

# Enhancing maritime data utilization through advanced machine learning techniques in AIS data analysis

---

**Baković, Josipa**

**Undergraduate thesis / Završni rad**

**2023**

*Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj:* **University of Rijeka, Faculty of Maritime Studies, Rijeka / Sveučilište u Rijeci, Pomorski fakultet**

*Permanent link / Trajna poveznica:* <https://um.nsk.hr/um:nbn:hr:187:525356>

*Rights / Prava:* [In copyright](#)/[Zaštićeno autorskim pravom.](#)

*Download date / Datum preuzimanja:* **2025-01-14**



**Sveučilište u Rijeci, Pomorski fakultet**  
University of Rijeka, Faculty of Maritime Studies

*Repository / Repozitorij:*

[Repository of the University of Rijeka, Faculty of Maritime Studies - FMSRI Repository](#)



**UNIVERSITY OF RIJEKA**  
**FACULTY OF MARITIME STUDIES**

**JOSIPA BAKOVIĆ**

**ENHANCING MARITIME DATA UTILIZATION  
THROUGH ADVANCED MACHINE LEARNING  
TECHNIQUES IN AIS DATA ANALYSIS**

**BACHELOR'S THESIS**

Rijeka, 2023.

**UNIVERSITY OF RIJEKA**  
**FACULTY OF MARITIME STUDIES**

**POBOLJŠANJE ISKORIŠTAVANJA POMORSKIH  
PODATAKA PUTEM NAPREDNIH TEHNIKA STROJNOG  
UČENJA U ANALIZI AIS PODATAKA**

**ENHANCING MARITIME DATA UTILIZATION  
THROUGH ADVANCED MACHINE LEARNING  
TECHNIQUES IN AIS DATA ANALYSIS**

**BACHELOR'S THESIS**

Course: Electronic Navigation Devices

Mentor: Prof. Irena Jurdana, PhD

Co-mentor: Assist. Prof. Nikola Lopac, PhD

Student: Josipa Baković

Study program: Marine Electronic Engineering and Information Technology

JMBAG: 0069076680

Rijeka, July 2023.

Student: Josipa Baković

Study program: Marine Electronic Engineering and Information Technology

JMBAG: 0069076680

DECLARATION OF INDEPENDENT PREPARATION OF THE BACHELOR'S  
THESIS

Whereby I declare that I have written the bachelor's thesis entitled Enhancing Maritime Data Utilization through Advanced Machine Learning Techniques in AIS Data Analysis, have independently prepared under the supervision of Prof. Irena Jurdana, PhD and Assist. Prof. Nikola Lopac, PhD.

In my thesis I have applied the methodology of writing a professional/scientific paper and used the literature listed at the end of the thesis. The knowledge, attitudes, conclusions, theories and laws of other people that I have reproduced in the final work, directly or paraphrasing, have been cited in the usual way and linked with footnotes and bibliographic units, and no part of the work violates the copyrights of others. The work is written in the spirit of the English language.

Student



Student name and surname

JOSIPA BAKOVIĆ

Student: Josipa Baković

Study program: Marine Electronic Engineering and Information Technology

JMBAG: 0069076680

STATEMENT OF THE STUDENT-AUTHOR ON THE PUBLIC ANNOUNCEMENT  
OF THE DEFENDED BACHELOR'S THESIS

I declare that as a student - the author of the bachelor's thesis, I allow the University of Rijeka, Faculty of the Maritime Studies to permanently publish it publicly and make it available to the public free of charge in full text in the online digital repository of the Faculty of the Maritime Studies.

For the purpose of supporting open access to bachelor's theses permanently published in the publicly available digital repository of the Faculty of the Maritime Studies, with this statement I grant the non-exclusive property right to exploit my bachelor's thesis without content, time, and space limitations as an author's work under the terms of the Creative Commons license CC BY Naming, according to the description available at <http://creativecommons.org/licenses/>

Student - author

A handwritten signature in black ink that reads "Josipa Bakovic". The signature is written in a cursive style with a large initial 'J'.

## **SUMMARY**

This bachelor's thesis reviews the literature on the application of diverse machine learning (ML) methods in the maritime industry, specifically focusing on automatic identification system (AIS) data. The review explores a wide range of ML techniques, such as neural networks, reinforcement learning, classification, regression, and clustering, and their utilization to enhance maritime operations, efficiency, and safety. Demonstrating positive results, various studies highlight the effectiveness of these methods as well as several shortcomings that are sought to be corrected in further research. Furthermore, the thesis identifies potential future applications of ML in fleet management, vessel maintenance, and navigational safety analysis. In summary, the thesis presents practical insights through the application of selected ML methods to real AIS data, yielding meaningful results for the maritime industry.

Keywords: AIS data, machine learning methods, maritime industry

## **SAŽETAK**

Ovaj završni rad pruža pregled literature o primjeni različitih metoda strojnog učenja u pomorstvu, fokusirajući se na podatke automatskog identifikacijskog sustava (AIS). Prikazane su različite primjene metoda strojnog učenja u područjima poput pomorske sigurnosti, praćenja brodova, optimizacije rute, detekcije anomalija i predikcije ponašanja brodova. Demonstrirajući pozitivne rezultate, različita istraživanja ističu učinkovitost tih metoda kao i nekolicinu nedostataka koji se nastoje ispraviti u daljnjim istraživanjima. Nadalje, rad identificira potencijalne buduće primjene metoda strojnog učenja u upravljanju flotom, održavanju brodova i analizi navigacijske sigurnosti. Ukratko, rad pruža praktične uvide kroz primjenu odabranih metoda na stvarnim AIS podacima, donoseći značajne rezultate za pomorsku industriju.

Ključne riječi: AIS podaci, metode strojnog učenja, pomorstvo

# TABLE OF CONTENTS

<b>SUMMARY .....</b>	<b>I</b>
<b>SAŽETAK .....</b>	<b>I</b>
<b>TABLE OF CONTENTS .....</b>	<b>II</b>
<b>1. INTRODUCTION .....</b>	<b>1</b>
<b>2. AUTOMATIC IDENTIFICATION SYSTEM .....</b>	<b>3</b>
2.1. AIS CLASSIFICATION.....	4
2.2. WORKING PRINCIPLE OF AIS .....	6
2.2.1. VHF DSC, VHF TDMA, and VHF transmitter .....	7
2.2.1.1. Maritime mobile service identities.....	8
<b>3. MACHINE LEARNING FUNDAMENTALS .....</b>	<b>10</b>
3.1. WORKING PRINCIPLE OF MACHINE LEARNING.....	10
3.2. MACHINE LEARNING CLASSIFICATION .....	11
<b>4. MACHINE LEARNING BASED ON AIS DATA IN MARITIME ....</b>	<b>14</b>
4.1. SUPERVISED LEARNING .....	14
4.1.1. Support vector machine.....	15
4.1.2. Artificial neural networks.....	18
4.1.3. Decision tree regressor .....	19
4.1.4. Memory-based learning .....	20
4.2. UNSUPERVISED LEARNING .....	23
4.2.1. DBSCAN.....	24
4.2.2. TREAD .....	26
4.3. REINFORCEMENT LEARNING.....	27
4.3.1. Deep reinforcement learning .....	28
<b>5. AIS TRAINING DATA FOR MACHINE LEARNING .....</b>	<b>31</b>
5.1. CNN METHOD .....	33
5.2. ANOMALIES OF AIS DATA .....	37
5.2.1. Methods for AIS anomalies detection.....	39
<b>6. SUMMARY TABLE OF USED ARTICLES.....</b>	<b>40</b>
<b>7. CHALLENGES AND FUTURE TRENDS .....</b>	<b>47</b>
<b>8. CONCLUSION.....</b>	<b>48</b>

<b>LITERATURE .....</b>	<b>49</b>
<b>ACRONYM INDEX.....</b>	<b>57</b>
<b>LIST OF FIGURES .....</b>	<b>60</b>
<b>LIST OF TABLES .....</b>	<b>61</b>



# 1. INTRODUCTION

The history of human civilization abounds with examples of technological progress that has influenced all spheres of life. Technology is developing expeditiously, and people sometimes fail to keep up with that pace. From the invention of tools and the wheel to modern technologies, humanity is perpetually putting in the effort to improve its living conditions and adapt more successfully to the environment. Since all development and progress, including technological ones, bring its own flaws and problems, they also appear in the development of the maritime industry. Nonetheless, this progress results in all kinds of challenges and also a few disadvantages. In the current landscape, the integration of machine learning (ML) presents tremendous possibilities and complexities across various facets of human endeavors. To put it simply, machine learning is a distinctive domain within the realm of artificial intelligence (AI) that endeavors to construct algorithms proficient in scrutinizing data and rendering well-informed judgments devoid of explicit programming directives. These methodologies are already finding practical employment in a multitude of domains. Among the noteworthy sectors embracing machine learning is the maritime industry, which holds a prominent status as a fundamental constituent of global commerce. As such an important component, it involves the utilization of automatic identification system (AIS) data to enhance various other objectives. This technology enables the automated analysis of vast quantities of data, enabling the identification of patterns and the ability to make predictions about the future.

In this technology that allows computers to learn from experience and enhance their efficacy, there are numerous advantages and disadvantages. Among the advantages are the optimized performance, accuracy, and speed of data analysis. As for the disadvantages, one of the drawbacks of AIS data is the potential for falsification, which leads to several additional issues. One potential example could be the lack of sufficient data for analysis. AIS knowledge insights are collected in real-time, but this can be limited in terms of time range or data quality. Therefore, collection and processing can be challenging. Another problem can be the obstacle in understanding and interpreting the obtained results of machine learning because it is a complex method that can be difficult to understand and interpret without adequate knowledge and experience within that domain. Among the aforementioned challenges, notable concerns include the absence of transparency in decision-making processes, the potential for discriminatory outcomes, and the diminishing

human control over decision-making. Consequently, it is important to research and understand how ML can impact maritime and apply it in a way that will provide the greatest benefits.

The study's research section involves the literature review related to certain machine learning techniques to a genuine AIS dataset, which demonstrates its effective use of it. The findings highlight the value of these results for diverse applications within the maritime sector. Briefly showing different articles was the purpose and goal of this thesis, each of which describes the way in which AIS data is being used and how it affects and applies to maritime affairs. The results and conclusions of these articles have been gathered into one interconnected unit. In this first introductory chapter, the problems and subjects of the thesis are presented, which then lead to the hypothesis and the purpose, namely the goal, of the research. The second part entitled 'Automatic Identification System', explains the basic characteristic of the considered device and communication system, as well as its main components. The device types are also mentioned, as well as the data it offers. It is important to process it first in order to facilitate a better understanding of the application of ML on its data later on. The main part of the thesis depicts the fundamentals of ML as well as their application. These sources provide a comprehensive overview of the practical uses in various maritime fields. The analysis highlights key insights and trends that have emerged in recent years and showcases how these developments have impacted industries. By examining these case studies, it is evident that the application of this method has led to revolution. The focus of the fifth section, 'AIS training data for ML', is data utilization and anomalies. Current challenges, and issues in the deployment, are described in the seventh section. This section also encompasses possible upcoming developments to enhance and broaden the subject. Overall conclusion of the thesis is addressed in the final section.

## 2. AUTOMATIC IDENTIFICATION SYSTEM

In the maritime industry, ensuring safety is paramount, and the invention of the AIS has played a crucial role in achieving this objective. AIS, widely recognized and utilized, has not only bolstered safety measures within the maritime sector but has also brought about a multitude of other advantages. AIS represents an advanced device used for live tracking of ships and vessels, providing crucial information, including ship category, call sign, flag country, speed, location, heading, and other relevant particulars. The primary objective behind developing AIS was to enhance safety at sea by enabling improved visibility among commercial vessels in various conditions and providing helmsmen with enhanced environmental information, thus serving as a valuable tool for collision avoidance. With that said, there are many other assets, such as better route planning and increased safety with the ability to quickly identify potential hazards as well as collect traffic data for further analysis. Overall, AIS has had a significant impact on improving navigation optimization and reliability for mariners, making it an essential tool in modern-day marine operations with a simple operating principle and several components within the device itself.

According to International Maritime Organization (IMO) regulations, every international ship with a gross tonnage exceeding 300t and every passenger ship must have an AIS transceiver installed. However, some vessels are not obligated to install them due to the large size of these transceivers, which can impede the receipt of navigational information from other boats and complicate their detection through passive systems. This robust instrument consists of very high frequency (VHF) transmitter, very high-frequency digital selective calling (VHF DSC) shipboard receiver, and very high-frequency time division multiple access (VHF TDMA) receivers. The range can vary, from 20 nautical miles in poor conditions to 350 nautical miles for transmissions under favorable atmospheric circumstances. The estimated average reception range of AIS receivers is approximately 40 nautical miles, considering the factors previously listed. In addition, AIS devices can be subject to various problems, including network overload and a lack of standardization in data collection. To improve the operation and design of AIS devices, advanced technologies, such as mentioned machine learning, are being applied [1,2]. Algorithms and statistical models are utilized by ML to analyze vast amounts of data and discover patterns that may be challenging for humans to discern visually. Applying this method to the analysis of AIS device operations can provide a deeper insight into the functioning of the device and uncover

hidden connections between different parameters, which could lead to improvements in performance.

## 2.1. AIS CLASSIFICATION

When speaking about the improvement, AIS transceivers can be divided into two classes (Classes A and B). These two types differ in terms of the number of reported data fields they transmit and the frequency at which they report this information [3]. Eleven distinct pieces of information can be gathered from a ship's Class A AIS transceiver, which can be further categorized into three distinct types: static data, voyage-related data, and dynamic data, as described by Kwang-il Kim et al. [4]. Static information includes permanent details about a ship, such as its name, call sign, type, and specifications. This data remains unaltered as time progresses. Voyage-related data pertains to a particular journey and encompasses information such as cargo specifics, ship draft, and the estimated arrival time (ETA). This data remains consistent throughout a singular voyage from the originating port to the destination port. Conversely, dynamic information offers up-to-the-minute situational particulars about a vessel, encompassing its velocity, course, and global positioning system (GPS) coordinates. Examples of static data, voyage-related data, and dynamic data are illustrated in Figure 1.



**Figure 1 Overview of AIS information types**

Source: Wakabayashi, N. and Jurdana, I. (2020) Maritime Communications and Remote Voyage Monitoring. Available at: <https://doi.org/10.1109/cobcom49975.2020.9174182>. [5]

The researchers [4,6,7] highlighted the different rates at which these types of information are transmitted through AIS. Static and voyage-related information is transmitted at regular intervals of 6 minutes, while dynamic information is transmitted at different rates depending on the ship's current navigation status. On the other hand, class B AIS is typically used on smaller vessels, such as recreational boats. Unlike Class A, which is mandatory for commercial vessels over 300 gross tons, Class B is not required by international regulations. Nevertheless, many boat owners choose to install Class B AIS devices to improve safety and situational awareness while navigating. Also, Class B devices operate on lower transmission power, meaning they have a shorter transmission range. Additionally, Class A typically reports information more frequently than Class B, which can bring Class B to some loss of real-time accuracy in tracking. The pivotal difference between these two classes is reflected in the transmission rate, as shown in Table 1, which is influenced by factors such as the size of the message and the available bandwidth.

