a frequency domain analysis was conducted using the Fourier Transform. The Fourier Transform is a mathematical technique that decomposes a time series into its constituent frequencies, allowing us to observe periodic components within the data. By converting time-domain data (turnaround times over time) into the frequency domain, we can identify dominant cycles that may reflect operational rhythms or systematic delays in the port's workflow.

In this frequency domain analysis, shown in Fig 8, the x-axis represents the period in hours, while the y-axis shows the amplitude of each periodic component. Peaks in amplitude indicate strong periodic components in the turnaround time data. The two peaks appear at 23.97 and 166.76 hours, indicating that truck turnaround time has strong periodic characteristics on daily and weekly basis. This periodic analysis not only confirms the presence of the previous figures, but also revealed that time signals can be transformed into "periods of a day" or "periods of a week" for interpretation to improve prediction accuracy.

4 Methodology

4.1 Mechanism of LSTM

The prediction model development phase is the core component of this research. Given the sequential nature of the data involved, we elaborate on the choice of employing time series-based prediction methods for forecasting truck turnaround times at sea terminals. Additionally, we discuss the reason behind selecting Long Short-Term Memory (LSTM) networks as the primary modeling technique.

The choice of LSTM is driven by the need to model complex sequential relationships between variables. As a type of recurrent neural network (RNN), LSTM was firstly introduced by Hochreiter and Schmidhuber, as a solution to the vanishing gradient problem commonly encountered in traditional RNNs[33].

As shown in fig 9, at time step t, the cell involves three inputs, three gate units and two outputs. Two of the inputs comes from the unit of previous time step, namely the cell state C_{t-1} and hidden state h_{t-1}. Another input is the current system variable vector x_t. The outputs of the cell are the updated cell state C_t and the hidden state h_t, which are passed to the next time step.

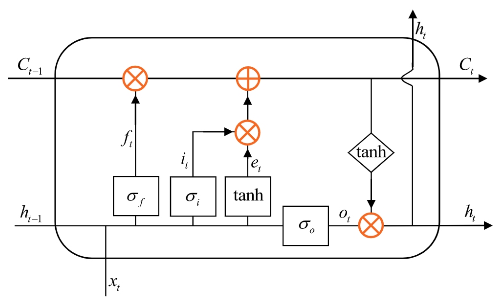


Figure 9: Typical repeating module in an LSTM[34]

Three gates are structured within a unit. The forget gate determines which information from the previous cell state C_{t-1} should be discarded.

where σ_f is the sigmoid activation function, W_f is the weight matrix, b_f is the bias vector, and $[h_{t-1}, x_t]$ represents the concatenation of the hidden state and input vector. The forget gate outputs a value $f_t \in [0, 1]$, where values close to 1 retain information and values close to 0 discard information.

The input gate controls which new information should be added to the cell state. The output gate determines what part of the cell state C_t should be output as the hidden state h_t and convert to next cell. Finally, the weights and biases will be updated through back propagation for minimizing the loss function. In this study, the loss function is determined by the prediction of the state.

4.2 Model development

In this research, a stacked-LSTM model is developed to predict the turnaround time. Existing studies have proved that a more complicated LSTM architecture is more probably able to build the representations of sequential data [35]. The stacked-LSTM model features a multi-layer architecture, allowing it to process sequential data hierarchically, capture both short-term and long-term dependencies, and build richer representations. In a stacked multi-layer LSTM architecture, the hidden layer's output serves as the input to the next layer in the sequence. This architecture allows for embedding multiple LSTM components to capture richer hidden features in the data.

The stacked LSTM model in this study employs a multi-layered architecture designed to effectively process and analyze complex sequential data, as illustrated in fig 10. The model integrates multiple LSTM components to capture both short-term and long-term temporal dependencies while extracting rich features from two primary data sources: container-related information and road traffic data. These two data streams are first processed independently through separate LSTM layers, allowing the model to focus on the unique characteristics and temporal patterns of each source.

The output from these initial LSTM layers is then combined and passed to a shared LSTM layer, where the model integrates the extracted features to understand the interdependencies between container operations and road traffic conditions. This consolidated information is further processed through a dense layer, which reduces dimensionality and refines the feature representation, before being fed into the output layer to generate predictions of truck turnaround time. By isolating and then integrating data from these two sources, the architecture ensures a comprehensive representation of the factors influencing turnaround time. The hierarchical design of stacked LSTM layers enhances its ability to capture intricate temporal relationships and dependencies, enabling more accurate and robust predictions.

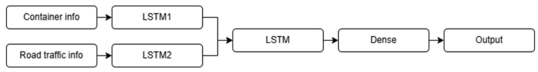


Figure 10: Architecture of stacked-LSTM model

The selected input parameters for the model encompass a range of operational and environmental factors essential for predicting truck turnaround time. These include truck arrival flow, container arrival volumes at terminals, lagged average turnaround times, and wind components such as speed and direction. To account for temporal patterns inherent in port operations, periodic features such as the day of the week and hour of the day are incorporated into the model. These features are transformed into continuous variables using sine and cosine functions, which preserve the cyclical nature of time and eliminate discontinuities in the data.

