Transportation Research Part E 183 (2024) 103426

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

Harnessing the power of Machine learning for AIS Data-Driven maritime Research: A comprehensive review

Ying Yang [a], Yang Liu [b,*], Guorong Li [c,*], Zekun Zhang [d], Yanbin Liu [e]

[a] School of Management, Shanghai University, Shanghai, China
[b] Department of Architecture and Civil Engineering, Chalmers University of Technology, Sweden
[c] State Key Laboratory of Intelligent Green Vehicle and Mobility, Tsinghua University, Beijing, China
[d] College of Transport & Communications, Shanghai Maritime University, Shanghai, China
[e] School of Vehicle and Mobility, Tsinghua University, Beijing, China

ARTICLE INFO

Keywords:
Maritime research
AIS data
Machine learning
Trajectory prediction
Collision avoidance
Anomaly detection
Energy efficiency

ABSTRACT

Automatic Identification System (AIS) data holds immense research value in the maritime industry because of its massive scale and the ability to reveal the spatial–temporal variation patterns of vessels. Unfortunately, its potential has long been limited by traditional methodologies. The emergence of machine learning (ML) offers a promising avenue to unlock the full potential of AIS data. In recent years, there has been a growing interest among researchers in leveraging ML to analyze and utilize AIS data. This paper, therefore, provides a comprehensive review of ML applications using AIS data and offers valuable suggestions for future research, such as constructing benchmark AIS datasets, exploring more deep learning (DL) and deep reinforcement learning (DRL) applications on AIS-based studies, and developing large-scale ML models trained by AIS data.

1. Introduction

Automatic Identification Systems (AIS) have become a critical component of the maritime industry, providing real-time information about vessel positions, speed, and other important navigational details. The pervasive application of AIS has enabled a massive amount of data to be generated, which presents both chances and challenges for maritime research. On the one hand, this wealth of data offers an unprecedented opportunity for improved decision-making and a better understanding of vessel behaviors. On the other hand, processing and analyzing such a vast amount of data is not easy and requires advanced data processing and analysis techniques. In the early days, traditional methods such as statistical analysis were widely used to investigate massive AIS data (Ristic et al., 2008). While these methods offer good interpretability, they are notably limited in terms of generalization and efficiency. With the exponential growth of data scale and the development of unmanned surface vehicles (USVs), traditional methods are struggling to meet the requirements. The sheer volume and complexity of the data generated by AIS and other sources have surpassed the capabilities of conventional techniques. Additionally, the emergence of USVs necessitates the development of novel methods capable of efficiently processing and analyzing large amounts of data to support real-time decision-making and autonomous functionalities.

Fortunately, advancements of artificial intelligence (AI) has paved the way for machine learning (ML) to emerge as a powerful and promising tool for unlocking valuable insights from AIS data. ML algorithms have revolutionized the analysis process, enabling

* Corresponding authors.
E-mail address: liuy@chalmers.se (Y. Liu).

https://doi.org/10.1016/j.tre.2024.103426
Received 5 October 2023; Received in revised form 18 January 2024; Accepted 20 January 2024
Available online 28 January 2024
1366-5545/

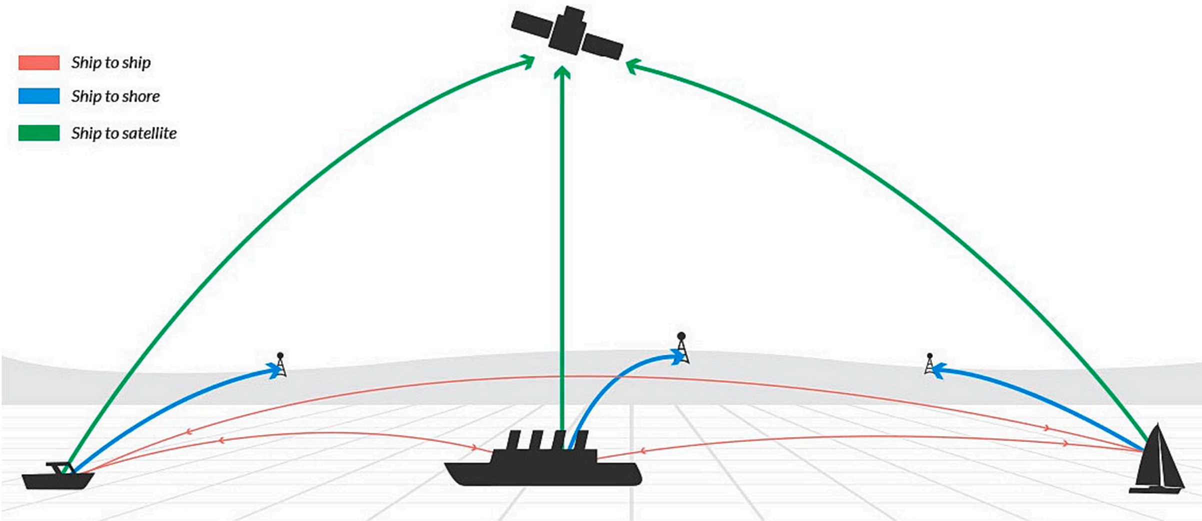

Fig. 1. AIS communication architecture. Sources: https://shipping.nato.int/nsc/operations/news/2021/ais-automatic-identification-system-overview.

researchers to extract a wealth of information that was previously inaccessible or time-consuming to obtain. By harnessing the capabilities of ML, researchers can delve into AIS data to uncover intricate vessel behavior patterns, identify collision risks, and evaluate the environmental impact of maritime activities. However, the majority of research in this field remains in the exploratory stage. Hence, it is crucial to conduct a comprehensive review of the existing relevant studies, which can help identify potential gaps and opportunities, thereby setting the groundwork for further exploration and innovation in this area.

To date, there have been several review papers that either provided a general overview of the use of AIS data or focused on specific topics in maritime research. For instance, Yang et al. (2019) conducted a review of studies regarding applications of AIS data and identified seven application fields. Similarly, Svanberg et al. (2019) offered a structured overview of how AIS data is utilized across ten identified areas. Furthermore, there are also review papers centered around specific research topics, including trajectory prediction (Gan et al., 2022; Zhang et al., 2022a; Cao et al., 2023; Choudhry & Qian, 2023; Dai et al., 2023), collision avoidance (Chai et al., 2022; Sarhadi et al., 2022; Fang et al., 2022; Xu et al., 2022; Rawson & Brito, 2023), anomaly detection (Yan & Wang, 2019; Wolsing et al., 2022), and energy efficiency (Yan et al., 2021a; Barreiro et al., 2022). Additionally, some researchers have paid attention to emerging approaches in maritime research (Tu et al., 2018; Munim et al., 2020; Yan et al., 2021b). These review papers have provided valuable insights and discussions on their respective topics. However, there are relatively fewer studies that comprehensively review the applications of ML on AIS data. Given the potential benefits of cross-topic learning and idea generation, it is important to conduct a comprehensive review that includes multiple ML applications on AIS data.

In this review paper, we explore the potential of ML in enhancing AIS data quality, and its applications for data-driven maritime research. Specifically, we examine various types of ML techniques that have been used in maritime research and their impact on AIS data processing and analysis. We also discuss challenges and limitations associated with the use of ML in the maritime domain and identify opportunities for future research. Overall, this review paper aims to provide a comprehensive overview of the potential benefits of ML for AIS data-driven maritime research and highlights the current state-of-the-art and future directions in this field.

2. AIS overview & application

With copious static and dynamic information of vessels, AIS data contain tremendous potential that can be applied in solving various scientific problems related to the maritime industry. This section brings a detailed interpretation of AIS data structure and mainstream applications in vessel trajectory prediction and upper-level topics about maritime safety & sustainability.

2.1. Automatic Identification system

Developed in the 1990 s, AIS is a short-range automatic tracking system designed to provide identification and positioning information for both vessels and shore stations. According to the Safety of Life at Sea (SOLAS) Convention, vessels weighing 300 gross tonnages (GT) or more involved in international voyages, cargo ships weighing 500 GT or more not involved in international voyages, and passenger ships of any size are obligated to have AIS installed onboard (International Maritime Organization, 2002). The specific purposes of equipped AIS include identifying ships, assisting in target tracking, assisting in search and rescue (SAR) operations, simplifying information exchange, and providing additional information to assist situational awareness (International Maritime Organization, 2015). Through the Very High Frequency (VHF) maritime band, the information from onboard AIS devices can be transmitted from ship to ship, ship to shore stations, and ship to satellite, as shown in Fig. 1.

Table 1
Main information included in AIS messages.

Field name	Generation	More information
Static (Updated every 6 mins or on request)		
MMSI, call sign, ship name	SOI	Might need amending when the ship changes ownership
IMO Number	SOI	Unique number for the ship
Length and beam	SOI	Might change if the ship size is changed
Ship type	SFPL	–
Dynamic (Depending on speed and course alteration)		
Position	AU	Longitude and Latitude, accuracy is approximately 10 m
Timestamp	AU	Timestamp for the position in UTC
COG, SOG, ROT	AU	Might not be available
Heading	AU	–
Navigational status	ME	–
Voyage-related (Updated every 6 mins or on request)		
Draught	ME	Amended as required
Destination and ETA	ME	Kept up to date as necessary

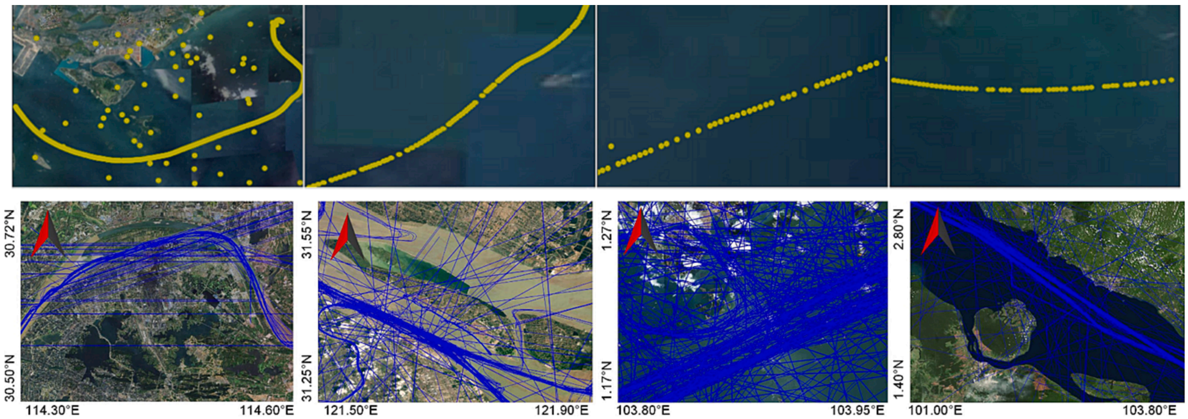

Fig. 2. Visualized raw AIS data with blank segments and outliers (Liu et al., 2021a).

