Center for Advanced Multimodal Mobility Solutions and Education

Project ID: 2017 Project 07

USE OF VESSEL AUTOMATIC INFORMATION SYSTEM DATA TO IMPROVE MULTIMODAL TRANSPORTATION IN AND AROUND PORTS

Final Report

by

Mehdi Azimi, PhD, PE (ORCID ID: https://orcid.org/0000-0001-5678-0323)
Assistant Professor, Department of Transportation Studies
3100 Cleburne Ave., Houston, TX 77004
Phone: (713) 313-1293; E-mail: mehdi.azimi@tsu.edu

Akintola Moses Aremu (ORCID ID: https://orcid.org/0000-0003-3361-2030)
Graduate Research Assistant
Dept. of Transportation Studies, Texas Southern University
3100 Cleburne St., Houston, TX, 77004
Phone: (713) 313-5696; E-mail: a.aremu3682@student.tsu.edu

Yi Qi, PhD (ORCID ID: https://orcid.org/0000-0002-6314-2626)
Professor and Chair
Department of Transportation Studies, Texas Southern University
3100 Cleburne Ave., Houston, TX 77004
Phone: (713) 313-6809; E-mail: yi.qi@tsu.edu

for

Center for Advanced Multimodal Mobility Solutions and Education
(CAMMSE @ UNC Charlotte)
The University of North Carolina at Charlotte
9201 University City Blvd.
Charlotte, NC 28223

September 2018

ACKNOWLEDGEMENTS

This project was funded by the Center for Advanced Multimodal Mobility Solutions and Education (CAMMSE @ UNC Charlotte), one of the Tier I University Transportation Centers that were selected in this nationwide competition by the Office of the Assistant Secretary for Research and Technology (OST-R), US Department of Transportation (US DOT), under the FAST Act. The authors would like to thank Houston Pilots for their comments and Brian W. Smith (US Coast Guard Vessel Traffic System, Houston/Galveston) and Oceaneering International, Inc. for their assistance in data collection.

DISCLAIMER

Table of Contents

List of Figures

List of Tables

EXCUTIVE SUMMARY

Multimodal transportation is an evolving system in supply chain management and an effective approach for facilitating the movement of cargo when different modes of transportation are available and involved. Since one mode of transportation is usually insufficient for door-to-door transportation of cargo, multimodal transportation has become an important concept. Hence, it becomes necessary to transfer goods from one mode of transportation to another. For an effective multimodal system, the different modes of transportation involved require close coordination and precise synchronization, especially in terms of arrival times and cargo allocation (synchro-modality).

The accuracy in arrival time of the vessel is most vital to imports, since it initiates the process of a multimodal transfer. Lack of certainty in estimated time of arrival (ETA) creates problems like delays and congestion at ports. It also leads to inadequate planning and resource management for port facilities and receiving modes of transportation. Vessel Automatic Information System (AIS) data provide vessels' voyage information, including the ETA as determined by the vessel's captain/operator. This information (the captain's ETA) is manually inputted into the system and thus is subject to errors. Furthermore, captains sometimes forget to update such information, which affects the results of the analysis. Hence, we propose a way to generate ETAs from a system that does not require the captain's ETA as input.

This research describes an approach that generates the ETA of vessels to the port terminals by using machine learning and AIS data. The results of the analysis show that near-exact predictions can be achieved without prior estimations by vessel captains. The results indicate that the farther from the destination, the more errors are made in prediction. This is also evident in the comparison of prediction errors between Bayport and Barbours Cut, two container terminals in the Port of Houston. The analysis shows that predictions made at the terminal level are more accurate than at the buoy level. The ETA predicted from this approach provides an adequate timeframe within which terminal and trucking companies can plan for the vessel's arrival.

1.1 Problem Statement

Multimodal transportation is an evolving system in supply chain management and movement of cargo. It has become an especially important concept because one mode of transportation is often insufficient for door-to-door transportation of cargo. Hence, it becomes necessary to transfer goods from one mode of transportation to another and to encourage the management of this process by the same carrier company for effective coordination. For an effective multimodal system, the various modes of transportation involved require synchronization, especially in terms of arrival times and cargo allocation (synchro-modality). Accuracy in estimating vessels' arrival time is most vital to imports, since it initiates the process of multimodal transfer. Lack of certainty in Estimated Time of Arrival (ETA) creates problems like delays and congestion at ports. It also leads to inadequate planning and resource management for port facilities and receiving modes of transportation. Vessel Automatic Information System (AIS) data provide vessels' voyage information, including the ETA as determined by the captain. This information is manually inputted into the system, and thus is subject to errors as captains sometimes forget to update such information, which affects the results of analysis. Hence, we propose a way to generate ETAs from a system that does not require the captain's ETA as input.

1.2 Objectives

The objectives of this report are to (1) identify and collect AIS data for ETA determination; (2) determine the ETA of vessels via a machine learning approach; and (3) evaluate the accuracy of the determined ETA.

1.3 Expected Contributions

To accomplish the objectives of the study, several tasks have been undertaken to develop a network for predicting vessels' time of arrival when the captain's ETA input to AIS is unavailable. This will make it possible for the operators of carriers without prior knowledge of the estimated duration of a given trip to generate ETAs for their vessels based on their current locations and other available parameters.

1.4 Report Overview

Chapter 2 of the report provides information on the different structures and components of a vessel-to-truck multimodal system and their interrelations. Stakeholders are identified, and the importance and challenges of a multimodal system are presented. With inaccurate ETAs identified as one of the problems of a multimodal system, Chapter 3 describes various approaches for deriving the ETA of vessels from data. This chapter also reviews previous studies on ETA determination, including those that employ machine learning to predict vessels' time of arrival to ports. Chapter 4 provides information on the Port of Houston, describes its activities, specific characteristics/facilities, and uniqueness as it relates to container cargo. Chapter 5 describes the methodology of this research as well as the data collection and approach for determining ETA from the data collected. Chapter 6 presents numerical results obtained from the analysis. Chapter 7 concludes with a summary and discussion of directions for future research.

Chapter 2. Literature Review

2.1 Introduction

This chapter provides an overview of the operations of a multimodal system for container cargo, while identifying its benefits and problems as well as ways of improving the system. The review identifies the root cause of a major problem and presents existing practices that have aimed to improve the system. This review also summarizes operations in the Port of Houston as pertains to container cargo.

2.2 Multimodal Transportation System

Multimodal transportation is often mistaken with intermodal transportation. These two terms are highly similar except that in multimodal transportation, the same carrier company is responsible for moving the shipment in all legs and modes of transportation employed. Hence, the whole transportation process is under a single contract or bill of lading. Multimodal "freight" transportation can thus be defined as the movement of "cargo" by the coordinated and sequential use of two or more modes of transportation under a single contract or bill of lading. Multimodal transportation has been proven to increase the supply chain productivity of shipment and the performance of distribution of cargo at large (Mokhtar and Shah, 2013).

Multimodal transportation can take various forms, depending on the elements (mode of transportation) involved. Table 2.1 delineates these elements.

Table 2.1: Elements of Multimodal Transportation

	Carriers	**Conveyance**	**Terminal**	**Infrastructure**
Ocean	Shipping lines	Ships and barges	Ports	Sea and inland waterways
Road	Motor carriers	Trucks	Truck terminals	Roadways
Air	Air cargo carriers	Airplanes	Airports	Airways
Rail	Railroads	Trains	Rail terminals	Railways

Figures 1-3 illustrate three basic multimodal cargo movements: truck-marine, truck-air, and truck-rail. It should be noted that more combinations of elements are possible in multimodal transportation than listed above.

2.2.1 Truck-Marine

A typical truck-marine cargo movement starts with the shipper or consignor loading the cargo into the container. A motor carrier picks up the container from the shipper and transports it to the seaport by road. When the container arrives at seaport, it is transferred to the vessel (ocean carrier) that transports it to an overseas port, where the container is transferred to the second motor carrier for delivery to the consignee.

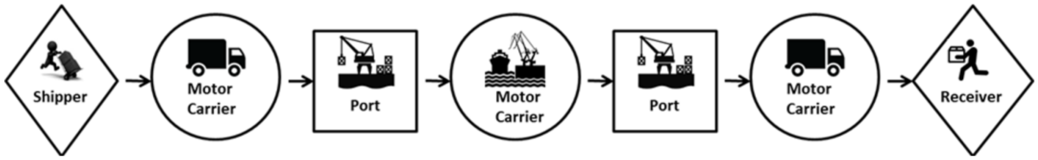


Figure 2.1: Truck-Marine Multimodal Transportation

2.2.2 Truck-Air

In a typical truck-air multimodal cargo movement, a motor carrier picks up the cargo from the shipper or consignor and transports it to an airport freight terminal. The cargo is then transferred to the airplane, which transports it to another airport, where a second motor carrier picks it up and delivers it to the consignee.

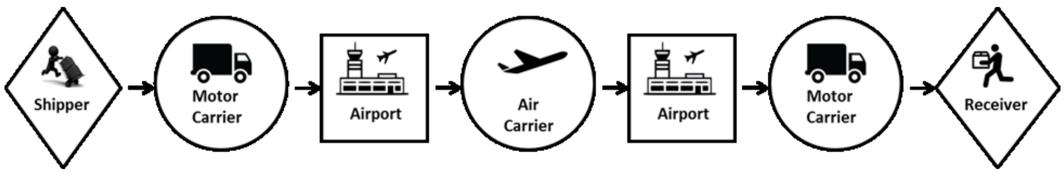


Figure 2.2: Truck-Air Multimodal Transportation

2.2.3 Truck-Rail

In a truck-rail combination, a motor carrier picks up the cargo from the shipper and transports it to the rail terminal, where it is transferred to a rail car. The cargo is transported by rail to another rail terminal, where the second motor carrier picks it up and delivers to the consignee.

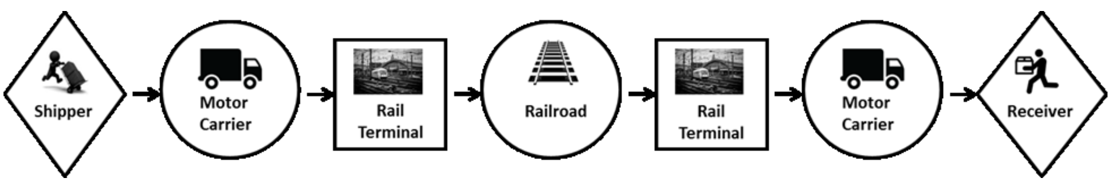


Figure 2.3: Truck-Rail Multimodal Transportation

Which mode of transportation to employ is a critical decision and highly dependent on the performance variables and availability. Apart from the logistic cost of moving freight, some of the most important factors to consider when selecting the mode of transportation, according to the Center for Urban Transportation Research (CUTR) at the University of South Florida, are schedule reliability and trip time. Shipment delays may affect the logistic costs, for example, by increasing the inventory cost or adding production cost, which in turn reduces the system's reliability (Notteboom, 2006).

2.3 Vessels in Multimodal Transportation

For the first time since 2010, the economic growth rate has outperformed expectations. In 2017, the GDP grew by 3.7%, and this trend is expected to continue through 2018 with a prediction of 4.0% growth (Hatzius et al., 2017). Growth in GDP, trade and seaborne shipments are interlinked and continue to move in tandem (Maritime-insight, 2015). The era of rapid economic and technical-technological development of modern production requires a transportation system that is well-organized and, above all, safe. Maritime transportation involves transportation of passengers and goods by sea, also known as “shipping trade,” which most often is cargo shipping. Samija (n.d.) stated that shipping operations are operated in accordance with their operational processes and quality control policies. These processes and policies are supervised by competent state institutions and international organizations for control of maritime navigation.

The basic function of maritime transportation is to physically transport cargo from the area of supply to the area of demand, following regulated procedures and policies that facilitate the activity. Essential for the movement of goods by maritime transportation are the following components:

- functional infrastructure, such as ports/terminals;
- means of transportation, such as ships and barges in good working condition; and
- organizational systems to ensure that ships and fixed infrastructure are used effectively and efficiently.

Maritime transportation has been highly relevant in multimodal transportation and transportation of cargo in general, due to its advantages in safety, energy efficiency, and environmental quality. Table 2.2 shows the advantages of maritime transportation over rail transportation. The capacity of cargo vessels is its biggest advantage over other modes, but this is also subject to vessel size (Tennessee Tombigbee Waterway).

Table 2.2: Comparing Vessel to Rail Car and 100-Car Train Unit

	Number of miles per gallon carrying one ton of cargo	**Hydrocarbons emitted (lbs/ton-mile)**	**Deaths per billion ton-miles**
Barge capacities	514	0.0009	0.01
One rail car	202	0.0046	0.84
100-car train unit	59	0.006	1.15

Source: Tennessee Tombigbee Waterway (2017)

Another great advantage of maritime transportation is its cost. Based on the data presented by Ballou (1998) and displayed in Table 2.3, maritime shipping has a lower cost per ton-mile than the rail and road transportation mode. This gives preference to the selection of vessels as a mode transportation, especially for large cargo.

Table 2.3: Cost Associated with Different Modes of Transportation (per ton-mile)

Mode	**Maritime**	**Rail**	**Road**
Cost (1995 USD)	1¢	3¢	25¢
Cost (2014 USD)	1.6¢	5.0¢	$3.88

Source: Ballou (1998)

2.4 Trucks in Multimodal Transportation

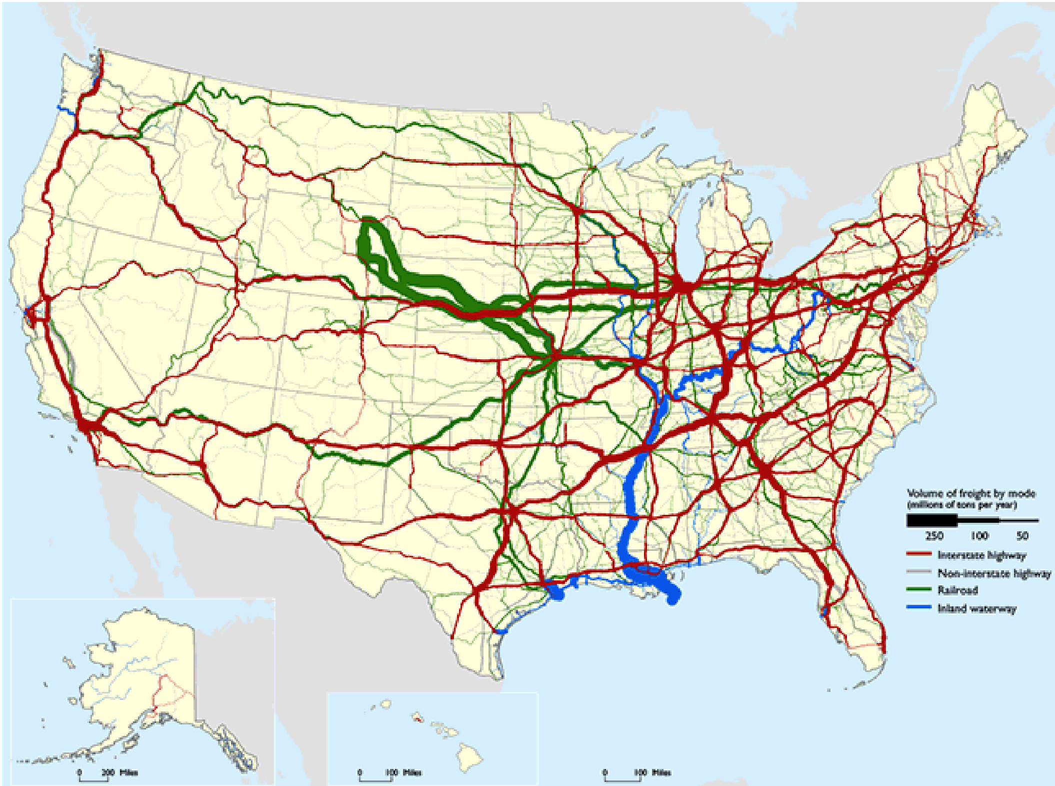
The importance of trucks to multimodal transportation cannot be over-emphasized. In most cases, a truck begins and ends the movement of freight either by rail, vessel or plane. Although the cost of transporting by road is higher than by other means, the accessibility and efficient network capacity of the road cannot be matched. The volume of freight moved by truck has grown in tandem with the increase in the marine, rail and air freight volume. Overall freight tonnage is expected to grow between 2016 and 2027 by 35%. Over that same period, the amount of freight moved by trucks is expected to grow by 27%. Truckload volumes will grow 2% annually between 2016 and 2022 and 1.6% per year thereafter until 2027 (American Trucking Associations 2017). Another publication by the USDOT (2018) forecasts a 44% increase in the tonnage moved by trucks between 2015 and 2045, which is approximately a 1.5% increase annually (see Table 2.4). This value is very close to figures stated by American Trucking Associations (2017). Therefore, as the years go by, trucks will likely be needed more.