Before training the model, several data preprocessing techniques are implemented in order to improve data quality and prediction accuracy. Specifically, three key steps are performed:

1. Outlier Removal: A significant portion of errors in container truck turnaround time data arises due to low sample sizes during certain time periods. Outlier predictions are handled through quantile clipping, where predictions falling outside a certain percentile range are adjusted or removed to improve the model's general accuracy. To address this, we first identify and clean these low-sample-size outlier periods, ensuring that the resulting dataset better represents the actual trends and patterns.

2. Masking Invalid Time Steps: After removing outliers, the cleaned data is combined with the original missing time steps to create a comprehensive mask.

3. Time-series forecasting models like LSTM can produce predictions that may fluctuate due to minor noise in the data, especially when dealing with complex operational environments like port terminals. In an LSTM-based prediction problem, the presence of missing or null values in the input time series poses a significant challenge, as the model cannot process null values during training. Simply imputing missing values with predefined values, such as zeroes, the mean of historical observations, or the last observed value, introduces bias into the model inputs. This bias can lead to inaccurate param-

eter estimation during the training process, ultimately affecting the model's predictive performance[36]. To address this, data quality improving techniques are applied to reduce these fluctuations and produce more stable predictions.

4.3 Hyperparameter selection

After resampling the data set into a data form characterized by time steps and preprocessing it, the finalized data are split into training data and test data. The first 80% of the data are used for model training and the remaining are used for test and validating. Different sequence lengths are applied to strike a balance between forecast accuracy, data availability and practical significance.

To standardize the input data, a min-max scaling technique is applied in order to prevent features with larger magnitudes from dominating the learning process and to accelerate model convergence during training. For a given feature x_t at time step t, the scaling is performed as follows:

$$x_t^{\text{scaled}} = \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

where $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the feature x_t in the dataset. The same scale is applied to both train and test set.

The performance of the stacked-LSTM model is significantly influenced by its hyperparameters, which are categorized into two groups: model architecture hyperparameters and training process hyperparameters.

Key model architecture hyperparameters include the number of LSTM layers, the number of neurons per layer, and the dropout rate. These parameters control the network's depth, capacity to capture patterns, and regularization to prevent overfitting. Input time steps are also critical, as they define the historical sequence length used for prediction, balancing the capture of long-term dependencies with potential noise. The `tanh` and `sigmoid` functions are employed as activation functions to handle non-linear temporal patterns, while the loss function and optimizer ensure efficient learning by minimizing prediction errors.

Training process hyperparameters include the batch size, learning rate, and number of epochs. The batch size determines the number of samples processed per iteration, affecting convergence speed and stability. The learning rate controls the step size of weight updates, requiring careful tuning to avoid divergence or slow convergence. Early stopping is implemented during training to prevent overfitting by halting the process if validation loss does not improve over several epochs.

Before constructing the model, it is essential to determine the model parameters. Some of the general parameters in the model are set based on recommended optimal parameter configurations, as shown in Table 2. All the RNN-based models are trained by minimizing the mean square error using the Adam optimization method. The early stopping mechanism is used to avoid over-fitting.

Table 2: Common settings of each layer.

Layers	**Common settings**
LSTM layer	Activation = Sigmoid
Dense layer	Activation = Linear
Training process	Optimizer = 'Adam'
	Learning rate = 0.001
	Loss = MSE
Dropout layer	Dropout rate = 0.2

In this study, hyperparameter tuning was performed to determine the optimal configuration for the proposed stacked-LSTM model. The hyperparameters considered for experimentation included time steps, neuron units in each layer, number of epochs, and batch size. The candidate values for each hyperparameter are as follows:

- ***time_steps***: 24, 72, 168
- ***units***: 32, 64, 128
- ***epochs***: 25, 50, 100
- ***batch_size***: 16, 32, 64, 128

For the epochs parameter, while longer training typically improves model performance, it comes with a significant increase in training time. However, due to the implementation of the early stopping mechanism, most experiments exhibited convergence within 50 epochs. As a result, 50 epochs were chosen as the default value for the training process.

The other three hyperparameters (time_steps, neuron units, and batch size) were evaluated using a grid search approach. Specifically, experiments were conducted by varying the combination of these parameters, and their effects on the model's performance were assessed. The mean absolute error (MAE) was selected as the evaluation metric for hyperparameter optimization. Figure 11 presents the 3D grid search results, highlighting the best hyperparameter set results. As shown in Fig 11, through experimental evaluation, it was

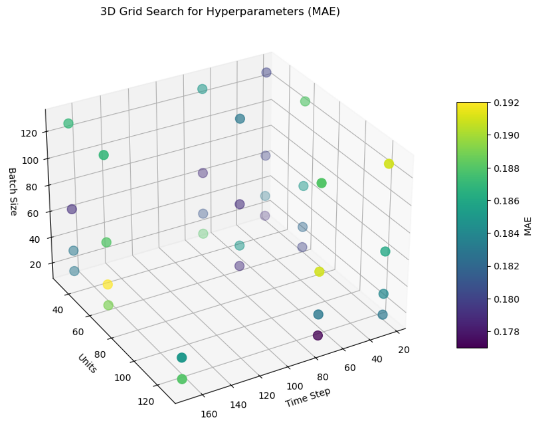


Figure 11: Hyperparameter tuning results for time_steps, neuron units, and batch size, evaluated using MAE.