Among all navigational aids, AIS is perceived as one of the most significant developments since radar was introduced. This system improves the performance of collision avoidance by continuously broadcasting vessel identity, position, speed, and course along with other relevant information to all other AIS-equipped vessels within reception range. Satellites equipped with AIS receivers are now able to receive AIS data worldwide, although the reception range of transmitted information was only limited to 10–20 nautical miles before 2008 (Yang et al., 2019). AIS not only enables vessels to perceive each other but also brings more valuable information to the Vessel Traffic Service (VTS), which enhances maritime surveillance for authorities to reduce navigational risk in coastal waters.

The main AIS information sent by ships is summarized in Table 1. These messages are mainly created in three ways: set on installation, automatically updated, and manually entered (International Maritime Organization, 2015). For instance, navigational status has to be manually entered by the officer of the watch (OOW), including underway by engines, underway by sail, at anchor, not under command, restricted in ability to maneuver, moored, constrained by draught, aground, engaged in fishing, etc. Additionally, update rates of AIS data may differ depending on the message type, movement status and shipborne equipment type. For example, static and voyage-rated information is updated every 6 min or upon request; for ships equipped with Class A shipborne equipment, dynamic information is automatically transmitted every 2–10 s while underway and every 3 min while at anchor and traveling at less than 3 knots; for ships equipped with Class B shipborne equipment, dynamic information is autonomously sent every 3–30 s while underway and every 3 min when traveling at less than 2 knots. Notably, Class B equipment is primarily installed on crafts not subject to the SOLAS carriage requirements. In the system, Class A equipment is given preference over Class B devices which operate at a lower reporting rate or when available time slots are open.

Note:

- Set on installation is represented as SOI
- Select from the pre-installed list is represented as SFPL
- Automatically updated is represented as AU
- Manually entered is represented as ME

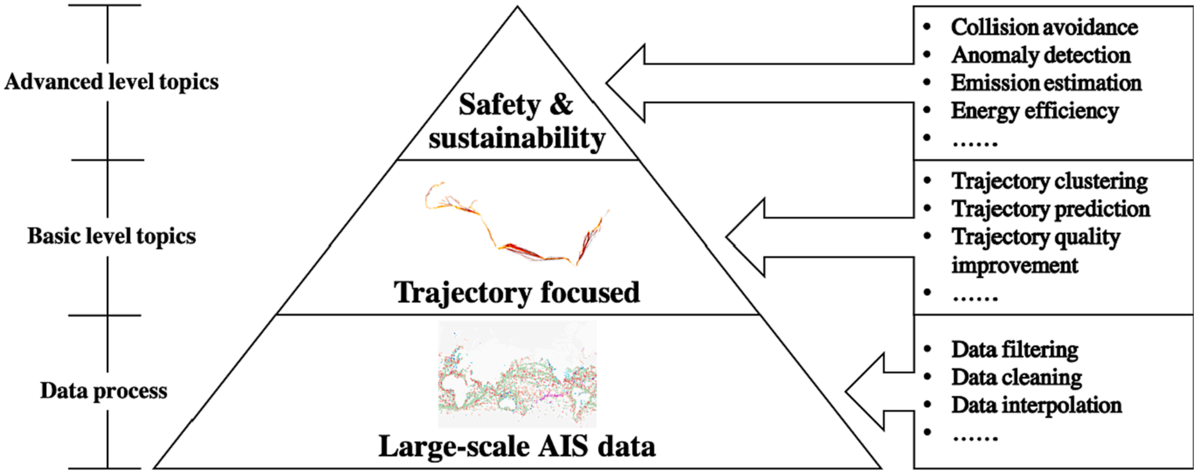

Fig. 3. Main research topics using AIS data.

2.2. Data quality consideration

Although AIS contributes a lot to the creation of convenient and powerful databases (e.g., MarineTraffic, VesselFinder, and Spire) for researchers and practitioners in the maritime industry, it is important to note that deletions, inaccuracies and errors still exist within the raw AIS data (Wolsing et al., 2022). For instance, some of the MMSI numbers, which should be unique to each vessel, are shared by multiple ships (Watson et al., 2015). In addition, numerous irregular blank segments and outliers can be found in AIS trajectories (Liu et al., 2021a), as shown in Fig. 2.

It is foreseeable that directly using such raw AIS data to train collision avoidance prediction models will greatly reduce the reliability of the prediction results. Most such studies need to pre-process raw AIS data before using it in training models (Zhang et al., 2023). In addition, the outliers may lead to biased spatial distribution when visualizing ship emissions based on AIS data (Weng et al., 2020b).

According to previous studies, quality issues in AIS data are generally caused by the following reasons:

- Manually entered information (e.g., draught, navigational status, destination, ETA) may have issues like ambiguous errors or inconsistent data formats.
- Information automatically updated by sensors might be unreliable if the position-fixing system is not functioning or is improperly connected to the AIS equipment (Zhang et al., 2022a).
- As a radio signal, parts of the AIS messages could get lost or scrambled affected by meteorology and magnetics.
- Although AIS transceivers should always be in operation when ships are underway or at anchor, there is a possibility that they might be intentionally turned off, resulting in missing data problems (Iphar et al., 2019).
- AIS signals can be easily spoofed and manipulated by attackers or parties willing to obscure their locations (Androjna et al., 2021).
- Update rates of AIS data vary from 2 s to 3 min, making the data too untidy to use.

Considering the inherent unreliability of AIS data, it is necessary to identify and deal with these data quality problems to ensure accurate and reliable results. Interpolation and resampling techniques are generally used to address these data quality problems, which can help to fill in gaps and reduce the impact of outliers or noisy data points (Capobianco et al., 2021; Ikonomakis et al., 2022). Regular expression as a popular text extraction tool is also helpful to extract text information from AIS data field with inconsistent formats, for instance, the destination of voyage.

2.3. AIS application

According to the Review of Maritime Transport 2022, there are more than 100,000 sea-going merchant vessels in the total fleet (UNCTAD, 2022). Assuming all these ships transmit AIS data every 10 s, approximately 300 billion AIS records can be generated across a year. Such vast quantities of global AIS data not only hold considerable research value, but bring many challenges. Research topics related to AIS data can be divided into two levels, as shown in Fig. 3. Studies directly focused on vessel trajectories can be treated as basic-level research topics. Advanced-level research topics contain further applications of AIS data, mainly related to maritime safety and marine sustainability. There are also studies utilizing AIS data on other topics, for instance, trade analysis, port efficiency analysis, etc (Yang et al., 2019).

2.3.1. Vessel trajectory

A vessel trajectory refers to the path or route followed by a vessel, which can be described by a group of scattered points in AIS data

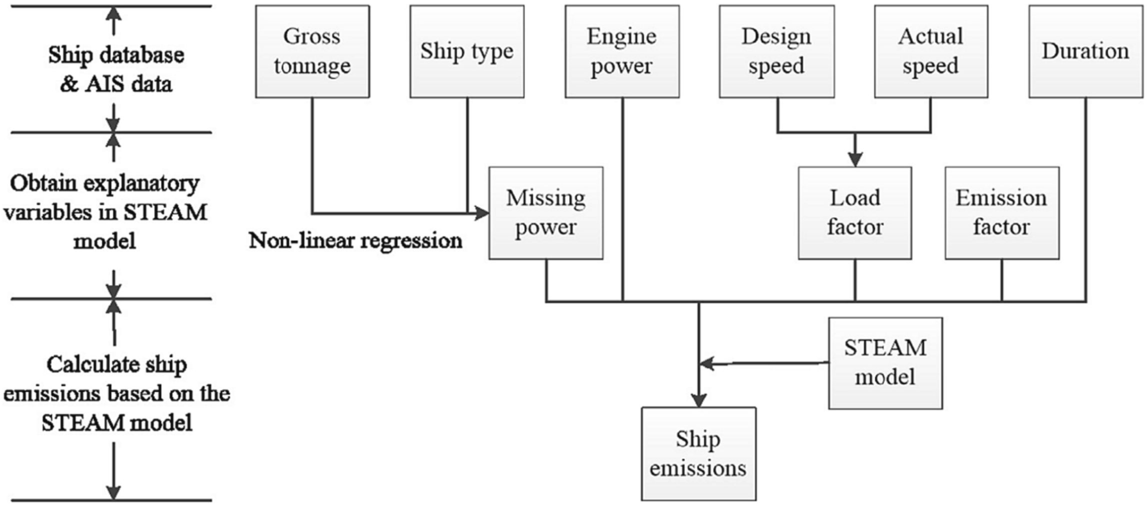

Fig. 4. A flowchart for AIS-based ship emission estimation by Weng et al., (2020b).

containing information such as position, speed, heading, and other relevant parameters about the ship and voyage. Vessel trajectories are fundamental units and play an essential role in a variety of research topics, including marine safety and sustainability. Most research topics concerning vessel trajectories revolve around areas like trajectory clustering, trajectory prediction, trajectory quality improvement. The dynamic and unpredictable nature of water currents and weather conditions lead to high complexity and irregular shapes of vessel trajectories. Therefore, clustering similar trajectories from messy AIS data could be a benefit for further studies in trajectory prediction, route planning, etc (Arguedas et al., 2018). By forecasting a short-term future trajectory of a ship, trajectory prediction not only helps to detect potential threats during the voyage but also improves the quality of historical trajectory datasets. Sometimes the basic data-cleaning process for AIS data is not enough to address all the data quality problems, which needs to be considered from a trajectory perspective (Liu et al., 2021a).