**Table 1 Transmit rates of Class A and Class B**

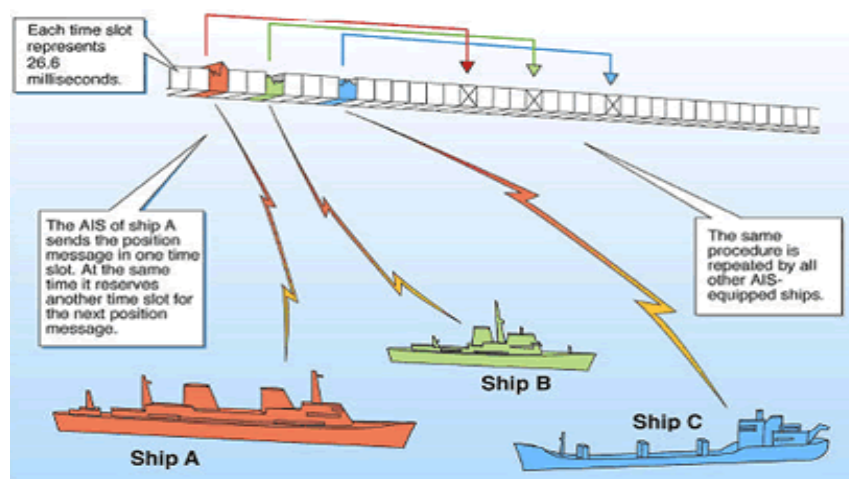
SHIP'S CONDITIONS	TYPE OF INFORMATION	CLASS A	CLASS B
Anchored or Moored ship	Static information	3 min	3 min
SOG 0-2 knots	Dynamic information	10 sec	3 min
SOG 2-14 knots	Dynamic information	10 sec	30 sec
SOG 2-14 knots with changing course	Dynamic information	3.3 sec	30 sec
SOG 14-23 knots	Dynamic information	6 sec	30 sec
SOG 14-23 knots with changing course	Dynamic information	2 sec	30 sec
SOG over 23 knots	Dynamic information	2 sec	30 sec
Voyage-related data	Static information	6 min	6 min

Source: made by student based on data from Passage Guardian (no date) Best Practice AIS For Bluewater Sailing Yachts. Available at: <https://passageguardian.nz/pages/best-practice-ais-for-bluewater-sailing-yachts>. (28<sup>th</sup> April 2023.) [8]

The frequency of sending dynamic information by Class A equipment depends on the vessel's speed and heading, ranging from every 2 to 10 seconds. During a ship's anchorage, a message is broadcasted at regular intervals of 3 minutes. Static information, including voyage-related data, is sent every 6 minutes [9]. On the other hand, Class B stations exhibit reduced transmission rates and are explicitly tailored for recreational watercraft [9].

## 2.2. WORKING PRINCIPLE OF AIS

AIS system utilizes VHF components because of its frequency range of 156.025MHz and 162.025MHz, with a bandwidth of 25kHz [10], which will be explained in the following paragraph. This range is subject to international regulations and is reserved for AIS communication. In the AIS system, each station creates its schedule for transmitting position reports, taking into account its traffic history and predictions of other stations' future actions. To avoid interference, the stations synchronize with each other continuously using 2250 time slots that are established every minute. The process of selecting time slots in the system is done randomly using a specific interval and an arbitrary timeout ranging from 0 to 8 frames. Whenever there is a change in a slot assignment by the station, it broadcasts the new timeout and location to guarantee that vessels within radio range are informed about the change, even if they have newly arrived. Figure 2 vividly illustrates the principle of transmitting AIS data within time slots.



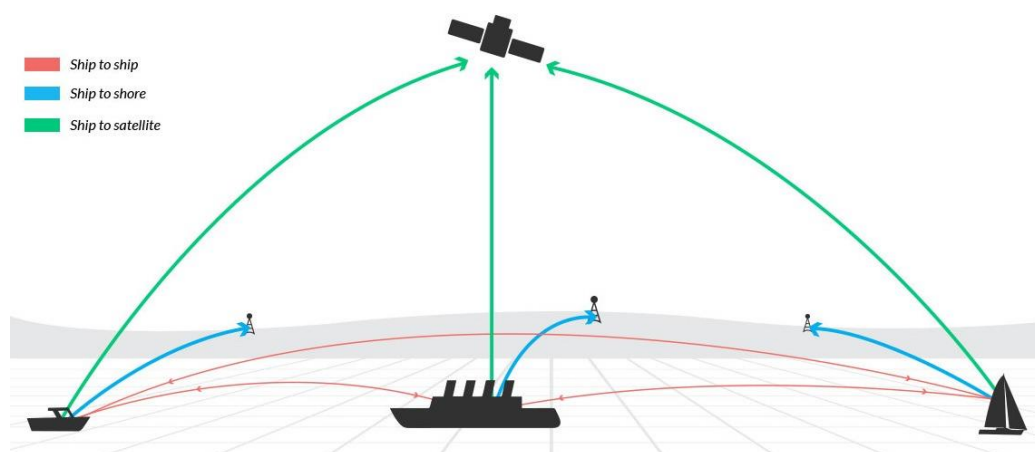
**Figure 2 AIS time slots**

Source: 2022., How AIS Works | Navigation Center (no date). Available at: <https://www.navcen.uscg.gov/howais-works>, (29<sup>th</sup> April 2023.) [11]

A minimum ship reporting capacity of 2000 time slots per minute is mandated by the IMO performance standard. Regardless, the system offers a total of 4500 time slots per minute. With the self-organized time-division multiple access (SOTDMA) broadcast mode, slots can be shared, leading to a potential overload of up to 500% [10]. If there is too much demand or excessive use, such that the system cannot handle it, closer targets receive priority, and only distant targets may experience dropouts. Additionally, the system is DSC-compatible, which enables Global Maritime Distress and Safety System (GMDSS) based on shore to identify vessels in a cost-effective manner and to form AIS operating channels.

### 2.2.1. VHF DSC, VHF TDMA, and VHF transmitter

The basic components of every AIS system are one VHF DSC, two VHF TDMA, and one VHF transmitter [11]. It works autonomously and continuously, regardless of the location - be it the open sea, shoreline, or land areas, as demonstrated in Figure 3. Transmissions use 9.6kb Gaussian minimum shift keying frequency modulation (GMSK FM) over 25 or 12.5kHz channels using high-level data link control (HDLC) packet protocols [12]. Individual stations receive and transmit via dual radio channels, even though only a single radio channel is essential.



**Figure 3 AIS - transmitting and receiving data**

Source: Passage Guardian (no date) Best Practice AIS For Bluewater Sailing Yachts. Available at: <https://passageguardian.nz/pages/best-practice-ais-for-bluewater-sailing-yachts>. (28<sup>th</sup> April 2023.) [61]

VHF DSC transmitter operates on the use of pre-selected digital channels, known as digital selective calling channels. These channels are used to send and receive distress calls, general calls, and other types of messages. To initiate communication, the sender selects the appropriate DSC channel and enters the maritime mobile service identities (MMSI) number of the recipient. Kraus et al. [9] conducted preprocessing of raw AIS messages to convert them into trajectories. They sorted the messages based on the unique MMSI and timestamp, creating trajectories with a maximum duration of 2 hours. To ensure data accuracy, they checked for errors by calculating the speed between consecutive messages and comparing it to the reported speed. Messages with significant speed discrepancies or invalid information were removed. Additionally, to handle the large volume of messages, a compression algorithm was applied to each trajectory for computational efficiency. Contrarily, VHF TDMA involves the use of time-division multiplexing to divide a dedicated VHF channel into time slots, which can share the same channel without interfering with each other, making VHF TDMA an efficient and reliable protocol for transmitting data within nautical surroundings. Moreover, the system is connected to various navigation sensors like global positioning system to guarantee the ship's precise location. The GPS is essential for the proper functioning of an AIS device as it supplies the required vessel position and movement data to be shared with other vessels and shore stations.

#### *2.2.1.1. Maritime mobile service identities*

Each AIS device is assigned a unique 9-digit number, known as MMSI, as an essential element of the AIS system. MMSI is employed to ascertain individual vessels, and without it, AIS would not be able to function as intended. When an AIS device is installed on a nautical craft, that number is programmed into the device's memory. The MMSI is transmitted as part of an AIS message that is broadcast by the device itself, allowing other AIS-equipped vessels and shore stations to identify the transmitting vessel. In addition to vessel identification, MMSI is used for a range of other purposes, such as facilitating distress alerts and communicating with other vessels. For example, if a vessel is in distress, the crew uses that 9-digit number to send a distress alert that will be received by nearby objects and shore stations that have AIS. Table 2 highlights the allocation of numbers for several



countries. In conclusion, the significance of the MMSI within the system cannot be overstated, particularly in the context of employing machine learning techniques.

**Table 2 Several countries and their Maritime Identification Digit**

<b>COUNTRY</b>	<b>MARITIME IDENTIFICATION DIGIT</b>
BAHRAIN	408
CROATIA	238
FRANCE	226,227,228
GERMANY	211
ITALY	247
MALTA	229
PORTUGAL	204
SPAIN	224,225
USA	366,367,368,369

Source: made by student based on data gathered from Table of Maritime Identification Digits (no date).  
Available at: <https://www.itu.int/en/ITU-R/terrestrial/fmd/Pages/mid.aspx>. [13]

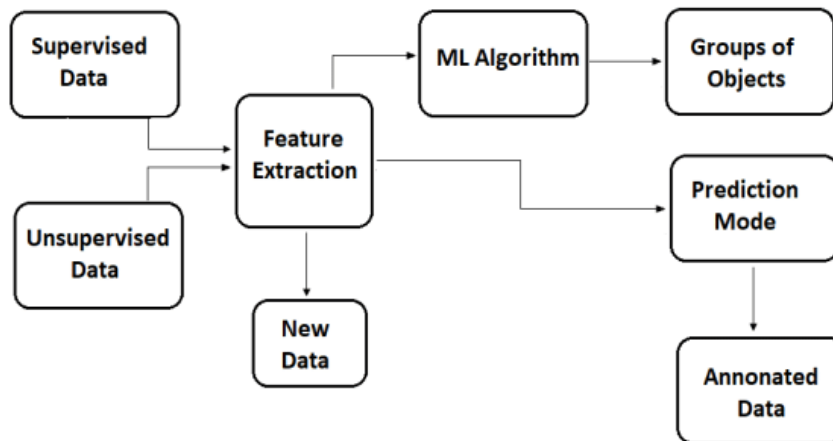
### **3. MACHINE LEARNING FUNDAMENTALS**

The first introduction of the term machine learning had indications in the 1950s when AI trailblazer Arthur Samuel constructed the initial self-teaching system for playing checkers. He observed that the system's performance improved as it played more games [14,15]. Then the human desire for progress, coupled with the advent of revolutionary ideas and technologies, has provided a solid foundation for the continued evolution of machine learning. Lately, there has been considerable focus on the rapidly evolving field of ML. It is a branch of AI that concentrates on creating computer algorithms capable of learning and improving on their own through experience, as said, without being specifically programmed to do so. As a result, this kind of approach is distinct from conventional programming, where a computer engineer creates a set of instructions to direct a computer on how to convert input data into the intended output. This process was simply described by researchers as if you were teaching machines to identify objects such as apples and pears. By presenting the machines with examples of different fruits, they gradually learn to label apples and pears without any assistance, given that they have been trained with suitable and precise examples [14].

#### **3.1. WORKING PRINCIPLE OF MACHINE LEARNING**

In order to maximize understanding of the remarkable capabilities and optimal utilization of this cutting-edge technology, it is imperative to grasp its operational mechanics. The whole process begins with inserting training data (in this case, AIS data) into the chosen algorithm. It needs to be noted that different types of data have an impact on the algorithm, which is mentioned before. Once new input data is incorporated into the machine learning algorithm, it is evaluated to ensure its proper functionality. The predictions and outcomes are subsequently cross-referenced to verify accuracy. In case the predictions and outcomes do not match, the algorithm undergoes multiple re-training sessions until the data achieves the desired outcome. That is how the ML algorithm is enabled to self-learn. Its ability to learn and improve from experience makes it a prominent resource for solving complex problems and automating tasks that would otherwise require human intervention

and are very frequent in the maritime industry. The general process of machine learning is depicted in Figure 4.



**Figure 4 Machine learning general process**

Source: Akanksha, E. et al. (2021) Review on Reinforcement Learning, Research Evolution and Scope of Application. Available at: <https://doi.org/10.1109/iccmc51019.2021.9418283>. [16]

### 3.2. MACHINE LEARNING CLASSIFICATION

Machine learning can be classified into three primary divisions, driven by the extensive range of data and diverse learning requirements [15]. This categorization is primarily determined by the nature of the learning process and the presence or absence of labeled data. Each category has its distinct characteristics and methodologies, offering tailored approaches to tackle various learning tasks and extract valuable insights from data. The division is as follows [17]:

1. Supervised learning
2. Unsupervised learning
3. Reinforcement learning.

Supervised learning encompasses training a model on data that has been labeled or annotated, where already in advance the correct answer is known. The goal is to teach this kind of model to make accurate predictions and generalized out of sample data. Furthermore, a substantial body of research confirms the effectiveness of semi-supervised learning

methodologies. This particular strategy harnesses the power of labeled and unlabeled data to enhance the precision of vessel trajectory classification, exemplifying its potential. Duan et al. [18] focused on utilizing the AIS data, which provides trajectory information of maritime vessels, for the task of vessel trajectory classification. Their objective was to tackle the challenge of scarce labeled data by proposing a semi-supervised deep learning approach. Furthermore, authors [18] utilized a blend of kinematic data (speed, heading, coordinates) and static information (vessel identification) extracted from AIS messages to derive vessel trajectories for classification objectives, as well as Ljunggren et al. [19]. To showcase the effectiveness of their approach, the researchers conducted extensive experiments utilizing a publicly available AIS dataset. Additionally, they analyzed the implications of their study within the realm of ocean management, emphasizing the potential advantages and practical applications of their findings in this field.

In contrast, unsupervised learning includes the training of a model on unlabeled data, where the correct outcome remains unknown [20]. The central objective is to unveil patterns and underlying structures within the data, leading to the grouping of similar entities together. A great example is the research in which authors focused on utilizing unsupervised learning and AIS data for vessel trajectory prediction. By employing trajectory clustering and classification techniques, Murray et al. [21] introduced a methodology for forecasting the future trajectory of a vessel by leveraging historical AIS data. They also proposed a dual linear autoencoder approach using historical AIS data to predict the future trajectory of a vessel for collision avoidance. Through the utilization of unsupervised learning and trajectory clustering, the method predicts the complete trajectory of the chosen vessel. It achieves this by estimating a latent distribution encompassing potential future trajectories. Through sampling from this distribution, numerous trajectories are produced, providing significant insights into the uncertainties linked to the anticipated positions of vessels. Schöller et al. [22] also used an encoder for their research, with a novel approach for anticipating the trajectories of marine vessels using AIS data. The method introduced a unique encoding technique that utilized AIS heatmaps generated through kernel density estimation (KDE). The study demonstrated that this method outperformed traditional approaches in marine trajectory prediction. By incorporating the AIS heatmap, the proposed method was able to accurately predict trajectories that could not be determined solely based on previous positions.

Finally, reinforcement learning (RL) comprises training a model through trial and error [20]. The model improves its performance by receiving rewards or punishments in the form of feedback based on its actions in given surroundings [14]. Reinforcement learning, despite its relatively lower adoption compared to other machine learning types, carries immense importance in the realm of maritime applications. A notable illustration of this can be observed in the research undertaken by Meyer et al. [23]. The researchers conducted a study focusing on reinforcement learning and artificial intelligence systems in the maritime domain. They trained an RL agent using deep reinforcement learning (DRL) techniques to create an autonomous vessel capable of avoiding collisions with other vessels while following a predetermined path. This kind of method will be analyzed later in the thesis. The agent relied on rangefinder sensors to measure distances to nearby obstacles such as shorelines and other vessels and made reactive control decisions based on this information. The researchers iteratively designed the reward function for the RL agent, incorporating additional features to improve its performance. Although increasing the complexity of the reward function was not desirable, it was necessary to achieve satisfactory results. The introduction of new reward function features immediately improved the agent's guidance behavior. One notable advantage of this approach is its ability to operate without prior knowledge of the vessel's internal dynamics, providing flexibility for customization of the performance measure. This flexibility makes it applicable to a wide range of vessel types with complex dynamics.

As a consequence of all of these referenced methods, ML can enhance its performance through acquired knowledge, similar to human learning. Nevertheless, unlike people, computers require a substantially larger number of learning examples to reach a comparable level of understanding.

## **4. MACHINE LEARNING BASED ON AIS DATA IN MARITIME**

In the forthcoming chapters, the aforementioned classifications will be elucidated in greater detail. Each classification encompasses several subtypes that will be examined based on their applications in the maritime domain. Each type holds its significance and has greatly contributed to the development of the infrastructure, as evident from the conclusions drawn by the authors who have employed these methodologies in their research articles.