observed that a time step of 72 hours, representing a prediction window of 3 days, achieved a favorable balance between predictive accuracy and computational requirements. Regarding batch size, smaller batch sizes consistently yielded better MAE values, suggesting that finer gradient updates were advantageous for this dataset despite the associated increase in training time. From the figure, it can be observed that the combination of $time\_steps = 72$, $neuron\ units = 128$, and $batch\ size = 16$ achieves the best performance with the lowest MAE. Based on the numerical experimental results and empirical insights, further adjustments to the model's hyperparameters were made. Under the finalized model architecture, the following settings were selected for subsequent training: the number of hidden layer units was set to 128, the number of epochs to 50, the activation function to Sigmoid, the dropout rate to 0.2, the time step to 72 hours, and the batch size to 16.

5 Model Results

5.1 Model Performance

Under the selected hyperparameter combination, the prediction results of the model are shown in the figure 12. The blue line in the figure represents the actual truck turnaround time, and the orange line represents the predicted turnaround time. The predicted values generally follow the fluctuation trend of the real values well. A clear pattern of fluctuations can be observed from the prediction results. Generally, the proposed LSTM model well captured the cyclical changes, and in most cases the predicted values align well with the true values at

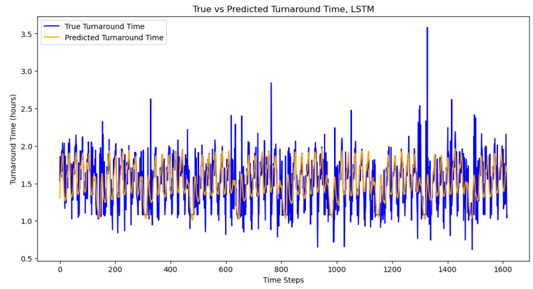


Figure 12: Prediction Results of the proposed stacked-LSTM model

the peaks and troughs in the time series.

Specifically, some local peaks, such as those around time steps 600 and 1400 in the validation set, were not accurately predicted. Abnormal fluctuations in peaks are often caused by short-term external events such as sudden terminal congestion, equipment failure, which may not be fully reflected in the obtained data and selected input features. The LSTM model captures long-term dependencies through memory units, which has a certain smoothing effect on noise and short-term extreme changes in the input sequence. When faced with severe fluctuations, models tend to predict values close to the overall trend rather than capturing these changes according to the set of the loss function. Additionally, the model fails to capture the deviation between some of the predicted troughs and true values.

Although the prediction errors at some time steps are large, the overall error distribution is relatively uniform, and there is no significant systematic overestimation or underestimation of the model's predicted values. This shows that the model has good generalization ability in the overall range and does not cause obvious bias problems.

5.2 Benchmarking

To evaluate the performance of the proposed model, several kinds of models for truck turnaround time prediction are compared with the prediction results. This study selected linear regression models, XGBoost models, random forest (RF) models, single-layer LSTM models, and multi-layer LSTM models based only on turnaround time for comparison.

Under the same dataset and sample division, we compared the performance of the proposed stacked LSTM model with other models to evaluate their advantages and disadvantages in predicting truck turnaround time within the context of this study. The performance of the models was measured using RMSE, MAE, and R-square metrics. The prediction comparison results are shown in the figure 13 and figure 15 in Appendix.

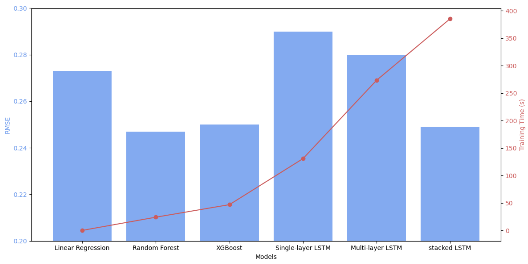


Figure 13: Result comparison of different models

Model	RMSE	MAE	R-square	(RMSE) GAP%
Stacked LSTM	0.249	0.173	0.418	-
Linear Regression	0.273	0.191	0.367	9.64%
Random Forest	0.247	0.168	0.420	-0.80%
XGBoost	0.250	0.172	0.414	0.40%
Single LSTM	0.290	0.213	0.290	16.47%
Multilayer LSTM	0.280	0.202	0.320	12.45%

Table 3: Comparison of Model Performance Metrics

Based on the Table 3, the proposed Stacked LSTM model demonstrates significant advantages in the given prediction scenario. It achieves an RMSE of 0.249 and an MAE of 0.173, which are comparable to the best-performing Random Forest model (RMSE = 0.247, MAE = 0.168) and slightly better than XGBoost (RMSE = 0.250, MAE = 0.172). Furthermore, the Stacked LSTM model shows a notable improvement in R-square (0.418), indicating its superior ability to explain the variance in the data compared to other LSTM-based models and Linear Regression.

While Random Forest slightly outperforms the Stacked LSTM in RMSE (-0.8% GAP), it is important to note that Random Forest's superior handling of masked data might have contributed disproportionately to its performance. In contrast, the Stacked LSTM model excels in capturing temporal dependencies, making it better suited for valid data predictions in this context. Additionally, the Stacked LSTM surpasses both Single-layer LSTM and Multilayer LSTM models by a significant margin in all metrics, highlighting the effectiveness of stacking layers in leveraging sequential patterns.