2.3.2. Marine safety

Collision avoidance: As the main application scenario of AIS, collision avoidance has been investigated by many researchers based on available AIS data. One basic concept for collision avoidance is ship domain which was first defined by Goodwin (1973) to ensure the safety of ships during encounters. AIS data can be helpful to construct effective ship domains. With AIS data in a specific water area, collision risk can be estimated based on the overlapping of ship domains (Qu et al., 2011; Chai et al., 2019). The ship domain is also a crucial factor used to evaluate the capacity of waterways for efficient vessel traffic flow and reduced collision risk (Liu et al., 2016; Kadarsa et al., 2017; Weng et al., 2020a). Christian & Kang (2017) identified different encounter situations (e.g., head-on, crossing, overtaking) and visualized waters with high collision risks based on AIS data.

Anomaly detection: Anomalies in AIS records can reveal important safety and security events, as noted by Wolsing et al. (2022). Therefore, anomaly detection using AIS data is crucial to ensure navigational safety, including VTS monitoring, risk evaluation, and SAR. Coastal waters typically require high-level maritime surveillance, making it advantageous for maritime authorities to utilize AIS anomaly detection to enhance surveillance efforts (Kowalska & Peel, 2012). Additionally, anomalies in AIS data can also signify illegal activities at sea, such as drug trade, piracy, and illegal fishing. Given that fishing vessels pose a significant threat to maritime safety and marine ecosystems, many researchers have been focusing on identifying illegal, unreported and unregulated (IUU) fishing activities from AIS data (Vespe et al., 2016; Ford et al., 2018).

2.3.3. Marine sustainability

The maritime industry has a significant environmental footprint primarily associated with both water and air pollution resulting from shipping activities. Water pollution, primarily caused by ship collisions leading to oil spills, constitutes a sporadic but significant hazard. Since vessels predominantly rely on fossil fuels for propulsion, they release a range of pollutants during the voyages, including sulfur oxides (SO_x), particulate matter, and greenhouse gases (GHGs) like nitrogen oxides (NO_x) and carbon dioxide (CO_2). These emissions have persistent adverse effects on public health and the environment. Efforts to address the environmental challenges stemming from maritime activities include the estimation of ship emissions and optimizing energy efficiency. Calculating ship emission inventories helps in gaining a comprehensive understanding of pollution sources and their magnitudes, thereby aiding in effective mitigation strategies. Simultaneously, energy efficiency optimization offer practical solutions to minimize pollutant emissions to the greatest extent possible.

Emission estimation: Containing both static and kinematics information of vessels, AIS data serves as one of the best sources for estimating emission inventories, enabling an accurate assessment of emissions. The "bottom-up" ship emission estimation method that relies on AIS data offers greater spatial and temporal resolution when compared to the "top-down" estimation method solely based on fuel consumption and fuel-based emission factors (Li et al., 2016). This is because the "bottom-up" method takes into consideration the dynamic information of vessels, such as velocity and location. For ship information not available through AIS data, supplementary details can be obtained from ship databases. A classic "bottom-up" approach for emission estimation is the Ship Traffic Emission Assessment Model (STEAM) developed by Jalkanen et al. (2009). The model computes pollutant emissions through the multiplication of engine power, load factor, operating time, and emission factor. Meanwhile, the emissions are also affected by the engine type, fuel type, and navigational status, most of which can be identified in AIS data. A flowchart by Weng et al., (2020b) demonstrated how AIS data can be used to estimate emission inventories, as shown in Fig. 4.

Energy efficiency: Aiming to reduce energy consumptions and emissions, AIS data could also be utilized to optimize sailing speed and enhance energy efficiency (Lee et al., 2018). Location, speed, and ETA from AIS data are the primary factors that need to be considered in energy efficiency studies. Reducing anchorage time is one way to improve energy efficiency (Jia et al., 2017). To this end, Watson et al. (2015) applied an integrated information system to optimize sailing speed for minimum anchorage time. Andersson and Ivehammar (2017a) proposed a speed adjustment method and validated it using AIS data from port waters. Route planning is another way to improve energy efficiency (Andersson and Ivehammar, 2017b; Praston, 2023). For instance, He et al. (2019) developed a route planning method based on historical routes from AIS data. Determining turning points is the basis of route planning. After turning point selection by clustering the historical routes, the Dijkstra algorithm and the ant colony algorithm were applied to generate and optimize the routes. Shorter routes with fewer turning points have been generated after route optimization, which can reduce energy consumption accordingly.

Although current studies still require a lot of effort to pre-process AIS data and extract effective trajectories, it is undeniable that those data have brought a lot of outstanding research works to the maritime industry, especially on collision avoidance, anomaly detection, emission estimation, and energy efficiency. However, as the explosive growth of data volume, the value of AIS data can hardly been maximized by traditional methods. Meanwhile, the blooming AI technologies has also put forward more intelligent requirements for the shipping industry. ML methods are becoming a reliable alternative solution for efficiently exploiting AIS data.

3. Machine learning-powered AIS application

In recent years, numerous ML-based solutions for classic topics in the maritime field have emerged to challenge traditional methods, especially in trajectory prediction, collision avoidance, anomaly detection, and energy efficiency. To get an in-depth understanding of current attempts, this section first brings detailed problem definitions for each topic, and then provide further discussions about different ML solutions in according subsections.

3.1. Vessel trajectory prediction

Among all the AIS applications, vessel trajectory prediction is the first to be discussed because it is considered to be one of the most essential topics in ensuring safety, intelligence and efficiency in maritime transportation. In this subsection, we first introduce the definition of trajectory and trajectory prediction. Following that, we review and introduce the ML methods mainly used in this topic. Finally, we further analyze what efforts the reviewed studies have made to improve their model performance.

3.1.1. Problem definition

A trajectory can be abstracted into a set of points containing spatial, temporal, and other information. Therefore, a series of tuples, $\{x_t, t \in T\}, x_t = (s, t, o)$, is generally used to express a vessel trajectory, where s refers to the geo-location of the point, t is the timestamp of the point, o represents other attributes of the point (e.g., SOG, COG, heading.), and T stands for the set of timestamps $\{1, 2, 3, \ldots, t\}$. It should be noted that s and t are necessary elements to represent a trajectory, while o is an optional one that depends on the need of the study. All these elements can be extracted from AIS data. Generally, given a fully observed trajectory from timestamp 1 to t, denoted by $X = \{x_1, x_2, x_3, \cdots, x_t\}$, trajectory prediction refers to predicting its following trajectory in a short-term future after timestamp t, denoted by $Y = \{x_{t+1}, x_{t+2}, x_{t+3}, \cdots, x_{t+a}\}$, where a stands for the prediction horizon. Besides predicting trajectories as point sets, some studies transform trajectories into grids or probability distributions for predictions (Nguyen et al., 2018; Dalsnes et al., 2018).

3.1.2. ML methods for trajectory prediction

3.1.2.1. Methods. Both basic ML and deep learning (DL) methods have been widely applied in traffic prediction (Liu et al., 2019a). Basic ML methods for trajectory prediction include Principal Component Analysis (PCA), k-Nearest Neighbors (k-NN), Support Vector Machine (SVM), Artificial Neural Network (ANN), Back-Propagation Neural Network (BPNN), and Extreme Learning Machine (ELM) (Zhang et al., 2022b). DL methods for trajectory prediction include Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Encoder-Decoder Architecture, Gated Recurrent Unit (GRU), Convolutional Neural Network (CNN), Transformer, Generative Adversarial Network (GAN), and Deep Neural Network (DNN).

Some of these studies explored the application of basic ML methods in this task. For example, PCA can be used to reduce the dimension of the features from AIS data that describe the trajectories (Murray & Perera, 2019). Thinking of vessel trajectory prediction as a classification problem, Duca et al. (2017) utilized a k-NN classifier to predict trajectories based on AIS data around Malta. The problem can also be regarded as a regression task, k-NN regressor (Virjonen et al., 2018), SVM and its variants can be employed to predict trajectories (Liu et al., 2019b; Liu et al., 2020). In addition, some researchers have investigated the performance of simple ANN (Volkova et al., 2021; Gan et al., 2016), BPNN (Zhou et al., 2019; Zhang et al., 2020) and ELM (Tu et al., 2020) for vessel trajectory prediction. Remarkably, ELM is considered to be a fast and robust ML algorithm with good generalization ability (Mao et al., 2018).

In order to explore methods with better performance in vessel trajectory prediction, more studies have leveraged state-of-the-art DL models. The architecture of DL models typically consists of multiple layers of nonlinear transformations, which together create a hierarchy of representations. Each level of representation is obtained by transforming the previous level into a higher and more abstract one. These transformations allow DL algorithms to learn complex relationships from observed trajectory samples (Fei et al., 2024; LeCun et al., 2015). Since trajectory samples contain timestamps, trajectory prediction can also be considered as time series prediction. Therefore, the LSTM network and its variants (e.g., GRU) are the most frequently used approaches (i.e., 73.68 %) for predicting vessel trajectory (Tang et al., 2019; Wang et al., 2020; Zhang et al., 2020). In addition, other researchers managed to compare the trajectory prediction task to the image inpainting/generation task. DL frameworks designed for image-related tasks, such as CNN and GAN, are also proven effective for vessel trajectory prediction (Kim & Lee, 2018a; Wang & He, 2021).

3.1.2.2. Highlights. This subsection summarizes the highlights of these related papers and identifies four main perspectives worthy of discussion.

Data input: In addition to the location and timestamps of ships, some studies tried to include more input features relevant to AIS data for their proposed trajectory prediction models. Among various features, SOG and COG are the most popular ones considered, while heading and ship type are sometimes also included for predicting vessel trajectories. To accurately predict trajectories at intersections, the destination of the voyage is also regarded as an important indicator (Capobianco et al., 2021). The popularity of these features may primarily depend on their correlation and integrity. Except for AIS data, some other studies also make use of supplementary data sources, for instance, shoreline shapefiles, meteorological data, and satellite images (Venskus et al., 2021; Mehri et al., 2021; Dijt & Mettes, 2020; Duca et al., 2017).