### **4.1. SUPERVISED LEARNING**

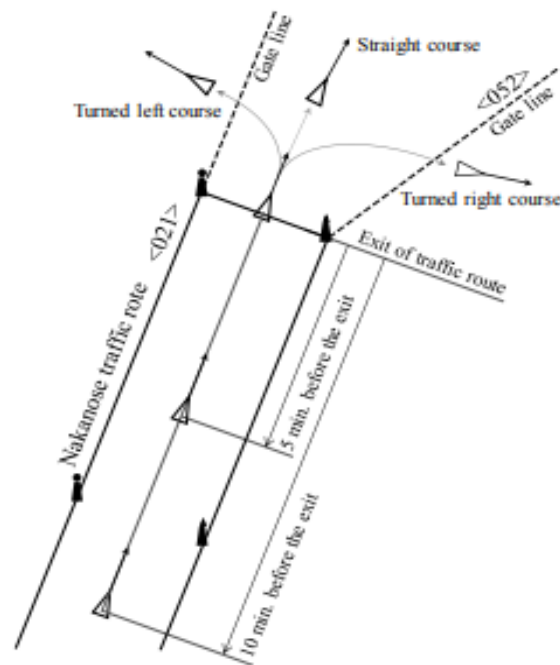
As previously stated in this thesis, there are abundant articles utilizing ML and AIS data in maritime infrastructure. By examining those articles, we can conclude that supervised learning stands out among these three groups. The fact that it includes the largest number of subgroups indicates its widespread adoption and versatility in various applications. This observation highlights the extensive application of supervised learning and its relevance in different domains. In recent years, several supervised learning techniques have gained significant traction. Particularly noteworthy, support vector machines (SVMs), artificial neural networks (ANNs), decision trees, memory-based learning, naive Bayes, random forests bagged trees, boosted trees, and boosted stumps have emerged as prominent methods in the field [24]. These methods have garnered attention for their effectiveness in various applications and their ability to handle complex datasets. One of many is energy-efficient shipping, highlighted in a research investigation carried out by Abebe et al. [20]. The study concluded that the developed models of supervised learning can accurately estimate the speed over ground (SOG) of ships, considering various factors such as load conditions, weather conditions, sailing distance/direction, and draughts, gathered from AIS. It was found that linear regression and polynomial models were not effective in predicting SOG due to their highly nonlinear nature over time [20]. In particular, the extra trees regressor (ETR), which is a bagging ensemble model, demonstrated high accuracy with low computational time in predicting SOG. As the remaining groups will be elaborated on in more detail due to their widespread applications, it is worth mentioning random forest trees (RFT) even though it is not as used as others. Snapir et al. [25] introduced an approach to categorize synthetic aperture radar (SAR) vessel detections into fishing and non-fishing classes. Their method

involved employing a random forest (RF) classifier with five input attributes: longitude, latitude, length, distance to the nearest shore, and time of measurement (AM or PM). The classifier underwent training and evaluation using labeled AIS data, and subsequently applied to S-1 (Satellite Sentinel-1) detections obtained from the SUMO detector. Drawing from their results, the researchers identified that the length, latitude, longitude, and distance to the nearest shore held the utmost importance as features for classification. During evaluations conducted on datasets featuring representative sample proportions (12.6% fishing, 87.4% non-fishing), the method achieved an overall classification accuracy of 91%. However, it should be noted that the precision for the fishing class yielded a modest score of 58%. The precision exhibited variation contingent on the location, particularly in regions where overlap between the two vessel categories and the non-fishing class prevailed. Extending the application of their method, they subsequently employed S-1 data from 2017 and observed a favorable agreement between their monthly fishing vessel count and the data obtained from global fishing watch (GFW) regarding the presence of fishing vessels in 2016. Although their vessel count may not provide an exact estimation of the precise number of fishing/non-fishing vessels, the outcomes indicated that their approach effectively captured the accurate temporal fluctuations in fishing activity. This methodology holds substantial value in identifying changes in fishing efforts, such as intensification or reduction, which play a pivotal role in addressing the global issue of overfishing.

#### **4.1.1. Support vector machine**

Support vector machine algorithms are potent supervised learning methods that portray training data as points in a multidimensional realm and establish stringent boundaries amidst clusters [24]. By mapping new input tuples into this space and measuring their distances from other categories, they are assigned to specific categories. SVM has been considered a state-of-the-art model in shallow machine learning approaches [26]. Despite having a small training set, SVM exhibits impressive generalization capabilities. In one study, Ichimura et al. [26] expected to achieve excellent performance with SVM, even when handling a limited number of data points per vessel. They aimed to address the issue of unreliable AIS signals by proposing a method to accurately classify the ship type using AIS data. They recognized that incorrect AIS settings could hinder the trustworthiness of the

signals. In order to support traffic management and enhance safety measures, the study focused on extracting distinctive features, referred to as "signature characteristics", from the AIS signals of ships. These features were then utilized in a machine learning model for ship type classification, enabling more precise identification of ships based on their AIS data. Furthermore, the objective of the study made by Nishizaki et al. [27] was to present a novel approach for optimizing the utilization of navigational big data within the advanced navigation system (ADNAVIS) to enhance the safety of ship navigation. The focus of this study revolves around the development of a prediction method that effectively determines the next course of a ship upon exiting a traffic route in Tokyo Bay. By employing SVM, the proposed method offers a reliable and accurate prediction mechanism, ensuring enhanced navigation safety in critical areas. Using AIS data, they focused on accurately understanding ship behaviors. AIS data, transmitted by each ship, consisted of static information (e.g., ship details) and dynamic information (e.g., position, speed, navigational status). Example of used learning data can be seen in Figure 5.



**Figure 5 Example of learning data for SVM**

Source: Nishizaki, C. et al. (2018) Development of Navigation Support System to Predict New Course of Ship. Available at: <https://doi.org/10.23919/wac.2018.8430436>. [27]



In order to train AI models, the researchers performed interpolation on the dynamic information of AIS data, ensuring that data points were available every second. The researchers performed test cases aimed at predicting whether a ship would anchor in the bay or proceed directly to the port, thereby classifying the associated AIS data into distinct categories. To assess the predictive capabilities concerning a ship's subsequent course, the authors conducted two specific test cases [27]. The outcomes demonstrated that the accuracy of predicting a ship's next course at the exit of the traffic route achieved an impressive rate of 87.9%. Furthermore, the accuracy for predictions made five minutes prior to the exit stands at 80.3%, while the accuracy for predictions made ten minutes prior to the exit is 77.0% [27]. However, for practical application, further improvements are required to enhance the prediction accuracy [27].

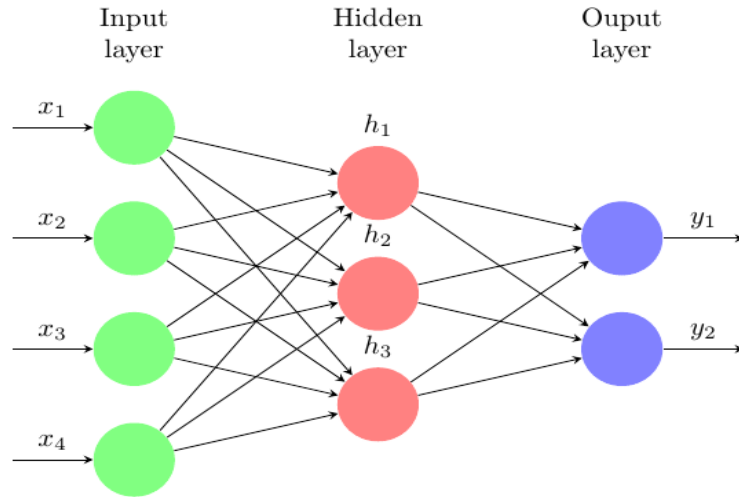
Many others used this technique with great results, such as Coleman et al. [28] for behavioral anomaly detection, and Ginoulhac et al. [29] for coastal radar target recognition by focusing on utilizing AIS data to devise a method for classifying targets based on their kinematic information. The researchers employed AIS data to acquire labeled trajectories, establishing a proof of concept that could potentially be applied to radar tracks in the future. Their proposed approach involved extracting statistical features from different temporal variables, including speed and acceleration. These extracted features were utilized as input for a gradient boosting classifier. The findings showcased that the method achieved an impressive classification accuracy of approximately 86% when dealing with seven distinct classes, comparable to the current state-of-the-art methods in the field. This suggests its potential applicability in radar systems for target classification based on trajectory information. Lang et al. [30] demonstrated the potential of leveraging AIS knowledge to improve SAR ship classification performance using one type of SVM classifier. They proposed an enhanced multiclass A-SVM with simpler naive geometric features (NGFs), conducting comprehensive experiments that showed significant improvements compared to traditional methods, especially with limited training samples. The proposed classifier's performance was further enhanced by selecting an optimized feature combination, and its effectiveness correlated weakly with the number of labeled samples. However, the method's limitation was its reliance on length and width as transferable features from AIS, suggesting the need for future research to incorporate additional features from SAR images. Even though Kim et al. [4] proposed the convolutional neural networks (CNN) method, which is a part of ANNs, they used SVM to compare gained results. This means that even though the

SVM method itself is not directly used, it still serves to improve and advance another utilized approach.

#### **4.1.2. Artificial neural networks**

Artificial neural networks are highly sophisticated machine learning tools inspired by the complex networks of neurons found in biological systems. They have gained prominence across various domains, offering valuable insights where the underlying functions are intricate and not fully understood. In the maritime industry, ANNs have been extensively applied to vessel activity identification, with a particular focus on fishing vessels. A notable study referenced by authors [24] delved into the identification of "métier," which refers to groups of vessels sharing similar fishing patterns, such as gear usage, fishing grounds, and target species, over extended periods. To tackle this challenge, the researchers employed ANNs, specifically utilizing a multilayer perceptron network (MPN), to accurately allocate fishing efforts at the individual trip level to specific métiers. Similarly, in other investigations, ANNs have been employed in conjunction with vessel monitoring system (VMS) data to differentiate between fishing and non-fishing points along vessel tracks or to infer fishing tactics based on daily landing data [24]. These applications exemplify the versatility and effectiveness of ANNs in the maritime domain, offering perspectives into vessel behavior and facilitating improved fisheries management.

Furthermore, ANNs have proven instrumental in diverse fields, including speech recognition [31] and target recognition, offering a comprehensive approach to complex data analysis through interconnected nodes arranged in input, hidden, and output layers [32]. These nodes are linked to weights that are determined using training data, enabling ANNs to approximate the connections between inputs and outputs, similar to conventional multivariate regression analysis [32]. A vivid depiction of such nodes is shown in Figure 6.



**Figure 6 Artificial neural network nodes**

Source: Cherukuri, H.P. et al. (2019) “Machining Chatter Prediction Using a Data Learning Model,” Journal of Manufacturing and Materials Processing, 3(2), p. 45. Available at: <https://doi.org/10.3390/jmmp3020045>. [33]

#### 4.1.3. Decision tree regressor

A regression method for supervised learning that does not rely on fixed parameters, under the name of decision tree Regressor (DTR), partitions features into a rectangular space and trains a simple model (tree) for each feature [34]. Its most common method is classification and regression trees (CART) in which the feature space is repeatedly partitioned into two regions until certain stopping criteria are met in this process [34]. To develop the model, for a given  $n$  number of dataset samples and  $d$  number of features,  $D\{(X_i, Y_i)\}$  ( $|D| = n, X_i \in R^d, Y_i \in R$ ) the feature space is assumed to be split into  $K$  – number of regions, called  $R_K$  and the prediction value of the model is obtained from the average number of observations [20]. The process of finding the optimal value is simple and can be calculated with formula (1), but it poses considerable difficulty to divide the region:

$$\hat{Y}_i = \text{ave}(y_i | x_i \in R_K) \quad (1)$$

In their approach [20], the researchers aimed to obtain the best-predicted values,  $\hat{Y}_i$ , by minimizing the least square error of the difference between the true values,  $y_i$ , and the predicted values. While calculating the optimal  $\hat{Y}_i$  values is relatively straightforward, the

challenge lies in determining the optimal splitting nodes within the region. To address this, they employed a greedy algorithm that iteratively determines the optimal splitting nodes until a stopping point is reached [20]. The stopping point is typically determined based on the hyperparameters and the complexity of the underlying problem. This iterative process allows for efficient and effective determination of the optimal splitting nodes, leading to improved prediction performance. In order to optimize the model, the proper hyperparameter values need to be selected. This is typically done through a process called grid searching, which involves evaluating all possible combinations of hyperparameters to find the best one. The hyperparameters that need to be specified include [20,35]:

- the maximum depth
- minimum number of samples needed to split an internal node
- minimum number of samples required at each leaf node
- the number of features to consider during the search for the optimal split.

However, the studies acknowledged certain limitations and challenges associated with AIS data. As described by Abebe et al. [20], data acquired for ML speed prediction for efficient shipping is AIS data as well as weather data obtained from 14 tankers and 62 cargo ships within 3 min intervals to estimate SOG. Static and dynamic information given by AIS is described in Paragraph 2 of this thesis, whose collection is more or less simple. The real problem one can encounter when using this method is weather data, which undoubtedly affects ship speed so it must be included. The most influential weather conditions are waves (causing the additional resistance), the viscosity of the water, and ocean currents (affecting ship heading and speed [20]). The chosen features, among those mentioned, are gross tonnage, wind, temperature, salinity, ship dimensions, dead weight, and course over ground (COG).

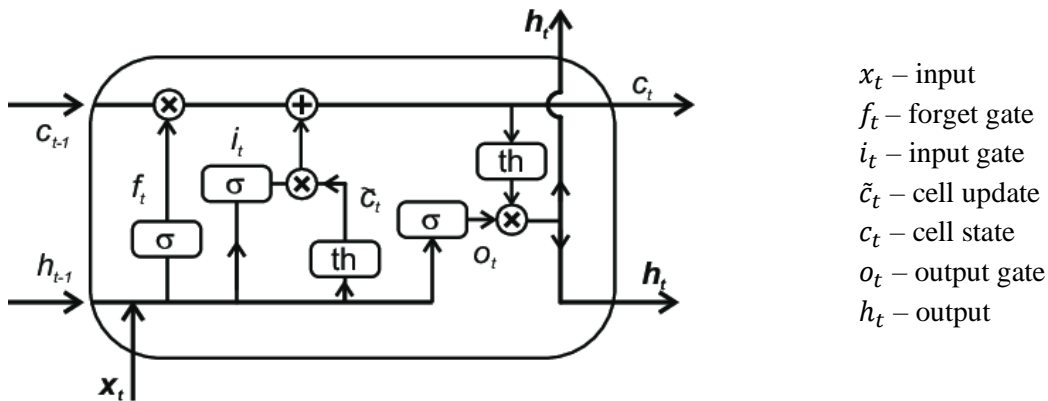
#### **4.1.4. Memory-based learning**

In the field of machine learning, memory-based learning, also known as instance-based learning [36], is a group of learning algorithms that make predictions or decisions by comparing new instances with previously observed ones stored in memory [37]. Unlike

traditional learning algorithms that perform explicit generalization, memory-based learning directly constructs hypotheses based on the training examples themselves. This flexibility allows the hypothesis complexity to expand alongside the data, enabling the model to adapt to previously unseen instances. One key advantage of memory-based learning is its ability to dynamically update the model by incorporating new instances or discarding old ones [38]. This "lazy" approach defers computation until a new instance is encountered, offering computational efficiency and the potential for effective adaptation to changing data. By leveraging the stored knowledge of past instances, instance-based learners can handle complex and evolving patterns in a wide range of applications. This was proven to be most effective in anomaly detection in AIS devices. The availability and utilization of AIS data posed both advantages and challenges in the studies explained in used articles. AIS messages, recorded via satellites with a sampling frequency of approximately one hour per vessel, served as the foundation for the majority of the researchers.