Table 2.4: Annual Tons of Freight Moved across the US and Projected Increase

	2015	**2045**	**Increase**
Truck	11.5 billion	16.5 billion	44%
Rail	1.7 billion	2.1 billion	24%
Water	835 million	1.2 billion	38%
Air	7 billion	24 billion	234%
Total	18 billion	25.3 billion	40%

Source: USDOT - Beyond Traffic 2045 (2018)

Road transportation is also highly relevant to freight movement, because it has the highest network connectivity, which in turn promotes end-to-end delivery of cargo. Figure 2.4 displays the network for 2011 commodity movement. Highways constituted a major network throughout the US. They even support the inland waterways; as they are mainly used along the Mississippi River and its tributaries (Dong et al., 2015).



Adopted from USDOT Bureau of Statistics, 2015

Figure 2.4: Tonnage on Highways, Railroads, and Inland Waterways in 2011

2.5 Containerization in Multimodal Transportation

Containerization is the generalized use of the container as a support for freight transportation. It has increasingly been adopted as a mode for supporting freight distribution since a growing number of transportation systems can handle these standardized containers. Efficiency in the movement of freight from one location to another using just one mode of transportation has always been limited. It has remained so due to the difficulties encountered when transferring goods from one mode of transportation to another. Shippers incur additional terminal costs and delays when load unit needs to be changed, which is common for bulk transportation (Rodrigue et al., 2006).

A shift to multimodal transportation was encouraged by the success achieved through the introduction of the container. The most obvious advantage of using shipping containers is the fact that it makes loading and unloading easier and enables rapid change from one mode to another (Broeze, 2002). Containerization has impacted the conventional transportation system in two distinct ways – spatially and organizationally. With the introduction of the container system, processes in the port have drastically changed in terms of equipment, manpower, and port charges. This has enhanced organized port operations although complex; however, the implementation of a container system achieves efficiency by promoting an organized structure at the terminals. Containerization has also encouraged transportation companies to embrace multimodal operations. The focus is now more on the organization of the transportation industry and the synchronization of an integrated logistical system (Carrese and Tatarelli, 2011).

To operate an efficient multimodal system, intensive co-operation and co-ordination among the various transportation modes are essential. Containers have the advantage of being used by several modes of transportation (i.e., maritime, rail and road) since these modes can handle containers smoothly. International Standardization for Organization (ISO) containers are 10, 20, 30 or 40 feet long. However, for measurement, the reference size container 20 feet long, 8 feet high and 8 feet wide, corresponding to the twenty-foot equivalent units (TEUs). The most common container that can be loaded on ship, truck or railcar is 40 feet long (Mokhtar and Shah, 2013).

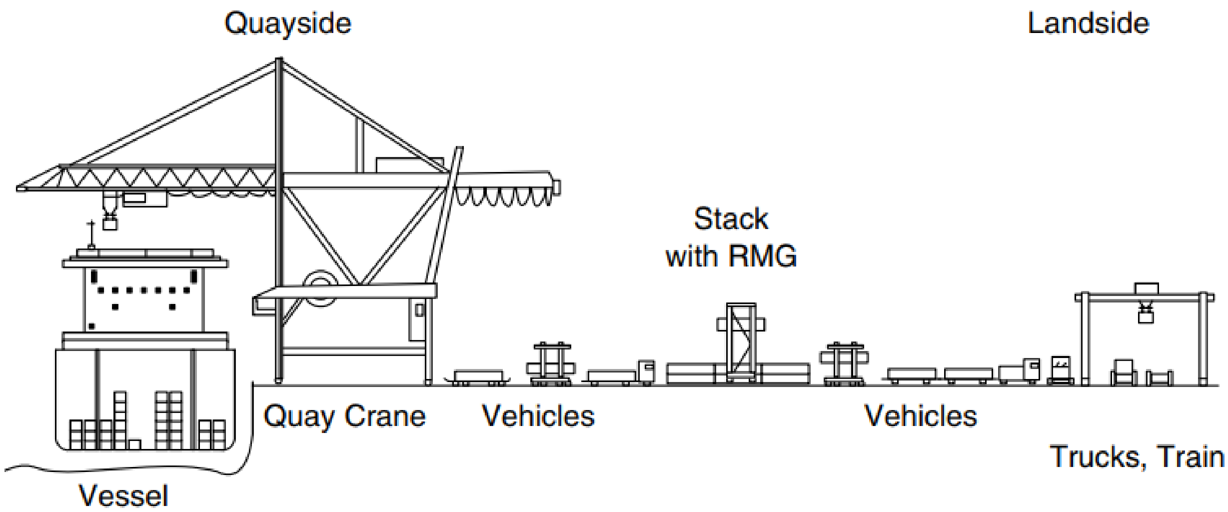
2.6 Stakeholders in Multimodal Transportation

2.6.1 Carriers

Carriers are recognized companies chartered by the consignor for the transportation of cargo from origin to destination. In the case of a marine-truck multimodal system, the carrier splits its activities into vessel transportation and truck transportation. To carry out this task, they operate a fleet of vessels that are suited for transporting the type of cargo intended. Vessel carriers can transport cargo in containers, tankers or other means. A carrier in a multimodal system is known as a Multimodal Transportation Operator (MTO). The type of cargo transported by the vessels determines the type of trucks that will be operated by the MTO. The MTO is liable for any losses or damage to the goods as well as any delays in the delivery while the goods are in their possession (Jashari, 2007).

2.6.2 Container Terminals

A container is a mode of cargo transportation that requires specialized ports and terminals for effective handling. The facilities required to transport containers between ships and shore include berths for docking the ship, land areas for container storage and handling equipment like cranes, which are basically used to load and offload containers to and from the vessels (Liu, 2010). A typical container terminal is represented in Figure 2.5. According to Steenken (2004), a container terminal is divided into quayside and landside operations. Quayside activity is responsible for transferring the containers between ship and shore. Quay cranes are the main equipment used in this part of the terminal. Landside operations involve the transfer of containers from the stacking area to modes of land transportation like trains and trucks.



Adopted from Steenken, 2004

Figure 2.5: Schematic Representation of Container Terminal

2.6.3 Importers

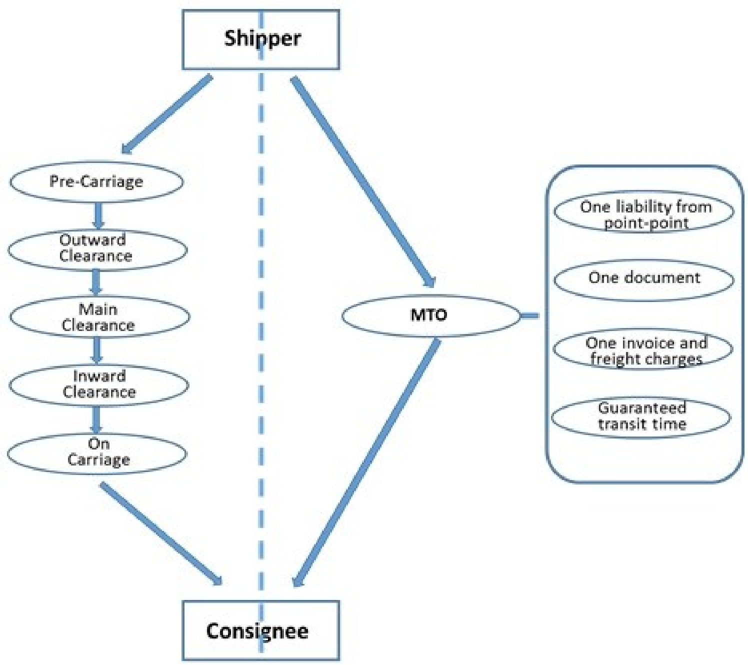
Importers are the buyers and receivers of cargo in transported containers and play a significant role in carrier selection as they are more interested in on-time and cost-efficient delivery of goods at the specified location. Importers take into account the convenience they can get from the use of a carrier. Importers are interested in receiving their goods in good shape and within the agreed time-frame at the destination of interest.

2.7 Advantages of Multimodal Transportation

The multimodal transportation system is considered a game-changer as it is quite effective in solving most cargo mobility issues. By combining more than one mode of transportation and properly managing the entire process, it facilitates the best rate and timely delivery. The introduction of multimodal transportation eliminates the need for lengthy processing, as shown in Figure 2.5. Multimodal transportation has been of great benefit to the movement of freight in the several ways:

a) Minimizes time loss at trans-shipment points – Due to continuous and unbroken communication links maintained by multimodal transportation operators, there is effective coordination of trans-shipment points, which avoids the time that would have been lost if the transportation processes were segmented.
b) Provides faster transit of goods – Multimodal transportation physically shortens the market as goods are transported quickly from one point to another. The disadvantages of distance from markets and the typing-up of capital are eliminated. It also reduces the distance between source materials and customers.

c) Reduces the burden of documentation and formalities – Multimodal transportation minimizes the burden of multiple documentation and other formalities connected with each segment of the transportation chain since one operator handles all modes.
d) Reduces cost of exports – The inherent advantages of a multimodal transportation system help to minimize the cost of exports and improve the competitive position of the MTO.
e) Establishes only one agency to deal with – Consignee just needs to transact with the MTO as far as transportation of goods is concerned. Multimodal transportation eliminates the need to establish a connection with all entities individually.



Adopted from Hayuth (1987)

Figure 2.6: Segmented Transportation versus Multimodal Transportation

2.8 Factors Affecting the Efficiency of Multimodal Transportation

The growth in multimodal freight transportation is associated with the pressure for improved performance. This warrants the identification of factors that could affect the efficiency and throughput of the system. A major factor affecting multimodal transportation, and particularly the truck-marine combination, is congestion. Congestion at the port could result from vessel operators' poor schedule reliability, inefficiency of the transportation infrastructure that links a marine terminal to roadways, and the time chosen by shippers or truckers to pick up their shipments (World Shipping Council, 2015). Congestion will not shut down ports and terminals but can have a devastating impact on reliance in multimodal services.

The efficiency of a transportation system is a function of its reliability. Reliability and accuracy of vessels' Estimated Time of Arrivals (ETAs) are some of the most important characteristics of freight transportation. In an era of just-in-time inventory systems, the weakest link of a multimodal system will not be the port or terminal operations but the reliability in arrival time of the vessel to the port. Therefore, meeting the time announced in schedules is significant to shipping lines. Unless the inaccuracy of ETAs is addressed, the multimodal system as a whole will be inefficient. Schedule reliability may be the factor that shippers consider most important when selecting a mode of transportation and planning their supply chains with realistic expectations of delivery time (Notteboom, 2006).

Chung and Chiang (2011) categorized shipping activities into port assignments and navigation by sea, which were further divided into factors and criteria that could influence schedule reliability. These factors are as follows: operating strategy of shipping lines, staff in shipping lines, process management in shipping lines, and port's condition. The factors and criteria are summarized in Table 2.5 below.

Table 2.5: Influential Factors on Schedule Reliability of Container Shipping Lines

Goal	Objective	Criteria	Statement of criteria
Influential Factors on Schedule Reliability	**Operating strategy of shipping lines**	Planning suitable ports	Shipping lines need to choose suitable ports according to port condition, cargo volume, etc.
		Chase strategy	Whether shipping lines execute the chase strategy or not
		Specialized terminal investment	Shipping lines have invested in specialized terminals
	Staff in shipping lines	Staff's sense of mission	Every staff member has strong sense of mission in their work
		Ability of staff to coordinate with external relations	Staff should coordinate well with market players (e.g., port authority and customs) to decrease waiting time and increase efficiency
		Control and management of staff in terminal	Shipping lines should effectively control and manage staff in the terminal to avoid strike or slow work pace
	Process management in shipping lines	Well-arranged time window	Shipping lines should plan the time window appropriately
		Planning the berth and warehouse beforehand	Before arriving to port, shipping lines should plan the berth and warehouse
		Trans-ship arrangement	Shipping lines should trans-ship properly to avoid delays in delivery
	Ports' condition	Free-flowing of ports' access roads	Access roads of a port are free-flowing
		Berth allocation	Berth allocation will influence operating time
		Terminal efficiency	Terminal efficiency will influence operating time

Source: Chung and Chiang, 2011

Assessing the importance of each criterion identifies the significance of factors. In terms of schedule reliability, results showed that ‘process management in shipping lines’ is the most influential factor and ‘well-arranged time window’ is the most important factor. A total of 81.5% of the criteria were split between the top five criteria – ‘well-arranged time window,’ ‘transship arrangement,’ ‘planning suitable ports,’ ‘planning the berth and warehouse beforehand,’ and ‘control and management of staff in the terminal.’ These results show that time-related factors are the top criteria; thus, the availability of a vessel’s accurate time of arrival to the port greatly impacts the system’s efficiency. With this in mind, perfecting the ETA provided to shippers and shipping companies is paramount to the industry’s growth.

2.9 Synchronizing Vessel and Truck Arrival Time

The most important performance measure for port operation is the turnaround time of trucks in the terminal (Esmer, 2008). Their time has not reached its optimum due to various factors, like the lack of synchronization in the time of arrival of vessels and the time of arrival of loading trucks. This creates congestion at the ports, especially when trucks have to wait a long time for the vessel’s arrival. Congestion is a major problem at ports and is becoming common at major US ports. In an attempt to maximize profit, ocean carriers have employed bigger vessels, which brings the benefit of providing more fuel efficiency and carrying more cargo. The introduction of large vessels into the system calls for more trucks during offloading, which in turn contributes to congestion at the terminals. Thus, a more synchronized logistic system should be employed to compensate for this upgrade and the congestion it causes. Synchronization between vessel arrival and the availability of trucks to pick up cargo can only be made possible if there is accuracy in vessel’s ETA coupled with ‘just in time’ availability of trucks through an effective operational plan.

For a synchronized system to be achieved in multimodal transportation, responsibilities must be performed effectively by the trucking company, arriving vessel and terminal. Hence, factors related to these different components need to be addressed.

2.9.1 Truck-Related Factors:

Trucking operation at the beginning or end of a multimodal process plays a very important role in enhancing the effectiveness and synchronization of the process. Some of the basic characteristics of the trucking operation that should be taken into consideration are traffic condition before arrival at the terminal, availability of trucks, and capacity of the trucks.

Traffic before arriving at port

The traffic condition of roads leading to the terminal plays a major role in the timeliness of the trucks. Upon a vessel’s arrival, it is very important that receiving trucks are available. The availability of these trucks will only be possible when the trucking company takes into consideration the traffic conditions of roads and any obstructions that could prevent them from arriving at the terminal on time.

Availability and capacity of trucks

A multimodal system will function well and have synchronized transfer of containers when there are enough trucks to receive and transport containers from arriving vessels. One of the problems that multimodal transportation seeks to eliminate is the use of storage capacity. Hence, it is very pertinent for the available trucks to have sufficient capacity to accommodate the containers carried by the arriving vessel.

2.9.2 Vessel-Related Factors

The activities of the vessels before arrival at the terminal are major determinants of the seamless transfer of container to and from the truck. The most important factor here is the effective communication between operators of the vessel, the terminal, and the trucks. Information about the vessel should be constantly updated as changes occur in the sailing activities. Information provided to truckers at the destination port should be up to date and accurate so that the trucks can prepare properly for the vessel's arrival. Information that should be provided includes any delays, whether deliberate or accidental. A deliberate delay may occur when the captain desires to arrive at the port at certain time of day. This might be influenced by existing policies at the port of call. Some ports operate on daylight restrictions for certain vessel. Hence, the ship captain may decide to lower his speed in order to arrive in the daytime so as to avoid waiting after arrival. In the event that an uncontrollable delay is experienced by the vessel due to damage or weather, such information should be provided to the trucking company ahead of time.

2.9.3 Terminal-Related Factors

The terminal greatly impacts how seamlessly and effectively shipments can be transferred from vessel to truck. Some basic characteristics of the terminal that determine this success are its structure and size, its equipment for handling containers, and its security initiatives.

Structure and size of terminal

The way a terminal is structured greatly determines how smoothly the process flows. For a terminal to be effective, three attributes must be considered in the design and arrangement of the facility: safety, free flow, and truck turnaround time. The size and arrangement of the terminal can influence the turnaround time of trucks entering and exiting while taking into consideration the free flow and safety of workers. The size and arrangement of the terminal must provide for easy and safe maneuvering of truck and terminal equipment. Truck turnaround time, according to Yoon (2007), is the time it takes a truck to complete a transaction – picking up an import container or dropping a container off. There are four activities that affect trucks' turnaround time at the terminal: inbound process, yard crane activities, inner transportation vehicle process, and outbound process. These activities will only be performed efficiently if the structural arrangement and size of the terminal fit the activities.

Container-handling equipment and facilities available

Available equipment and facilities for container handling play a vital role in determining the efficiency of terminal activities. The major equipment in a terminal are cranes of different types and capacity. To improve truck turnaround time, Huynh and Walton (2005) examined a measure of increasing yard cranes. They developed methods to help terminal operators evaluate and apply this measure in making decisions for crane purchase and determining how many to purchase. They also studied the availability of cranes versus truck turnaround time. Results indicate that having more road cranes generally lowers truck turnaround time. Another container terminal factor that enhances efficient operation is the size of the berth, which is always considered when adding more equipment like wharf cranes to the existing set of equipment.