Clustering: Clustering is a popular choice for assisting trajectory prediction, which can be used either to remove irrelevant data and extract routes (Liu et al., 2021a), or to group trajectories that share similar characteristics (Murray & Perera, 2020; Murray & Perera, 2021). Extracting routes with the clustering technique is beneficial for acquiring high-quality training data. On the one hand, as mentioned in section 2.2, AIS data may contain error messages (e.g., wrong position) due to various reasons, making it difficult to identify through regular data cleaning processes. On the other hand, trajectory data generated when ships are at anchor or exhibiting irregular movement patterns are not ideal high-quality input samples, which should also be filtered out. Moreover, grouping and training trajectories by similar characteristics can significantly ease model training and enhance the robustness of trajectory prediction. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is a well-known clustering method that can effectively extract trajectories and captures key characteristics of vessel movement patterns (Li et al., 2019). It can be applied to eliminate unnecessary data or group trajectories with similar patterns for prediction.

Model complexity: Amongst all the efforts to enhance model performance, the most common way is to explore more complex model architectures. For example, variants of LSTM such as bidirectional LSTM (Bi-LSTM) and bidirectional GRU (Bi-GRU) are widely used to improve prediction accuracy, due to their ability to consider the effectiveness from both historical and future time-series data (Gao et al., 2018; Wang et al., 2020). The attention mechanism can be incorporated to allow different focuses on the different parts of hidden information of the Bi-LSTM network (Liu & Guo, 2019). In recent years, generative models are also developed for trajectory prediction. A GAN architecture consists of a generator and a discriminator that are trained to compete with each other (Tedjopurnomo et al., 2022), where the role of a generator network is to generate a data distribution while the role of a discriminator is to determine whether this generated sample is from the real distribution or not (Goodfellow et al., 2014). For trajectory prediction tasks, the generator is usually formed by an LSTM encoder-decoder. According to relevant studies, generative models could also provide remarkable predictions compared with Sequence to Sequence models (Wang & He, 2021).

Loss function: The loss function is a crucial component in evaluating model performance and updating model parameters during the training process. Which loss function to use depends on the type of data and problem being addressed. In the case of trajectory prediction, it is typically chosen as the geographical distance between predicted and ground-truth trajectories. The AIS coordinates commonly expressed in Spherical Coordinate System (SCS) employ different distance measures such as Haversine distance, Vincenty distance, and Equirectangular distance (Zhang et al., 2019; Murray & Perera, 2018; Sekhon & Fleming, 2020). Some studies have adopted Cartesian Coordinate System (CCS), whereas others still use non-geographical distance measures to define the loss function. Since trajectory prediction is usually regarded as a regression problem, metrics like Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE) are frequently used to evaluate performance.

3.2. Vessel collision avoidance

Collisions have received the most attention among all types of maritime accidents due to the high frequency and severe consequences. The deployment of USVs can significantly decrease the dependence on human operators by replacing human decision-making with an automated intelligent collision avoidance system. Therefore, state-of-the-art ML methods have been adopted for vessel collision avoidance in recent years. With reliable AIS data, emerging techniques are mainly applied to ship encounter recognition, ship domain learning, collision risk assessment, collision avoidance decision-making, etc. This subsection first defines the problem of collision avoidance and then reviews the ML methods proposed to avoid ship collisions.

3.2.1. Problem definition

To fully understand collision avoidance and the contribution of ML in this area, it is necessary to clarify the following basic concepts first:

- **Ship domain**: The fundamental idea for collision avoidance is the ship domain, which refers to the minimum safety zone surrounding a ship, enabling navigators to promptly respond and prevent any potential collisions. The size and shape of the ship domain are mainly affected by ship size, SOG & COG, traffic density, current & weather, etc (Tu et al., 2018). Considering the dynamic complexity of determining appropriate ship domains, ML could be a valuable tool to generate ship domains in real time.
- **Encounter situation:** The International Regulations for Preventing Collisions at Sea 1972 (COLREGs) have defined three basic types of encounter situations: crossing, head-on and overtaking, and crossing includes large-angle crossing and small-angle crossing (Chai et al., 2019). ML could help to classify encounter situations from a large amount of AIS data, and identify non-accident events such as near misses with ship domains.
- **Closest Point of Approach (CPA) analysis:** The concept of the CPA pertains to the proximity between two vessels assuming they maintain identical speed and course, which can be calculated based on AIS data. In collision risk evaluation, the distance to CPA (DCPA) and the time to CPA (TCPA) serve as fundamental parameters (Huang et al., 2020). Additional details on computing DCPA and TCPA can be found in the paper by Sarhadi et al. (2022).

3.2.2. ML methods for collision avoidance

3.2.2.1. Subtasks-oriented model. This part introduces ML applications for the subtasks of collision avoidance, for instance, ship encounter recognition, ship domain learning, collision risk assessment. Each subtask plays a vital role in successfully developing an automated intelligent collision avoidance system. The capability of automatically identifying encounter situations of ships is of paramount importance. Chen et al. (2021) proposed a Semi-Supervised Convolutional Encoder-Decoder Network (SCEDN) for ship encounter situation classification based on AIS data. The reason for adopting a semi-supervised learning method is that it can share training parameters with unlabelled data, which accounts for a large proportion of the training dataset. In their study, relevant distance, relevant speed, TCPA, and DCPA were selected as the features to train the encounter situation classification model. Accuracy and F1 score were chosen as the performance evaluation metrics similar to other classification tasks. Ship encounter recognition could efficiently recognize and label encounter situations for large amounts of unlabelled AIS data, which is beneficial to ship domain learning and collision risk assessment.

Empirical ship domain development tends to lack generalization because it can only consider several predefined situations. Learning ship domain with ML methods could be more flexible and self-adaptive, which can consider different influencing factors. Rawson & Brito (2021) applied a Random Forest (RF) algorithm to predict the size and shape of the ship domain by taking advantage of encounters from historical AIS data. Nine features were selected to describe the encounters for domain distance prediction, including bearing, encounter type, distance from shore, SOG, vessel size, vessel type, near traffic separation scheme (TSS), near port, wind speed, and day/night. The domain shape was then determined by the distance predicted in each direction. The learned ship domain could thus be applied to measure collision risk.

ML methods have been found to have a trade-off between accuracy and responsiveness for vessel collision risk assessment. Tritsarolis et al. (2022) proposed a Multi-Layered Perceptron (MLP) model to evaluate collision risk using a large-scale AIS dataset. Their model outperformed the traditional kinematic-based approach. Furthermore, AIS data can also be used to model non-accident events, such as near misses, which are useful in expressing collision risks. Kim & Lee (2018b) developed an MLP architecture to predict near-collision risks of ships, which can be integrated into VTS systems to reduce the workload of officers. In addition to collision risk assessment, Zhang et al. (2022b) proposed a risk evaluation model for grounding accidents. The model incorporated not only basic features extracted from AIS data but also ship drafts, an essential element for grounding risk assessment. The study also used hydrometeorological data and bathymetry data. For model construction, K-means clustering, dynamic time warping, and Douglas-Peucker (DP) algorithms were employed.

3.2.2.2. Decision-making oriented model. Other than the aforementioned subtasks of collision avoidance where ML has a role to play, there are also studies focused on decision-making in collision avoidance. One example is using ML techniques to learn effective collision avoidance strategies from extensive historical AIS data. For instance, Shi & Liu (2020) developed a double GRU-RNN model to identify collision avoidance patterns in ship encounter AIS data and generated appropriate collision avoidance decisions for various navigational scenarios. The DP algorithm was utilized to prepare the AIS data for training the model, and collision risk was assessed based on the predefined ship domain. This ML-based approach has potential applications in autonomous collision avoidance for USVs.

To enhance real-time collision avoidance capabilities for USVs, deep reinforcement learning (DRL) is commonly employed for decision-making. Zhao & Roh (2019) developed a DRL algorithm for collision avoidance among multiple USVs. The model utilized a two-layer MLP with a proximal policy optimization (PPO) learning algorithm. The reward function considers reaching the goal, heading error, cross-tracking error, drift, collision to obstacles, and COLREGs. Meyer et al. (2020) likewise applied a DRL algorithm with PPO for controlling an autonomous agent (i.e., USVs) to follow a predetermined path while avoiding collisions with other vessels in accordance with the COLREGs. AIS data from an inlet of the Norwegian Sea were used to evaluate the agent's ability to navigate challenging marine terrain and realistic vessel encounters. The proposed reward function can be divided into two parts: path following, influenced by cross-tracking error and speed, and collision avoidance, considered separately for static and dynamic obstacles. Jiang et al. (2022) proposed an attention-based DRL method for human-like collision avoidance, with a collision risk assessment module and a motion planning module included. Simulation scenarios, including static obstacles, dynamic multi-ship encounters, and dynamic and static obstacle coexistence, were designed to validate the effectiveness of the method.

3.3. Vessel anomaly detection

As AIS data becomes increasingly accessible, there is a rising demand for identifying abnormal AIS data. However, manually extracting suspicious ship activities from massive AIS data is impractical. According to recent studies on vessel anomaly detection, ML approaches have gained more popularity in automatically detecting various abnormal ship tracks. This subsection first introduces the problem definition of anomaly detection and then reviews the ML methods used to detect anomalies.

3.3.1. Problem definition

Anomalies in AIS tracks refer to behaviors that deviate from the norm or are unexpected during typical operations. In the context of vessel traffic, this may include sudden changes in ship velocity or uncommon travel routes that are not in line with standard practices. General forms of anomalous behaviors of ships include deviation from normal routes, discontinuous trajectories, unexpected port arrivals, close approach anomalies, and zone entry anomalies (Lane et al., 2010):

- **Deviation from normal routes:** A deviation from a straight route in open seas or from straight paths between predefined waypoints in complex waters may indicate an anomaly.
- **Discontinuous trajectories:** Intentional on–off switching of AIS equipment or other intentional behaviors may cause AIS signal loss, which could potentially obscure malicious actions.
- **Unexpected port arrivals:** Vessels may arrive at unexpected ports due to various illegal reasons (e.g., illegal fishing). These anomalies can be identified using voyage-related information from AIS data, such as destination port, ETA, and corresponding waypoints.
- **Close approach anomalies:** Close approaches between vessels that last a long time are uncommon, except in emergency situations. Such motion patterns may suggest illegal activities at sea (e.g., exchange of contraband or drugs).
- **Zone entry anomalies:** Vessels entering a restricted area (e.g., marine protected areas) for a considerable period should be considered an abnormal activity.

Ship anomaly can also be classified into three types from a kinematical perspective (Tu et al., 2018):

- **Position anomalies:** A ship appears in an unexpected position (e.g., restricted area, forbidden area).
- **Speed anomaly:** The speed of a ship is significantly higher or lower than normal status for a long time.
- **Time anomaly:** The visiting time of a ship is unexpected.