Another approach is by utilizing location-based and data-driven approaches [28]. The primary objective was to present a behavioral model for anomaly detection in AIS using machine learning techniques [28]. The study that dealt with the development of that method aims to address the challenge of identifying anomalous behavior in AIS data, which can also be useful for maritime security and safety purposes. Also, it is a great example of the mentioned disadvantage of the device – falsifying AIS data, because the automatic identification system lacks encryption and message authentication. This means that any message can be broadcast by any vessel. Considering there are over 400 000 installments of the device, with the fact that any message can be sent as if it is from another vessel [28], this defect needs to be taken seriously. A machine learning-based anomaly detection scheme proved to be an effective solution. The aforementioned two types of data were used. Location-based approaches analyzed the current location and vessel trajectory. For this dataset, they collected ideas and opinions from other experts and researchers to arrive at the best solution. The first mentioned was Liang et al. [39] who proposed a two-step supervised learning approach using long short-term memory (LSTM) to reconstruct a ship's trajectory in the event of lost AIS location data. This can occur when the AIS transceiver is unable to transmit due to a shortage of available time slots, resulting in resource starvation and also due to unfavorable weather conditions. The AIS system plays a crucial role in ship tracking and providing position reports, resulting in a wealth of data that can be applied in different maritime applications, including enhancing security and improving transportation

efficiency. LSTM overcomes this problem by incorporating a series of ‘memory cells’ that can selectively retain or forget information over time. The memory cell of the model incorporates multiple gates that regulate the ingress and egress of data, as shown in Figure 7. These gates can be thought of as filters that allow the cell to process information based on its relevance to the current task.



**Figure 7 LSTM cell with its internal structure**

Source: Hrnjica, B. and Bonacci, O. (2019) “Lake Level Prediction using Feed Forward and Recurrent Neural Networks,” *Water Resources Management*, 33(7), pp. 2471–2484. Available at: <https://doi.org/10.1007/s11269-019-02255-2>. [40]

Another usage was for Data-Driven Trajectory Quality Improvement for Promoting Intelligent Vessel Traffic Services in 6G-Enabled Maritime IoT Systems [41]. Vessel trajectories are a type of time-series data that consists of spatial points with timestamps. It is crucial to acknowledge that the sequential timestamps within a singular vessel trajectory are interdependent owing to the vessel's consistent stability during navigation [41]. In order to handle these trajectories and overcome the obstacles of long-term time-series data, the study explored the use of recurrent neural networks (RNNs) and specifically focused on the application of long short-term memory networks. LSTM was also introduced by Hochreiter and Schmidhuber as a solution to capture long-term dependencies in recurrent networks. The architecture of the LSTM memory block was employed to effectively learn and reconstruct vessel trajectories, overcoming issues such as gradient exploding or vanishing that simple RNNs often encounter when dealing with long-term data. On the other hand, Tang et al. [42] introduced a long short-term memory model specifically designed for probabilistic ship

position prediction. In their research, the authors focused on the prediction of vessel trajectories in port regions for autonomous navigation systems. They proposed a model using LSTM and sequence prediction techniques to accurately forecast vessel trajectories. The model was trained and tested using real-world AIS data from the port of Tianjin, China. The experimental results confirmed the model's ability to predict ship trajectories accurately, showcasing its potential for enhancing autonomous navigation systems. Another study, similar to the previous one, trained the model using AIS data and aimed to estimate the positional density at a specified point in the future [43]. By predicting the variance, mean, and covariance of a bivariate Gaussian distribution [43], the LSTM model provided insights into the likelihood of ship positions. Nevertheless, a drawback of this method was its inability to generate a comprehensive trajectory as it could solely anticipate the future position for a solitary time step. Moreover, it necessitated the predicted point to align with the timeframe of the input data, implying that a 10-minute input data was mandatory to forecast a position precisely 10 minutes in advance [22]. Liu et al. [41], proposed a two-phase vessel trajectory reconstruction framework that utilizes the bidirectional long short-term memory (BLSTM) model. This particular model was selected for its capability to proficiently restore deteriorated timestamped points within vessel trajectories by capturing intricate non-linear patterns and executing reliable pattern extraction. By leveraging the self-similarities present in vessel trajectories, the BLSTM model aimed to improve the quality of the reconstructed trajectories. The effectiveness of the proposed framework was assessed by conducting experiments on both synthetic and real-world trajectory datasets, demonstrating its ability to enhance the quality of vessel trajectories. Furthermore, by accurately detecting outliers and pinpointing their positions, the BLSTM model was utilized to capture intricate sequential patterns from extensive historical vessel trajectories, leveraging a data-centric approach.

## **4.2. UNSUPERVISED LEARNING**

Murray used unsupervised learning to enhance trajectory clustering and classification, incorporating a set of historical AIS trajectories clustered together [44]. They aimed to enhance the ship navigators' high-level situation awareness by emulating the cognitive process of human situation awareness using machine learning techniques.

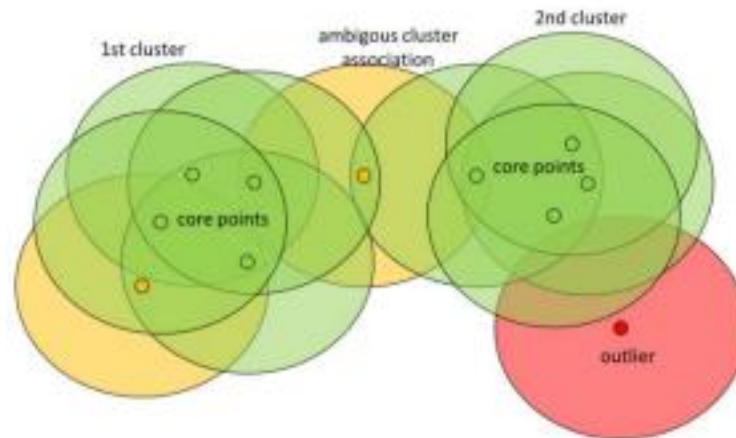
Historical automatic identification system data was leveraged to develop categorization functions and generate meaningful representations of ship behavior patterns. The researchers also developed methods for pattern matching and predicting future ship behavior within specific clusters, intending to facilitate proactive collision avoidance actions. Two approaches, eigendecomposition-based and deep learning-based, were utilized successfully to discover behavior clusters, classify trajectory segments, and predict future ship behavior. The deep learning models, although requiring more training time, provided almost instantaneous predictions once trained. Conversely, this kind of method can be used for anomaly detection just like supervised learning but with the difference of supervised learning methods relying on labeled training data to construct the learning model [45]. Among the different types, unsupervised learning methods stand out, where unlabeled training data is employed to build the learning model through clustering algorithms [45]. These distinct approaches offer unique advantages and are applied in various scenarios based on the availability of labeled or unlabeled data, contributing to the effectiveness and versatility of machine learning. Through an in-depth analysis of the selected articles, we can draw the following conclusion: Two of the most popular subgroups of unsupervised learning used in maritime are density-based spatial clustering of applications with noise (DBSCAN) and trajectory representation, exploration, and analysis for discovery (TREAD).

#### 4.2.1. DBSCAN

Density-based spatial clustering of applications with noise is an exceptionally efficient clustering algorithm developed on the principle that clusters demonstrate dense areas in space, with gaps separating them in regions of lower density. This algorithm effectively identifies densely packed data points and groups them as a single cluster. This approach allows for the detection of clusters of varying shapes and sizes, making it particularly useful in scenarios where the data may have irregular or non-linear structures. This method was used for the detection of search and rescue (SAR) vessel patterns [24]. Also, in an article about enhancing maritime surveillance [46], the authors gave an excellent explanation of how exactly this method is used. They concluded that DBSCAN relies on two parameters:  $\epsilon$  and  $min_{pts}$ . Points with more than  $min_{pts}$  within their  $\epsilon$ -neighborhood are designated as core points. The algorithm then proceeds to determine the connection



components on the neighborhood graph. The remaining points are connected through a connectivity component if they fall within an  $\epsilon$ -neighborhood of core points [46]. Points that do not meet these criteria are considered outliers, as shown in Figure 8. By employing the DBSCAN algorithm and these sequential steps, it becomes possible to effectively cluster routes and identify points of interest within the data.



**Figure 8 Schema of DBSCAN**

Source: Dästner, K. et al. (2018) Machine Learning Techniques for Enhancing Maritime Surveillance Based on GMTI Radar and AIS. Available at: <https://doi.org/10.23919/irs.2018.8447961>. [46]

Different extensions of the DBSCAN algorithm exist to address specific clustering requirements. Instances of such techniques encompass the utilization of ordering points to ascertain the clustering structure (OPTICS), hierarchical density-based spatial clustering of applications with noise (HDBSCAN), and subspace clustering (SUBCLU), which focus on hierarchical clustering or subspace clustering [46]. These improvements are important as they allow for a certain level of parallelization in the clustering process. Once points of interest have been identified, the subsequent stage in the analysis involves clustering the routes connecting them. There are various techniques available for clustering routes, each requiring specific metrics to quantify the level of similarity. One widely used method is dynamic time warping (DTW). DTW compares two series, P and Q, which represent the plots of separate routes. The algorithm actively seeks a wrapping path, which is a sequence of points that initiates from the starting point of P and Q and terminates at their respective

endpoints. This wrap path ensures a valid transformation between the routes, with monotonically increasing components. The DTW distance, calculated as the minimum sum of distances between corresponding points along the wrapping path, can be efficiently computed using a recursive formula and backtracking. Alternatively, other approaches such as Hausdorff Frechet, one-way, or symmetrised-segment-path-distance can be utilized in place of DTW [46]. The approach utilized by Dästner [46] involves dividing AIS plots into partitions based on similar course vectors. Plots sharing the same course domain are grouped, and a DBSCAN algorithm is utilized to geo-cluster segments exhibiting similar courses. While representing the identified route clusters, it is common practice to employ an averaging route. However, finding an exact solution to this problem poses challenges, as non-linear alignment and average filter (NLAAF) methods are solely applicable to two sequences, while approximate methods are employed for more than two sequences. These varied clustering approaches offer flexibility in analyzing and representing the data according to the specific requirements of the given problem.

Nonetheless, its widespread application is evident in various domains, as shown by Martinčič et al. [47]. In this context, DBSCAN was utilized similarly as previously described. By leveraging validated AIS data, DBSCAN was employed to cluster and analyze vessel and port efficiency metrics. This allowed for the identification of patterns and anomalies in the data, enabling informed decision-making and optimization of vessel and port operations. The versatility of DBSCAN makes it a valuable tool in diverse applications where spatial clustering and analysis are required.

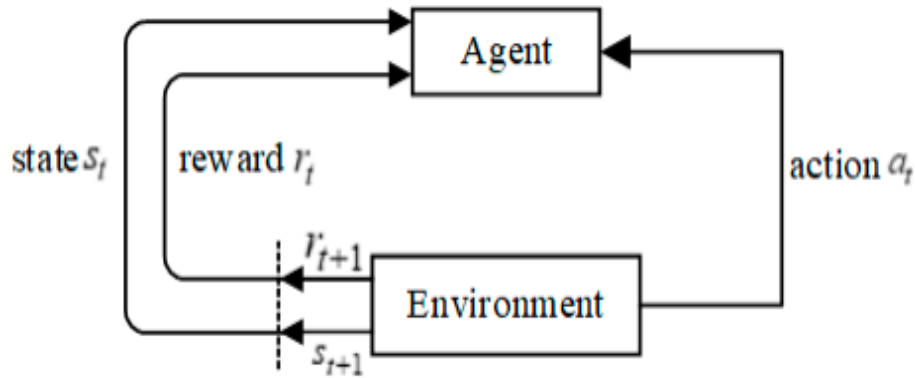
#### **4.2.2. TREAD**

For the second group, it is important to mention trajectory representation, exploration, and analysis for discovery. TREAD is an unsupervised learning approach specifically designed for maritime trajectory data analysis. It aims to uncover patterns and extract valuable insights from large-scale vessel trajectory datasets. TREAD employs a two-step process: trajectory representation and trajectory exploration. In the trajectory representation step, TREAD transforms raw trajectory data into a meaningful and compact representation. It accomplishes this by employing dimensionality reduction methods, such as principal component analysis (PCA) or dynamic time warping, similar to the approach

used in DBSCAN, to capture the essential characteristics of the trajectories while reducing computational complexity. Once the trajectories are represented, the trajectory exploration step begins. TREAD employs various clustering algorithms, based on density-based spatial clustering of applications with noise, to group similar trajectories based on their spatial and temporal properties [48]. This allows for the identification of significant maritime patterns, such as common routes, ports of call, or areas of high traffic density. Furthermore, TREAD provides visualization tools to analyze and interpret the discovered patterns. It allows users to interactively explore the clusters and individual trajectories, enabling the identification of outliers, anomalies, and other interesting phenomena. Overall, TREAD offers a comprehensive framework for unsupervised learning in maritime trajectory analysis. It aids in uncovering actionable intelligence, understanding vessel behavior, optimizing maritime operations, and supporting decision-making processes in the maritime domain. [45]

### **4.3. REINFORCEMENT LEARNING**

Reinforcement learning, another branch of machine learning, focuses on maximizing cumulative rewards to learn optimal strategies for achieving goals. RL allows agents to directly interact with the environment and evaluate action values, reducing the reliance on extensive training data. The agent observes the environment state, takes actions based on a policy, receives feedback rewards, and sequentially transitions to new states to maximize long-term rewards [49]. Reinforcement learning was introduced by Guo et al. [49] for intelligent route planning of unmanned ships in unfamiliar environments. Although RL has multiple subdivisions depending on the field of application, when it comes to its utilization in maritime settings, DRL is by far the most significant. Other types are not as widely adopted or used effectively. Therefore, we will focus exclusively on DRL in this thesis which is covered in the following subsection. The visual representation of RL working principal is shown in Figure 9.



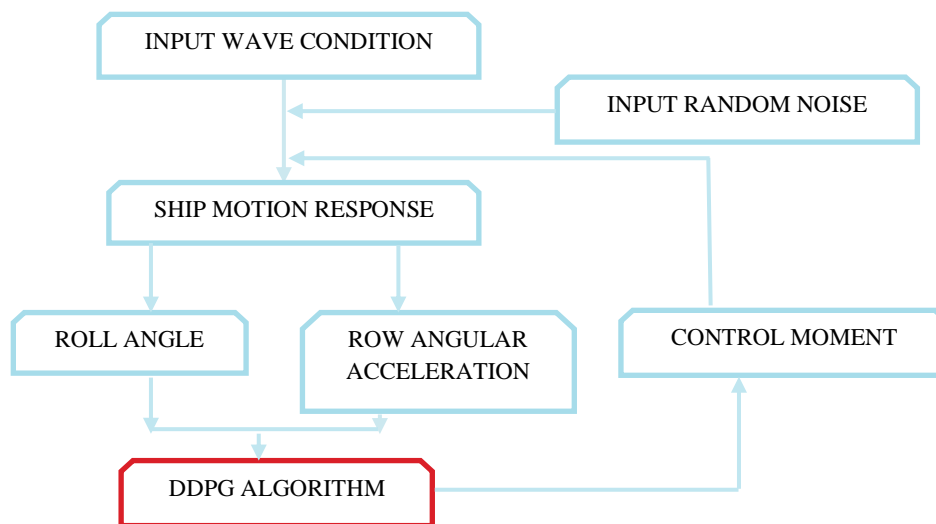
**Figure 9 The Principle of Reinforcement learning**

Source: Guo, S. et al. (2020) “An Autonomous Path Planning Model for Unmanned Ships Based on Deep Reinforcement Learning,” *Sensors*, 20(2), p. 426. Available at: <https://doi.org/10.3390/s20020426>. [49]

#### 4.3.1. Deep reinforcement learning

DRL combines DL (Deep Learning) and reinforcement learning, leveraging DL's perception capabilities and RL's decision-making abilities. It has demonstrated notable advancements in continuous motion control and effectively tackles the limitations associated with conventional unmanned ships in terms of path planning [49]. Through the progressions made in technologies such as artificial intelligence, big data, and cloud computing, ships have undergone a notable enhancement in their intelligence. As a result, the future trend of unmanned ships has arisen. Unmanned surface vehicles (USVs) are small, intelligent vessels used for specific tasks such as maritime rescue, surveys, and hazardous goods monitoring [8]. Those kinds of ships were the main focus of research by Shi et al. [50] where they established a theoretical foundation for developing intelligent collision avoidance systems for USVs. The researchers introduced a novel approach called the double gated recurrent unit-recurrent neural network (GRU-RNN), which effectively learned the decision-making process for collision avoidance based on successful case data. By leveraging two GRU-RNN networks, they accomplished successful processing of both their own ship trajectory data and target ship trajectory data. This method demonstrated the ability to handle trajectories of varying lengths, offering enhanced flexibility and generating more accurate collision-avoidance decisions. The proposed algorithm involved extensive training with big data and

exhibited wide-ranging applications and versatility, enabling the generation of timely strategies and responses for improved collision avoidance security. The researchers specifically employed the deep deterministic policy gradient (DDPG) algorithm, a type of DRL algorithm suitable for continuous action spaces, for an anti-roll control system on a ship, as shown in Figure 10 [51]. DDPG utilizes a deep network structure and policy gradient algorithm to select actions based on learned strategies and outputs specific action values [52]. DDPG has been successfully applied in the transportation sector, particularly in driverless cars and unmanned ships [51]. In order to ensure safety, these vehicles often necessitate continuous motion control, strict adherence to traffic rules, and the expertise of human operators with operational experience [53]. DDPG's self-learning and function-fitting abilities make it a promising approach in the maritime domain, offering the potential for further expansion and application.