Security initiative of the terminal

One major security program at container terminals is the Container Security Initiative (CSI) announced in 2002. The primary purpose of CSI is to protect the global trading system and the trade lanes between CSI ports. The activities of the CSI in US ports are performed by a team of officers deployed to work with their counterparts in the host nation to target all containers that are potential threats (Container Security Initiative, n.d.). The effectiveness of security checks at originating ports can greatly reduce the turnaround time of trucks picking up containers at the terminal. Thus, it is important to understand the security initiative operated by the terminal and the extent to which it may delay the readiness of containers for pickup.

2.10 Estimated Time of Arrival (ETA)

ETA is the anticipated time when a vehicle, ship, aircraft, or cargo is expected to arrive at a certain place. Knowledge of ETA is a crucial aspect of transportation as it promotes effective operation planning and prevents unnecessary panic about location of cargo.

2.10.1 ETA Relevance to Stakeholders in a Multimodal System

Stakeholders in a multimodal system have different uses for a vessel's ETA. From the moment one decides to transport containers from the origin until it gets delivered at the destination, stakeholders are interested in the arrival time of the vessel. The benefits and applications of this information as it concerns individual stakeholders are highlighted.

Carriers

Carriers are most concerned with meeting the deadlines set for transportation of cargo to the destination. In a bid to avoid paying penalties on late deliveries, shippers provide a feasible arrival time to shippers. The carrier considers economical speed of travel that will minimize fuel consumption. Hence, ETA information is valuable to carriers as it assists them to know whether they are on track with meeting their deadline with the shipper. This also helps carriers make decisions about speed during the trip. If they are behind schedule, they can increase the speed to meet the planned time; if they are ahead of schedule, they may choose to slow down in order to save fuel.

Container terminal

The competitiveness of a container terminal is a product of its efficiency in container handling and its preparedness for arriving vessels. Knowledge of a vessel's arrival time at the terminal helps ports to plan ahead; it assists in allocation of berthing, equipment and personnel. When accurate ETA data is available, terminal operations can be planned in such a way as to prevent congestion as plans can be made for subsequent arriving vessels to promote smooth operations at the terminal. Personnel shifts can be properly planned and in an event that extra hands will be needed at the terminal, the operations manager can plan for a shift change or double-shift as needed.

Importers

Importers are also the most concerned about their cargo's arrival. The vessel's arrival time determines the kind of commitment they have with their customers. When an ETA is accurate, it increases the customers' confidence in their reliability.

2.10.2 ETA Determination

Fast and accurate calculation of ETA is of great importance in several areas of the ocean shipping industry. Different techniques have been used to determine ETA. There is a general concept which is applicable to all modes of transportation – air, water or road. It revolves around calculating the distance between the cargo and its destination; afterwards, the ETA of that vessel can be estimated by dividing the distance by the sailing speed (Fagerholt, 2000).

ETA can also be determined by referring to historical data for that particular itinerary. When data of a trip from point A to point B are collected, one can predict the time required for the same type of trip involving the same points in subsequent trips. Most journeys traveled by vessels feature a series of stops. If one assumes a trip from point A to point C via point B, then the time spent on this journey could be split into time spent from point A to point B and from point B to point C. The time spent at point B should also be noted in this determination. Distances between these points can be based on historical data by following ship locations throughout the distance it covers and saving this data at a number of set points together with the time spent at each point. A proper estimate can be made in the future since data for the same trip is available. By repeating this process, one can build a database that will be referred to as "HistoricalLeg." As time passes, this database grows, allowing one to base estimates on more historical data, thereby providing more accurate estimations (Parolas, 2016).

Heywood et al. (2009) present the standard method for determining ETA and recommend an easier way by analyzing a particular route's historical shipment data and segmenting each trip into legs. Each historical leg thus represents one trip from the point at which location data were received to the next transfer point. Over time, the table grows until the commonly traveled path from a given point A to point B is littered with start points for historical leg rows. This method provides more accurate ETA than simply using initial and terminal points. To further increase the reliability of the calculation, legs with similar start- and endpoints (within a critical radius of the actual leg's start- and/or endpoints) are also used.

When requesting the ETA of in-transit cargo traveling currently at point P between points A and B, the general concept of this recommendation is that the algorithm will query the historical leg table for a similar point P with the same destination B. The system will then use the mean of the elapsed time for each row to calculate the estimated time from the current location to B. The general equation for determining ETA by this method is as follows:

$$ETA = t_{PB} + t_{BZ} + t_{transfer} \tag{1}$$

derived from

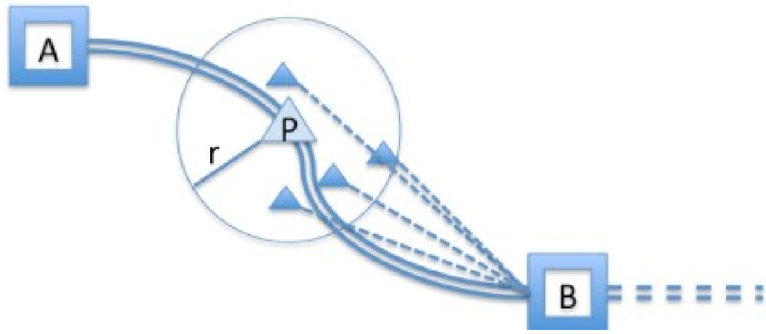
$$t_{BZ} = t_{BC} + t_{CD} + \ldots\ldots + t_{YZ} \tag{2}$$

$$t_{transfer} = t_B + t_C + \ldots\ldots + t_Z \tag{3}$$

$$t_{PB} = \frac{\sum_{i=1}^{n} t_i}{n} \tag{4}$$

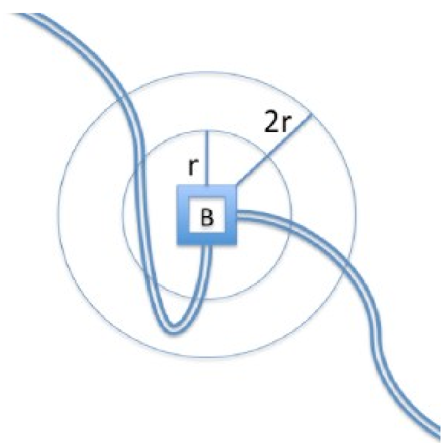
where
t_{bz} = time to move from b to z
$t_{transfer}$ = time spent at each port
t_{bp} = time required to reach first port, considering the historic legs

One limitation of this technique is that some segments of the path traveled will yield poor GPS data, due to sparse availability of data in that location. These legs with little location data will lead to errors in ETA calculations. The introduction of a sensitivity radius (see Fig. 2.7a) around the current location of cargo can be the solution. The sensitivity radius was decided using the formula below.



$r = 0.05 * AB$

2.7.a



2.7 b

Figure 2.7: Sensitivity Radius for Points (Locations of Cargo) and Transfer Points

Another limitation of this technique is the unpredictability of the time spent at transfer points. This was corrected by introducing an arrival circle of radius *r* and a departure circle of radius 2r, as shown in Figure 2.7b. The cargo is considered to have arrived at the transfer point when the inner sensitivity circle is breached, and the cargo is considered to have departed from the transfer point when it leaves the outer circle.

Veldhuis (2015) highlighted steps for developing an automated solution for ETA definition of long-distance shipping. In his work, he created an automated version of the existing process by analyzing the existing way of determining vessels' ETA which basically entailed the collection of shipping schedules and verification of departure times. Based on the departure time, ETA data is collected from websites (marintraffic.com, apmtrotterdam.nl) and updated in the system. Heywood et al. (2009) reinforced the importance of providing an accurate estimated time of departure (ETD) in order to achieve the best results in ETA determination.

> The automation process for data collection is known as "Web scraping," where computer software and programs are used to collect information from the Internet such as Cloudscrape, RapidMiner and WebHarvest. Veldhuis also achieved some predictive results by using historical data to model a ship's route and splitting the route into parts between various ports. ETA was determined by the time needed to cover the distance between ports and the time spent at these ports. By combining these two methods, more reliable results were achieved.

2.11 Automatic Identification System (AIS) Data as a Tool for ETA Determination

Technology has steadily found its way into operations of different industries and institutions, and multimodal transportation is no exception. The future of multimodal transportation systems lies in the application of new technologies. It is a known fact that technology (information and communication) is the nervous system of multimodal transportation. It comes with many benefits like providing real-time information (visibility and data exchange) about shipments, and gives transporting organizations flexibility when reacting to unforeseen changes (Harris et al., 2015). In marine transportation, AIS is the one of the most impressive technologies that has helped provide real-time information about vessels in transit. The most important part of this system is its collection, transmitting, interpretation and implementation of data. The availability of this data, if properly implemented, is expected to improve the safety and effectiveness of multimodal transportation.

AIS is an automated and standalone system that has the ability to exchange navigational information between vessels and shore stations equipped with compatible systems that can understand its messages. A basic benefit of AIS is that ship-to-ship and ship-to-shore communication enhances vessel traffic services, monitoring and safety. AIS serves as a broadcasting system onboard the ship. Operating like a radar transponder in the VHF maritime band, it uses VHF broadcast technology to send vessel movement data. The system is capable of handling over 4,500 reports per minute, which are updated as often as every two seconds. At the basic level, any AIS system requires two inputs and one output in order to function effectively. One of the inputs is the GPS feed, which is responsible for position identification, and the second is the VHS feed, which receives incoming AIS signals from other vessels. The output is also a VHF connection, which is necessary for transmitting the position and core information of the vessel. When satellites are used to detect AIS signatures, the term 'satellite-AIS' (S-AIS) is used.

The most important part of AIS is the transponder, which serves as the receiver and transmitter of feeds. There are three types of AIS transponder: (a) class A, (b) class B, and (c) "receive only." The question of which transponder to install is based on the type of vessel and type of information to be transmitted and received. Class A is the higher specification of transponders and is mandated for commercial vessels, while class B is the lower classification, and receiver-only transponders are for smaller, mostly leisure, vessels.

Class A

> Under international Safety of Life at Sea (SOLAS) regulations, class A transponders are mandated on all international ships with a gross tonnage of 300 tons or more, and on all passenger ships regardless of size. Class A units must have the ability to send the ship's

information to other ships and to shore. They must also be able to receive and process information from other sources, including other ships. These transponders have a horizontal range of up to 40 nm and transmit continuously at 12.5 watts. The transponder uses Self-Organized Time Division Multiple-Access (SOTDMA) technology so that each transmission is automatically adjusted to avoid interfering with others in range. In areas with high-density shipping, the system also shrinks the area of coverage when necessary to ensure that the system is not overloaded.

Class B

Class B transponders were developed to provide smaller vessels (usually recreational vessels and small fishing boats) with voluntary access to the AIS system benefits enjoyed by the larger vessels. These transponders' horizontal range is around 7 nm, and they transmit every 30 seconds at 2 watts. They use Carrier Sense Time Division Multiple Access (CSTDMA) technology, which checks for Class A transmissions before sending its own signal. Class B information is only broadcast when there is sufficient space on the AIS channel.

Receive-only

The third option for a small vessel is to just receive AIS transmissions from other vessels and display them. This was initially used by small vessels before Class B transponders entered the market and came to be favored over the receive-only transponder. Having a receive-only transponder means that you can see other vessels but they cannot see you.

Two channels – 87 B (161.975 MHz) and 88 B (162.025 MHz) – in the marine VHF allocation are reserved primarily for AIS transmission. Like the normal VHF, the range depends on antenna height although the AIS signal is more rugged, and hence has longer range. It can typically pick up transmission from a large ship up to 20 miles away. To accommodate many vessels transmitting on the limited channels, a Self- Organizing Time Division Multiple Access (SOTDMA) system is used. This works through a principle where a time period is divided into about 4,500 slots. When a Class A or B transponder switches on, the system looks for a vacant time slot and reserves it. Once this slot is filled, other sets in range will avoid it and select another vacant slot. Precise timing is needed to ensure that all vessels are synchronized, and this is derived from a GPS receiver that is present in both class A and B equipment. It is important to delve into the exact type of information that each class of AIS transponders provides. We should also keep in mind that only class A and B transponders transmit information, while class B transponders transmit only static information about the vessel. Class A transponders provide three types of information: static, dynamic, and voyage-related.

Static information

This information is entered into the AIS system upon installation. It only changes if there is a major change in the ship’s characteristics, such as name or ship type. The static information is verified periodically. Such information includes: 1) Maritime Mobile Service Identity (MMSI); 2) call sign and name of vessel; 3) IMO number; 4) length and beam; 5) type of ship; and 6) location of position-fixing antenna.

Dynamic information

Apart from navigational status information, dynamic information is automatically updated by the ship sensors connected to the AIS. They broadcast every few seconds. For proper and accurate operation of AIS, it is important to properly install and confirm operation of connected sensors. Dynamic information includes 1) ship's position; 2) position time stamp in Coordinated Universal Time (UTC); 3) course over ground (COG); 4) speed over ground (SOG); 5) heading; 6) navigational status; and 7) rate of turn (ROT).

Voyage-related information

This information is manually entered and updated based on trip conditions. Voyage-related information includes 1) ship's draught; 2) hazardous cargo (type); 3) destination and ETA; and 4) route plan.

It is important that the navigation status of vessels underway be updated throughout the course of a voyage as the system broadcasts every 2-10 seconds. When vessels are moored or at anchor, they broadcast every 3 minutes. Due to the frequency of broadcast, voyage-related information uses up a significant amount of bandwidth, which may affect the response time when first responders require such information. Sometimes collecting information at coast stations is impossible when the vessels are out of the coastal AIS range. It is now possible to receive AIS information through the use of satellite which has the capacity to receive AIS data from out-of-range vessels, transmit it to off-shore stations, and even make it available globally through the Internet. Figure 9 shows how AIS messages and other relevant messages (GPS and satellite data) are exchanged within the system. Several providers (companies) offer the information thereof, which can be accessed normally through subscription and received by fleet operators.

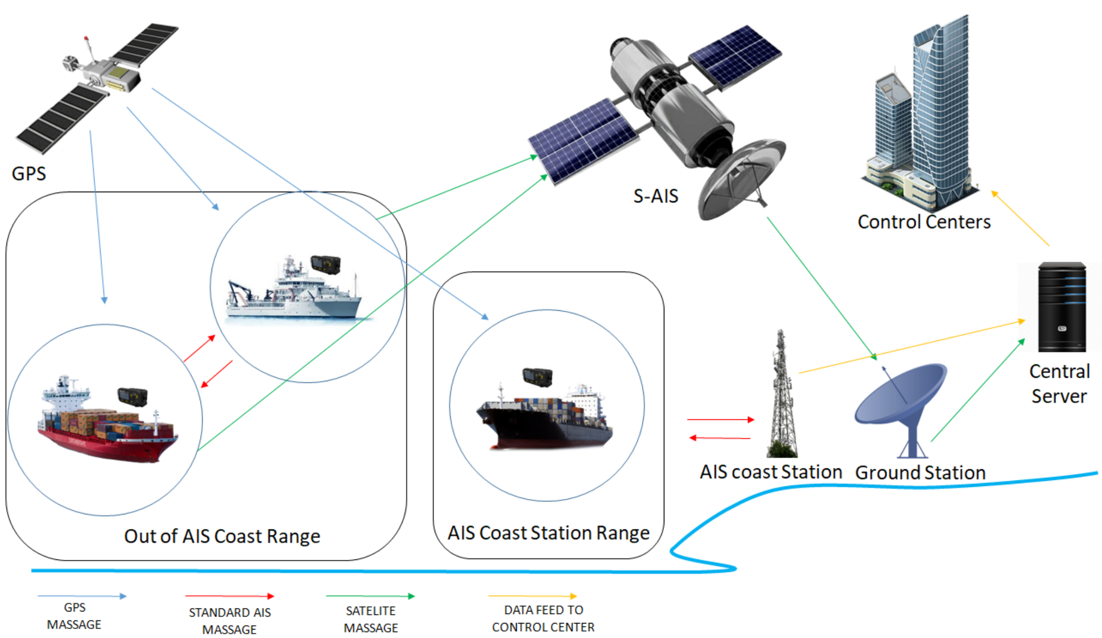


Figure 2.8: Data Transfer in AIS

There are a number of websites and Internet applications that permit stakeholders to view vessels' AIS data, such as PortVision (www.portvision.com) and Marine Traffic (www.marinetraffic.com). These work by taking data from numerous receiving points around the coast and aggregating it to create an overall picture. There are also some applications that will transmit AIS data from a tablet or smart phone, such as (http://www.marinetraffic.com/en/p/mais). Provided that Internet connectivity is available, the vessel's positions will start getting reported once the mobile AIS application is activated on the mobile device. It is important to note that they are not transmitting on VHF but sending data directly to the Internet over the phone data system. This means that the information will not show up on a normal VHF AIS receiver on nearby vessels but only on the website associated with that mobile application (29, 30).

2.11.1 AIS Data Description

Data collected from AIS each have their own relevance, especially for different applications. Relevance of some of the information collected from AIS data are highlighted below.

IMO number: The International Maritime Organization (IMO) number is a unique identification for vessels and registered owners/companies. This number was introduced to improve safety and security for vessels. It is linked to a vessel for its useful life, regardless of any change of name or ownership. This number is relevant for unique identification of a vessel, especially in tracking its time of arrival.