3.3.2. ML methods for anomaly detection

According to the reviewed papers, 90 % of the studies focus on deviation anomalies. These studies also often incorporate machine learning methods for various other types of anomalies. Therefore, the anomaly type discussed in the following subsections is deviation anomalies by default. Studies focused on other anomaly types will also be mentioned in the review.

3.3.2.1. Clustering-based model. Clustering is highly effective for distinguishing abnormal and normal trajectories from unlabeled AIS data. DBSCAN is a particularly useful clustering algorithm for detecting anomalies in AIS data because it can effectively recognize key characteristics of vessel movement patterns, as mentioned in 3.1.2.2. Sometimes K-means clustering is also applied to detect anomalies in AIS data (Guo et al., 2021). For example, Pallotta et al. (2013) developed a method called Traffic Rout Extraction and Anomaly Detection (TREAD) using DBSCAN. TREAD enabled unsupervised learning of a statistical model from AIS data and the extraction of valuable information for decision-making. In addition to the position of the ships, kinematical features like SOG and COG can be incorporated as additional features for anomaly detection. Hierarchical reasoning was employed to further improve the TREAD by identifying off-route vessels with only positional information and detecting anomalies of on-route vessels with heading and speed information (Pallotta & Jousselme, 2015). Later, TREAD was further applied by Arguedas et al. (2018) for developing a Maritime Traffic Knowledge Discovery and Representation System, which contributes to real-time traffic monitoring, anomaly detection, and situation prediction.

In addition, there is another way to cluster ship tracks considered kinematical features. To identify ship-moving & stopping areas, Liu et al. (2014) proposed an extension of DBSCAN, termed DBSCANSD, which incorporates SOG and COG as non-spatial features. To improve detection accuracy, the model was further refined by extending it into three division distances based on position, speed, and direction (Liu et al., 2015). On the basis of DBSCANSD, Wang et al. (2014) developed an anomaly detection framework that combined both unsupervised learning and supervised learning methods. They applied DBSCANSD to pre-cluster the data points as an initial step in anomaly detection. Normal and abnormal clustered ship tracks were then distinguished and labeled leveraging expert knowledge. Finally, a parallel *meta*-learning (PML) algorithm was trained based on the labeled data to detect anomalies automatically. Other types of features sometimes could also be included into specific anomaly detection schemes. For example, Radon et al. (2015) proposed a DBSCAN-based framework to filter false alarms in anomaly detection. Contextual information like weather information was also accounted for in the framework. Massive AIS data from U.S. Coast Guard were used to demonstrate that the proposed method can adapt to new contextual information.

3.3.2.2. Neural Network-based model. Unlike the clustering-based methods that are more suitable for anomaly detection from historical AIS data, neural network-based methods are more apt for real-time anomaly detection. In earlier studies, a Fuzzy ARTMAP neural network that took real-time AIS data as input could learn the motion patterns of vessels and detect anomalies in real-time (Bomberger et al., 2006). To simplify the motion pattern, this paper categorized COG into four directions (i.e., north, south, east, and west) and SOG into three levels (i.e., slow, medium, and fast). A grid neural network consisting of nodes (i.e., junctions) and synaptic connections (e.g., grid edges) was established. The future position of the ship is predicted by the weight of connections that stem from its current location. Any deviation of the route from the expected paths will end up being identified as anomalous. Although it is a dynamic unsupervised learning model that can detect anomalies from real-time unlabeled AIS data, it requires a large amount of historical data and appropriate grid size to learn the weights of the network and achieve satisfactory prediction accuracy.

In recent years, some researchers have been trying to explore the possibility of using DL methods for anomaly detection. Singh et al. (2020) developed a multi-class ANN-based anomaly detection framework to identify intentional and non-intentional AIS on–off switching anomalies. A four-dimensional vector consisting of latitude, longitude, SOG and COG was extracted from AIS data as the training and validation set. A bi-class (i.e., normal, abnormal) and a tri-class (i.e., normal, power outage, abnormal) anomaly detection models were trained separately and achieved an overall accuracy of 99.9 %. Nguyen et al. (2021) proposed a DL scheme called GeoTrackNet for detecting anomalies from AIS streams. The model comprises a variational RNN for identifying complex and heterogeneous motion patterns of vessels, and a contrario detector designed for evaluating the likelihood of an AIS trajectory. More than 4.2 million AIS records were utilized to train this model. The aforementioned DBSCAN-based method TREAD was chosen as the baseline in this study. The proposed GeoTrackNet has been proven to outperform the baseline model and could monitor AIS trajectories deviating from maritime routes—something DBSCAN-based models struggle with.

3.4. Vessel energy efficiency

As society becomes more stringent in reducing the energy consumption of ships, traditional methods of measuring energy efficiency are insufficient to meet the requirements considering the complex and dynamic nature of navigation conditions. In addition, the advancement of USVs necessitates automatic optimization of energy efficiency. Thus, with the aid of extensive AIS data, ML methods have emerged as an effective approach for optimizing energy efficiency. This subsection begins by introducing the related definitions of energy efficiency and then examines how ML applications can improve it.

3.4.1. Problem definition

In the marine sector, energy efficiency refers to the measurement of how efficiently a ship utilizes energy to transport goods or people by sea. Energy efficiency can be improved through a variety of means, including the use of more efficient engines, the optimization of ship design and operations, and the adoption of alternative fuels and propulsion systems. Improving energy efficiency in shipping can reduce fuel consumption, decrease greenhouse gas emissions, and improve the sector's overall sustainability.

The International Convention for the Prevention of Pollution from Ships (MARPOL) has put forth many indexes to measure energy efficiency, which can be used to optimize energy efficiency both at the design stage and the operational stage. These indexes include Energy Efficiency Design Index (EEDI), Energy Efficiency Operational Index (EEOI), Ship Energy Efficiency Management Plan (SEEMP). New amendments expected to enter into force in 2023 include Energy Efficiency Existing Ship Index (EEXI), and Carbon Intensity Indicator (CII). For detailed definitions of these indexes, we direct the reader to the paper by Barreiro et al. (2022).

Understanding estimation of energy consumption and emission is quite helpful before delving into energy efficiency. In general, three crucial elements for calculating fuel consumption and emissions are engine power, fuel type, and ship dynamic data obtained from AIS records. The majority of a ship's engine power can be obtained from Lloyd's database. Gross tonnage serves as a critical indicator that can help estimate missing engine power (Weng et al., 2020b). Fuel types are closely associated with the type of engines used. Typically, the main engine primarily consumes residual oil, while the auxiliary engine predominantly uses marine distillate oil. More detailed information regarding fuel types can be found in the relevant policies specific to particular maritime areas. When estimating emissions, it is essential to apply emission factors that indicate the relationship between ship dynamics and emissions per unit of various pollutants. For a deeper understanding of emission estimation, please refer to subsection 2.3.3.

3.4.2. ML methods for energy efficiency

To achieve better energy efficiency at the operational stage, the majority of the studies focus on vessel fuel consumption, often regarded as the basis of speed optimization and route planning. With the development of state-of-the-art techniques, ML has become a mainstream solution for fuel consumption prediction. In addition, ML methods have also been applied to develop Energy Management Systems (EMS) for hybrid vessels.

3.4.2.1. Fuel consumption prediction. ML models have gained popularity in predicting fuel consumption under different environments, owing to the challenging task that involves intricate interrelations among ship factors (e.g., engine, hull), kinematical factors (e.g., SOG, COG), and environmental factors (e.g., wind, wave, current, and temperature). According to the reviewed papers, the most popular ML model for energy consumption prediction is ANN (e.g., Du et al., 2019; Farag & Ölçer, 2020; Le et al., 2020). Other ML methods applied to this topic include tree-based models, Least Absolute Shrinkage and Selection Operator (LASSO) regression, Ridge regression, SVM regressors, Self-Organizing Maps (SOM), and Gaussian Mixture Model (GMM) (Barreiro et al., 2022). Apart from ML models, DL models like RNN, LSTM, and Elman Neural Network (ENN) (e.g., Panapakidis et al., 2020; Yuan et al., 2021) have also been employed in some studies.

While the aforementioned studies have proposed various solutions to predict energy consumption rates, it remains unclear which methods perform better or which data sources utilized lead to better prediction accuracy. + To address this question, Du et al., (2022a) evaluated the performance of eleven ML models widely used for energy consumption prediction, using nine specifically constructed datasets. These datasets were constructed based on voyage report data, AIS data, and meteorological data related to eight mega containerships. The eleven selected ML models include ANN, SVM, Ridge, LASSO, Decision Tree (DT), Extremely randomized Trees (ET), RF, AdaBoost (AB), Gradient tree Boosting (GB), XGBoost (XG), and LightGBM (LB). To measure the performance of these models, R^2, RMSE, and MAE were adopted. According to their experimental results, the combination of the three mentioned data sources resulted in notable benefits for model prediction. More importantly, tree-based models, including ET, AB, GB, and XG, were found to have superior performance over the best datasets. To deep dive into how data fusion and ML methods can help predict ship fuel efficiency, they also attempted to bring more data sources in their other studies, for instance, sensor data, which offered more possible ML solutions on this topic (Du et al., 2022b; Li et al., 2022). Nonetheless, due to the absence of a benchmark dataset, it is not possible to validate and compare the performance of fuel consumption prediction models across various studies. Having publicly accessible high-quality AIS data that includes concrete evidence of fuel consumption as training/testing data could facilitate the establishment of a standard for assessing the performance of future models. Once the model has been trained using the benchmark dataset, its performance can be compared with that of other studies, which also employ the same benchmark dataset for training their models.