**Figure 10 Iterative training process of the ship anti-roll control with the DDPG algorithm**

Source: made by a student based on Figure 3 from Lin, J. et al. (2022) "Intelligent ship anti-rolling control system based on a deep deterministic policy gradient algorithm and the Magnus effect," *Physics of Fluids*, 34(5), p. 057102. Available at: <https://doi.org/10.1063/5.0089697>. [51]

Based on the conducted research and analysis, it can be concluded that deep reinforcement learning proves to be a remarkably effective approach in tackling continuous control problems, thereby establishing its popularity as a preferred choice in diverse

domains, including path planning. In the context of unmanned ships operating in unknown environments, an autonomous path-planning model based on DRL has been developed [49]. This model utilizes the mentioned deep deterministic policy gradient algorithm and acquires optimal action strategies by continuously interacting with the environment and leveraging historical experience data [53]. To guarantee the integrity and reliability of the intended trajectory, regulations pertaining to navigation and encounter scenarios are converted into zones with navigational restrictions. The training process employs vessel data extracted from the automatic identification system. Moreover, an improved version of deep reinforcement learning is achieved by combining the deep deterministic policy gradient algorithm with the concept of artificial potential fields. This fusion results in a path-planning model that seamlessly integrates with an electronic chart platform for thorough testing and evaluation. Comparative experimental evaluations unequivocally demonstrate that the upgraded model empowers autonomous path planning with the commendable speed of convergence and unwavering stability. These compelling findings underscore the substantial prospects offered by DRL-based methodologies in advancing the efficiency and safety of unmanned maritime operations. To conclude, the researchers [49] explored the concept of DRL and its application in the field of unmanned ships. Deep learning has been utilized to extract accurate feature samples from input data, leading to significant advancements in object recognition, detection, and voice recognition. DL also makes use of deep neural networks to automatically learn high-dimensional data features through iterative processes such as backpropagation and training [53].

## 5. AIS TRAINING DATA FOR MACHINE LEARNING

Training data refers to a collection of instances used to train a machine learning algorithm. The main purpose of this sort of data is to enable the algorithm to learn patterns and correlations from the data and leverage the acquired knowledge to formulate predictions. Important considerations when selecting AIS training data for ML are [20]:

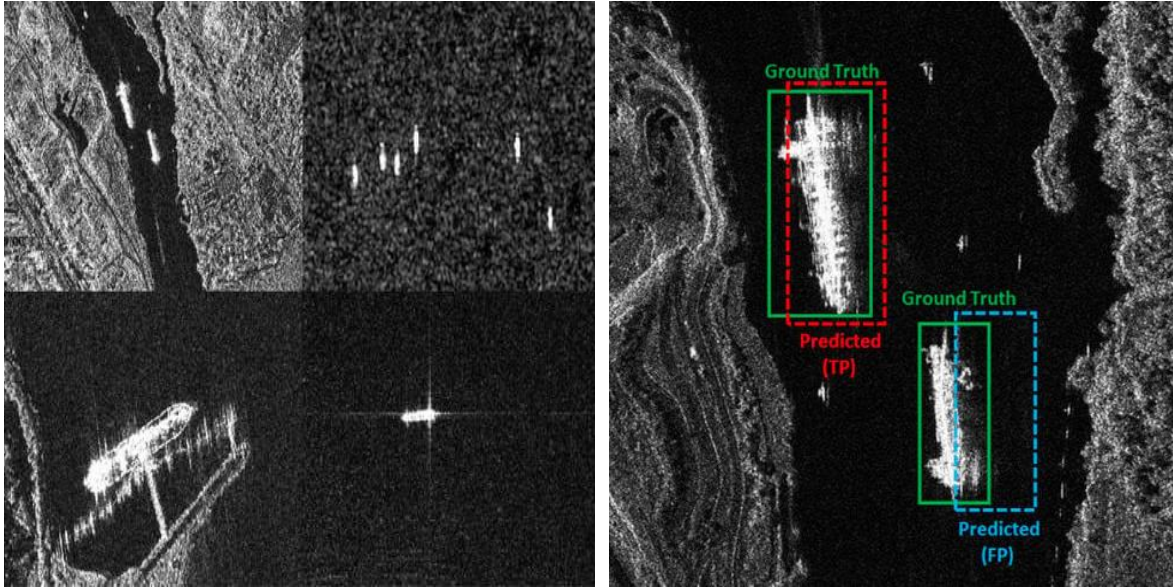
1. Data quality: AIS data can contain noises and defects. It is important to clean and preprocess the data before using it for ML.
2. Data volume: ML techniques entail a large amount of data for making accurate predictions and patterns.
3. Data diversity: Collected data should be diverse enough to cover a range of vessel types, sizes and behaviors.
4. Data privacy: AIS data can contain sensitive information such as vessel cargo and routes.
5. Data labeling: It refers to relevant information.

AIS training data should be of high quality, diverse, voluminous, appropriately labeled, and privacy compliant. The study conducted by Kim et al. [54] places significant emphasis on the specific aspect mentioned. The researchers specifically address two challenges encountered in implementing fully connected neural networks (NN) for speed regime modeling: data quality and accessibility. They discuss these challenges in detail and propose techniques to mitigate them. Understanding the advantages and limitations of fully connected NN is crucial for harnessing its predictive capabilities to enhance safety, emergency response, and transport planning in Arctic voyages. Without it, the algorithm would not have sufficient information to learn from and could produce inaccurate or biased results. As with everything else, precise and proper preparation results in maximum beneficial output, and the same goes for ML.

In previous studies, obtaining training data for ship analysis through visual interpretation has proven challenging. Most studies utilized SAR images in combination with convolutional neural networks containing limited coverage and a small number of vessels, which turned out to be the most effective way [55]. To address this, researchers started extracting training data by interpolating the AIS positions to precisely match the

ship's position at the time of image acquisition. Preprocessing the training data can also enhance the efficiency and relevance of the machine learning technique while addressing the issue of blind learning. Even though this data is necessary for any ML methods, the most attention was garnered in articles that focused on ship detection [56], convolutional neural network-based gear type identification [28], ship classification [39], trajectory predictions [57], and for classifying regional ship collision risk levels [58]. Different approaches for collecting training data are employed depending on their suitability for the specific method being used. In the following subsections, an analysis of methodologies used for analyzing that data will be described, drawing insights from relevant articles where they have been successfully applied. When considering these methods, it is important to recognize that although they all utilize training data with a focus on dealing with data problems, they diverge in terms of their underlying algorithms, objectives, and specific applications within the maritime domain. It's important to choose the most appropriate method based on the specific problem and requirements at hand and to be able to use additional help like SAR images, heatmaps, and others. Figure 11 shows samples of SAR images from Chang et al. [59] where they were used for evaluating the performance of the YOLOv2 deep network architecture for vessel detection. Their experiments showed that YOLOv2 outperformed existing technologies in terms of accuracy and real-time performance, especially in complex scenarios. They compared it to faster region-based convolutional neural networks (R-CNN), which had lower accuracy in specific areas. YOLOv2 demonstrated faster detection speed and better accuracy than recent methods. However, considering the need for real-time systems, the authors introduced YOLOv2-reduced, which achieved even faster detection times without compromising accuracy. The proposed method showed significant improvement over YOLOv2 on a specific GPU, offering faster detection by 2.5 times while maintaining competent accuracy on the SAR ship detection dataset (SSDD).





**Figure 11 Sample images from the SAR ship detection**

Source: Chang, Y.-L. et al. (2019) "Ship Detection Based on YOLOv2 for SAR Imagery," *Remote Sensing*, 11(7), p. 786. Available at: <https://doi.org/10.3390/rs11070786>. [59]

## 5.1. CNN METHOD

Convolutional neural networks are powerful mathematical models used for various applications [60,61], including target recognition, image classification, and ship classification [62], but most importantly, for processing AIS training data. Salem et al. [62] introduced an enhanced model architecture for ship classification in inland waterways. Their approach incorporated transfer learning and ensemble learning techniques, leveraging optimized convolutional neural network models. This methodology aimed to improve the accuracy and performance of ship classification in inland waterway scenarios. The model achieved a success rate of 98.38% and outperformed other algorithms. The system was trained using the "Game of Deep Learning" dataset, consisting of cargo, military, carrier, cruise, and tanker ships. The best model architecture was improved through hyperparameter optimization using the particle swarm optimization (PSO) algorithm [63], and a new classification block was added for better performance. The proposed method was validated on the MARVEL dataset with an accuracy of 96.36% and demonstrated superior performance compared to existing algorithms. Future work will focus on improving the method for different weather conditions and noisy images to meet the requirements of a

modern vessel classification system. As already said, ANNs consist of interconnected nodes organized in layers, where weights are determined through the utilization of training data to correspond the given inputs with desired outputs [58]. CNNs, a more advanced version of ANNs, excel in complex classification tasks, particularly image recognition. Images are represented as 3-dimensional volumes in CNNs, where different layers perform neuron activations to interpret and classify images. CNNs have demonstrated impressive performance compared to other deep learning architectures, effectively addressing challenges such as large weight and gradient divergence. In the maritime domain, CNNs are also employed to interpret images specifically for assessing ship-ship collision risk [32]. In this case, AIS data fields are used to construct inputs for CNN collision risk analysis, as shown in Table 3.

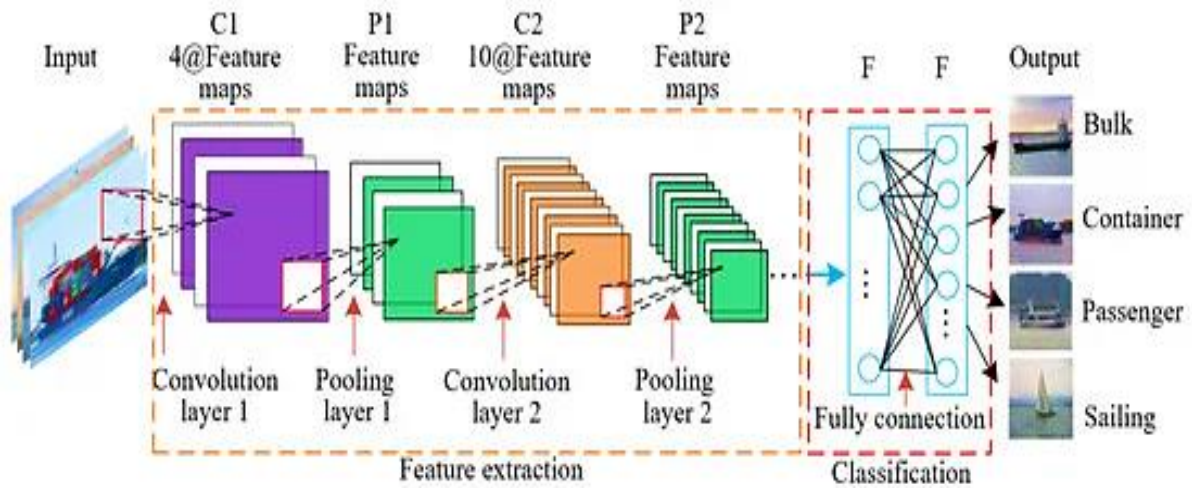
**Table 3 AIS data fields applied in constructing inputs for CNN collision risk analysis mode**

<b>DATA FIELD</b>	<b>UNIT</b>	<b>EXPLANATION</b>
MMSI number	[ = ]	9-digit code for identifying a vessel
Timestamp	[ S ]	Time at which the message is recorded
Position	[ = ]	Longitude and latitude of the transmitted message
Ship type	[ = ]	2-digit code identifying the vessel type
Ship length	[ m ]	Ship dimension from bow to stern
Ship speed	[ kn ]	Speed over ground
Ship course	[ ° ]	Course over speed

Source: made by student based on Table 2. from Zhang, W. 2020., ‘Towards a Convolutional Neural Network model for classifying regional ship collision risk levels for waterway risk analysis, Reliability Engineering and System Safety’ [32]

The core method of CNN involves three aspects: local receptive field, weight sharing, and pooling [32]. The local receptive field connects hidden layers to specific parts of an image, reducing dimensionality and computational complexity. Weight sharing ensures identical weights within the same convolution kernel, reducing training parameters and enabling image padding. Pooling layers summarize neighboring neuron outputs, compressing feature maps and extracting main image features. While training data significantly affects CNN performance, previous studies have primarily focused on enhancing the model's architecture. In ship detection, machine learning techniques like ANN, support vector machine, and CNN has gained prominence. Many studies have aimed to improve CNN parameters for high performance and efficiency. One of them was the study by Do Nascimento et al. [64], in which an application using convolutional neural network was developed for image recognition in the field of maritime safety. The goal was to identify the type of ship and assess the risk of accidents using images of vessels and plimsoll disks. Six different CNN architectures were evaluated, including variations in the number of filters, convolutional layers, and the use of transfer learning with the VGG-16 network (CNN that is 16 layers deep) and fine-tuning. The results showed promising outcomes, indicating the potential of CNN as a method for accurately identifying vessel images and assessing accident risks based on plimsoll disk analysis. This research highlights the application of AI and deep learning techniques to improve maritime safety.

On the other hand, the main objective of Song et al. [55] was to verify the enhanced detection capabilities offered by automatic identification system-assisted training data in contrast to manually curated training data. Through the utilization of a conventional convolutional neural network-based object detector and a comprehensive evaluation of both datasets, the study accomplished this objective [55]. Even though this process can be costly and laborious because of the large number of potential combinations, the CNN method is also employed to construct potential ML models for predicting ship velocity [65] and for ship classification [66]. Example of CNN structure for ship classification is shown in Figure 12.



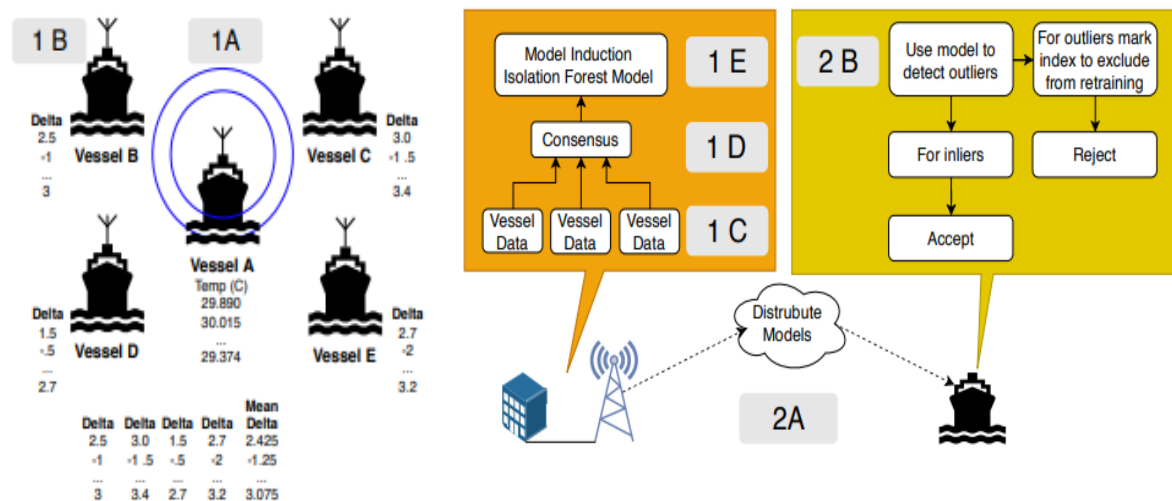
**Figure 12 Typical convolutional neural network structure for ship classification**

Source: Ren, Y.-M. et al. (2019) "Multi-Feature Fusion with Convolutional Neural Network for Ship Classification in Optical Images," *Applied Sciences*, 9(20), p. 4209. Available at: <https://doi.org/10.3390/app9204209>. [66]

Kang et al. [65] investigated to explore the potential of utilizing SAR imagery and a convolutional neural network to automatically estimate ship velocity. They discovered a proportional relationship between the azimuth offset, which represents the difference in direction between ships and wakes, and the Doppler shift effect observed in SAR. This effect is indicative of the radial velocity of a moving target, such as a ship. Based on these findings, the authors proposed a method that utilizes a CNN to detect ship wakes from TanDEM-X data, enabling the automated estimation of ship velocity. Similar to that, but still in a way that differs to some extent, Ren et al. [66] introduced a novel ship classification approach using a fusion of multi-features with a CNN method. They tested the proposed method on visible and infrared spectrums (VAIS) dataset and created their ship dataset from various sources. The improved CNN model had shallow layers and fewer parameters, reducing computational complexity. Handcrafted features, histogram of oriented gradients (HOG) and local binary patterns (LBP), were used to compensate for CNN's limitations in capturing local features, resulting in a robust ship classification model. Experimental results showed high classification accuracies and outperformance compared to other methods. The authors acknowledged the need for further improvements, including expanding the dataset and exploring transfer learning techniques.