Call sign: A call sign is allocated to a vessel when first issued a ship radio license. It uniquely identifies vessels within the International Maritime Mobile Service. Call signs are used solely for search-and-rescue purposes. When there is a change in vessel ownership, call signs may be kept with the vessel or the new owner will be directed to obtain a new call sign.

Maritime Mobile Service Identity (MMSI): A MMSI is a unique nine-digit number associated with VHF installations that serves as a vessel's digital "call sign." They are sent over a radio frequency channel to identify stations. It can be used by telephone or telex subscribers connected to the general telecommunications network to call ships.

Length and beam: These are basic characteristics for describing a vessel's size. They determine how long and wide a vessel is and provide information on the vessel's capacity. This information is important when describing a vessel's maneuverability and its ability to sail and turn through ship channels.

Type of ship: This data provides information on the vessel's function. Vessel types include cargo, tanker, and passenger vessels. This information can be used to filter and identify vessels of interest.

Ship's position: The position of vessels, presented in latitude and longitude, shows a specific set of numbers that represents the vessel's specific location. Latitude and longitude are a common choice of coordinate system and are relevant in determining vessels' ETA because knowledge of the current location and destination has a great impact on such determination.

Position time stamp: Just as location is an important factor, so is the time at which the location is identified. This serves as a benchmark for determining the time of arrival.

Heading: This is the compass direction in which the vessel is pointed. This information is useful in determining time of arrival because it specifies the vessel's direction of travel and nearness to its destination.

Course over ground (COG): This is the actual direction of final destination between two points, with respect to the surface of the earth. Heading may differ from course, due to route taken or the effects of wind and current. The COG is relevant in determining time of arrival as it provides information about direction towards destination.

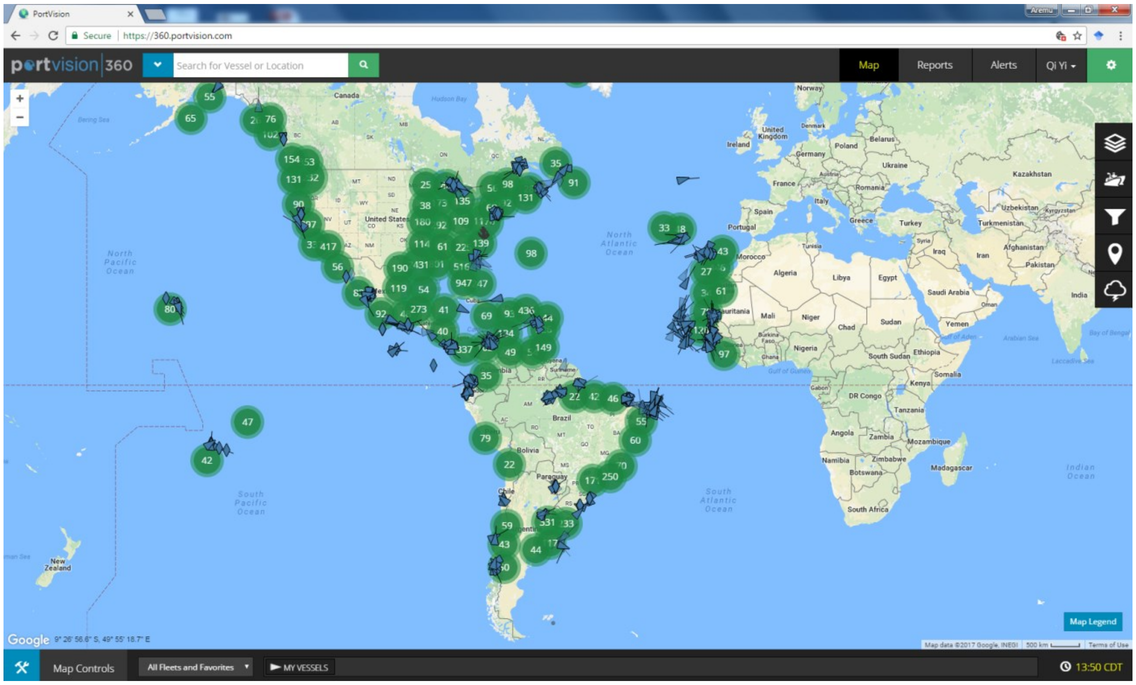
Speed over ground (SOG): This is the speed of the vessel relative to the earth's surface. It identifies how much distance is covered by the vessel in a given period. This factor is important for determining time of arrival as the greater the SOG, the more the distance covered at a given time and lesser time of arrival, provided other factors are unchanged

Ship's draught: This is the vertical distance between the waterline and the bottom of the hull. It determines the minimum depth of water a ship can safely navigate. It is also used to determine the weight of cargo carried by the vessel.

2.11.2 PortVision as a Source of AIS Data

One of the well-known providers of AIS data is PortVision by Oceaneering International, Inc. PortVision is a Web-based service that provides real-time and historical transit data of maritime vessel operations in ports, inland waterways and oceans. PortVision also supports location reporting for vessels at sea through satellite-based tracking. First deployed in early 2007, PortVision has facilitated a compelling increase in efficiency, cost savings, and safety and security of waterways. PortVision has also provided visibility and transparency of all vessel activities to all stakeholders; a noticeable change in waterway culture.

With PortVision, users can leverage AIS transmissions to support their businesses and experience more efficient business practices. PortVision records these transmissions at certain intervals then uses the data to locate vessels and determine vessel movements and tracks. It is now possible to deploy an information system that provides real-time vessel locations and recorded vessel movements for all commercial ship traffic along the waterway. PortVision currently provides service for major seaports in the US and North America and over 60 international ports, including Africa, Asia, Europe and South America. Figure 10 shows a typical display of the PortVision home page with numbers representing the counts of AIS-supported vessels in each location.



Adopted from www.portvision.com

Figure 2.9: PortVision Display Page

PortVision offers a range of maritime services that includes access to relevant information on vessel movements, terminal/port arrival and departure, and creating and monitoring key points of interests on the portal. Ability to set up notifications of arrival and departure of vessels in zones of interest through e-mails and mobile text messages is a good functionality of the system. Another interesting feature is the provision of animated playback and historical reporting, which allows users to analyze past events and generate documentation suitable for demurrage analysis, vendor and partner compliance, negotiation and litigation, etc. It provides a platform where agents, ship owners, and terminal personnel can collaborate to more effectively schedule dock resources. The platform can also be used for document exchange between agents and terminal personnel. Through this, dock scheduling information is made visible to other stakeholders to drive efficiency, while strictly maintaining confidentiality within the system. Most importantly, all information can be accessed from anywhere through a standard Web browser.

2.11.3 ETA Forecasting Tools

One of the crucial factors to consider when deciding which mode of transportation to employ is the reliability and acceptable time of arrival that the system can provide. With this in mind, it becomes necessary to be well informed about the ETA of any means of transportation to be considered. As the word “estimated” implies, it is a rough calculation of the value, number, quantity, or extent of something. The ETA of any mode of transportation does not provide an exact time of arrival; there is room for error. The main purpose of performing research in this area is to limit the degree of error and uncertainty in the estimation.

Many research efforts have been made to determine and optimize the ETA of different modes of transportation. The different methods developed, regardless of the mode of transportation considered, can be applied to other modes of transportation because all modes share a common ground in terms of the distance traveled and the speed of travel. These efforts have resulted in the application of three different models that can be applied when determining ETA: (a) models based on historical data, (b) multi-linear regression models, and (c) machine learning. The first type infers the current and future travel times of a means of transportation based on historical travel times for the same itinerary. In general, this model is reliable only when the traffic pattern in the area of interest is relatively stable. One major limitation of historical data models is that they require an extensive data set, which may not be available in practice, especially when the traffic pattern varies significantly over time. The second type is a mathematical model that predicts expected travel times between stops and then the ETA at individual stops. This type of model is usually established by regressing travel times against a set of independent variables. This approach has limitations in that it is only reliable when a regression equation for the operation can be established, which may not be possible for application environments where many of the system variables are not correlated. The third approach is the application of a machine learning technique to predict ETA. This technique is capable of capturing the complex nonlinear relationships that are typically seen in transportation application environments.

2.12 Machine Learning

Machine learning generally falls into three categories: (1) supervised learning, (2) unsupervised learning, and (3) reinforcement learning (Chao, 2011).

2.12.1 Unsupervised Learning

Unsupervised learning is the most common learning process in the brain, which makes it very important. According to Dayan (2008), this process studies how systems can learn to represent particular input patterns in a way that reflects the statistical structure of the overall collection of input patterns. In unsupervised learning, target outputs or environmental evaluations are not associated with each input. The system uses prior biases to determine what aspects of the structure of the input should be captured in the output.

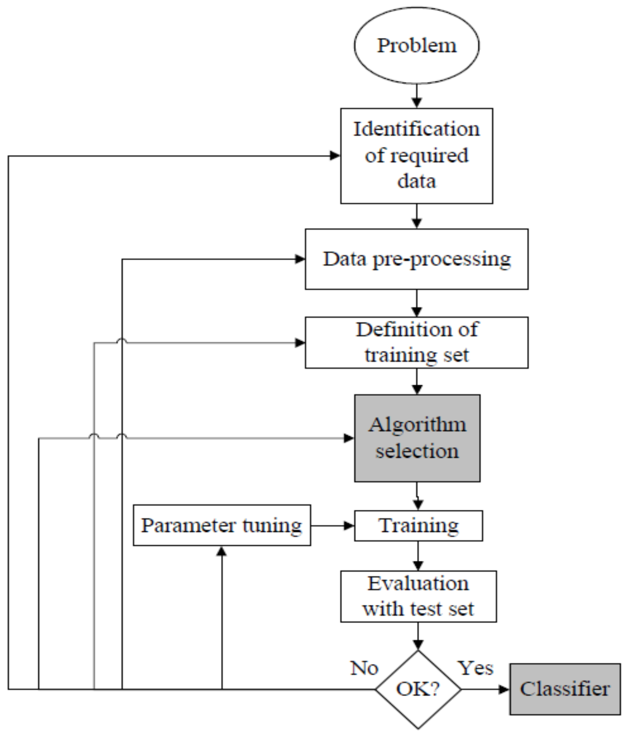
2.12.2 Reinforcement Learning

According to Sutton and Barto (2017), reinforced learning is a kind of learning that discovers which action will yield the greatest reward. It is basically characterized by trial-and-error searching. The computer is simply given a goal to achieve, and – through the trial and error of interacting with its environment – learns how to achieve that goal. Harmon and Harmon (2001) explain reinforcement learning as an approach to machine intelligence that combines two disciplines to solve problems that neither discipline can address on their own. It is an appealing approach for researchers due to its generalizability.

2.12.3 Supervised Learning

Supervised learning is the most important methodology in machine learning and is central to the processing of non-linear data**.** According to Cunningham et al. (2011), supervised learning generally entails mapping a set of input variables X and an output variable Y and applying this mapping to predict the outputs for future data.

Of these categories of machine learning, supervised learning has been used by researchers for predicting ETA. It has leveraged the availability of historical trip data where there is information of input variables and the actual time of arrival. When applied to real-world problems, it follows the steps described in Figure 2.9.



Adopted from Kotsiantis (2007)

Figure 2.10: Process of Supervised Machine Learning

As explained above, supervised learning is the most important method of machine learning. Delving into a few of the algorithms that operate in this way will provide a better understanding of the method. Kotsiantis (2007) stated that amongst supervised learning algorithms, the multilayered perceptron also known as Artificial Neural Network (ANN) and Support Vector Machines (SVMs) tend to perform much better when dealing with multi-dimensions and continuous features. Some of the characteristics shared by ANN and SVMs that make them unique are as follows:

- Large sample size is required to optimize their prediction accuracy;
- They perform better when multicollinearity is present and a nonlinear relationship exists between input and output;
- They require more training time as they learn more slowly than other supervised learning algorithms;
- Memory space for execution is usually smaller than the training space;

- They have more parameters than other techniques; and
- They have poor interpretability, which makes their principle of operation hard to understand.

Neural Network

Khajanchi (2003) defined a Neural Network (NN) as an information-processing technique developed from the concept of biological nervous systems. Unlike traditional statistical methods, a neural network has the ability to model non-linear problems and perform predictive analysis where relationships are not constant. Neural networks can identify complex trends that are difficult or impossible for humans or other computer techniques to detect. Neural networks derive their strength from their ability to recognize the relationship between input and output data. Some special functionalities of neural networks are adaptive learning, self-organization, and real-time operation fault tolerance via redundant information coding. One of the most important strengths of neural network models is their ability to learn from series of iteration input data and the resulting outputs. Understanding their architecture sheds more light on the concept. Their three-layered architecture (Fig. 2.10) consists of an input layer, a hidden layer and an output layer.

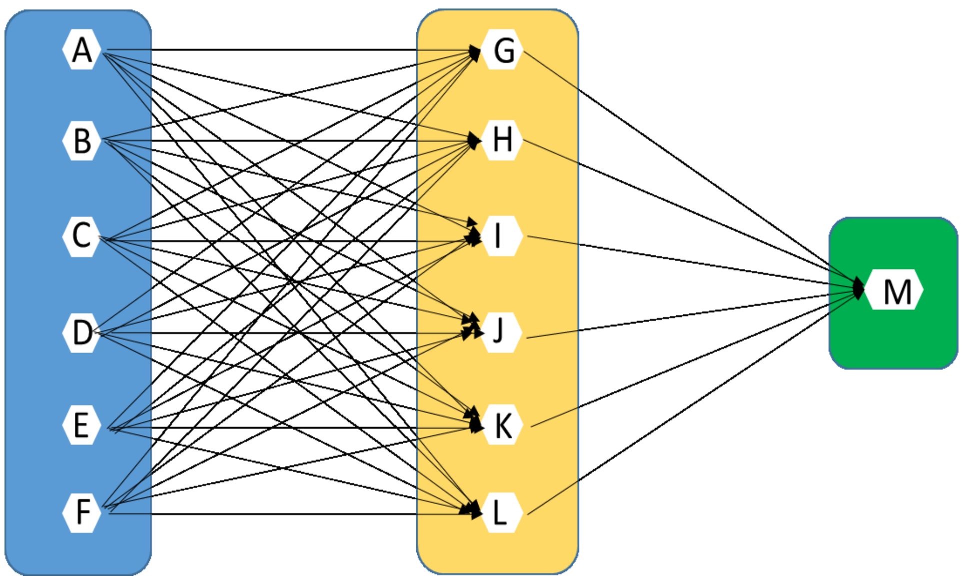


Figure 2.11: Architecture of a Three-layered Neural Network

Variables A-F are input neurons representing input variables. Variables G-L are neurons in the hidden layers which capture relationships between the input and output layers. Variable M is the output neuron representing the model output.

Stergiou and Siganos (1996) categorized the learning process and response of the network into associative mapping and regularity detection. In associative mapping, when there is any addition or distortion to the existing input, the network learns to produce a new pattern in response. In regularity detection, the network response is based on particular properties of the input pattern. A neural network can be a fixed or an adaptive network. When the learning method used for an adaptive network is supervised, the output unit is told by an external teacher how to respond to the input signal. This minimizes any error between the desired output and the computed value.

ANN models require extensive training to reduce error in the results (Sun et al. 2007). It was also discovered that the neural network method performed better than other machine learning techniques. USDOT.BOS (2015) showed that neural networks performed better than linear regression models for predicting purposes. Calculating the time of arrival of different modes of transportation is a non-linear process, and major irregularities are present. Therefore, the neural network's flexibility, non-linearity and arbitrary function make it preferred over linear regression models. Jeong (2004) also developed a historically based model, regression models, and artificial neural network (ANN) model to predict bus arrival time of Automatic Vehicle Location (AVL) systems. It was found that the ANN models outperformed both historical data-based models and multi-linear regression models. It was hypothesized that the ANN achieves an advantage over other models through its ability to identify the complex non-linear relationships between travel time and the independent variables. Carbonneau et al. (2008) applied machine learning techniques to supply chain demand forecasting. He compared machine learning techniques (neural network, recurrent neural network and support vector machine) and traditional methods (naïve forecasting, trend, moving average, multiple linear regression, and time series). The analysis showed that machine learning techniques, especially the neural network, performed better than other techniques.

2.13 Variable Importance

Variable importance represents the statistical significance of each variable present in the data with respect to its effect on the model generated; it quantifies which input variables are more influential than others. Variable importance is the predictor ranking of each variable based on its contribution to the model. Identifying variable importance helps data analysts to eliminate any variables that are contributing little or nothing to the model but increase prediction processing time. Commonly used methodologies for quantifying variable contributions in ANNs include the connection weight approach (Olden's algorithm), Garson's algorithm, partial derivatives, input perturbation, sensitivity analysis, forward stepwise addition, backward stepwise elimination, and improvised stepwise selection. The most popular methods for constructing variable importance are Garson's algorithm (Garson 1991) and Olden's algorithm.

2.13.1 Garson's Algorithm

Garson's algorithm is used to determine variable importance by calculating the weighted connections between nodes of interest. Garson's approach partitions hidden-output connection weights into components associated with each input neuron using absolute values of connection weights.