5.3 Factor Sensitivity Analysis

In order to explore the impact of different input features on the average turnaround time prediction accuracy, this section decomposes the input factors of the model. These input factors include historical average turnaround time, wind speed, container arrivals at each terminal, and truck arrivals. To systematically evaluate the contribution of each feature, the experiments were designed to start with a baseline model containing all input features and gradually eliminate individual features or specific combinations to observe the resulting changes in prediction performance. This elimination method enables the identification of critical features whose removal significantly impacts the model's accuracy.

The experiments were conducted using the Stacked LSTM model with consistent architecture, training and testing datasets, and training parameters across all experiments. The design includes tests where each individual feature is removed (e.g., historical average turnaround time, wind speed, container arrivals, or truck arrivals) to analyze its standalone impact on the predictions. In addition, specific feature combinations were excluded to evaluate the influence of the combined effect of internal and external factors, also to observe the inner connection of certain features. For example, historical turnaround time and truck arrivals were removed together to assess the effect of excluding traffic-related dynamics, while wind conditions and container arrivals were excluded to test the significance of terminal-side operational factors. The experimental design for this sensitivity analysis is shown in Table 4.

Table 4: Experimental Design for Sensitivity Analysis Based on Input Features

Experiment Number	TTT	Wind	Container Count	Truck Arrival	Periodicity
1	✓	✓	✓	✓	✓
2	-	✓	✓	✓	✓
3	✓	-	✓	✓	✓
4	✓	✓	-	✓	✓
5	✓	✓	✓	-	✓
6	✓	✓	✓	✓	-
7	✓	-	-	✓	✓
8	✓	✓	-	-	✓
9	-	✓	✓	-	✓

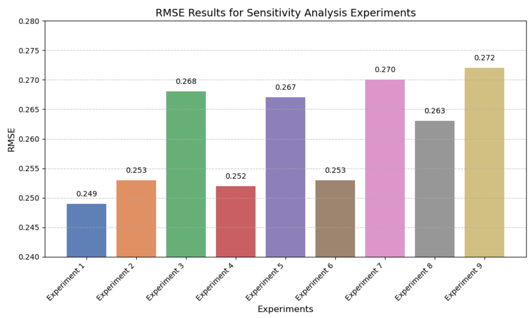


Figure 14: Result of Sensitivity Experiments

The experimental results of sensitivity analysis are shown in the figure 14. In the single-feature analysis, removing any feature negatively impacted prediction accuracy. Among them, current wind speed (environmental factor) and truck arrivals had the most significant influence, with their removal causing a noticeable increase in RMSE. From an operational perspective, adverse wind conditions can render terminal operations infeasible or significantly inefficient, leading to simultaneous shutdowns on the landside and quayside. The absence of wind condition information makes it difficult for the model to capture unexpected shutdowns. This parameter is crucial for all terminals in the Maasvlakte area. Similarly, truck arrivals have a clear real-world correlation with average truck turnaround times. A higher queue of trucks inside the port directly impacts the turnaround time of subsequent trucks.

In contrast, container arrival information and periodicity did not significantly affect prediction accuracy. While previous analyses have shown that port turnaround times exhibit notable daily or weekly periodicity, the model appears capable of capturing these temporal patterns from the time-series data itself. In the double-feature combination experiments, the removal of any of the three designed feature combinations also led to a significant increase in RMSE. The results of Experiments 2 and 9 highlight the substantial contribution of truck arrivals to improving prediction accuracy. Meanwhile, Experiment 7 performed slightly worse than Experiment 3, suggesting that container arrival counts, as a workload indicator, can complement other features to enhance the model's performance. The experimental results demonstrate that removing any single feature or feature combination results in predictions that are less accurate than those obtained by using all features together.

5.4 Implications

This chapter focuses on the training of the proposed model architecture and compares the results obtained using different numerical imputation strategies and parameter combinations. Based on the optimal numerical imputation strategy and parameter configuration, the model was trained to obtain prediction results and conduct performance analysis. Furthermore, a factor sensitivity analysis was performed on the input features of the model, providing insights into parameter importance. Finally, k-fold cross-validation was applied to the training results, demonstrating the model's relatively reliable performance in predicting truck turnaround times.

Several findings emerged from this chapter.

- All tested models struggled to accurately predict the sharp peaks in the dataset. This limitation is largely due to the quality and scope of the original dataset, which may lack the detailed information needed to explain these extreme variations. Improving data quality and adding more informative features could help address this issue in the future.
- In terms of the prediction results of each terminal, the model shows different performance, showing the different operational characteristics and operational differences of the terminal. It also shows that in actual operations and practical application of the model, it is necessary to adjust accordingly and make decisions based on the particularity of each case.
- Sensitivity analysis and k-fold cross-validation confirmed the model's effectiveness and reliability. The sensitivity analysis provided helpful insights into the relationship between input features and truck turnaround times. However, due to the limitations of neural networks, the model's ability to explain these relationships in detail is somewhat restricted.