3.4.2.2. Energy management system. For the development of green energy-powered ships, a basic need is to construct EMS to manage energy consumption from fuel and battery. To achieve near-optimal average voyage cost, Wu et al. (2020) proposed a Double Deep Q-Network (DDQN)-based EMS to mitigate the high cost of energy consumption following cost-effective energy management strategies. The double Q agent was trained on operational ship data from a coastal ferry, including both propulsive and auxiliary loads, and was validated using a dataset collected from another period. The results showed that these strategies can achieve near-optimal cost performance (96.9 %) without any advance information regarding future power demands, as opposed to those obtained through dynamic programming using the same state space resolution. According to the DRL-based EMS, the authors developed a holistic optimization approach for solving the power source sizing issue in the system, utilizing constrained mixed-integer multi-objective optimization in the outer layer (Wu & Bucknall, 2020). They conducted simulations replicating previous journeys and proved that using the hybrid system for case studies could result in a minimum of 65 % reduction in GHG emissions. In a follow-up study, Wu et al. (2021) introduced another DDQN approach to enhance the performance of energy management, resulting in an additional 5.5 % decrease in costs and a 93.8 % reduction in training time. A further EMS utilizing ML techniques was developed by Planakis et al. (2022). They suggested hierarchical clustering to identify the loading patterns of ships and used a Feedforward Neural Network (FNN) to anticipate future engine speed references for the EMS. In the experiment, the use of the developed EMS resulted in a 6 % reduction in fuel consumption and an 8.5 % reduction in NOx emissions for the ship based on a specific hybrid powertrain.

4. Challenges & future directions

After conducting a thorough review of existing research about ML applications on AIS data, the following perspectives from data usage and each research topic are further discussed to identify gaps and potential future directions.

4.1. Data usage

- **Benchmark datasets:** As noted in subsection 2.2, AIS data can suffer from a variety of quality issues for due to different reasons. While the majority of previous studies have typically conducted initial pre-processing on the raw AIS data, there is an absence of a unified standard to guide the cleaning process. Consequently, utilizing such varied data for ML applications could result in subjective assessments of their applicability. Moreover, although existing research on ML methods has shown promising outcomes, these studies are confined to specific datasets, such as particular water areas or ship types. This limitation reduces transparency and hinders the ability to determine the superiority of one ML method over another (Du et al., 2022a). Evaluating methods based on data from different water areas or with different pre-processing standards fails to establish a meaningful comparison. Therefore, it is crucial for future research to create a sufficiently large and high-quality AIS database. This will allow researchers to more confidently assess which ML methods are appropriate for particular tasks, even across different research teams. To accomplish this goal, it is highly recommended to create standardized AIS datasets and guidelines for AIS data pre-processing, similar to the practices observed in other fields (e.g., MNIST, ImageNet). These standardized benchmarks will provide an objective framework for ML tasks in the maritime sector.
- **Data fusion:** Significant efforts has been made to explore the use of multi-source data in the maritime field, rather than relying solely on AIS data. These varied data sources include satellite imagery, radar and Electronic Navigation Chart (ENC) images, meteorological data, hydrological data, and other data collected from shipboard sensors. The aim of incorporating these data sources is to enhance the performance of ML tasks, such as vessel trajectory prediction and energy consumption prediction. However, the identification and integration of these multi-modal data sources pose a critical challenge in supporting specific ML tasks. Each task may require the fusion of AIS data with different data sources, each with its own unique characteristics. Consequently, the complexity of the problem escalates. For example, when combining data from various sources, it is necessary to align them based on the lowest resolution among the diverse data sources. Techniques such as aggregation, resampling, and interpolation may be employed during the data fusion process. Furthermore, future research should provide guidelines for effectively fusing AIS data with other data sources, promoting standardized practices in this area.

4.2. Trajectory prediction

Firstly, the utilization of DL approaches for vessel trajectory prediction has experienced a notable surge, as evident from our reviewed papers. This growing adoption can be attributed to the successful application of forecasting techniques in land transportation, sparking interest in maritime applications. Emerging techniques like DRL show potential in enhancing the reliability of trajectory prediction (Zhang et al., 2022a), making them worthy of being pursued. Secondly, the main challenge with AIS data lies in the presence of irregular gaps in trajectory data, as discussed in subsection 2.2. While mainstream research on ship trajectory prediction predominantly focuses on forecasting future trajectories, there is limited exploration of utilizing this technology for addressing missing data issues. However, in the realm of image inpainting, researchers have already investigated the use of partial convolution methods to handle irregularly missing image data (Liu et al., 2018). It is thus anticipated that similar DL methods will be developed to address irregularly missing issues and reconstruct trajectory data, enabling the generation of higher-quality vessel trajectories for advanced-level studies.

4.3. Collision avoidance

Firstly, the ship domain plays a critical role in ensuring collision avoidance. In most studies, predefined fixed-shape domains, such as circular or elliptical, are commonly utilized due to technical and data constraints. However, with the advancement of ML techniques, there is growing interest in exploring learning-based domains as a promising solution for the next generation of ship domains. These learning-based domains have the potential to incorporate various influencing factors and adapt dynamically (Tu et al., 2018). To enhance the accuracy and effectiveness of ship domains, it would be more meaningful to leverage real-time AIS data along with other relevant data sources, such as meteorological data. By integrating these diverse datasets, we can train ML models to learn the dynamic shape of ship domains, allowing them to adapt to changing environmental conditions instead of assuming a constant domain shape that disregards the impact of environmental factors. Only a limited number of studies have contributed to learning-based domains according to our review. Secondly, DRL is found to be a prevailing approach to support decision-making for collision avoidance. This popularity could be attributed to its strong compatibility with control tasks. However, one major obstacle to utilizing DRL is the design of the reward function. Defining the reward and determining the appropriate weights for its indices can become a complex and intricate task (Sarhadi et al., 2022).

4.4. Anomaly detection

Firstly, the reviewed papers primarily focus on offline anomaly detection, utilizing historical AIS data for training and testing detection models. However, with the rise of USVs in maritime transportation, integrating AIS track anomaly detectors directly into vessels on sail becomes increasingly crucial for improving safe automated navigation (Wolsing et al., 2022). Despite the fact that some studies have made efforts to develop online anomaly detection models using AIS data streams, there's still a need for more effective real-time anomaly detection algorithms and systems capable of identifying online vessel abnormal behaviors. Secondly, many existing anomaly detection models primarily focus on the kinematic attributes of vessels, such as SOG, COG, and location. However, it is also important to recognize that static information, such as vessel type, size and flag, as well as environmental factors like traffic density and weather conditions, can also influence vessel behaviors (Yan & Wang, 2019). Incorporating these factors in the training of anomaly detection models could enhance the accuracy of detection.

4.5. Energy efficiency

Firstly, current fuel consumption prediction models lack generalizability as they are mainly tailored to specific ships. Developing unified models applicable across different vessels would overcome this limitation. Although tree-based models like RF show superior performance, they may require adjustments for direct fuel consumption prediction due to their discontinuous output. Secondly, the issue of endogeneity is often overlooked in energy consumption predictions. The complexity of vessel propulsion and environmental factors lead to endogeneity in fuel consumption models. In addition, ML-based fuel consumption prediction models rely heavily on feature engineering, which involves selecting valid features, constructing new ones, encoding and processing features, and identifying feature importance (Yan et al., 2021a,b). As a result, the prediction step is often treated as a black box where all processed features are fed into ML models, with little attention given to endogeneity. Future research should focus on exploring potential strategies for addressing endogeneity in fuel consumption prediction models.

4.6. Ais-based large ML model

The advent of Chat-GPT has captivated global attention due to its exceptional capabilities (OpenAI, 2023). As the popularity of Chat-GPT soars, large-scale ML models are emerging as the solution for the next generation of generative AI (Liu et al., 2023; Qu et al., 2023). Among various domains that stand to benefit, the maritime industry holds great potential for constructing a valuable and promising AIS data-based large ML model. Our review highlights a variety of significant research avenues that can be explored using AIS data as a foundation. Given its capacity to analyze extensive historical AIS data, a large ML model can reveal valuable insights and facilitate diverse predictive analytics mentioned before, thereby supporting decision-making processes within the maritime sector. Nevertheless, several crucial factors must be carefully addressed when building such a large ML model, including issues related to data quality, data privacy & security, sufficient computational resources, and expert domain knowledge.

5. Concluding remarks

This paper presents a comprehensive review of machine learning applications grounded in AIS data in the maritime industry. The review encompasses multiple aspects, starting with an overview of AIS data and its utilization in maritime research. Existing research topics related to AIS data and the ML approaches developed to tackle relevant issues are summarized. In this review, AIS-related research is classified into two levels: the basic level focusing on vessel trajectory, and the advanced level that delves into maritime

Table A1
Abbreviations.

Abbreviation	Full name
Maritime-related	
AIS	Automatic Identification System
CII	Carbon Intensity Indicator
COG	Course Over Ground
COLREGs	International Regulations for Preventing Collisions at sea
CPA	Closest Point of Approach
DCPA	Distance to Closest Point of Approach
EEDI	Energy Efficiency Design Index
EEOI	Energy Efficiency Operational Index
EEXI	Energy Efficiency Existing Ship Index
EMS	Energy Management System
ENC	Electronic Navigation Chart
ETA	Estimated Time of Arrival
GT	Gross Tonnage
IMO	International Maritime Organization
IUU	Illegal, Unreported, and Unregulated
MARPOL	International Convention for the Prevention of Pollution from Ships
MMSI	Maritime Mobile Service Identity
OOW	Officer of the Watch
ROT	Rate of Turn
SAR	Search and Rescue
SEEMP	Ship Energy Efficiency Management Plan
SOG	Speed Over Ground
SOLAS	Safety of Life at Sea
TCPA	Time to Closest Point of Approach
TSS	Traffic Separation Scheme
USV	Unmanned Surface Vehicle
VHF	Very High Frequency
VTS	Vessel Traffic Service
ML-related	
AB	AdaBoost
AI	Artificial Intelligence
ANN	Artificial Neural Network
Bi-GRU	Bidirectional Gated Recurrent Unit
Bi-LSTM	Bidirectional Long Short-Term Memory
BPNN	Back-Propagation Neural Network
CNN	Convolutional Neural Network
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree
DP	Douglas-Peucker
DRL	Deep Reinforcement Learning
ELM	Extreme Learning Machine
ENN	Elman Neural Network
ET	Extremely randomized Trees
FNN	Feedforward Neural Network
GAN	Generative Adversarial Network
GB	Gradient tree Boosting
GRU	Gated Recurrent Unit
LASSO	Least Absolute Shrinkage and Selection Operator
LB	LightGBM
LSTM	Long Short-Term Memory
k-NN	k-Nearest Neighbors
MAE	Mean Square Error
ML	Machine Learning
MLP	Multi-Layered Perceptron
MSE	Mean Square Error
PCA	Principal Component Analysis
PML	Parallel Meta-Learning
PPO	Proximal Policy Optimization
RF	Random Forest
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SOM	Self-Organizing Maps
SVM	Support Vector Machine
XG	XGBoost
Others	

(*continued on next page*)

Table A1 (*continued*)

Abbreviation	Full name
CCS	Cartesian Coordinate System
GHG	Greenhouse Gas
SCS	Spherical Coordinate System

safety and sustainability. Within these levels, four main topics are explored in depth: trajectory prediction, collision avoidance, anomaly detection, and energy efficiency.