Figure 2.12 shows a typical neural network with one input layer, one hidden layer, and an output layer. $W_{i,j}$ indicates the connection weight of the input-hidden layer. $W_{o,j}$ also represents the connection weight between the hidden layer and the output layer.

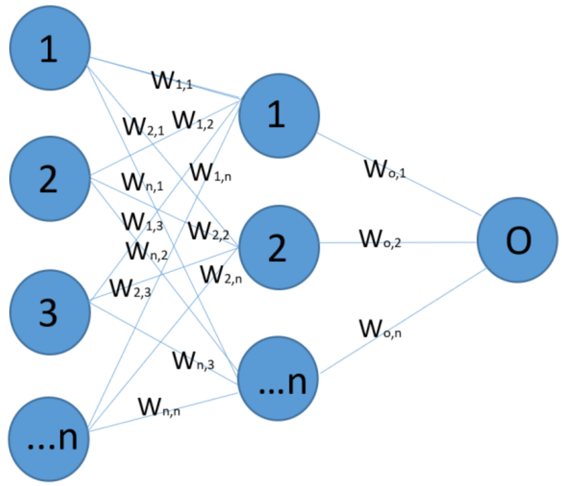


Figure 2.12: Typical Neural Network with Input-hidden and Hidden-output Weights

Garson's algorithm is represented by the steps in Tables 2.6-2.9. These steps show the approach employed in quantifying the importance of variables involved in developing the neural network model in Figure 2.12.

Table 2.6: Model's Input-hidden and Hidden-output Weights

	Hidden 1	Hidden 2	…Hidden n
Input 1	$W_{1,1}$	$W_{2,1}$	$W_{n,1}$
Input 2	$W_{1,2}$	$W_{2,2}$	$W_{n,2}$
Input 3	$W_{1,3}$	$W_{2,3}$	$W_{n,3}$
…Input n	$W_{1,n}$	$W_{2,n}$	$W_{n,n}$
Output	$W_{o,1}$	$W_{o,2}$	$W_{o,n}$

Contribution of each input neuron to output ($C_{i,j}$) = $W_{i,j}$ * $W_{o,i}$

Table 2.7: Contribution of Each Input Neuron to Output Neuron

	Hidden 1	Hidden 2	…Hidden n
Input 1	$C_{1,1}$	$C_{2,1}$	$C_{n,1}$
Input 2	$C_{1,2}$	$C_{2,2}$	$C_{n,2}$
Input 3	$C_{1,3}$	$C_{2,3}$	$C_{n,3}$
…Input n	$C_{1,n}$	$C_{2,n}$	$C_{n,n}$

Relative contribution of each input neuron to output ($R_{i,j}$) = $C_{i,j} / \sum C_{i,j}$

Table 2.8: Relative Contribution of Each Input Neuron to Output Neuron

	Hidden 1	Hidden 2	…Hidden n	Sum
Input 1	$R_{1,1}$	$R_{2,1}$	$R_{n,1}$	S_1
Input 2	$R_{1,2}$	$R_{2,2}$	$R_{n,2}$	S_2
Input 3	$R_{1,3}$	$R_{2,3}$	$R_{n,3}$	S_3
…Input n	$R_{1,n}$	$R_{2,n}$	$R_{n,n}$	S_n

Relative importance $(R_j) = (S_j/\sum S_j)*100$

Table 2.9: Garson's Relative Importance of Input Variables to Model

	Importance
Input 1	R_1
Input 2	R_2
Input 3	R_3
…Input n	R_n

2.13.2 Olden's Algorithm

Olden's algorithm calculates the product of the raw input-hidden and hidden-output connection weights between each input neuron and output neuron and sums the products across all hidden neurons. Similar to Garson's algorithm, Tables 2.10-2.13 highlight the steps followed using Olden's algorithm to quantify the variable importance of the neural network model in Figure 2.12.

Table 2.10: Model's Input-hidden and Hidden-output Weights

	Hidden 1	Hidden 2	…Hidden n
Input 1	$W_{1,1}$	$W_{2,1}$	$W_{n,1}$
Input 2	$W_{1,2}$	$W_{2,2}$	$W_{n,2}$
Input 3	$W_{1,3}$	$W_{2,3}$	$W_{n,3}$
…Input n	$W_{1,n}$	$W_{2,n}$	$W_{n,n}$
Output	$W_{o,1}$	$W_{o,2}$	$W_{o,n}$

Contribution of each input neuron to output $(C_{i,j}) = W_{i,j} * W_{o,i}$

Table 2.11: Contribution of Each Input Neuron to Output

	Hidden 1	Hidden 2	…Hidden n
Input 1	$C_{1,1}$	$C_{2,1}$	$C_{n,1}$
Input 2	$C_{1,2}$	$C_{2,2}$	$C_{n,2}$
Input 3	$C_{1,3}$	$C_{2,3}$	$C_{n,3}$
…Input n	$C_{1,n}$	$C_{2,n}$	$C_{n,n}$

Connecting weight importance of each input variable $(CW_j) = \sum C_{,j}$

Table 2.12: Connecting Weight Importance of Each Input Variable

	Importance
Input 1	CW_1
Input 2	CW_2
Input 3	CW_3
…Input n	CW_n

Relative importance (R) = $(CW_i/\sum CW_i)*100$

Table 2.13: Olden's Relative Importance of Input Variables to Model

	Importance
Input 1	R_1
Input 2	R_2
Input 3	R_3
…Input n	R_n

According to Greenwell et al. (2018), Olden's algorithm has outperformed Garson's method in various simulations. Olden et al. (2004) also found that Olden's algorithm outperformed all other approaches and provided the best result by accurately quantifying variable importance.

2.14 Optimization of ETA from AIS Data

This section reviews different applications of machine learning to determine either ETA or delays of vessels to ports and terminals. Pani et al. (2015) employed a regression approach in machine learning by using logistic regression, classification tree and random forest to predict the delay or early arrival of vessels. AIS and weather data were used as inputs. This method was used as it can be explained and interpreted more intuitively. The authors employed algorithms provided a qualitative estimate of the delay/advance by knowing whether or not an incoming vessel was likely to arrive before or after the scheduled ETA. Random forest outperformed the other algorithm.

Fancello et al. (2010) predicted the ETA of vessels to optimize container handling at Cagliari's container terminal by examining the calibration of a neural-network-based simulation model. Neural network was developed with different numbers of variables selected based on previous knowledge from previous works. In order to identify the best fit for the system, numerous network-varying characteristics were tested for, which included trying out different learning algorithms, learning parameters (e.g., learning cycles, learning rate), and numbers of hidden nodes. The analysis found that three variables with two hidden layers having 4 nodes and 1 output layer gave the best results with the least error in the predicted time of arrival.

Pani et al. (2014) used a data-mining approach to predict the level of daily alarm related to late arrivals. They categorized the delay level into clusters and, using the Ward's method, identified the best cluster to use to analyze the delay rankings as variables. Three different machine learning models (naïve Bayes, decision trees, and random forests) were used to predict the delay alarm level for each day and tested. Predictive power of the algorithms was determined by comparing the predicted and observed levels of delay. The best results were obtained with the random forest algorithm, which yielded a relatively low absolute error.

Parolas (2016) predicted vessels' ETA at the port of Rotterdam using a neural network and support vector machine. The effect of weather conditions was also analyzed. The following variables were used as inputs with 10 hidden layers and one output layer as ETA.

> **AIS Data:** latitude (degrees), longitude (degrees), distance to be covered (km/h), current speed of vessel (km/h), change in speed over last 3 hours (km/h), average speed over last 12 hours (km/h), time used for calculating average speed (hours), length of ship (meters), **b**readth of ship (meters), and ETA of ship's agent (number of days).
>
> **Weather Data:** current U-component (m/s), current V-component (m/s), wind U-component (m/s), wind V-component (m/s), peak wave period (s), peak wave direction (degrees), and significant wave height (m). Results obtained from the prediction models were compared to the vessel's actual time of arrival. Mean absolute error (MAE) and root mean squared error (RMSE) methods were used to evaluate the model performance of the two machine learning methods. Results showed that both the SVMs and NN gave more accurate predictions than the current situation based on the ETA provided by the shipping agent. Furthermore, SVM outperformed the NN for every point in the examined time-horizon. In regard to the influence of weather on the ETA, it was found that it does not play a crucial role in estimating ETA for the examined route.

2.15 Port of Houston

The Port of Houston is a major port with over 150 public and private facilities. The complex is about 25 miles long and very important due to its large tonnage-handling capacity and economic impact. In international waterborne tonnage handled, the Port of Houston is ranked first in the United States. Also, it is ranked second and fifth in terms of total cargo tonnage handled and busiest port in the world, respectively (Qu, 2012). Efficiency of any port depends on the availability of adequate dockside infrastructures such as berth space, cranes, wharfs, and channel depth. The quality and quantity of such infrastructure goes a long way in determining the reliability of the service provided or to be expected by the terminals. The Maritime Administration (MARAD) highlighted some factors that influence the quality of service provided by ports when making a choice of port of call.

Navigability: For a port to attract the largest (Neopanamax) vessels, it needs a channel deep and wide enough for effective navigation. MARAD recommends channels 47.6 to 50 feet deep.

Air draft restrictions: Some container vessels carry large stacks of containers well above water. Bridges over the channels must be high enough to accommodate such vessels.

Terminal capacity: For timely handling of large container vessels, it is necessary for the port to have adequate yard size, labor, cranes and other terminal equipment.

Landside connectivity: Ports and container terminals are associated with huge amounts of truck traffic. Therefore, transfer facilities and entrance/exit routes must be properly designed. The reliability of port and off-port facilities greatly depends on their ability to move shipments in and out of the port and through metropolitan areas, which greatly impacts the port's attractiveness.

These characteristics of the Port of Houston have factored largely in its competitiveness. This study's profile of the Port of Houston focuses on the container operations at its Barbours Cut and Bayport container terminals, which together handled more than 2 million TEUs in recent years. These container terminals have handled approximately 67% of container traffic for the Gulf Coast and 95% of all container traffic for ports in Texas (Payson et al. 2017).

2.15.1 Bayport Container Terminal

Bayport Container Terminal is recognized as the most modern and environmentally sensitive container terminal in the US with a capacity to handle 2.3 million TEUs annually. Its electronic data exchange capability and computerized inventory makes it efficient in tracking the status and location of individual containers (Bayport Container Terminal). Around 2,500 transactions are conducted daily. This includes receiving of import containerized shipments, delivery of export containerized shipments, receiving and delivery of empty containers, and receiving and delivery of chassis. About 65% of trucks visiting Bayport perform dual transactions in each visit (Bierling et al., 2015).

2.15.2 Barbours Cut Container Terminal

Completed in 1977, Barbours Cut Container Terminal has grown to become a leading container-handling facility in the US Gulf of Mexico. Located at the mouth of Galveston Bay, it is comprised of six berths, roll-on/roll-off platforms, a lash dock and 230 acres of paved marshaling area. With a current capacity for 1.2 million TEUs, the terminal is expected to increase its capacity to 2 million TEUs by the end of the ongoing modernization program (Barbours Cut Container Terminal). A summary of the characteristics of both container terminals is presented in Table 2.14.

Table 2.14: Characteristics of Bayport and Barbours Cut Container Terminals

	Barbours Cut Terminal	**Bayport Terminal**
Berthing Docks	6 docks	3 docks
	6,000-ft berths	3,300-ft berths
Equipment	13 wharf cranes	9 wharf cranes
	42 RTG yard cranes	39 RTG yard cranes
Capacity	190 acres of loaded container storage	165 acres of loaded container storage
	390 acres of total terminal acreage	230 acres of total terminal acreage
	1.4 million TEUs annual throughput	1.2 million TEUs annual throughput
	36,00 TUEs static capacity	32,000 TUEs static capacity
Accessibility	Access to all major highways	Access to all major highways
	Access to two major rail lines	
	Terminal gates operating from 7 am to 7 pm weekdays	Terminal gates operating from 7 am to 11 pm weekdays
	Automated gate system with 14 inbound and 12 outbound lanes	Automated gate system with 28 inbound and 12 outbound lanes
Ship arrival rate (47)	2.10 per day	1.53 per day
Ship stay duration (47)	Triangular distribution with minimum of 3.71 hours	Triangular distribution with minimum of 8.92 hours
	Average of 28.2 hours	Average of 22.54 hours
	Maximum of 222.42 hours	Maximum of 105.33 hours

The first arrival station for container vessels designated for either the Bayport or Barbours Cut container terminals is the Galveston sea buoy. At this stage, the vessels are handed over to operators of Port of Houston who navigate the vessel to its designated container terminal. Such vessels experience wait time at the buoy before and after arrival of the allotted operators.

2.16 Port of Houston Operations

The Port of Houston operations for container vessels are categorized into activities at the sea buoy and activities at the container terminal. These activities are managed by the joint operations between Houston pilots, the US Coast Guard, US Customs and Border Protection, and the Port Authority.

2.16.1 Buoy Operations

As stated previously, every vessel entering the Houston Ship Channel is required to stop at the Galveston buoy, where a pilot takes over the sailing to the terminals. Pilots are available 24/7 and the buoy services up to 60 vessels daily. Although congestion at the buoy is unlikely, unforeseen circumstances like oil spillage and fog may cause congestion and backlog of vessels to be serviced. Information on vessels' arrival time to the buoy is provided to pilots by the shipping agent prior to the vessel's arrival. With this information, a pilot is made available to sail the vessel into the port. Operations at the buoy are sometimes influenced by other activities, like dredging. Beam restriction of the ship channel size makes it impossible for inbound and outbound vessels to meet in the ship channel when one of the vessels has a length greater than 1,000 ft or width greater than 138 ft. Such vessels are also restricted to daylight sailing even if they arrive at the buoy at night. Apart from delays experienced due to the presence of a big vessel, weather conditions and daytime restriction, there is a permissible delay of not more than an hour in situations where efforts are made to optimize manpower. When a vessel arrives and the available pilot is on a return trip that will take more than an hour to get back to the buoy, a new pilot is allocated to the arriving vessel. Low sailing speed is recommended through the ship channel to reduce the wake generated by the vessel and prevent damage to tugboats. Dredging is also a factor that can cause channel closure if piping crosses the sailing path.

2.16.2 Terminal Operations

Discussion of the Port of Houston's operations would be incomplete without exploring what goes on at the terminal and its gate. This is important in order to identify the Port of Houston's potential for multimodal systems. Terminal activity is divided into berth operations and gate operations.

Berth operations

Berth operations at the Port of Houston include basic activities performed by berth operators. Berth operators allot a berth area to arriving vessels, load and unload containers, and store containers. Operations here include basic transfer of containers from ship to shore, which requires quay cranes and RTG yard cranes for transporting the containers to the stacking area or transporting vehicles as needed.

Gate operations

Bayport and Barbours Cut Container Terminal have similar operational stages:
Stage 1 – Inbound Optical Character Recognition (OCR) and ticket generation
Stage 2 – Scaling and activity
Stage 3 – Outbound OCR and Customs and Border Protection (CBP) inspection
Stage 4 – Outbound

Optical Character Recognition (OCR): At this stage, images of the truck and container are captured from different angles as well as the license plate, chassis number and container number. These images can be accessed from the Port of Houston website by simply inputting the container number. Drivers proceed through one of the 14 gates for Barbour's Cut Terminal (BCT) or 28 gates for Bayport Terminal, scan their tickets and are processed based on their mission (drop-off, pickup, or both). After this pre-check, which takes about 10-15 seconds, drivers proceed for Transportation Worker Identification Credential (TWIC) verification then enter the terminal.

Scaling and activity: This stage is an automated process where the truck's weight is calculated. A pick-up or drop-off ticket is then generated. Inside the terminal, trucks drop off and pick up containers, or both, depending on their mission. The trucks then approach the outbound gates, passing through 4 CBP rpm (x-rays) for inspection.

Outbound OCR and Customs and Border Protection (CBP) inspection: This stage is composed of an outbound OCR system with 8 lanes at the BCT and 6 lanes at the Bayport terminal. Cameras take images of exiting trucks. The OCR used here is important for damage inspection and allows for automated transaction completion.

Outbound: At this stage, drivers scan their ticket to exit the terminal.

2.17 Summary

Preceding sections comprehensively review a multimodal transportation system, ways of determining ETA, and the role played by accurate ETA in the system's efficiency.

Chapter 3. Solution Methodology

3.1 Introduction

To predict vessels' time of arrival to port terminals, this study examined this mode of transportation's characteristics and facilities. Based on availability of data, existing techniques and a literature review, the most efficient method was selected.

Machine learning is an algorithm that can learn from data without relying on rules-based programming. Due to the complexity and irregularity observed in the collected data and the fact that traditional statistical forecasting models have limitations in estimating the complexity of a real system (Zhang et al., 1998; Kotsiantis 2007), it became necessary to opt for a neural network. This selection was also reinforced by works in the literature like Fancello (2011) and Parolas (2007) as well as these works' results.

3.2 Steps for Method Execution

The structuring and execution of this approach involved the following steps: a) choice of predictive approach; b) choice of paradigm; c) choice of input variables; d) variable normalization; e) choice of network architecture; f) choice of number of hidden layers and nodes; g) training, validation and testing of the network; h) second leg analysis; and i) interpretation of results.