## 5.2. ANOMALIES OF AIS DATA

As for the data-driven approaches, Balduzzi et al. [58] described and categorized various attacks. The first type refers to implementation-specific vulnerabilities in software, while the second type pertains to protocol-specific vulnerabilities in AIS radio transponders. At the software layer, there existed the potential to spoof another ship's maritime mobile service identity and impersonate that particular vessel [39]. For radio attacks, which are the second type, an individual can manipulate the message transmitted by a physical vessel, enabling them to change the real-time location of the vessels in transit. The contribution of this work was mainly the development of a machine learning method for the detection of anomalies at vessels. Their approach was presented through a set of consecutive phases, as depicted in Figure 13 [28].



**Figure 13 Behavioral model anomaly detection**

Source: Coleman, J., Kandah, F. and Huber, B. (2020) Behavioral Model Anomaly Detection in Automatic Identification Systems (AIS), 2020 10th Annual Computing and Communication Workshop and Conference (CCWC). Available at: <https://doi.org/10.1109/ccwc47524.2020.9031248>. [19]

The maritime vessel communication protocol AIS is vulnerable to manipulation due to a lack of encryption and authentication. Behavior modeling using ML models can detect abnormal vessel behavior and enable appropriate mitigation strategies in real-time, thereby identifying additional threats. Even vessels lacking the necessary equipment can benefit

from implementing behavior modeling. Some of the studies aimed to exemplify the application of ML in modeling vessel behavior by utilizing temperature sensor data transmitted through AIS. As part of future research, the integration of supplementary sensors [67] could be explored to enhance the precision and reliability of vessel models. Additionally, ML models have the potential to impact trust networks within the maritime sector, particularly in vessel-to-vessel communication. By facilitating the maintenance or elevation of trust ratings for reliable vessels and the reduction of trust ratings for abnormal vessels, ML can play a significant role in shaping robust trust networks in the maritime industry [28].

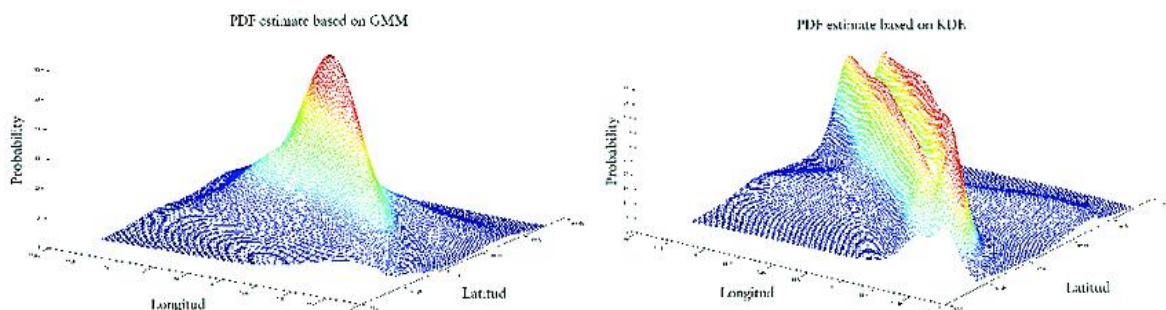
Riverio et al. [68] conducted tests to evaluate their approach to anomaly detection. They collaborated with Saab Microwave Systems, which provided synthetic test data with intentionally hidden anomalies. Using their application, the authors were able to identify three different types of anomalies within the test data:

1. Abnormal speed values: One of the vessels exhibited speed values that deviated significantly from the patterns observed in the training data.
2. Fishing boat behavior: An HF vessel displayed behavior that was typical of a fishing boat, which was different from its expected behavior.
3. Coastal approach: Another HF vessel demonstrated unprecedented behavior by approaching the coast. This behavior had not been observed in the analyzed area before, and the heading values of the vessel were considered abnormal.

However, detecting the vessel approaching the coastline required significantly reducing the threshold value, which led to a high number of false positives in the results. The preliminary results of their research showed that their system was capable of detecting anomalies related to single attributes, indicating satisfactory outcomes. Nonetheless, it's important to note that the experiments conducted solely evaluated the performance of the system in terms of anomaly detection. They have not conducted user tests involving domain experts at this stage, but they plan to carry out such evaluations in the future. Additionally, the authors highlighted that while their application was designed for maritime anomaly detection, it has broader applications in other domains, such as network intrusion detection and other areas where anomaly detection is necessary.

### 5.2.1. Methods for AIS anomalies detection

Gaussian mixture model (GMM) and kernel density estimation are both utilized for ship anomaly detection [45]. Some researchers proposed a fast and efficient algorithm based on KDE for predicting ship location and velocity [45], which can also be applied to detect abnormal ship behavior. Another study also introduced two unsupervised clustering models using GMM and the expectation-maximization (EM) algorithm to capture normal vessel traffic patterns, considering momentary velocities and positions [69]. Several studies by Laxhammar [70] have compared the performance of KDE and GMM in ship anomaly detection, evaluating adaptive kernel density estimation and Gaussian mixture model using a novel performance measure. The results indicated that while KDE excelled in modeling ship normalcy, there was no notable distinction between kernel density estimation and Gaussian mixture model in terms of anomaly detection [70]. Similarly, Anneken et al. [20] employed GMM and KDE for predicting anomalies, with Gaussian mixture model representing a model that finds a set of Gaussian distributions to explain the dataset, while kernel density estimation estimates the probability density function of a random variable without assuming a specific distribution. Both methods have also shown effectiveness in vessel trajectory prediction using historical AIS data [57,21]. In the case of AIS data, clustering of forward trajectories is accomplished through GMM clustering using the expectation maximization algorithm [21]. The comparison between GMM and KDE based on the probability density functions (PDF) is graphically presented in Figure 14.



**Figure 14 GMM and KDE probability density functions in the position space, modeling two parallel vessel trajectories**

Source: Riveiro, M., Falkman, G. and Ziemke, T. (2008) Visual Analytics for the Detection of Anomalous Maritime Behavior. Available at: <https://doi.org/10.1109/iv.2008.25>. [68]

## 6. SUMMARY TABLE OF USED ARTICLES

**Table 4 Review of scientific papers reviewed in this bachelor’s thesis**

REFERENCE	DESCRIPTION OF APPLICATION	CONCLUSION
[4] Kim2020	This research developed a deep-learning method to classify fishing gear types using AIS data and environmental factors. The proposed approach showed good performance in estimating fish catch, detecting illegal fishing, and monitoring unregistered ships. Further improvements and data expansion were identified as future opportunities for enhancing the model.	In conclusion, future work should focus on refining the model, including reducing the sliding window size, expanding the number of fishing gear types, and collecting larger datasets for improved performance.
[9] Kraus2018	This paper presents a vessel trajectory classification model for the German Bight, achieving 97.51% accuracy by incorporating specialized features and ship dimensions. High-resolution radar observations are crucial for accurate classification when AIS data is unavailable.	Ship-specific characteristics and regular routes play a significant role in classification accuracy in the German Bight. Future work will explore this effect and compare the proposed algorithm with deep-learning approaches.
[18] Duan2022	The researchers employed the SSL-VTC method, which integrates both labeled and unlabeled data to improve the classification of vessel trajectories. This approach tackles the issue of scarce labeled data availability.	The SSL-VTC framework effectively extracts feature representations from AIS messages and outperforms traditional supervised learning approaches. The integration of labeled and unlabeled data significantly improves classification accuracy, offering valuable insights for maritime applications and ocean management.
[19] Ljunggren2018	This paper applies deep learning to sea surveillance for ship-type classification based on motion trajectories. By training neural networks using a large AIS dataset, the study demonstrates the effectiveness of this approach compared to conventional methods relying on reported ship types.	The use of deep learning enables accurate ship classification based on motion trajectories, even when AIS information is unreliable. This approach enhances sea surveillance capabilities and offers a more robust ship classification system.
[20] Abebe2020	This research introduces a data-centric approach that leverages AIS and weather data to precisely predict ship speed over ground across various conditions, ship categories, and routes.	The findings demonstrate the effectiveness of the methodology, achieving high model accuracies of over 96% and providing valuable insights for route optimization and performance enhancement in shipping operations.



[21] Murray2020	This study uses a linear autoencoder to accurately predict complex trajectories with computational efficiency. By clustering historical trajectories and interpolating them in the latent space, new trajectories can be generated. The method improves prediction quality by focusing on individual trajectory clusters.	The proposed method improves vessel trajectory prediction, enhancing maritime safety and situational awareness. It can be extended to autonomous vessels for intelligent decision-making. Future work includes exploring advanced deep-learning methods for better navigation and collision risk evaluation.
[22] Schöller2021	This paper presents a novel method for marine vessel trajectory prediction using AIS data and AIS heatmaps generated through kernel density estimation. It outperforms traditional methods by accurately predicting trajectories based on both historical data and positional information.	The proposed method improves marine trajectory estimation by leveraging AIS heatmaps and demonstrates the potential of using AIS data for more accurate predictions in marine applications.
[23] Meyer2020	The research demonstrated that a DRL-based autonomous vessel can avoid collisions while following a predetermined path, using rangefinder sensors to detect obstacles. The agent's training in a simulated environment successfully generalized to real-world scenarios, indicating its potential in complex guidance tasks.	Enhancing DRL agent acceptance in safety-critical systems involves refining reward functions, ensuring explainability, and stability analysis, accounting for environmental disturbances, and exploring multi-agent environments for COLREG compliance.
[24] Chatzikokolakis2021	In this study, machine learning algorithms are evaluated for detecting SAR patterns in noisy AIS data. Decision trees, random forests, and gradient boosted trees are tested using a substantial amount of historical AIS data, resulting in high accuracy in identifying SAR activity. Among the algorithms, random forests exhibit slightly superior performance and speed compared to the others.	This study employs machine learning algorithms to independently detect SAR patterns in AIS data, achieving a high level of accuracy. Among the algorithms tested, random forests exhibit superior performance and speed. These findings contribute to enhancing the effectiveness and autonomy of SAR pattern detection.
[25] Snapir2019	The scientists devised an approach that utilizes an RF classifier with five input characteristics to differentiate SAR vessel detections into fishing and non-fishing categories. The method attained a general classification accuracy of 91%, although the precision for the fishing class stood at 58%. The precision varied depending on the location and the dominance of class overlap areas.	While the vessel count may not accurately estimate the number of fishing/non-fishing vessels, the method effectively captures temporal variations in fishing activity, addressing the global overfishing problem.

[26] Ichimura2019	The study highlights accurate ship type identification using time series and position data, considering factors such as seasons and weather. Defining training data for abnormal routes and suspicious vessel movements is vital, utilizing valuable feature values.	The study finds accurate ship type identification through specific data and time series analysis. Future work will incorporate more factors and create new features for improved identification.
[27] Nishizaki2018	The paper proposes an innovative method for marine vessel trajectory prediction using AIS data. It leverages AIS heatmaps generated through kernel density estimation to encode long-term historical data. The approach surpasses traditional methods, utilizing the AIS heatmap to accurately predict trajectories that cannot be inferred from previous positions alone.	The study demonstrates the superiority of the proposed method for marine vessel trajectory prediction, utilizing AIS heatmaps and encoding long-term data. It outperforms traditional approaches, accurately predicting trajectories missed by using only previous positions, emphasizing the importance of AIS heatmaps in trajectory prediction.
[28] Coleman2020	The article highlights the limitations of the AIS protocol in maritime vessel communication and proposes a machine learning-based solution to model vessel behavior and identify abnormal activities in real time. By leveraging temperature sensor data transmitted via AIS, the study demonstrates the potential of this approach for enhancing threat detection and benefiting vessels without advanced communication equipment.	Machine learning models can enhance vessel communication security in the maritime domain by improving behavior modeling. Integrating additional sensors and features improves accuracy. Future work should focus on integrating models into trust networks for reliable communication and strengthened security measures.
[29] Ginoulhac2019	The application focuses on proposing an algorithm for classifying 2D tracks, with the potential to handle 3D tracks as well. The algorithm is designed to classify tracks in real-time and update its decisions as more data becomes available.	The proposed algorithm for track classification has shown promising performance. Despite already achieving satisfying results, authors claim that there is room for improvement. Future work will involve testing the algorithm on larger and more balanced datasets, as well as incorporating additional statistics and temporal variables to enhance the classifier's performance.
[30] Lang2018	The implementation presents an enhanced multi-class A-SVM classifier that incorporates simplified NGFs to bolster the classification of SAR ships, utilizing AIS information. By conducting experiments, it showcases the efficacy of leveraging AIS knowledge, particularly when dealing with a limited number of training samples.	The research confirms that incorporating AIS knowledge greatly improves SAR ship classification, especially with limited training samples. It highlights the potential of leveraging AIS knowledge to enhance classification accuracy. Future work should expand the methodology to include additional SAR image features.