The paper further highlights the challenges and opportunities of data usage and each research topic.

- Regarding limitations on data usage, the review highlights deficiencies in the quality of data employed in contemporary studies. To promote the advancement of ML-driven solutions within the maritime sector, it is imperative to focus on two critical and challenging areas: the creation of benchmark AIS datasets for training ML models and the integration of AIS data with other informative data sources.
- In the context of trajectory prediction, one of the primary challenges lies in dealing with irregular patterns of missing data in ship trajectories. To enhance the training of effective prediction models, it is recommended to employ techniques designed for handling irregularly missing image data, such as partial convolution, which can be applied to reconstruct ship trajectories.
- For collision avoidance, a notable constraint is the incapability of current static ship domains to assess real-time risks. A promising way is the development of dynamic ship domains that adapt based on learning. This endeavor might involve the integration of AIS data and other real-time data sources, such as meteorological information. The paper also proposes the exploration of DL models, including DRL techniques, to aid in decision-making for collision avoidance. This endeavor may require meticulous design of reward functions to be effective.
- It is recommended to enhance the efficiency of anomaly detection algorithms and align them with real-time detection needs. Furthermore, it is advisable to incorporate ship-related information and environmental factors into the anomaly detection process to elevate its overall performance.
- Previous studies have acknowledged that endogeneity is a primary limitation in predicting fuel consumption. Addressing this issue may require a meticulous selection of valid features. Furthermore, it is advisable to enhance the generalizability of fuel consumption prediction models to ensure they can effectively accommodate various vessel types.
- The advent of large ML models trained on AIS data holds promise as a transformative solution for revolutionizing intelligent maritime management. Such models have the potential to provide all-encompassing solutions for a wide range of research areas based on AIS data.
- In addition to the studies reviewed by this paper, emerging AI technology from other fields should also be noticed to boost the development of intelligent maritime solutions (Liu et al., 2021b; Liu et al., 2022; Lin et al., 2023; Qin and Liao, 2022; Shen et al., 2023; Wang et al., 2023; Zeng et al., 2023; Zheng et al., 2023).

The insights provided in this review are intended to benefit researchers, practitioners, and policymakers in the maritime industry, with the aspiration of inspiring further research and development in this field.

CRediT authorship contribution statement

Ying Yang: Writing – original draft, Conceptualization. **Yang Liu:** Writing – original draft, Conceptualization. **Guorong Li:** Writing – review & editing, Writing – original draft, Conceptualization. **Zekun Zhang:** Formal analysis, Data curation. **Yanbin Liu:** Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

This work was supported by the National Key Research and Development Program of China under grant number 2023YFB4301800, the National Natural Science Foundation of China (NSFC) under grant numbers 52220105001, 52221005, 72322002, 72361137001, and 72331001, and the International Cooperation Project (21160710800) of Shanghai Science and Technology Committee.

Appendix

(See Table A1.).

References

Andersson, P., Ivehammar, P., 2017a. Green approaches at sea – The benefits of adjusting speed instead of anchoring. Transportation Research Part d: Transport and Environment 51, 240–249.

Andersson, P., Ivehammar, P., 2017b. Dynamic route planning in the Baltic Sea Region – A cost-benefit analysis based on AIS data. Maritime Economics & Logistics 19 (4), 631–649.

Androjna, A., Perkovič, M., Pavic, I., et al., 2021. AIS Data Vulnerability Indicated by a Spoofing Case-Study. Applied Sciences 11, 5015.

Arguedas, V.F., Pallotta, G., Vespe, M., 2018. Maritime traffic networks: From historical positioning data to unsupervised maritime traffic monitoring. IEEE Transactions on Intelligent Transportation Systems 19 (3), 722–732.

Barreiro, J., Zaragoza, S., Diaz-Casas, V., 2022. Review of ship energy efficiency. Ocean Engineering 257, 111594.

Bomberger, N., Rhodes, B.J., Seibert, M., et al., 2006. Associative learning of vessel motion patterns for maritime situation awareness. In Information Fusion, 2006 9th International Conference on. IEEE, 1–8.

Cao, L., Luo, Y., Wang, Y., Chen, J., He, Y., 2023. Vehicle sideslip trajectory prediction based on time-series analysis and multi-physical model fusion. Journal of Intelligent and Connected Vehicles 6 (3), 161–172.

Capobianco, S., Millefiori, L.M., Forti, N., et al., 2021. Deep Learning Methods for Vessel Trajectory Prediction based on Recurrent Neural Networks. arXiv: 2101.02486v2.

Chai, T., Weng, J., Li, G., 2019. Estimation of vessel collision frequency in the Yangtze River estuary considering dynamic ship domains. Journal of Marine Science & Technology. 25, 964–977.

Chai, C., Zhou, Z., Yin, W., Hurwitz, D.S., Zhang, S., 2022. Evaluating the moderating effect of in-vehicle warning information on mental workload and collision avoidance performance. Journal of Intelligent and Connected Vehicles 5 (2), 49–62.

Chen, X., Liu, Y., Achuthan, K., et al., 2021. A semi-supervised deep learning model for ship encounter situation classification. Ocean Engineering 239, 109824.

Choudhry, A., Qian, S., 2023. Inferring truck activities using privacy-preserving truck trajectories data. Journal of Intelligent and Connected Vehicles 6 (1), 16–33.

Christian, R., Kang, H.G., 2017. Probabilistic risk assessment on maritime spent nuclear fuel transportation (Part II: ship collision probability). Reliable Engineering & System Safety 164, 136–149.

Dai, C., Zong, C., Zhang, D., Li, G., Chuyo, K., Zheng, H., Gao, F., 2023. Human-like lane-changing trajectory planning algorithm for human-machine conflict mitigation. Journal of Intelligent and Connected Vehicles 6 (1), 46–63.

Dalsnes, B.R., Hexeberg, S., Flaten, A.L., et al., 2018. The neighbor course distribution method with Gaussian mixture models for AIS-based vessel trajectory prediction. In: In Proceedings of the 21st International Conference on Information Fusion, pp. 580–587.

Dijt, P., & Mettes, P., 2020. Trajectory prediction network for future anticipation of ships. in Proc. Int. Conf. Multimedia Retr., Jun. 2020, 73–81.

Du, Y., Meng, Q., Wang, S., et al., 2019. Two-phase optimal solutions for ship speed and trim optimization over a voyage using voyage report data. Transportation Research Part b: Methodological. 122, 88–114.

Du, Y., Chen, Y., Li, X., et al., 2022a. Data fusion and machine learning for ship fuel efficiency modeling: Part II Voyage report data, AIS data and meteorological data. Communications in Transportation Research 2, 100073.

Du, Y., Chen, Y., Li, X., et al., 2022b. Data fusion and machine learning for ship fuel efficiency modeling: Part III-Sensor data and meteorological data. Communications in Transportation Research 2, 100072.

Duca, A.L., Bacciu, C., Marchetti, A., 2017. A K-nearest neighbor classifier for ship route prediction. OCEANS, 2017. Aberdeen. Aberdeen, UK, pp. 1–6.

Fang, L., Guan, Z., Wang, T., Gong, J., Du, F., 2022. Collision avoidance model and its validation for intelligent vehicles based on deep learning LSTM. Journal of Automotive Safety and Energy 13 (1), 104–111.

Farag, Y.B., Ölçer, A.I., 2020. The development of a ship performance model in varying operating conditions based on ANN and regression techniques. Ocean Engineering. 198, 106972.

Fei, Y., Shi, P., Li, Y., Liu, Y., Qu, X., 2024. Formation control of multi-agent systems with actuator saturation via neural-based sliding mode estimators. Knowledge-Based Systems 284, 111292.

Ford, J.H., Peel, D., Kroodsma, D., et al., 2018. Detecting suspicious activities at sea based on anomalies in Automatic Identification Systems transmissions. PLoS ONE 13, e0201640.

Gan, S., Liang, S., Li, K., et al., 2016. Ship trajectory prediction for intelligent traffic management using clustering and ANN. 2016 UKACC 11th International Conference on Control.

Gan, N., Zhang, M., Zhou, B., Chai, T., Wu, X., Bian, Y., 2022. Spatio-temporal heuristic method: a trajectory planning for automatic parking considering obstacle behavior. Journal of Intelligent and Connected Vehicles 5 (3), 177–187.

Gao, M., Shi, G., Li, S., 2018. Online prediction of ship behavior with automatic identification system sensor data using bidirectional long short-term memory recurrent neural network. Sensors 18 (12), 4211.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al., 2014. Generative adversarial nets. In Proceedings of Advances in Neural Information Processing Systems 27, 1–6.

Goodwin, E., 1973. A statistical study of ship domains. The Journal of Navigation, 26(1), 130–130.

Guo, S., Mou, J., Chen, L., et al., 2021. An Anomaly Detection Method for AIS Trajectory Based on Kinematic Interpolation. Journal of Marine Science and Engineering 9, 609.

He, Y., Zhang, D., Zhang, J., et al., 2019. Ship Route Planning Using Historical Trajectories Derived from AIS Data. The International Journal on Marine Navigation and Safety of Sea Transportation 13 (1), 69–76.

Huang, Y., Chen, L., Chen, P., et al., 2020. Ship collision avoidance methods: State-of-the-art. Safety Science 121, 451–473.

Ikonomakis, A., Nielsen, U.D., Holst, K.K., et al., 2022. Validation and correction of auto-logged position measurements. Communications in Transportation Research 2, 100051.

International Maritime Organization (IMO), 2002. SOLAS Chapter V—1/7/02 Safety of Navigation. London, UK, IMO.

International Maritime Organization (IMO), 2015. Resolution A.1106(29) Revised Guidelines for the Onboard Operational Use of Shipborne Automatic Identification Systems (AIS). IMO: London, UK.