3.2.1 Choice of Predictive Approach

Neural network was used as a predictive approach that could be trained to recognize patterns and relationships between independent input variables and the output (time of arrival).

3.2.2 Choice of Paradigm

The back-propagation algorithm in ANN was employed. It consists of multiple neuron layers, each of which is fully connected to the next. Neurons in the input layer represent the input data with all other neurons mapping the inputs to the output by a linear combination of weight and bias. Its steps consist of 1) feeding forward the values, 2) calculating the error, and 3) propagating it back to earlier layers.

3.2.3 Choice of Input Variables

Selection of input variables was achieved using prior knowledge from previous related works, and the uniqueness of the port of study. Twelve input variables were chosen: voyage ID, vessel's International Maritime Organization (IMO) number, length of vessel, beam of vessel, speed, average speed, heading, course, latitude, longitude, distance to buoy, and distance to destination (terminal). The variables were selected due to their relevance to determining ETA. All identified variables have different effects on ETA and possess their own level of importance with regards to their impact on the predictive power of the network. These variables were selected due to their relevance in determining time of arrival, as will be explained in Section 3.3

3.2.4 Variable Normalization

This step was accomplished after removing outliers from the records. To increase consistency, and for easier data mapping, it was necessary to normalize the variables, appropriately scaling them to the transfer function used. In this case, min-max normalization was used to present the data in a [0,1] range.

3.2.5 Choice of Network Architecture

Database records were divided into training and testing sets at 80% and 20%, respectively. Once the network was trained, testing for prediction accuracy was evaluated on the test set.

3.2.6 Choice of Number of Hidden Layers and Nodes

To prevent over- and under-fitting of the network, it was important to use the best number of hidden layers and neurons.

3.2.7 Training, Validating and Testing of the Network

Training, validating, and testing of the network was conducted with the aid of R software and the Neuralnet, a library that trained the neural networks using back-propagation. The designed network was trained with the training dataset and then tested to confirm the network's ability to predict the output based on inputs from the testing dataset.

3.2.8 Second Leg Analysis

The study area has its uniqueness and thus requires a special approach for trips from the pilot point (Sea buoy) to the respective terminals. With this in mind, steps a-g were repeated for data relevant to this portion of the trip, and collected results were summarized. This section aims to reduce any error that may have resulted from the analysis. This portion of the vessel's trip possesses different attributes because the trip is performed in a controlled environment (ship channel) with many restrictions.

3.3 Data

As mentioned in Section 2, the variables in this analysis are available in AIS databases. They include voyage ID, IMO, length of vessel, beam of vessel, speed, cumulative average speed, heading, course, latitude, longitude, distance to buoy, distance to terminal, arrival time at buoy, and berthing time. Each identified variable has its own effect on ETA and possesses its own level of importance in regard to its impact on the network's predictive power. All variables are processed to useable datasets to allow the network map relationship between them and to improve accuracy of results obtained.

Data used for this analysis were consolidated from the AIS and USCG databases. AIS historical records for container vessels called to the Bayport and Barbours Cut Container Terminals in the Port of Houston were collected. Information retrieved from the AIS records include the static information of the vessel and data point information, such as length of vessel, beam of vessel, current speed, heading, course, and location showing the latitude and longitude. These records were tied to information collected from the USCG database, which provided the information on the actual time of arrival of vessels to the pilot point at the Galveston sea buoy as well as the actual berthing time of the vessel at the container terminal. Records from USCG were preferred for the actual time of arrival and berth time because they portray real-life arrival times at the buoy and terminal. Times recorded in the USCG database were also compared with times received from AIS data for the same vessel. Cumulative average speed, distance to buoy and distance to terminal were used as variables in this analysis. They were calculated from the collected AIS and USGC datasets.

In total, 237 trip records were randomly selected from 2016 to 2018. Collected data were based on the difference between record stamp time and arrival time. Records were categorized into timeframes that consider both medium and short time horizons. The medium time horizon was identified as 5 days prior to the vessel's arrival, and aims to assist planning activities for port operators and other stakeholders. The short time horizon considers 24 hours prior to the vessel's arrival. The categories of collected data are summarized in Table 3.1.

Table 3.1: Summary of Data Used for Analysis

Category	Duration of data	Frequency	Data volume
1	5 days to arrival	Hourly	14,473
2	1 day to arrival	5 minutes	17,855
3	Buoy to terminal	5 minutes	988

The different parameters are described below and categorized into input and output variables.

3.3.1 Input Variables

The input variables for this analysis are highlighted below.

Voyage ID

The voyage ID provides the uniqueness for each identified trip. This parameter is useful to the network as it ties together all data specific to a vessel and throughout a given trip. It is the basis for identifying patterns created in the trip as it relates different data points for a specific trip together. In the dataset, voyage IDs are unique numbers that are only repeated when the same trip is undertaken again.

Length of vessel

Vessel length is one of the static data collected for each data point. This measure is a great variable for determining the vessel's size. It is relevant to the network for ETA prediction since a vessel's size determines its speed and maneuverability. Length of a vessel in this dataset is expressed in meters.

Beam of vessel

Similar to the length of the vessel, the beam is also a measure of a vessel's size. Coupled with vessel length, beam describes the vessel's platform area, expressed in meters.

IMO number

The International Maritime Organization (IMO) number is a unique ID for a particular vessel; however, in the data points collected, there was repetition of an IMO for every trip by the same vessel with a different voyage ID.

Speed

A vessel's speed is a major factor that determines its time of arrival to port. Generally, the greater the speed, the lower the time required to complete a trip, other variables kept constant. This makes the speed of travel an indispensable parameter in this analysis.

Average speed

It was necessary to calculate a vessel's average speed from a time previous data point. This variable gives the network a better view of previous speeds by the vessel. This compensates for any sudden drop or increase in the vessel's speed throughout the trip.

Heading

The heading provides the network with the vessel's direction at that point. The network tries to understand and be trained on the direction taken by the vessel at such location.

Course

The course of the vessel provides the network with the direction of its final destination from its current location. Expressed in degrees, it improves the network's directional sense.

Latitude and longitude

Latitude and longitude are the basis for determining a vessel's location at different data points recorded. It is relevant to the network as it is a known fact that the closer you are to destination, the faster you can reach it. Hence, the longitude and latitude expressed in decimal format were included in this analysis.

Distance to buoy

The linear distance between a vessel's current location and the sea buoy was an important variable in this analysis. We used the Haversine distance, which is based on a spherical model of the earth, to calculate this distance. Haversine distance is defined as follows:

$$d((x_1, y_1), (x_2, y_2) = \frac{2r_0\sqrt{sin^2\left(\frac{y_2-y_1}{2}\right)\cos(y_1)\cos(y_2)sin^2\left(\frac{x_2-x_1}{2}\right)}}{1609} \tag{5}$$

where

x_1 and x_2: first and second longitude values, respectively

y_1 and y_2: first and second latitude values, respectively

$2r_0$: approximate radius of Earth (6,378,137 m)

Although the linear distance does not represent the actual path taken by the vessel, it provides a rough estimate of the vessel's distance from the destination. Linear distance for this analysis is expressed in miles.

Distance terminal

Similar to the distance to the buoy, the distance to the terminal is calculated using the latitudes and longitudes of the current location and destination.

3.3.2 Output Variables

Buoy arrival time

This is the difference between the data point generation time and the vessel's actual time of arrival at the buoy, expressed in minutes. This is the output parameter that the network learns to predict. Buoy arrival time was determined by the following equation:

$$ATA = ((D_{ATA} - D_{DGT})\times 1440)+((H_{ATA}-H_{DGT})\times 60)+(M_{ATA}-M_{DGT}) \quad (6)$$

where

D_{DGT} : data point generation date (mm/dd/yyyy)

D_{ATA} : ATA date (mm/dd/yyyy)

H_{DGT} : data point generation hour

H_{ATA} : ATA hour

M_{DGT} : data point generation minute

M_{ATA} : ATA minute

Berthing time

Similar to actual time of arrival, the berth time is also the difference between the data point generation time and the actual berth time of the vessel at the terminal. This serves as the output parameter for the second leg of the trip from the buoy to the container terminal. Berth time is derived from the expression below.

$$\text{Berth Time} = ((D_{BT}-D_{DGT})\times 1440)+((H_{BT}-H_{DGT})\times 60)+(M_{BT}-M_{DGT}) \quad (7)$$

where

D_{DGT} : data point generation date (mm/dd/yyyy)

D_{BT} : ATA date (mm/dd/yyyy)

H_{DGT} : data point generation hour

H_{BT} : ATA hour

M_{DGT} : data point generation minute

M_{BT} : ATA minute

3.4 Procedure

The first step was to normalize each dataset. Normalization is essential to avoid the large impact that some variables can have on the prediction variable due to its scale. Min-max normalization was employed. Using the index variable, we created training and test datasets; we used 80% of each dataset as the training dataset and the remaining 20% as the testing dataset.

There is no fixed rule on how many hidden layers or hidden neurons to use in a network. Alice (2015) stated that the number of neurons should be between the input layer size and the output layer size, usually 2/3 of the input layer size. The number of hidden layers and nodes in this analysis were selected through trial and error. In this section, we borrowed ideas from Vishwakarma (1994), who recommended comparing node size for one hidden layer and two hidden layers to identify the best option. Hence, it was decided to compare networks with one and two hidden layers when selecting the number of hidden nodes to consider. Through trial and error, the optimal number of hidden layers and neurons was selected. Table 3.2 represents the error levels upon which selection was made.

Table 3.2: Basis for Architecture Selection

1 Hidden Layer		2 Hidden Layers					
Neurons	**Error**	**Neurons**	**Error**	**Neurons**	**Error**	**Neurons**	**Error**
10	2.0019	5,5	2.0411	6,4	9.4283	7,3	2.9491
9	2.8078	5,4	2.7014	6,3	3.0238	7,2	2.0763
8	2.3875	4,4	6.4157	5,3	2.1679	6,2	2.2155
7	3.1467	4,3	2.6130	5,2	3.1834	6,1	3.2063

The network that was built using one hidden layer with 10 neurons produced the least network error. This made the structure of the network [11, 10, 1], representing 11 input neurons, one hidden layer with 10 neurons, and 1 output neuron (Fig. 3.1).

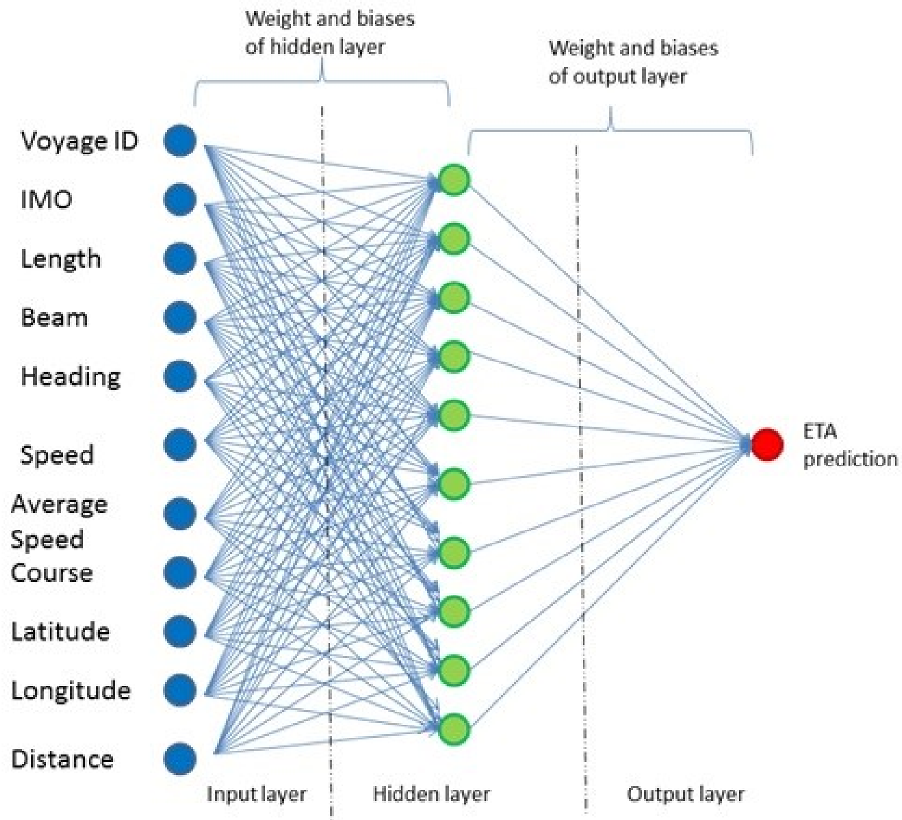


Figure 3.1: Architecture of Neural Network Developed for ETA Prediction

The output layer of the NN is represented using a single neuron, which is aimed at predicting an ETA as close as possible to the actual time of arrival (ATA).

3.4.1 Training Phase

During training, 80% of records that included the input variables alongside the output variable (actual time of arrival) were provided to the NN. The aim of this training phase is for the network to identify patterns and relationships between input and output variables by finding the optimal weights that connect the NN layers. The network was trained through back-propagation.

3.4.2 Testing Phase

The testing phase is when the error of the NN is determined. It was the phase responsible for determining the developed network's accuracy. Here, another dataset without an output variable was provided to the network, and the network was made to predict the output (vessel arrival time).

The study area was unique and thus required a special approach for trips from the pilot point (sea buoy) to the respective terminals. With this in mind, the above-mentioned procedures were repeated using data relevant to this portion of the trip. This section considers the complete trip up to the terminals, where transfer of modes occurs. This portion of the vessel trip possesses different attributes, because the trip is performed in a controlled environment (ship channel) with many restrictions. The output parameter for this leg of the trip was the berthing time. The distance applied to the network for this leg was the distance from the buoy to the terminals. The NN developed was used to predict the arrival time to terminals at different locations (Bayport and Barbours Cut).

3.4.3 Error Metrics Used for Evaluating ETA Predictions

The model developed for determining the vessel's ETA to the Port of Houston was evaluated using two error metrics: mean absolute error (MAE) and root mean square error (RMSE). The MAE metric is a representation of the average error in minutes, whereas the RMSE supplies the variance of the prediction errors and is always greater than the MAE. The MAE is determined as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |\mathrm{X_i} - \mathrm{Y_i}|}{\mathrm{n}} \tag{8}$$

In a similar pattern, the RMSE was determined using the formula:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\mathrm{X_i} - \mathrm{Y_i})^2}{\mathrm{n}}} \tag{9}$$

where

X_i : actual time of arrival
Y_i : predicted time of arrival from the model
n : the number of observations

3.5 Results

This section presents the results obtained by applying the described methodology. The focus is more on the errors of the methodology, which is a representation of how well the model captured and learned from the presented data. It identifies the level of discrepancy between the ETA and the ATA at the Port of Houston. For the medium time horizon, errors were estimated at different time intervals, between 5 days and one day to arrival. Figure 3.2 shows trends in the accuracy of the predictions. This result was based on hourly data points as input to the NN. The accuracy of the ETA improved as the vessel approached the pilot point. Day 5 to arrival records had prediction MAE and RMSE of 1055 and 1501 minutes, respectively. Prediction accuracy improved by 41.6% for results obtained at 4 days to arrival. This improvement extended to the third day and second day to arrival by 25.3%. At one day to arrival, the prediction MAE and RMSE were at 246 and 345 minutes, respectively.

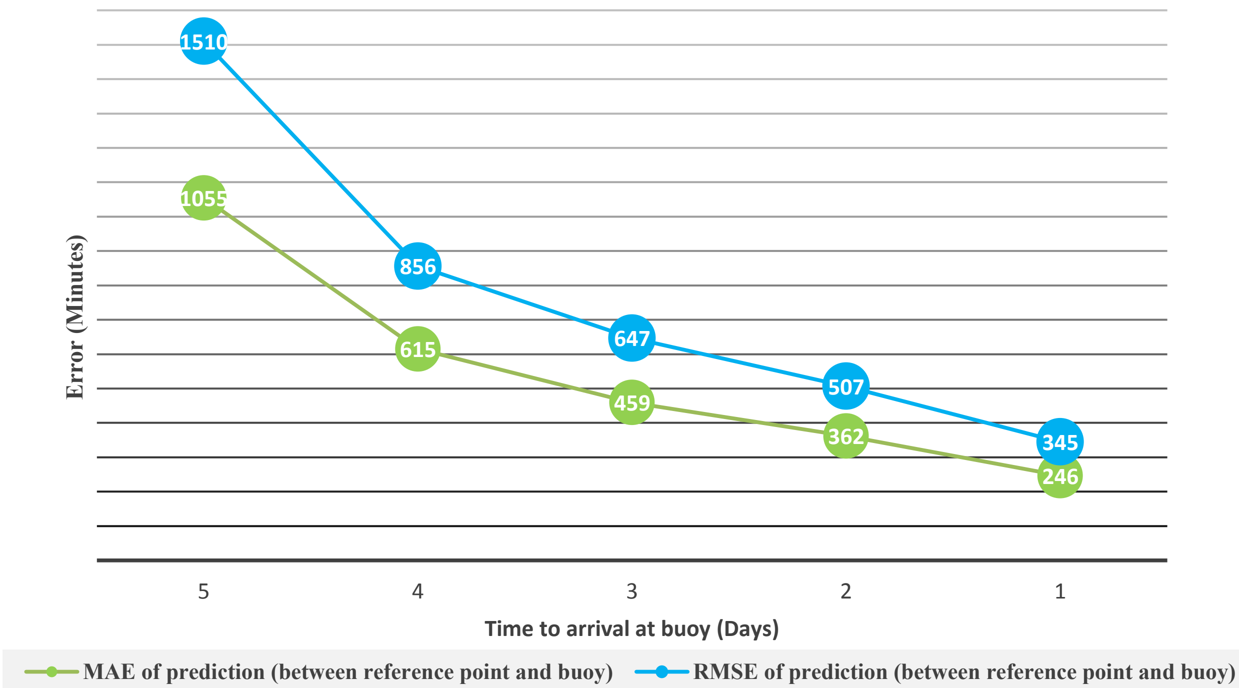


Figure 3.2: Mean Absolute and Root Mean Square Errors on ETA Predictions from Developed Neural Networks for the Last Five Days before Arrival at Buoy

Predictions made in the last 24 hours to arrival also followed the same trend as the last 5 days readings. There was a steady decrease in prediction error as the vessel approached the final hour of the trip to the buoy. The trend in the error level at different hours is presented in Figure 3.3. At 24 hours to arrival, the discrepancy in the predicted time of arrival was 246 and 345 minutes for MAE and RMSE, respectively. As the vessel approached the destination and within the last two hours before arrival at buoy, there was an approximate 93% drop in the errors with 15 minutes MAE and 21 minutes RMSE.