[32] Zhang2020	<p>In this paper, a novel method for ship encounter classification is introduced, leveraging convolutional neural networks. The approach involves transforming AIS data into encounter images and incorporating ship navigational information to assess collision risk levels using CNN-based classification techniques.</p>	<p>The effectiveness of CNNs in classifying ship encounter situations based on collision risk levels is demonstrated by the proposed method. By integrating AIS data with navigational information, precise risk assessments can be conducted, thereby improving safety in maritime operations.</p>
[39] Liang2019	<p>The article introduces a two-step approach that utilizes random forest and LSTM-based supervised learning techniques to reconstruct vessel trajectories by addressing missing AIS data. The objective is to enhance the quality and completeness of the trajectories.</p>	<p>The experimental outcomes validate the efficacy and resilience of the proposed approach in reconstructing vessel trajectories, presenting promising advantages in bolstering maritime security and optimizing transportation efficiency.</p>
[41] Liu2020	<p>The paper proposes a framework for reconstructing vessel trajectories in 6G-enabled maritime IoT systems. It utilizes outlier detection and bidirectional RNNs to enhance the reliability of vessel traffic services by removing outliers and restoring corrupted points. The framework improves trajectory quality compared to other methods.</p>	<p>In conclusion, their two-phase approach effectively restored degraded trajectories. Future research aims to leverage GPU computing and diverse data for enhanced computation.</p>
[42] Tang2019	<p>The paper presents an LSTM-based model for precise vessel trajectory prediction in port regions, benefiting autonomous navigation systems. Experimental validation using real AIS data from Tianjin Port, China, verifies the model's effectiveness in trajectory prediction, demonstrating its relevance to autonomous navigation applications.</p>	<p>The study emphasizes precise trajectory prediction for autonomous vessels in ports using an LSTM model and AIS data. The findings support its integration into navigation systems for improved collision avoidance and proactive navigation.</p>
[43] Zhang2022	<p>The study focuses on predicting vessel locations using a combined model and AIS data. They employ the K-nearest neighbors (KNN) method to enhance precision, particularly in short-term predictions. The model considers motion parameters and predicts ship positions within a designated time range, taking into account ship speed and prediction duration.</p>	<p>The model excels in short-term ship location identification but struggles with long-term predictions due to simplified factors and fixed parameter weights. Future research should prioritize accuracy by considering diverse vessel types, analyzing trajectories, and incorporating environmental factors and AIS data for improved ship movement predictions.</p>
[44] Murray2021	<p>This study explores using deep learning for proactive collision avoidance in maritime transportation. The proposed framework predicts vessel trajectories quickly by decomposing regional ship behavior into local models based on historical AIS data.</p>	<p>The suggested framework successfully predicts global-scale trajectories and aims to improve uncertainty estimation, classification accuracy, and weather parameter integration in future research.</p>

[45] Yan2019	The paper reviews statistical and machine learning models used for anomaly detection in the maritime domain based on AIS data. Hybrid models and interactive systems are explored to enhance detection efficiency and accuracy.	Authors highlighted significance in enhancing situational awareness and maritime safety. Future research can focus on integrating expert knowledge and advancing data-driven techniques for improved anomaly detection in maritime operations.
[46] Dästner2018	Global localization data from AIS and automatic dependent surveillance–broadcast (ADS-B) networks enable advanced analytics in surveillance systems using machine learning. It allows for extracting vessel points, classifying traffic, and generating route maps. Handling the large data volume requires distributed data processing with tools like Spark, Hadoop, Kafka, or Flink.	AIS and ADS-B data analytics enhance GMTI radar and sensor-based processing. They enable route simulation, improve target tracking, and enable trajectory predictions. Integrating these analytics enhances surveillance capabilities and advances sensor-related processing.
[47] Martinčič2020	The proposed approach was used to detect and correct erroneous data, ultimately improving the accuracy and validity of the AIS data. Additionally, a metric is introduced to assess business and environmental efficiency based on validated AIS data.	The demonstrated tool showcases the effectiveness of the proposed solutions in optimizing logistic chains and providing valuable insights for vessel operators and ports.
[49] Guo2020	The paper presents an autonomous path-planning method for unmanned ships using DDPG. It combines ship data, COLREGS, and crew experience for effective path planning and adaptation in unfamiliar environments.	Proposed unmanned ship path planning using DDPG and improved DRL outperforms classical methods. The future research is expected to enhance effectiveness by incorporating the ship's motion model and real-world verification.
[50] Shi2020	The aim of this study was to establish a theoretical foundation for the development of intelligent collision-avoidance systems for unmanned surface vehicles. The research introduced a novel technique called double GRU-RNN, which effectively acquired collision-avoidance decision-making skills through the analysis of successful case data.	The proposed algorithm, with its extensive training in big data and remarkable versatility, presents a wide array of applications and strengthens collision-avoidance security by facilitating the timely generation and response of strategies in unmanned surface vehicles. The utilization of the double GRU-RNN approach demonstrates great potential in designing advanced and intelligent collision-avoidance systems.
[51] Lin2022	This study develops a dynamic model and an efficient anti-rolling controller for ship anti-rolling systems. Using deep reinforcement learning, the controller achieves effective control under diverse conditions, laying the foundation for future advancements.	The DDPG-based intelligent ship anti-rolling control method achieves over 95% efficiency and rapid convergence, addressing system limitations and enabling advanced anti-rolling controllers for full-scale ships.

[54] Kim2020	The study uses neural networks to predict vessel speeds on the Northern Sea Route. By considering location, time, vessel characteristics, and ice conditions, accurate speed forecasts are achieved. Validated with enhanced AIS data, the model addresses data challenges and offers reliable predictions.	Neural networks predict vessel speeds in Arctic voyages, improving safety and transport planning. Data challenges must be addressed to optimize their use. Understanding the model's capabilities is crucial for effective implementation in Arctic navigation.
[55] Song2020	This research automated ship-bounding box extraction from SAR images using AIS information, achieving accurate results without human intervention. The proposed methodology, integrating the Kalman filter and satellite orbit restoration, produced precise bounding boxes that can be applied to SAR and optical image ship detection.	The study demonstrated the effectiveness of the automated method for ship bounding box extraction using AIS information in SAR images, especially in coastal areas. The methodology can be further extended to ship detection in optical images, with the potential for ship classification using AIS information in future research.
[56] Song2020	This article aimed to enhance vessel detection and classification using SAR imagery by integrating AIS data. High-quality training data was obtained from Sentinel-1 SAR images, along with AIS information, to improve accuracy.	The authors automated ship-bounding box extraction from SAR images using AIS information, demonstrating its efficacy for ship detection, classification, and addressing challenges like illegal fishing and marine conservation while optimizing shipping operations.
[59] Chang2019	The paper evaluates YOLOv2 deep network architecture for vessel detection in SAR images. YOLOv2 outperforms existing technologies in accuracy and near real-time performance, especially in complex scenarios.	YOLOv2 shows improved accuracy and speed but lacks the real-time capability. Introducing YOLOv2-reduced, a new architecture with better detection time on an NVIDIA TITAN X GPU. Evaluation of the SSDD dataset demonstrates significantly faster detection with competent accuracy.
[62] Salem2023	This article introduces an improved model structure for classifying ships in inland waterways. By incorporating optimized CNN models, transfer learning, and ensemble learning methods, the classification performance is significantly enhanced. The ensemble learning approach demonstrates exceptional effectiveness, achieving an impressive success rate of 98.38% on the 'Game of Deep Learning' dataset."	The proposed classification method outperforms existing algorithms, demonstrating its effectiveness in ship classification tasks. It exhibits high accuracy on datasets such as MARVEL and performs well in categorizing various ship classes in inland waterways. Further improvements will focus on addressing weather conditions, enhancing processing speed, handling small data samples, and improving image preprocessing for noisy and low-contrast images.

[64] Nascimento2023	A tool utilizing deep learning and a convolutional neural network classifies vessel types and loading conditions based on the plimsoll disk, enhancing navigation safety, and reducing accidents. Its capabilities can extend beyond image analysis, and despite a limited database, the convolutional neural network achieves satisfactory results in classification.	The CNN successfully classified vessel types and loading conditions despite limited data, with future work focusing on real-time image analysis, expanding the database, and utilizing neural networks for port security. Overcoming data challenges can be addressed using synthetic data, highlighting the broad applications of neural networks in the field.
[65] Kang2019	This research presents a novel approach for ship wake detection and ship velocity estimation using deep learning and image analysis techniques. By utilizing Convolutional Neural Networks, the method accurately identifies wakes and extracts relevant data for analysis. The proposed approach shows effectiveness in estimating ship velocity, particularly in low wind conditions, by incorporating Radon transforms and edge filtering.	The study validates the effectiveness of the proposed method, combining deep learning-based wake detection and ship velocity estimation using azimuth offsets. The strong correlation with AIS data confirms its accuracy, making it a reliable solution for estimating ship velocity in open sea conditions.
[66] Ren2019	The authors introduce a novel ship classification approach, utilizing a multi-feature fusion CNN method. Their evaluation, conducted using the VAIS dataset along with their own ship dataset, demonstrates the method's performance. By combining handcrafted features such as HOG and LBP with CNN, the proposed method achieves remarkable classification accuracy, surpassing other state-of-the-art methods.	The proposed method achieves high ship classification accuracy, outperforming other deep networks, with rates of 97.50% and 93.60% on the self-built and VAIS datasets, respectively. Evaluation metrics validate the effectiveness of the improved CNN and feature extraction. Further improvements can be achieved by expanding the dataset and exploring transfer learning techniques.

## **7. CHALLENGES AND FUTURE TRENDS**

As ML continues to advance and AIS data becomes more readily available, there are both challenges and exciting possibilities on the horizon. One major challenge is still of course the quality and reliability of the AIS data itself. Ensuring the accuracy and completeness of the data is crucial for training ML models effectively and obtaining reliable insights. Additionally, the sheer volume and velocity of AIS data pose significant challenges for processing and analysis. Developing efficient algorithms and infrastructure to handle large-scale data processing in real-time is essential for harnessing the full potential of ML in maritime applications. Another challenge lies in the interpretability of ML models applied to AIS data. As ML algorithms become more complex, understanding how they arrive at their predictions becomes increasingly important, especially in safety-critical domains. Building transparent and interpretable ML models can enhance trust and facilitate the adoption of ML-based solutions in the maritime industry.

Looking ahead, there are promising future trends for ML and AIS data. One notable trend that is already being analyzed is the integration of ML with other rising technologies, such as Internet of Things (IoT). This kind of merging will allow many advantages like improving response times, leading to enhancements and effectiveness in maritime operations. Furthermore, the incorporation of ML into anomaly detection systems based on AIS data holds great potential. ML algorithms can learn patterns and detect unusual behaviors or potential threats, enhancing maritime security and safety. By continuously learning from new data, these systems can adapt to evolving risks and provide early warnings for potential incidents. Lastly, the integration of ML and AIS data with other maritime data sources, such as weather conditions, ocean currents, and port activities, can enable holistic and intelligent decision support systems. By leveraging multiple data streams, ML algorithms can provide valuable insights for optimizing vessel routes, improving fuel efficiency, reducing emissions, and enhancing overall operational efficiency in the maritime sector. In summary, while there are challenges to overcome, ML and AIS data offer immense opportunities for the maritime industry. By addressing data quality, interpretability, and scalability issues, and by embracing emerging trends, we can unlock the full potential of ML in transforming maritime operations, sustainability, and most importantly – safety.

## 8. CONCLUSION

Through the detailed analysis of recent scientific literature, this thesis explores the wide range of applications, techniques, advantages, and limitations associated with the utilization of machine learning and AIS data in the maritime domain. These articles collectively demonstrate the remarkable advancements made possible by ML when combined with AIS data. Despite the seemingly ordinary nature of AIS devices for vessel identification, they have proven to be crucial in the methodologies examined. From vessel classification and trajectory prediction to anomaly detection and maritime security, the integration of ML and AIS data consistently enhances various aspects of maritime operations. The significance of feature engineering is emphasized, as researchers have successfully extracted vital attributes from AIS data, including vessel identifiers, positional information, time stamps, and kinematic variables. These features play a pivotal role in training robust ML models capable of accurately classifying vessels, predicting future trajectories, and identifying anomalous behaviors.

Furthermore, the inclusion of supplementary data sources such as satellite imagery, historical data, and weather information has further augmented the performance and dependability of ML models. Despite the numerous advantages, challenges related to limited labeled training data, data quality, and integrity persist. The need for interpretable and explainable models is identified as an area requiring additional attention and research. The mentioned limitations merely underscore the importance of understanding the principles behind both ML and AIS devices. Ultimately, ML, in conjunction with AIS data, is poised to shape the future of the maritime industry, as evidenced by this thesis, which represents only a fraction of the remarkable applications of ML and AIS data in this domain.



## LITERATURE

- [1] Liu, H., Jurdana, I., Lopac, N., Wakabayashi, N. (2022) “BlueNavi: A Microservices Architecture-Styled Platform Providing Maritime Information,” *Sustainability*, 14(4), p. 2173. Available at: <https://doi.org/10.3390/su14042173>.
- [2] Lopac, N., Jurdana, I., Wakabayashi, N. & Liu, H. (2022) “Application of Advanced Digital Technologies for AIS Data Utilization”, *15th Baška GNSS Conference: Technologies, Techniques and Applications Across PNT and The 2nd Workshop on Smart, Blue and Green Maritime Technologies*, Baška, Croatia, p. 145-148.
- [3] Zhang, J., Ran, F. and Liu, D. (2019) “Maritime cognitive radio spectrum sensing based on multi-antenna cyclostationary feature detection,” *International Journal of Electronics* [Preprint]. Available at: <https://doi.org/10.1080/00207217.2019.1692373>.
- [4] Kim, K.S. and Lee, K.J. (2020) “Convolutional Neural Network-Based Gear Type Identification from Automatic Identification System Trajectory Data,” *Applied Sciences*, 10(11), p. 4010. Available at: <https://doi.org/10.3390/app10114010>.
- [5] Wakabayashi, N. and Jurdana, I. (2020) *Maritime Communications and Remote Voyage Monitoring*. Available at: <https://doi.org/10.1109/cobcom49975.2020.9174182>.
- [6] Jurdana, I., Lopac, N., Wakabayashi, N., Liu, H. (2021) “Shipboard Data Compression Method for Sustainable Real-Time Maritime Communication in Remote Voyage Monitoring of Autonomous Ships,” *Sustainability*, 13(15), p. 8264. Available at: <https://doi.org/10.3390/su13158264>.
- [7] Lopac, N., Jurdana, I., Wakabayashi, N. & Liu, H. (2023), “A Data Compression Approach to Reducing Demands on Maritime Communication Systems”, *16th Baška GNSS Conference: Technologies, Techniques and Applications Across PNT and The 3rd Workshop on Smart, Blue and Green Maritime Technologies*, Baška, Croatia, p. 105-109.
- [8] Passage Guardian (no date) *Best Practice AIS For Bluewater Sailing Yachts*. Available at: <https://passageguardian.nz/pages/best-practice-ais-for-bluewater-sailing-yachts>. (28<sup>th</sup> April 2023.)

- [9] Kraus, P.J., Mohrdieck, C. and Schwenker, F. (2018) *Ship classification based on trajectory data with machine-learning methods*. Available at: <https://doi.org/10.23919/irs.2018.8448028>.
- [10] 2014.,  *AIS AUTOMATIC IDENTIFICATION SYSTEM FOR VESSELS AT SEA*, Available at: [https://www.bluebird-electric.net/AIS\\_automatic\\_satellite\\_ship\\_identification\\_system.htm](https://www.bluebird-electric.net/AIS_automatic_satellite_ship_identification_system.htm)., (30<sup>th</sup> April 2023.)
- [11] 2022., *How AIS Works / Navigation Center* (no date). Available at: <https://www.navcen.uscg.gov/how-ais-works>, (29<sup>th</sup> April 2023.)
- [12] Gucma, M. (2008) “Low Cost Ais System for Safe Navigation,” *Journal of Konbin*, 6(3), pp. 235–246. Available at: <https://doi.org/10.2478/v10040-008-0070-2>.
- [13] *Table of Maritime Identification Digits* (no date). Available at: <https://www.itu.int/en/ITU-R/terrestrial/fmd/Pages/mid.aspx>.
- [14] Chai, W. (2020) “A Timeline of Machine Learning History,” *WhatIs.com* [Preprint]. Available at: <https://www.techtarget.com/whatis/A-Timeline-of-Machine-Learning-History>., (5<sup>th</sup> May 2023.)
- [15] Archana, V. 2023., ‘Machine Learning in Modern World’, *Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN)* ISSN: 2799-1172 3.01, 39-46.
- [16] Akanksha, E. *et al.* (2021) *Review on Reinforcement Learning, Research Evolution and Scope of Application*. Available at: <https://doi.org/10.1109/iccmc51019.2021.9418283>.
- [17] Reddy, Hhintaparathi Hema Kiran (2021) "A Hassle-Free & Intelligent Web Based Attendance System using Single Page Application.", Available at: [https://sist.sathyabama.ac.in/sist\\_naac/documents/1.3.4/b.e-cse-batchno-166.pdf](https://sist.sathyabama.ac.in/sist_naac/documents/1.3.4/b.e-cse-batchno-166.pdf)
- [18] Duan, H. *et al.* (2022) “A semi-supervised deep learning approach for vessel trajectory classification based on AIS data,” *Ocean & Coastal Management*, 218, p. 106015. Available at: <https://doi.org/10.1016/j.ocecoaman.2021.106015>.
- [19] Ljunggren, H. (2018) *Using Deep Learning for Classifying Ship Trajectories*. Available at: <https://doi.org/10.23919/icif.2018.8455776>.