6 Conclusions

Container terminals serve as crucial hubs in international maritime logistics, acting as transshipment points for goods transported globally. Container trucks are vital for connecting terminals with inland logistics, and accurately predicting truck turnaround time can significantly enhance terminal operations. This study proposes a machine learning-based framework for turnaround time prediction, utilizing multi-source data integration and advanced neural network models. By improving turnaround time prediction accuracy, this research contributes to optimizing terminal operations, reducing congestion, and supporting better decision-making for logistics companies, terminal operators, and port authorities.

The research of this article mainly focuses on Bluetooth trip identification, prediction model construction and performance comparison. First, by filtering and integrating Bluetooth detection data, an efficient itinerary identification method was developed, achieving an identification accuracy of more than 90% under double verification. In the construction of the prediction model, this article designed and implemented a stacked long short-term memory (stacked LSTM) network. This model uses multi-source data to capture the temporal dependencies in port operations and improves the prediction accuracy through the optimization of hyperparameters. In the sensitivity analysis, this paper found that container truck arrival volume and wind speed are the variables that have the greatest impact on turnaround time, further verifying the advantages of multi-source data integration. In the sensitivity analysis, this paper found that container truck arrival volume and wind speed are the variables that have the greatest impact on turnaround time, further verifying the advantages of multi-source data integration.

Compared with traditional forecasting research, this paper not only relies on historical traffic flow and terminal operation data, but also introduces environmental variables and container dynamic information, thereby comprehensively capturing the complexity of terminal operations. At the same time, the use of stacked LSTM models breaks through the limitations of traditional regression models and can more effectively model nonlinear characteristics and time series patterns of data. In addition, this article innovatively introduces a masking mechanism to deal with the problem of missing data, which significantly improves the robustness of the model in scenarios with incomplete data quality.

The research in this article has important implications for managers. The prediction results can not only guide logistics companies to adjust the arrival time of container trucks to avoid congestion during peak periods, but also provide a decision-making basis for terminal managers to optimize resource allocation. Compared with the previous method of forcibly allocating container truck arrival times through quotas, this paper's method is more flexible. It guides container truck companies to spontaneously adjust scheduling strategies by publishing forecast information, effectively balancing terminal operation efficiency and flexibility. In addition, the results of the sensitivity analysis emphasize the importance of traffic and environmental variables in prediction, and managers should make full use of these data to improve port operation efficiency.

Future research directions include expanding the dataset to introduce more turnaround time-related features, such as equipment operating status and intermodal connection information; further optimizing the model to improve prediction accuracy and enhance interpretability; and studying logistics companies through behavioral logic, and deeply exploring the impact of predictive information on decision-making. At the same time, verifying and promoting the model in this article in other ports and exploring its universality and scalability will provide new ideas and methods for smart logistics management in global ports.

References

[1] E. van Asperen, B. Borgman, and R. Dekker, "Evaluating impact of truck announcements on container stacking efficiency", *Flexible Services and Manufacturing Journal*, vol. 25, no. 4, pp. 543–556, 2013, ISSN: 1936-6590. DOI: 10.1007/s10696-011-9108-1. [Online]. Available: https://doi.org/10.1007/s10696-011-9108-1.

[2] O. Sharif, N. Huynh, and J. M. Vidal, "Application of el farol model for managing marine terminal gate congestion", *Research in Transportation Economics*, vol. 32, no. 1, pp. 81–89, 2011, Container Transport Operations, ISSN: 0739-8859. DOI: https://doi.org/10.1016/j.retrec.2011.06.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S073988591100014X.

[3] R. Namboothiri and A. L. Erera, "Planning local container drayage operations given a port access appointment system", *Transportation Research Part E: Logistics and Transportation Review*, vol. 44, no. 2, pp. 185–202, 2008, Selected Papers from the National Urban Freight Conference, ISSN: 1366-5545. DOI: https://doi.org/10.1016/j.tre.2007.07.004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1366554507000701.

[4] Y. Du, Y. Zhao, and D. Gao, "Predictive simulation of external truck operation time in a container terminal based on traffic big data", *Journal of Shanghai Jiaotong University (Science)*, 2022, ISSN: 1995-8188. DOI: 10.1007/s12204-022-2415-8. [Online]. Available: https://doi.org/10.1007/s12204-022-2415-8.

[5] R. Wang, J. Li, and R. Bai, "Prediction and analysis of container terminal logistics arrival time based on simulation interactive modeling: A case study of ningbo port", *Mathematics*, vol. 11, no. 15, 2023, ISSN: 2227-7390. DOI: 10.3390/math11153271. [Online]. Available: https://www.mdpi.com/2227-7390/11/15/3271.

[6] S. Van der Spoel, C. Amrit, and J. Hillegersberg, "A benchmark for predicting turnaround time for trucks at a container terminal", May 2016.

[7] M. Sidarta, "Predicting truck cycle time in earthworks using a machine learning approach", M.S. thesis, TU Delft, http://resolver.tudelft.nl/uuid:0f3c5ea9-c2b5-447d-b282-d6d25843f97d, 2021.

[8] M. Torkjazi, N. Huynh, and S. Shiri, "Truck appointment systems considering impact to drayage truck tours", *Transportation Research Part E: Logistics and Transportation Review*, vol. 116, pp. 208–228, 2018, ISSN: 1366-5545. DOI: https://doi.org/10.1016/j.tre.2018.06.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1366554517309195.