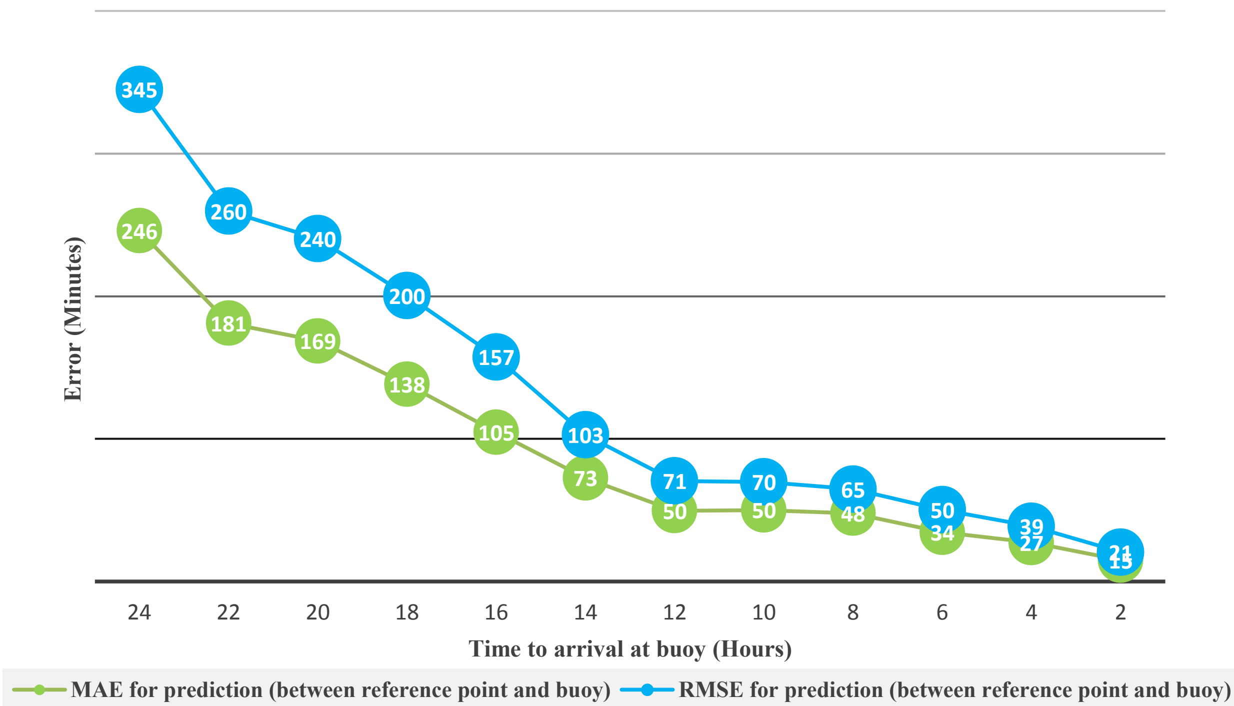


Figure 3.3: Mean Absolute and Root Mean Square Errors on ETA Predictions from Developed Neural Networks for the Last 24 Hours before Arrival at Buoy

Prediction error data was collected for trips between the pilot point (buoy) and terminals. Results for Barbours Cut and Bayport terminals were estimated separately. Data points for predictions on both legs of the trip differ in quantity; hence, the expressions used for calculating the total prediction errors are as follows:

$$MAE_{total} = \frac{n_1 MAE_1 + n_2 MAE_2}{n_1 + n_2} \quad (5)$$

$$RMSE_{total} = \sqrt{\frac{n_1 {RMSE_1}^2 + n_2 {RMSE_2}^2}{n_1 + n_2}} \quad (6)$$

where

n_1 : number of prediction records for trips between reference point and buoy
n_2 : number of prediction records for trips between buoy and terminal
MAE_1 : mean absolute error for trips between reference point and buoy
MAE_2 : mean absolute error for trips between buoy and terminal
$RMSE_1$: root mean square error for trips between reference point and buoy
$RMSE_2$: root mean square error for trips between buoy and terminal

Figures 3.4 and 3.5 show the MAE and RMSE of prediction for trips to Barbours Cut Container Terminal when the reference point is between the last 5 days and last 24 hours to pilot point.

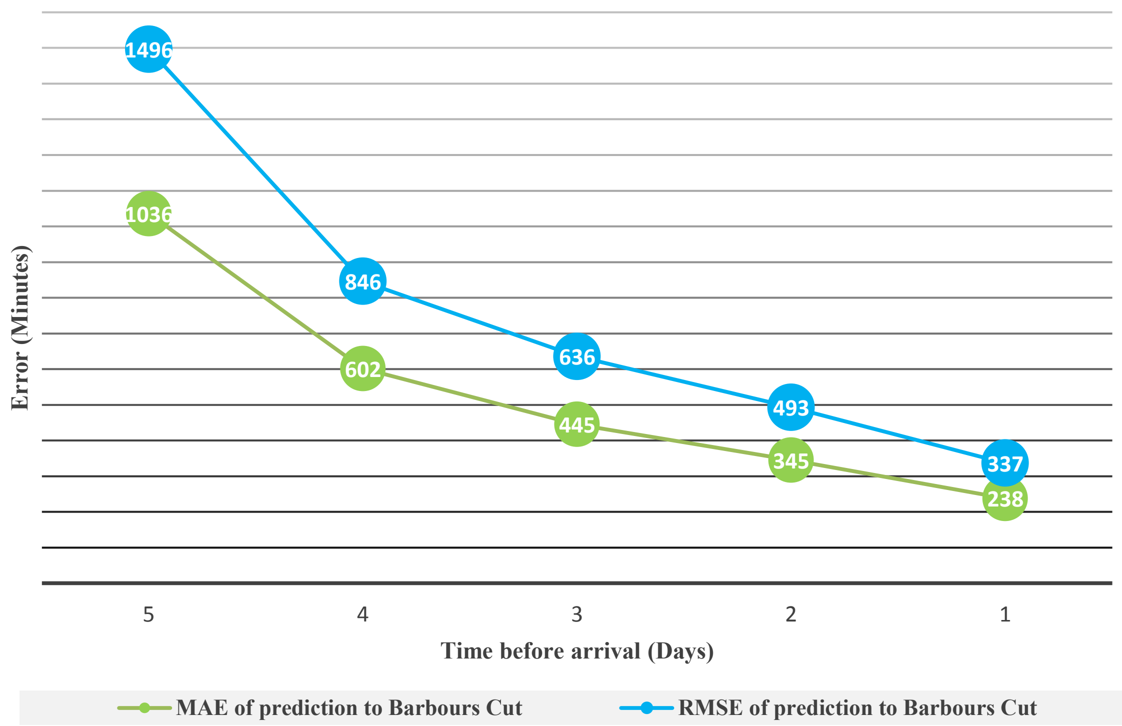


Figure 3.4: Mean Absolute and Root Mean Square Errors on ETA Predictions from Developed Neural Networks for Trips to Barbours Cut Terminal Considering Five Days before Arrival at Buoy

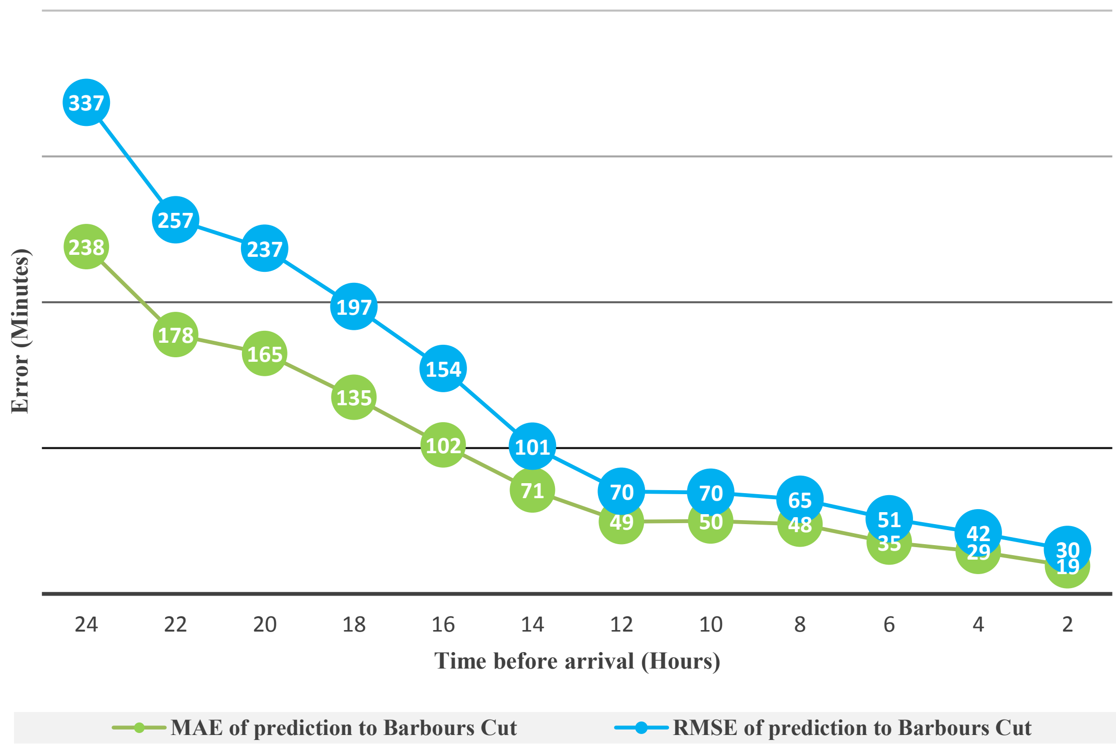


Figure 3.5: Mean Absolute and Root Mean Square Errors on ETA Predictions from Developed Neural Networks for Trips to Barbours Cut Terminal Considering 24 Hours before Arrival at Buoy

For trips to Barbours Cut Terminal, the model's prediction strength improved beyond what were experienced when only trips to the sea buoy were analyzed. These predictions followed the same trend of reduction in error as the vessels approached the terminal. Comparing these results to those obtained for trips to the sea buoy shows that the predictive strength improved. Five days before arrival at sea buoy, the MAE and RMSE were at 1055 and 1510 minutes, respectively, for the predictions made to the sea buoy. The errors dropped to 1036 and 1495 minutes at the Barbours Cut terminal level analysis. This reduction in error level was similar for all time periods leading up to the final two hours before arrival at the buoy. Similarly, Figures 3.6 and 3.7 show the MAE and RMSE, respectively, of predictions for trips to Bayport Container Terminal when the reference point is between the last five days and the last 24 hours to the pilot point.

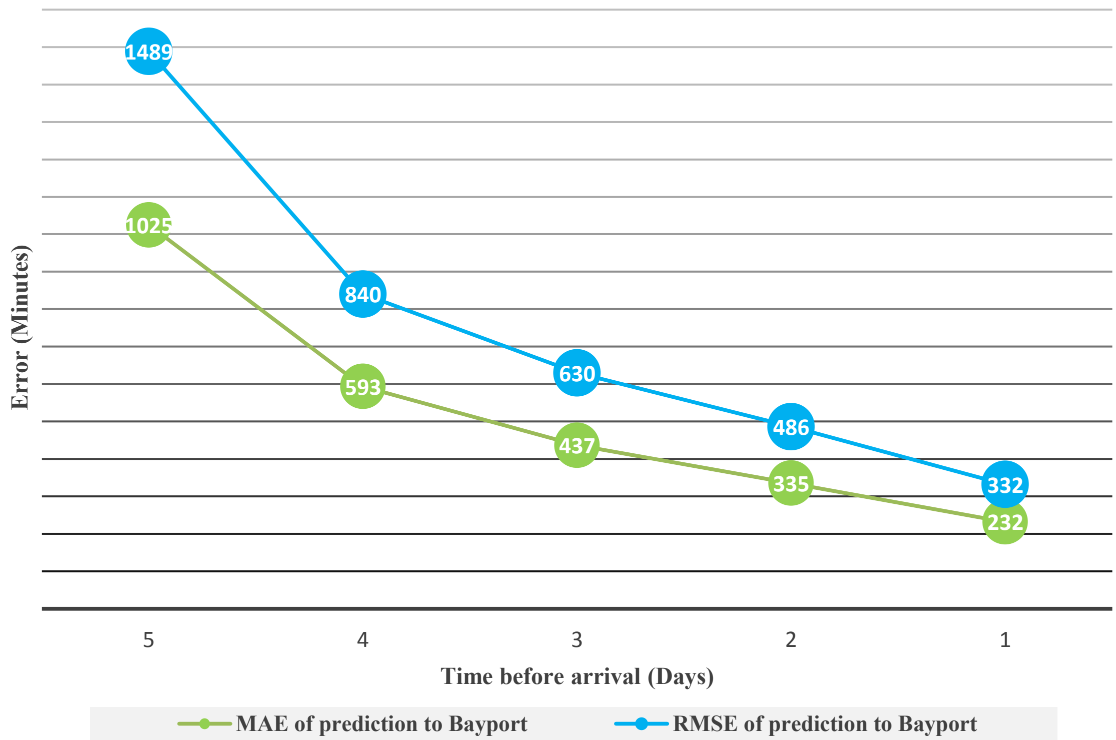


Figure 3.6: Mean Absolute and Root Mean Square Errors on ETA Predictions from Developed Neural Networks for Trips to Bayport Terminal Considering Five Days before Arrival at Buoy

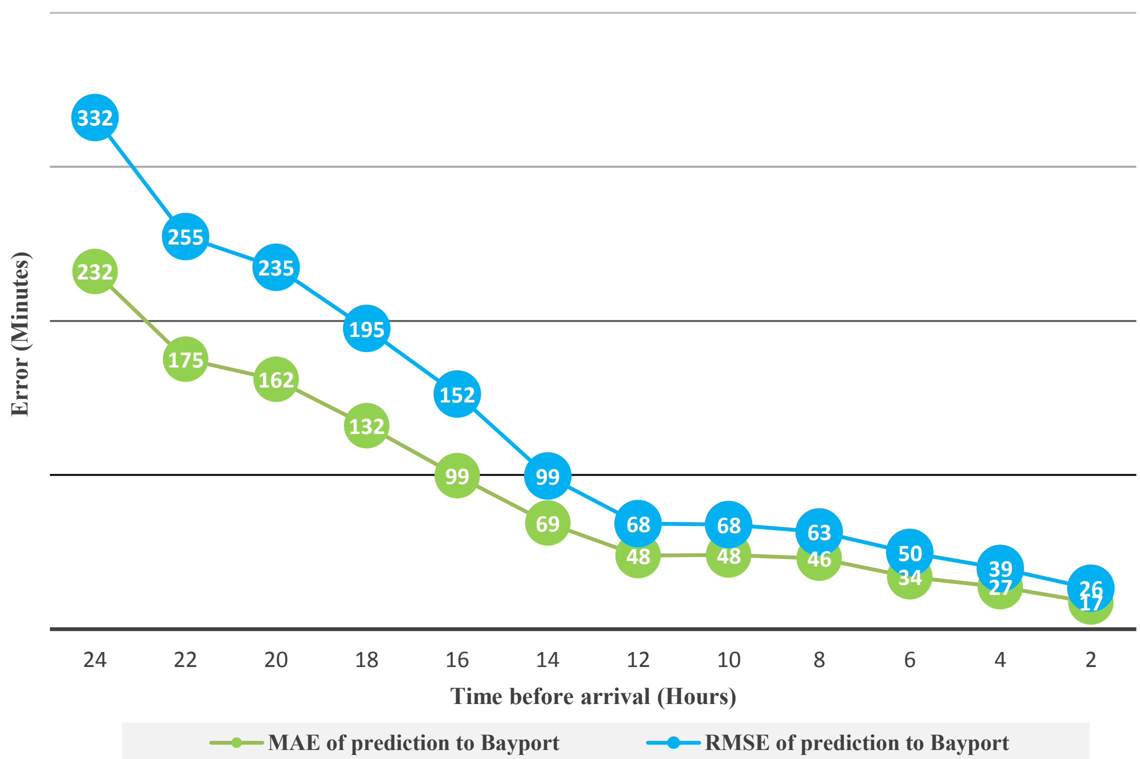


Figure 3.7: Mean Absolute and Root Mean Square Errors on ETA Predictions from Developed Neural Networks for Trips to Bayport Terminal Considering 24 Hours before Arrival at Buoy

At the Bayport terminal level, the NN produced up to a 5% improvement (5% reduced MAE) in predictive strength as compared to predictions made at the buoy level. In terms of the RMSE, there was an improvement of about 4% in the predictive strength of the network when comparing predictions made at the buoy and Bayport terminal level.