- [20] Abebe, M. *et al.* (2020) “Machine Learning Approaches for Ship Speed Prediction towards Energy Efficient Shipping,” *Applied Sciences*, 10(7), p. 2325. Available at: <https://doi.org/10.3390/app10072325>.
- [21] Murray, B.J. and Perera, L.P. (2020) “A dual linear autoencoder approach for vessel trajectory prediction using historical AIS data,” *Ocean Engineering*, 209, p. 107478. Available at: <https://doi.org/10.1016/j.oceaneng.2020.107478>.
- [22] Schöller, F.E.T. *et al.* (2021) “Trajectory Prediction for Marine Vessels using Historical AIS Heatmaps and Long Short-Term Memory Networks,” *IFAC-PapersOnLine*, 54(16), pp. 83–89. Available at: <https://doi.org/10.1016/j.ifacol.2021.10.077>.
- [23] Meyer, E. *et al.* (2020) “COLREG-Compliant Collision Avoidance for Unmanned Surface Vehicle Using Deep Reinforcement Learning,” *IEEE Access*, 8, pp. 165344–165364. Available at: <https://doi.org/10.1109/access.2020.3022600>.
- [24] Chatzikokolakis, K. *et al.* (2019) “A comparison of supervised learning schemes for the detection of search and rescue (SAR) vessel patterns,” *Geoinformatica*, 25(4), pp. 601–622. Available at: <https://doi.org/10.1007/s10707-019-00365-y>.
- [25] Snapir, B., Waine, T.W. and Biermann, L. (2019) “Maritime Vessel Classification to Monitor Fisheries with SAR: Demonstration in the North Sea,” *Remote Sensing*, 11(3), p. 353. Available at: <https://doi.org/10.3390/rs11030353>.
- [26] Ichimura, S. and Zhao, Q. (2019) *Route-Based Ship Classification*. Available at: <https://doi.org/10.1109/icawst.2019.8923540>.
- [27] Nishizaki, C. *et al.* (2018) *Development of Navigation Support System to Predict New Course of Ship*. Available at: <https://doi.org/10.23919/wac.2018.8430436>.
- [28] Coleman, J., Kandah, F. and Huber, B. (2020) *Behavioral Model Anomaly Detection in Automatic Identification Systems (AIS), 2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*. Available at: <https://doi.org/10.1109/ccwc47524.2020.9031248>.
- [29] Ginoulhac, R. *et al.* (2019) *Coastal Radar Target Recognition Based On Kinematic Data (AIS) With Machine Learning*. Available at: <https://doi.org/10.1109/radar41533.2019.171262>.

- [30] Lang, H., Wu, S. and Xu, Y.-J. (2018) "Ship Classification in SAR Images Improved by AIS Knowledge Transfer," *IEEE Geoscience and Remote Sensing Letters*, 15(3), pp. 439–443. Available at: <https://doi.org/10.1109/lgrs.2018.2792683>.
- [31] Coucheiro-Limeres, A. *et al.* (2018) "Resource2Vec: Linked Data distributed representations for term discovery in automatic speech recognition," *Expert Systems With Applications*, 112, pp. 301–320. Available at: <https://doi.org/10.1016/j.eswa.2018.06.039>.
- [32] Zhang, W. 2020., 'Towards a Convolutional Neural Network model for classifying regional ship collision risk levels for waterway risk analysis, *Reliability Engineering and System Safety*'
- [33] Cherukuri, H.P. *et al.* (2019) "Machining Chatter Prediction Using a Data Learning Model," *Journal of Manufacturing and Materials Processing*, 3(2), p. 45. Available at: <https://doi.org/10.3390/jmmp3020045>.
- [34] Geurts, P., Ernst, D. and Wehenkel, L. (2006) "Extremely randomized trees," *Machine Learning*, 63(1), pp. 3–42. Available at: <https://doi.org/10.1007/s10994-006-6226-1>.
- [35] Beatrice, G., (2018) "Detection of fake profiles in Online Social Networks (OSNs)." *Master's degree in Applied Telecommunications and Engineering Management (MASTEAM)*
- [36] Jin, X. *et al.* (2006) *Spam email filtering with bayesian belief network: using relevant words*. Available at: <https://doi.org/10.1109/grc.2006.1635790>.
- [37] De Haro-García, A., García, G.C. and García-Pedrajas, N. (2019) "Instance selection based on boosting for instance-based learners," *Pattern Recognition*, 96, p. 106959. Available at: <https://doi.org/10.1016/j.patcog.2019.07.004>.
- [38] Kesavaraj, G. and Sukumaran, S. (2013) *A study on classification techniques in data mining, 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*. Available at: <https://doi.org/10.1109/icccnt.2013.6726842>.
- [39] Liang, M. *et al.* (2019) *Neural Network-Based Automatic Reconstruction of Missing Vessel Trajectory Data*. Available at: <https://doi.org/10.1109/icbda.2019.8713215>.

- [40] Hrnjica, B. and Bonacci, O. (2019) “Lake Level Prediction using Feed Forward and Recurrent Neural Networks,” *Water Resources Management*, 33(7), pp. 2471–2484. Available at: <https://doi.org/10.1007/s11269-019-02255-2>.
- [41] Liu, R.W. *et al.* (2021) “Data-Driven Trajectory Quality Improvement for Promoting Intelligent Vessel Traffic Services in 6G-Enabled Maritime IoT Systems,” *IEEE Internet of Things Journal*, 8(7), pp. 5374–5385. Available at: <https://doi.org/10.1109/jiot.2020.3028743>.
- [42] Tang, H., Yin, Y. and Helong, S. (2019) “A model for vessel trajectory prediction based on long short-term memory neural network,” *Journal of Marine Engineering and Technology*, 21(3), pp. 136–145. Available at: <https://doi.org/10.1080/20464177.2019.1665258>.
- [43] Zhang, M. *et al.* (2022) “Short-Term Trajectory Prediction of Maritime Vessel Using k-Nearest Neighbor Points,” *Journal of Marine Science and Engineering*, 10(12), p. 1939. Available at: <https://doi.org/10.3390/jmse10121939>.
- [44] Murray, B. (2021). Machine Learning for Enhanced Maritime Situation Awareness: Leveraging Historical AIS Data for Ship Trajectory Prediction.
- [45] Yan, Ran, and Shuaian Wang., 2019. "Study of data-driven methods for vessel anomaly detection based on AIS data." *Smart Transportation Systems 2019*. Springer Singapore
- [46] Dästner, K. *et al.* (2018) *Machine Learning Techniques for Enhancing Maritime Surveillance Based on GMTI Radar and AIS*. Available at: <https://doi.org/10.23919/irs.2018.8447961>.
- [47] Martinčič, T. *et al.* (2020) *Vessel and Port Efficiency Metrics through Validated AIS data, Global Oceans 2020: Singapore – U.S. Gulf Coast*. Available at: <https://doi.org/10.1109/ieeecnf38699.2020.9389112>.
- [48] Li, X. *et al.* (2021) “Toward a crowdsourcing solution to identify high-risk highway segments through mining driving jerks,” *Accident Analysis & Prevention*, 155, p. 106101. Available at: <https://doi.org/10.1016/j.aap.2021.106101>.

- [49] Guo, S. *et al.* (2020) “An Autonomous Path Planning Model for Unmanned Ships Based on Deep Reinforcement Learning,” *Sensors*, 20(2), p. 426. Available at: <https://doi.org/10.3390/s20020426>.
- [50] Shi, J. and Liu, Z. (2020) “Deep Learning in Unmanned Surface Vehicles Collision-Avoidance Pattern Based on AIS Big Data with Double GRU-RNN,” *Journal of Marine Science and Engineering*, 8(9), p. 682. Available at: <https://doi.org/10.3390/jmse8090682>.
- [51] Lin, J. *et al.* (2022) “Intelligent ship anti-rolling control system based on a deep deterministic policy gradient algorithm and the Magnus effect,” *Physics of Fluids*, 34(5), p. 057102. Available at: <https://doi.org/10.1063/5.0089697>.
- [52] Serrano, W. (2019) “Deep Reinforcement Learning Algorithms in Intelligent Infrastructure,” *Infrastructures*, 4(3), p. 52. Available at: <https://doi.org/10.3390/infrastructures4030052>.
- [53] *Internet of Things and Artificial Intelligence in Transportation Revolution* (2021) MDPI eBooks. Available at: <https://doi.org/10.3390/books978-3-0365-0311-0>.
- [54] Kim, E., Smestad, B.B. and Asbjørnslett, B.E. (2020) “Predicting Ship Speeds in the Arctic Using Deep Learning on Historical AIS Data,” *OnePetro* [Preprint]. Available at: <https://onepetro.org/ISOPEIOPEC/proceedings-abstract/ISOPE20/All-ISOPE20/446449>.
- [55] Song, J.-Y., Kim, D.-J. and Kang, K.-M. (2020) “Automated Procurement of Training Data for Machine Learning Algorithm on Ship Detection Using AIS Information,” *Remote Sensing*, 12(9), p. 1443. Available at: <https://doi.org/10.3390/rs12091443>.
- [56] Song, J.-Y. and Kim, D.-J. (2020) *Fine Acquisition of Vessel Training Data for Machine Learning from Sentinel-1 SAR Images Accompanied by AIS Information*. Available at: <https://doi.org/10.1109/igarss39084.2020.9324387>.
- [57] Anneken, M., Fischer, Y. and Beyerer, J. (2015) *Evaluation and comparison of anomaly detection algorithms in annotated datasets from the maritime domain*. Available at: <https://doi.org/10.1109/intellisys.2015.7361141>.
- [58] Balduzzi, M., *AIS Exposed. Understanding Vulnerabilities & Attacks PDF Free Download* (no date). Available at: <https://docplayer.net/31826083-Ais-exposed-understanding-vulnerabilities-attacks-2-0.html>.

- [59] Chang, Y.-L. *et al.* (2019) “Ship Detection Based on YOLOv2 for SAR Imagery,” *Remote Sensing*, 11(7), p. 786. Available at: <https://doi.org/10.3390/rs11070786>.
- [60] Li, Z. *et al.* (2022) “A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects,” *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), pp. 6999–7019. Available at: <https://doi.org/10.1109/tnnls.2021.3084827>.
- [61] Lopac, N., Hrzić, F., Vuksanović, I. P., Lerga, J. (2022) “Detection of Non-Stationary GW Signals in High Noise From Cohen’s Class of Time–Frequency Representations Using Deep Learning,” *IEEE Access*, 10, pp. 2408–2428. Available at: <https://doi.org/10.1109/access.2021.3139850>.
- [62] Salem, M.H. *et al.* (2023) “A Transfer Learning and Optimized CNN Based Maritime Vessel Classification System,” *Applied Sciences*, 13(3), p. 1912. Available at: <https://doi.org/10.3390/app13031912>.
- [63] Lopac, N., Jurdana, I., Lerga, J., Wakabayashi, N. (2021) “Particle-Swarm-Optimization-Enhanced Radial-Basis-Function-Kernel-Based Adaptive Filtering Applied to Maritime Data,” *Journal of Marine Science and Engineering*, 9(4), p. 439. Available at: <https://doi.org/10.3390/jmse9040439>.
- [64] Nascimento, A.V.D. *et al.* (2023) “Convolutional Neural Network (CNN) Applied to the Risk Analysis of Accidents in Vessels Navigating the Amazon Rivers,” *International Journal of Advanced Engineering Research and Science*, 10(2), pp. 064–074. Available at: <https://doi.org/10.22161/ijaers.102.6>.
- [65] Kang, K.-M. and Kim, D.-J. (2019) “Ship Velocity Estimation From Ship Wakes Detected Using Convolutional Neural Networks,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(11), pp. 4379–4388. Available at: <https://doi.org/10.1109/jstars.2019.2949006>.
- [66] Ren, Y.-M. *et al.* (2019) “Multi-Feature Fusion with Convolutional Neural Network for Ship Classification in Optical Images,” *Applied Sciences*, 9(20), p. 4209. Available at: <https://doi.org/10.3390/app9204209>.
- [67] Lopac, N., Jurdana, I., Brnelić, A., Krljan, T. (2022) “Application of Laser Systems for Detection and Ranging in the Modern Road Transportation and Maritime Sector,” *Sensors*, 22(16), p. 5946. Available at: <https://doi.org/10.3390/s22165946>.

[68] Riveiro, M., Falkman, G. and Ziemke, T. (2008) *Visual Analytics for the Detection of Anomalous Maritime Behavior*. Available at: <https://doi.org/10.1109/iv.2008.25>.

[69] Qu, X. *et al.* (2019) *Smart Transportation Systems 2019*. Springer.

[70] Laxhammar, R., Falkman, G., & Sviestins, E. (2009, July). Anomaly detection in sea traffic-a comparison of the gaussian mixture model and the kernel density estimator.

In *2009 12th international conference on information fusion* (pp. 756-763). IEEE.



## **ACRONYM INDEX**

ADNAVIS – Advanced Navigation System

ADS – B – Automatic Dependent Surveillance–Broadcast

AI – Artificial Intelligence

AIS – Automatic Identification System

ANN – Artificial Neural Networks

BLSTM – Bidirectional Long Short-Term Memory

CART – Classification and Regression Trees

CNN – Convolutional Neural Networks

COG – Course Over Ground

DBSCAN – Density-Based Spatial Clustering of Applications with Noise

DDPG – Deep Deterministic Policy Gradient

DL – Deep Learning

DSC – Digital Selective Calling

DTR – Decision Tree Regressor

DTW – Dynamic Time Warping

EM – Expectation Maximization

ETA – Estimated Time of Arrival

ETR – Extra Trees Regressor

GMDSS – Global Maritime Distress and Safety System

GMM – Gaussian mixture model

GMSK FM – Gaussian Minimum Shift Keying Frequency Modulation

GPS – Global Positioning System

GRU-RNN – Gated Recurrent Unit - Recurrent Neural Network

HDBSCAN – Hierarchical Density-Based Spatial Clustering of Applications with Noise

HDLC – High-level Data Link Control

HOG – Histogram of Oriented Gradients

IMO – International Maritime Organization

IoT – Internet of Things

KDE – Kernel density estimation

kHz – Kilo Hertz (frequency)

KNN – K-Nearest Neighbors

LBP – Local Binary Patterns

LSTM – Long Short-Term Memory

MHz – Mega Hertz (frequency)

ML – Machine Learning

MMSI – Maritime Mobile Service Identities

MPN – Multilayer Perceptron Network

NGF – Naive Geometric Features

NLAAF – Non-Linear Alignment and Average Filter

OPTICS – Ordering Points To Identify the Clustering Structure

PCA – Principal Component Analysis

R-CNN – Region-based Convolutional Neural Networks

RL – Reinforcement Learning

RNNs – Recurrent Neural Networks

S-1 – Sentinel 1

SAR – Search And Rescue

SAR – Synthetic Aperture Radar

SOG – Speed Over Ground

SOTDMA – Self-Organized Time-Division Multiple Access

SSDD – SAR ship detection dataset

SUBCLU – SUBspace CLUstering

SVM – Support Vector Machines

TREAD – Trajectory Representation, Exploration, and Analysis for Discovery

USVs – Unmanned Surface Vehicles

VAIS – Visible And Infrared Spectrums

VHF – Very High Frequency

VHF DSC – Very High Frequency Digital Selective Calling

VHF TDMA – Very High Frequency Time Division Multiple Access

VMS – Vessel Monitoring System

## LIST OF FIGURES

<b>Figure 1 Overview of AIS information types</b> .....	4
<b>Figure 2 AIS time slots</b> .....	6
<b>Figure 3 AIS - transmitting and receiving data</b> .....	7
<b>Figure 4 Machine learning general process</b> .....	11
<b>Figure 5 Example of learning data for SVM</b> .....	16
<b>Figure 6 Artificial neural network nodes</b> .....	19
<b>Figure 7 LSTM cell with its internal structure</b> .....	22
<b>Figure 8 Schema of DBSCAN</b> .....	25
<b>Figure 9 The Principle of Reinforcement learning</b> .....	28
<b>Figure 10 Iterative training process of the ship anti-roll control with the DDPG algorithm</b> .....	29
<b>Figure 11 Sample images from the SAR ship detection</b> .....	33
<b>Figure 12 Typical convolutional neural network structure for ship classification</b> ....	36
<b>Figure 13 Behavioral model anomaly detection</b> .....	37
<b>Figure 14 GMM and KDE probability density functions in the position space, modeling two parallel vessel trajectories</b> .....	39

## LIST OF TABLES

<b>Table 1 Transmit rates of Class A and Class B .....</b>	<b>5</b>
<b>Table 2 Several countries and their Maritime Identification Digit.....</b>	<b>9</b>
<b>Table 3 AIS data fields applied in constructing inputs for CNN collision risk analysis mode.....</b>	<b>34</b>
<b>Table 4 Review of scientific papers reviewed in this bachelor's thesis.....</b>	<b>40</b>