[9] X. Zhang, Q. Zeng, and Z. Yang, "Optimization of truck appointments in container terminals", *Maritime Economics and Logistics*, vol. 21, no. 1, pp. 125–145, Apr. 2018. DOI: 10.1057/s41278-018-0105-0. [Online]. Available: https://doi.org/10.1057/s41278-018-0105-0.

[10] 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, ISSN: 1366-5545. DOI: https://doi.org/10.1016/j.tre.2011.04.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1366554511000433.

[11] K. G. Murty, Y. Wan, J. Liu, *et al.*, "Hongkong International terminals gains elastic capacity using a Data-Intensive Decision-Support system", *Interfaces*, vol. 35, no. 1, pp. 61–75, Feb. 2005. DOI: 10.1287/inte.1040.0120. [Online]. Available: https://doi.org/10.1287/inte.1040.0120.

[12] W. Zhao and A. V. Goodchild, "The impact of truck arrival information on container terminal rehandling", *Transportation Research Part E: Logistics and Transportation Review*, vol. 46, no. 3, pp. 327–343, 2010, ISSN: 1366-5545. DOI: https://doi.org/10.1016/j.tre.2009.11.007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1366554509001409.

[13] X. Zhang, Q. Zeng, and W. Chen, "Optimization model for truck appointment in container terminals", *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 1938–1947, 2013, Intelligent and Integrated Sustainable Multimodal Transportation Systems Proceedings from the 13th COTA International Conference of Transportation Professionals (CICTP2013), ISSN: 1877-0428. DOI: https://doi.org/10.1016/j.sbspro.2013.08.219. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877042813023458.

[14] 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, ISSN: 0191-2615. DOI: https://doi.org/10.1016/j.trb.2016.01.006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0191261516000151.

[15] F. Schulte, E. Lalla-Ruiz, R. G. González-Ramírez, and S. Voß, "Reducing port-related empty truck emissions: A mathematical approach for truck appointments with collaboration", *Transportation Research Part E: Logistics and Transportation Review*, vol. 105, pp. 195–212, 2017, ISSN: 1366-5545. DOI: https://doi.org/10.1016/j.tre.2017.03.008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1366554516300801.

[16] Z. Yang, G. Chen, and D. R. Moodie, "Modeling road traffic demand of container consolidation in a chinese port terminal", *Journal of Transportation Engineering*, vol. 136, no. 10, pp. 881–886, 2010. DOI: 10.1061/(ASCE)TE.1943-5436.0000152. eprint: https://ascelibrary.org/doi/pdf/10.1061/%28ASCE%29TE.1943-5436.0000152. [Online]. Available: https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29TE.1943-5436.0000152.

[17] G. Chen and Z. Yang, "Optimizing time windows for managing export container arrivals at Chinese container terminals", *Maritime Economics and Logistics*, vol. 12, no. 1, pp. 111–126, Feb. 2010. DOI: 10.1057/mel.2009.21. [Online]. Available: https://doi.org/10.1057/mel.2009.21.

[18] G. Chen and L. Jiang, "Managing customer arrivals with time windows: a case of truck arrivals at a congested container terminal", *Annals of Operations Research*, vol. 244, no. 2, pp. 349–365, Mar. 2016. DOI: 10.1007/s10479-016-2150-3. [Online]. Available: https://doi.org/10.1007/s10479-016-2150-3.

[19] Z. Chu, R. Yan, and S. Wang, "Evaluation and prediction of punctuality of vessel arrival at port: a case study of Hong Kong", *Maritime Policy and Management*, pp. 1–29, May 2023. DOI: 10.1080/03088839.2023.2217168. [Online]. Available: https://doi.org/10.1080/03088839.2023.2217168.

[20] J. Yu, G. Tang, X. Song, *et al.*, "Ship arrival prediction and its value on daily container terminal operation", *Ocean Engineering*, vol. 157, pp. 73–86, 2018, ISSN: 0029-8018. DOI: https://doi.org/10.1016/j.oceaneng.2018.03.038. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801818303032.

[21] S. Filom, A. M. Amiri, and S. Razavi, "Applications of machine learning methods in port operations – a systematic literature review", *Transportation Research Part E: Logistics and Transportation Review*, vol. 161, p. 102 722, 2022, ISSN: 1366-5545. DOI: https://doi.org/10.1016/j.tre.2022.102722. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1366554522001132.

[22] Z. Chu, R. Yan, and S. Wang, "Vessel turnaround time prediction: A machine learning approach", *Ocean and Coastal Management*, vol. 249, p. 107 021, 2024, ISSN: 0964-5691. DOI: https://doi.org/10.1016/j.ocecoaman.2024.107021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0964569124000061.

[23] L. R. Abreu, I. S. Maciel, J. S. Alves, L. C. Braga, and H. L. Pontes, "A decision tree model for the prediction of the stay time of ships in brazilian ports", *Engineering Applications of Artificial Intelligence*, vol. 117, p. 105 634, 2023, ISSN: 0952-1976. DOI: https://doi.org/10.1016/j.engappai.2022.105634. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197622006248.