After presenting the result obtained from the neural network, understanding the contribution of the variables in predictions made at different time-frames will provide a better explanation of the knowledge extracted from the data. Identifying the contribution of the variables means quantifying the importance of the variable used in the network.

Table 3.3 presents the results obtained from calculating the variable importance over the last 24 hours of the vessel's arrival time predictions. Olden's algorithm was employed to quantify and identify variables at different hours.

Table 3.3: Variable Importance over the Last 24 Hours of the Vessel's Arrival Time Prediction

	Last Hours											
	2	**4**	**6**	**8**	**10**	**12**	**14**	**16**	**18**	**20**	**22**	**24**
Voyage ID	3.1	1.3	5.7	3.4	0.4	1.8	0.2	0.3	1.1	0.2	1.3	0.6
IMO	22.1	10.2	0.9	3.3	0.9	7.7	0.7	1.2	2.5	0.0	16.3	0.8
Length	10.6	1.9	0.1	0.2	0.5	3.1	0.0	1.0	2.3	0.6	5.6	0.1
Beam	2.5	4.3	9.4	1.7	4.2	5.4	91.2	16.8	56.7	96.0	59.3	71.0
Heading	1.5	1.0	22.1	14.6	14.0	17.0	1.0	2.5	11.6	1.5	1.7	9.3
Speed	20.8	4.2	4.3	11.5	3.0	2.7	2.4	1.4	1.7	0.3	3.5	0.3
Average Speed	10.8	29.1	9.5	25.3	0.1	11.4	1.5	1.1	1.5	0.1	0.4	4.3
Course	0.1	2.2	0.6	27.3	0.1	0.0	0.3	0.0	0.3	0.1	0.7	0.1
Latitude	1.9	5.4	0.1	0.4	1.1	4.4	0.1	3.5	0.0	0.4	1.6	5.1
Longitude	18.4	24.7	23.2	4.3	17.1	38.6	1.7	16.9	2.7	0.4	0.8	1.6
Distance to Buoy	8.3	15.6	24.2	8.0	58.7	7.8	0.9	55.4	19.6	0.5	8.7	6.8

The importance of vessel length in the last two hours of the prediction stood out. The result in Table 3.3 shows a significant increase in the importance of length when compared to earlier hours of the trip. There was a 457% increase from the last level of relative importance for length, from 1.9% to 10.6%. This increased importance is due to the available sailable area during these periods. As the vessel approaches a region of higher vessel density, the effects of vessel length on speed and maneuverability begin to manifest. In this time interval, more attention must be paid to the vessel's speed and maneuvering pattern to avoid collisions.

Highest speed importance was also observed at this time interval as greater caution is required in this region, which tends to affect sailing speed. Sailing speed during this segment of the trip is usually at its lowest. In the last two hours of the trip, speed was the most important variable. It should also be noted that the importance of speed in the model rose by about 395% compared to the previous 4 hours of the trip speed, from 4.2 to 20.8% of relative importance. Hence, the influence of speed as a variable developed for determining arrival time of a vessel to port is at the highest when vessels are close to port, a region where congestion is high.

Latitude and longitude as factors that identify the location of a vessel have been identified as important variables for predicting vessel arrival time as they also determine the distance between the vessel's current location and destination. Due to the location of the Port of Houston and the path followed, the majority of trips made by vessels heading to the port are northward. Hence, these movements are along the longitude which justifies greater values recorded in importance level of longitudes as compared to latitudes. Less east or west orientation of the trips in these final hours was observed.

3.6 Summary

A description of the approach employed in determining the ETA was presented. Types of data used were described, and the procedure was explained. Results of this approach were also presented.

Chapter 4. Summary and Conclusions

4.1 Introduction

A neural network was applied to predict the ETA of vessels. The developed network produced interpretable results, which require a summary based on its applicability. Results of the analysis were presented in Section 3.5 with brief interpretations. The rest of this chapter is organized as follows. Section 4.2 provides a summary of the results, and concludes with the author's views. Section 4.3 details directions for further research to improve ETA prediction using machine learning.

4.2 Summary and Conclusions

This research described a neural network approach that can be used to generate the ETA of vessels to port terminals. From the results collected, we found that there is great potential in the use of neural networks in this pursuit. Our findings show that near exact predictions can be achieved even without prior estimations by vessel captains. The results indicated that the farther from the destination, the greater the error in prediction. This is also evident in the comparison of prediction errors between Bayport and Barbours Cut Container Terminals. These results follow a trend similar to that in the work by Parolas (*10*), where a neural network was applied. Another observation in the results of this analysis is that predictions made at the terminal level were more accurate than those made at the buoy level. Although there are inaccuracies in the prediction, the ETA generated by this approach provides a timeframe within which the terminal and trucking companies can plan ahead for arriving vessels. It should also be noted that the results from the variable importance analysis will assist in the selection of useful variables in future predictions.

4.3 Directions for Future Research

For further studies, improvements can be made by exploring a larger dataset. Considering other machine learning algorithms will also help to reveal possibilities for improvement.

References

1. Mokhtar, K., and Shah, M. (2013), "Efficiency of Operations in Container Terminals: A Frontier Method," *European Journal of Business and Management*, 5(2), pp. 91-106.
2. The Center for Urban Transportation Research (CUTR) at the University of South Florida. "Analysis of Freight Movement Mode Choice Factors". Report for Florida Department of Transportation Rail Planning and Safety. Available at http://www.fdot.gov/rail/Publications/Studies/Planning/ModeChoiceFactors.pdf
3. Notteboom, T. (2006), "The Time Factor in Liner Shipping Services," *Maritime Economics and Logistics*, 8(1), pp. 19-39.
4. Hatzius J., Stehn S. J., Fawcett N., and Chaudhary M. (2017), "As good as it gets," *Global Economics Analyst.*
5. Maritime-insight (2015), "Study on the Analysis and Evolution of International and EU Shipping". Available at https://ec.europa.eu/transport/sites/transport/files/modes/maritime/studies/doc/2015-sept-study-internat-eu-shipping-final.pdf
6. Samija, S., Maritime Transport Management for The Purpose of Efficiency and Safety of Shipping Services," Doctoral thesis, University of Rijeka. Available at http://www.geocities.ws/icts_papers/Papers/Samija.pdf
7. Tennessee Tombigbee Waterway, "Cargo Capacity of Different Transportation Mode." Accessed on September 17, 2017. Available at http://business.tenntom.org/why-use-the-waterway/shipping-comparisons/
8. Ballou, R. (1998), *Business Logistics Management*, 4th edition, Prentice Hall, Saddle River, NJ.
9. American Trucking Associations: Latest U.S. Freight Transportation Forecast Shows Continued Growth for Trucking. Accessed December 12, 2017 http://www.trucking.org/article/Latest-U.S.-Freight-Transportation-Forecast-Shows-anContinued-Growth-for-Trucking
10. Dong, J., Makaiwi, M., Shafrieirad, N., and Huang, Y. (2015), Modeling Multimodal Freight Transportation Network Performance under Disruptions. A Cooperative Research Project sponsored by U.S. Department of Transportation-Research and Innovative Technology Administration.
11. Rodrigue, J., Comtois, C., and Slack, B. (2006), *The Geography of Transport Systems,* Routledge Taylor and Francis Group, London and New York, pp. 114, 440.
12. Broeze, F. (2002), "The globalization of the oceans: containerization from the 1950s to the present," *International Maritime Economic History Association*. 23, pp. 279-285.
13. Carrese, S. and Tatarelli, L. (2011). "Optimizing the Stacking of the Intermodal Transport Units in an Inland Terminal: An Heuristic Procedure," *Procedia Social and Behavioral Science*, 20, pp. 994-1003.
14. Hayuth, Y. (1987). *Intermodality: Concept and Practice*, Lloyd's of London Press, London. Available at https://trid.trb.org/view.aspx?id=247406
15. World Shipping Council (2015), Some Observations on Port Congestion, Vessel Size and Vessel Sharing Agreements. http://www.worldshipping.org/industry-issues/transportation-infrastructure/Observations_on_Port_Congestion_Vessel_Size_and_VSA_May_28_2015.pdf
16. Jashari, N. (2007), "Liability of the Multimodal Transport Operator – The Case of Kosovo" *Scientific Journal of Maritime Research*, 31, pp. 84-88.

17. Liu, Q. (2010), "Efficiency Analysis of Container Ports and Terminals," Doctoral thesis, University College London (UCL).
18. Steenken, D., Voss, S., and Stahlbock, R. (2004), "Container Terminal Operation and Operations Research-A Classification and Literature Review," *OR spectrum*, 26(1), pp. 3-49.
19. Chung, C. and Chiang, C. (2011). "The Critical Factors: An Evaluation of Schedule Reliability in Liner Shipping," *International Journal of Operations Research*, 8(4), pp. 3-9.
20. Esmer, S. (2008), "Performance Measurements of Container Terminal Operations," *DokuzEylul University Journal of Graduate School of Social Sciences*, 10, pp. 238-255.
21. Yoon, D. (2007), "Analysis of Truck Delays at Container Terminal Security Inspection Stations," Doctoral Thesis, New Jersey Institute of Technology. Available at http://archives.njit.edu/vol01/etd/2000s/2007/njit-etd2007-059/njit-etd2007-059.pdf
22. Huynh, N. N. and Walton, C. M. (2005), "Methodologies for Reducing Truck Turn Time at Marine Container Terminals," Report for Southwest Region University Transportation Center, Center for Transportation Research, The University of Texas at Austin.
23. Container Security Initiative: just the facts, *Port Technology International*, 37. Available at https://www.porttechnology.org/technical_papers/container_security_initiative_just_the_facts
24. Fagerholt, K., Heimdal, S. I., & Loktu, A. (2000), "Shortest path in the presence of obstacles: An application to ocean shipping," *Journal of the operational research society*, 51, pp. 683-688.
25. Veldhuis, H. D. (2015), "Developing an automated solution for ETA definition concerning long distance shipping," Bachelor's Thesis, University of Twente.
26. Heywood, C., Connor, C., Browning, D., Smith, M. C., and Wang, B. (2009), "GPS tracking of intermodal transportation: System integration with delivery order system," *Systems and Information Engineering Design Symposium 2009. SIEDS '09,* pp. 191-196.
27. Parolas I. (2016), "Prediction of Vessels' Estimated Time of Arrival (ETA) Using Machine Learning - A Port of Rotterdam Case Study." Revised paper submitted for presentation at the 96th Annual Meeting of the Transportation Research Board January 2017 and publication in Transportation Research Record.
28. Harris, I., Wang, Y., Wang, H. (2015), "ICT in multimodal transport and technological trends: Unleashing potential for the future," *International Journal of Production Economics*, 159, pp. 88-103.
29. Philip John. Intelligent Maritime Navigation with the Automatic Identification System. Assessed at: http://ctrf.ca/wp-content/uploads/2014/07/Intelligent-Maritime-Navigation.pdf
30. bigoceandata global vessel management solutions. The Definitive AIS Handbook. https://www.marineinsight.com/wp-content/uploads/2016/11/AiS-Whitepaper.pdf
31. Chao, W. (2011), Machine Learning Tutorial. Graduate Institute of Communication Engineering, National Taiwan University.
32. Cunningham, P., Cord, M., and Delany, S. J., (2008). Machine Learning Techniques for Multimedia case studies on Organization and Retrieval Cord, p. 21. Retrieved from https://www.springer.com/cda/content/document/cda_downloaddocument/9783540751700-c2.pdf?SGWID=0-0-45-501205-p173779120
33. Kotsiantis, S. B. (2007), "Supervised Machine Learning: A Review of Classification Techniques," *Informatica*, 31, pp. 249-268.
34. Dayan, P. (1999). Unsupervised learning. In R. A. Wilson & F. Keil (Eds). The MIT encyclopedia of the cognitive sciences.

35. Sutton, R. S. and Barto, A. G. (2017), Reinforcement Learning: An Introduction, p. 1. The MIT Press Cambridge, MA. Retrieved from http://web2.qatar.cmu.edu/~gdicaro/15381/additional/SuttonBarto-RL-5Nov17.pdf
36. Harmon, M. E. and Harmon, S. S. (1996), Reinforcement Learning: A Tutorial, Retrieved from http://www.cs.toronto.edu/~zemel/documents/411/rltutorial.pdf
37. Khajanchi, A. (2003), "Artificial Neural Networks: The next intelligence," Retrieved from www.globalriskguard.com/resources/market/NN.pdf
38. Stergiou C. and Siganos D. "Neural Networks," Imperial College of London, Computer Department. https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Architecture
39. Sun D, Luo H., Fu L., Liu W., Liao X., and Zhao M. (2007) Predicting Bus Arrival Time on the Basis of Global Positioning System Data. *Journal of the Transportation Research Board*, No. 2034, pp. 62-72.
40. United States Department of Transportation. Bureau of transportation statistics. 2015. https://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/data_and_statistics/by_subject/freight/freight_facts_2015/chapter3/fig3_4
41. Jeong R. (2004). The Prediction of Bus Arrival Time using Automatic Vehicle Location Systems Data. A Dissertation Submitted to Texas A&M University.
42. Stergiou C. and Siganos D. (1996). "Neural Networks," Imperial College of London, Computer Department, https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html#Architecture
43. Carbonneau R., Laframboise K., & Vahidov R. (2008), "Application of machine learning techniques for supply chain demand forecasting," E*uropean Journal of Operational Research,* 184, pp. 1140-1154.
44. Fancello G., Pani C., Pisano M. Serra P., and Fadda P. (2010), "Prediction of Arrival Times and Human Resources Allocation for Container Terminal," 12th WCTR, Lisbon, Portugal
45. Parolas I. (2016). "Prediction of Vessels' Estimated Time of Arrival (ETA) Using Machine Learning – A Port of Rotterdam case study. Revised paper submitted for presentation at the 96th Annual Meeting of the *Transportation Research Board* January 2017 and publication in *Journal of Transportation Research Record*.
46. Pani C., Vanelslander. T., Fancello, G., and Cannas, M. (2015), "Prediction of late/early arrivals in container terminals: A qualitative approach," *European Journal of Transport and Infrastructure Research*, 15(4), pp. 536-55.
47. Pani, C., Fadda, P., Fancello, G., Frigau, L., Mola, F. (2014), "A Data Mining Approach to Forecast Late Arrivals in a Transhipment Container Terminal," *Transport,*29(2), pp. 175-184.
48. Qu X., Jackson, M. A., Davis L. B. (2012), "Simulation-based Evaluation of Port Emergency Evacuation Plans for Predictable Natural Disasters," *Journal of Homeland Security and Emergency Management*, 9(2), 4
49. Payson A. and Porter C. (2017) "Texas Port and the Panama Canal: Commodities and Infrastructure," Center for Transportation Research. Available at https://library.ctr.utexas.edu/ctr-publications/5-6690-01/prp2.pdf
50. Bayport Container Terminal. Accessed on February 21, 2018. Available at http://porthouston.com/container-terminals/bayport-container-terminal/

51. Barbours Cut Container Terminal. Accessed on February 21, 2018. Available at http://porthouston.com/container-terminals/barbours-cut-container-terminal/
52. Bierling, D. H., Kruse, C. J., Voigt, T. P., Vickich, M. J., Tydlacka, J. M., and Skinner, R. C. (2015), "Test and evaluation of Bluetooth technology for measuring freight facility traffic flows: Bayport case study," Texas A&M Transportation Institute.
53. Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998), "Forecasting with artificial neural networks: The state of the art," *International Journal of Forecasting*, 14, pp. 35-62.
54. Vishwakarma, K.P. (1994), "A neural network to predict multiple, economic time series," *IEEE International Conference on Neural Networks*, 6, pp. 3674-3679.
55. Alice, M. (2015) Fitting a Neural Network in R; neuralnet package. Available at https://datascienceplus.com/fitting-neural-network-in-r/
56. Greenwell, B. M., Boehmke, B. C., and McCarthy, A. J. (2018), "A Simple and Effective Model-Based Variable Importance Measure." Available at Cornell University Library. https://arxiv.org/abs/1805.04755
57. Garson, G. D. (1991), Interpreting neural-network connection weights. *Artif. Intell. Expert,* 6, pp. 47-51.