[24] C. A. Sjoerd van der Spoel and J. van Hillegersberg, "Predictive analytics for truck arrival time estimation: A field study at a european distribution centre", *International Journal of Production Research*, vol. 55, no. 17, pp. 5062–5078, 2017. DOI: 10.1080/00207543.2015.1064183. eprint: https://doi.org/10.1080/00207543.2015.1064183. [Online]. Available: https://doi.org/10.1080/00207543.2015.1064183.

[25] B. N. Araghi, L. T. Christensen, R. Krishnan, J. H. Olesen, and H. Lahrmann, "Use of low-level sensor data to improve the accuracy of bluetooth-based travel time estimation", *Transportation Research Record Journal of the Transportation Research Board*, vol. 2338, no. 1, 29–34, Jan. 2013. DOI: 10.3141/2338-04. [Online]. Available: https://journals.sagepub.com/doi/abs/10.3141/2338-04.

[26] M. Margreiter, F. Busch, and C. Carstensen, "The evolution of bluetooth detection rates", en, in *Transportation Research Board*, 2019.

[27] B. N. Araghi, R. Krishnan, and H. Lahrmann, "Mode-specific travel time estimation using bluetooth technology", *Journal of Intelligent Transportation Systems*, vol. 20, no. 3, 219–228, May 2015. DOI: 10.1080/15472450.2015.1052906. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/15472450.2015.1052906.

[28] C. Bachmann, M. J. Roorda, B. Abdulhai, and B. Moshiri, "Fusing a bluetooth traffic monitoring system with loop detector data for improved freeway traffic speed estimation", *Journal of Intelligent Transportation Systems*, vol. 17, no. 2, 152–164, May 2012. DOI: 10.1080/15472450.2012.696449. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/15472450.2012.696449.

[29] F. Garrido-Valenzuela, S. Raveau, and J. C. Herrera, "Bayesian route choice inference to address missed bluetooth detections", *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1865–1874, 2022. DOI: 10.1109/TITS.2020.3028804.

[30] S. Sharma, H. Van Lint, L. Tavasszy, and M. Snelder, "Estimating route choice characteristics of truck drivers from sparse automated vehicle identification data through data fusion and bi-objective optimization", *Transportation Research Record Journal of the Transportation Research Board*, vol. 2676, no. 12, 280–292, Jun. 2022. DOI: 10.1177/03611981221095089. [Online]. Available: https://journals.sagepub.com/doi/10.1177/03611981221095089#bibr15-03611981221095089.

[31] F. Crawford, D. Watling, and R. Connors, "Identifying road user classes based on repeated trip behaviour using bluetooth data", *Transportation Research Part a Policy and Practice*, vol. 113, 55–74, Apr. 2018. DOI: 10.1016/j.tra.2018.03.027. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S096585641730352X.

[32] A. Wijbenga, J. Bac, and S. Boerma, "Een nieuwe manier van verkeerswaarneming: Bluetooth", *B50*, 2011.

[33] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory", *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. eprint: https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735.

[34] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Stacked bidirectional and unidirectional lstm recurrent neural network for forecasting network-wide traffic state with missing values", *Transportation Research Part C: Emerging Technologies*, vol. 118, p. 102 674, 2020, ISSN: 0968-090X. DOI: https://doi.org/10.1016/j.trc.2020.102674. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X20305891.

[35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[36] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values", *Scientific Reports*, vol. 8, no. 1, 2018. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85045746406&doi=10.1038%2fs41598-018-24271-9&partnerID=40&md5=b304485b031bc178a30bd0c008b7c619.

[37] C.-H. Wu, J.-M. Ho, and D.-T. Lee, "Travel-time prediction with support vector regression", *IEEE transactions on intelligent transportation systems*, vol. 5, no. 4, pp. 276–281, 2004.

[38] Y. Hollander and R. Liu, "Estimation of the distribution of travel times by repeated simulation", *Transportation Research Part C: Emerging Technologies*, vol. 16, no. 2, pp. 212–231, 2008.

[39] A. Khosravi, E. Mazloumi, S. Nahavandi, D. Creighton, and J. Van Lint, "A genetic algorithm-based method for improving quality of travel time prediction intervals", *Transportation Research Part C: Emerging Technologies*, vol. 19, no. 6, pp. 1364–1376, 2011.

[40] R. Lederman and L. Wynter, "Real-time traffic estimation using data expansion", *Transportation Research Part B: Methodological*, vol. 45, no. 7, pp. 1062–1079, 2011.

[41] I. Abualhaol, R. Falcon, R. Abielmona, and E. Petriu, "Data-driven vessel service time forecasting using long short-term memory recurrent neural networks", in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 2580–2590. DOI: 10.1109/BigData.2018.8622626.

[42] A. Balster, O. E. Hansen, H. Friedrich, and A. Ludwig, "An ETA prediction model for intermodal transport networks based on machine learning", *Business and Information Systems Engineering*, vol. 62, no. 5, pp. 403–416, May 2020. DOI: 10.1007/s12599-020-00653-0. [Online]. Available: https://doi.org/10.1007/s12599-020-00653-0.

[43] J. Pruyn, A. Kana, and W. Groeneveld, "Chapter 4 - analysis of port waiting time due to congestion by applying markov chain analysis", in *Maritime Supply Chains*, T. Vanelslander and C. Sys, Eds., Elsevier, 2020, pp. 69–94, ISBN: 978-0-12-818421-9. DOI: https://doi.org/10.1016/B978-0-12-818421-9.00005-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128184219000057.

[44] T. Antamis, C.-R. Medentzidis, M. Skoumperdis, *et al.*, "Ai-supported forecasting of intermodal freight transportation delivery time", in *2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*, 2021, pp. 1–6. DOI: 10.1109/ITMS52826.2021.9615330.

[45] D. Y. K. F. Y. Wenhao Peng Xiwen Bai and J. Wu, "A deep learning approach for port congestion estimation and prediction", *Maritime Policy & Management*, vol. 50, no. 7, pp. 835–860, 2023. DOI: 10.1080/03088839.2022.2057608. eprint: https://doi.org/10.1080/03088839.2022.2057608. [Online]. Avail-

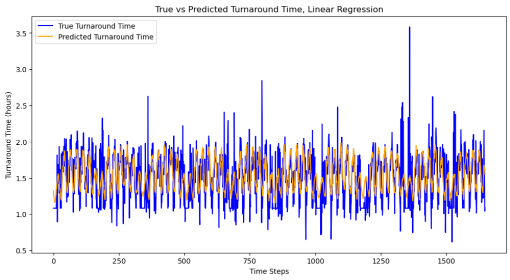
able: https://doi.org/10.1080/03088839.2022.2057608.

A Appendix A: Literature Table

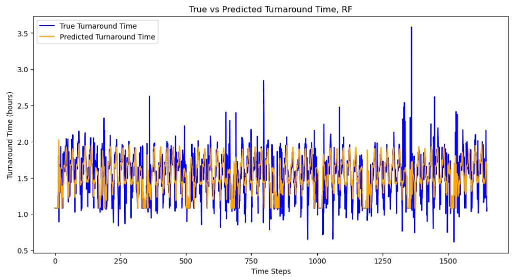
Table 5: Influencing factors involved in each paper

Reference	Year	Method	Influencing factors						
			Weather	Traffic Condition	Time period	Cumulative Flow	Cargo info	Vehicle/vessel info	Port congestion
Wu et al. [37]	2004	SVR	√	√		√			
Hollander and Liu [38]	2008	Microsimulation	√			√			
Khosravi et al. [39]	2011	Genetic algorithm	√		√	√			
Lederman and Wynter [40]	2011	Data expansion	√	√					
Sjoerd van der Spoel [24]	2017	Data mining	√			√			
Abualhaol et al. [41]	2018	LSTM						√	
Balster et al. [42]	2020	ML	√		√	√	√		
Pruyn et al. [43]	2020	MCA				√			
Antamis et al. [44]	2021	ML regression			√		√		
Wang et al. [5]	2023	Data mining	√		√	√	√		
Du et al. [4]	2022	System dynamics				√			√
Peng and Wu [45]	2023	LSTM						√	√
Chu et al. [19]	2023	Random forest						√	
Chu et al. [22]	2024	XGBoost			√			√	
This study		LSTM	√	√		√		√	

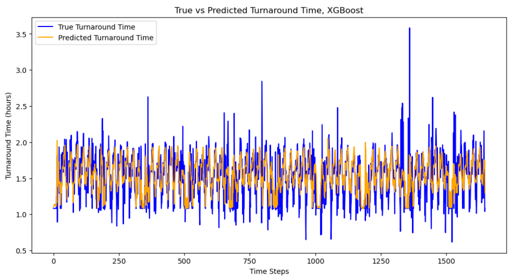
B Appendix B: Model Comparisons



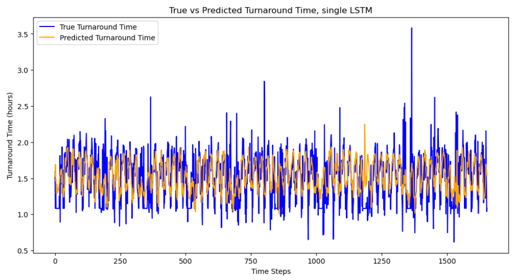
(a) Linear Regression



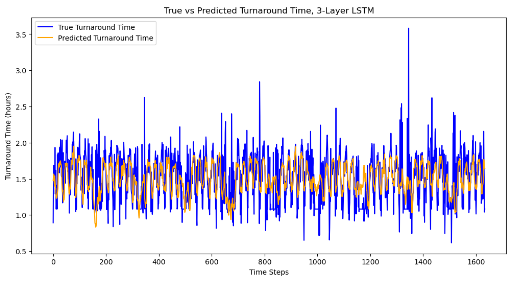
(b) Random Forest



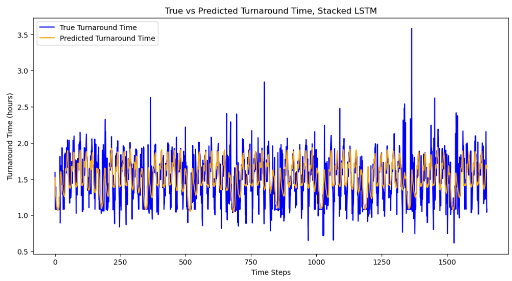
(c) XGBoost



(d) Single-layer LSTM



(e) Multi-layer LSTM



(f) Stacked LSTM

Figure 15: Comparison of results for different models