Iphar, C., Ray, C., Napoli, A., 2019. Uses and Misuses of the Automatic Identification System. In: In Proceedings of the OCEANS 2019, pp. 1–10.

Jalkanen, J.P., Brink, A., Kalli, J., et al., 2009. A modelling system for the exhaust emissions of marine traffic and its application in the Baltic Sea area. Atmospheric Chemistry and Physics 9 (23), 9209–9223.

Jia, H., Adland, R., Prakash, V., et al., 2017. Energy efficiency with the application of virtual arrival policy. Transportation Research Part d: Transport and Environment 54, 50–60.

Jiang, L., An, L., Zhang, X., et al., 2022. A human-like collision avoidance method for autonomous ship with attention-based deep reinforcement learning. Ocean Engineering 264, 112378.

Kadarsa, E., Lubis, A.R.S., Sjafruddin, A., et al., 2017. Fairway traffic capacity in Indonesia. Procedia Engineering 171, 1443–1453.

Kim, K.I., Lee, K.M., 2018a. Deep learning-based caution area traffic prediction with automatic identification system sensor data. Sensors 18 (9), 3172.

Kim, K.I., Lee, K.M., 2018b. Context-Aware Information Provisioning for Vessel Traffic Service Using Rule-Based and Deep Learning Techniques. International Journal of Fuzzy Logic and Intelligent Systems 18 (1), 13–19.
Kowalska, K., Peel, L., 2012. In: Maritime Anomaly Detection Using Gaussian Process Active Learning. IEEE, pp. 1164–1171.
Lane, R.O., Nevell, D.A., Hayward, S.D., 2010. Maritime anomaly detection and threat assessment, et al., 2010. 13th International Conference on Information Fusion. Edinburgh, UK, pp. 1–8.
Le, L.T., Lee, G., Park, K.S., et al., 2020. Neural network-based fuel consumption estimation for container ships in Korea. Maritime Policy & Management 47 (5), 615–632.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Lee, H., Aydin, N., Choi, Y., et al., 2018. A decision support system for vessel speed decision in maritime logistics using weather archive big data. Computers & Operations Research 98, 330–342.
Li, X., Du, Y., Chen, Y., et al., 2022. Data fusion and machine learning for ship fuel efficiency modeling: Part I-Voyage report data and meteorological data. Communications in Transportation Research 2, 100074.
Li, C., Yuan, Z., Ou, J., et al., 2016. An AIS-based high-resolution ship emission inventory and its uncertainty in Pearl River Delta region, China. Science of the Total Environment 573, 1–10.
Li, W., Zhang, C., Ma, J., et al., 2019. Long-term vessel motion predication by modeling trajectory patterns with AIS data. In: 2019 5th International Conference on Transportation Information and Safety (ICTIS), pp. 1389–1394.
Lin, H., Yan, Y., Cheng, Q., 2023. Future role of artificial intelligence in advancing transportation electrification. Journal of Intelligent and Connected Vehicles 6 (3), 183–186.
Liu, B., de Souza, E.N., Matwin, S., et al., 2014. Knowledge-based clustering of ship trajectories using density-based approach. In Proceedings of IEEE 2014 International Conference on Big Data, 603–608.
Liu, B., de Souza, E. N., Hilliard, C., et al., 2015. Ship movement anomaly detection using specialized distance measures. In Proceedings of IEEE 2015 18th International Conference on Information Fusion (Fusion), 1113–1120.
Liu, G., Reda, F.A., Shih, K. J., et al., 2018. Image inpainting for irregular holes using partial convolutions. In Proceedings of the European conference on computer vision, 85–100.
Liu, X., He, W., Xie, J., et al., 2020. Predicting the trajectories of vessels using machine learning, in Proc. 5th Int. Conf. Control, Robot. Cybern. (CRC), 66–70.
Liu, G., Guo, J., 2019. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. Neurocomputing 337, 325–338.
Liu, Y., Liu, Z., Jia, R., 2019a. DeepPF: A deep learning based architecture for metro passenger flow prediction. Transportation Research Part c: Emerging Technologies 101, 18–34.
Liu, Y., Lyu, C., Zhang, Y., Liu, Z., Yu, W., Qu, X., 2021b. DeepTSP: Deep traffic state prediction model based on large-scale empirical data. Communications in Transportation Research 1, 100012.
Liu, Y., Jia, R., Ye, J., et al., 2022. How machine learning informs ride-hailing services: A survey. Communications in Transportation Research 2, 100075.
Liu, W.R., Nie, J., Garg, S., et al., 2021a. Data-Driven Trajectory Quality Improvement for Promoting Intelligent Vessel Traffic Services in 6G-Enabled Maritime IoT Systems. IEEE Internet of Things Journal 8 (7), 5374–5385.
Liu, Y., Wu, F., Liu, Z., et al., 2023. Can language models be used for real-world urban-delivery route optimization? The Innovation 4 (6).
Liu, J., Zhou, F., Li, Z., et al., 2016. Dynamic ship domain models for capacity analysis of restricted water channels. The Journal of Navigation 69 (3), 481–503.
Liu, J., Shi, G., Zhu, K., 2019b. Vessel trajectory prediction model based on AIS sensor data and adaptive chaos differential evolution support vector regression (ACDE-SVR). Applied Sciences 9 (15), 2983.
Mao, S., Tu, E., Zhang, G., et al., 2018. An automatic identification system (AIS) database for maritime trajectory prediction and data mining. In Proceedings of ELM, Switzerland: Springer, 241–257.
Mehri, S., Alesheikh, A.A., Basiri, A., 2021. A contextual hybrid model for vessel movement prediction. IEEE Access 9, 45600–45613.
Meyer, E., Heiberg, A., Rasheed, A., et al., 2020. COLREG-Compliant Collision Avoidance for Unmanned Surface Vehicle Using Deep Reinforcement Learning. IEEE Access 8, 165344–165364.
Munim, Z.H., Dushenko, M., Jimenez, V.J., et al., 2020. Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. Maritime Policy & Management 47 (5), 577–597.
Murray, B., & Perera, L.P., 2018. A data-driven approach to vessel trajectory prediction for safe autonomous ship operations. In Proceedings of the 2018 Thirteenth International Conference on Digital Information Management, 240–247.
Murray, B., & Perera, L.P., 2019. An AIS-based multiple trajectory prediction approach for collision avoidance in future vessels. In Proceedings of the ASME 2019 38th International Conference on Ocean, Offshore and Arctic Engineering, 7B.
Murray, B., Perera, L.P., 2020. A dual linear autoencoder approach for vessel trajectory prediction using historical AIS data. Ocean Engineering 209, 107478.
Murray, B., Perera, L.P., 2021. An AIS-based deep learning framework for regional ship behavior prediction. Reliable Engineering & System Safety 215, 107819.
Nguyen, D.D, Van, C.L, & Ali, M.I., 2018. Vessel Trajectory Prediction using Sequence-to-Sequence Models over Spatial Grid. In Proceedings of the 12th ACM International Conference on Distributed and Event-based Systems. Association for Computing Machinery, New York, NY, USA, 258–261.
Nguyen, D., Vadaine, R., Hajduch, G., et al., 2021. GeoTrackNetA Maritime Anomaly Detector using Probabilistic Neural Network Representation of AIS Tracks and A Contrario Detection. IEEE Transactions on Intelligent Transportation Systems 23 (6), 5655–5667.
OpenAI, 2023. GPT-4 Technical Report. arXiv:2303.08774.
Pallotta, G., & Jousselme, A.L., 2015. Data-driven detection and context-based classification of maritime anomalies. In Proceedings of IEEE 2015 18th International Conference on Information Fusion, 1152–1159.
Pallotta, G., Vespe, M., Bryan, K., 2013. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. Entropy 15 (6), 2218–2245.
Panapakidis, I., Sourtzi, V.M., Dagoumas, A., 2020. Forecasting the fuel consumption of passenger ships with a combination of shallow and deep learning. Electronics 9 (5), 1–25.
Planakis, N., Papalambrou, G., Kyrtatos, N., 2022. Ship energy management system development and experimental evaluation utilizing marine loading cycles based on machine learning techniques. Applied Energy 307, 118085.
Praston, L.F., 2023. Assessing impacts to maritime shipping from marine chokepoint closures. Communications in Transportation Research 3, 100083.
Qin, J., Liao, F., 2022. Space-time prisms in multimodal supernetwork-Part 2: Application for analyses of accessibility and equality. Communications in Transportation Research 2, 100063.
Qu, X., Meng, Q., Li, S., 2011. Ship collision risk assessment for the Singapore Strait. Accident Analysis & Prevention 43 (6), 2030–2036.
Qu, X., Lin, H., Liu, Y., 2023. Envisioning the future of transportation: Inspiration of ChatGPT and large models. Communications in Transportation Research 3, 100103.
Radon, A.N., Wang, K., Glässer, U., et al., 2015. Contextual verification for false alarm reduction in maritime anomaly detection. In Proceedings of IEEE 2015 International Conference on Big Data, 11231133.
Rawson, A., Brito, M., 2021. Developing contextually aware ship domains using machine learning. The Journal of Navigation 74 (3), 515–532.
Rawson, A., Brito, M., 2023. A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis. Transport Reviews 43 (1), 108–130.
Ristic, B., La Scala, B., Morelande, M., & Gordon, N., 2008. Statistical analysis of motion patterns in AIS Data: Anomaly detection and motion prediction. 2008 11th International Conference on Information Fusion, Cologne, Germany, 2008, 1–7.
Sarhadi, P., Naeem, W., Athanasopoulos, N., 2022. An Integrated Risk Assessment and Collision Avoidance Methodology for an Autonomous Catamaran with Fuzzy Weighting Functions. In: In 2022 UKACC 13th International Conference on Control. IEEE, Plymouth, UK, pp. 228–234.

Sekhon, J., & Fleming, C., 2020. A spatially and temporally attentive joint trajectory prediction framework for modeling vessel intent. In Proceedings of the 2nd Conference on Learning for Dynamics and Control, 2020, 318–327.
Shen, Z., Li, S., Liu, Y., Tang, X., 2023. Analysis of Driving Behavior in Unprotected Left Turns for Autonomous Vehicles using Ensemble Deep Clustering. IEEE Transactions on Intelligent Vehicles. https://doi.org/10.1109/TIV.2023.3345892.
Shi, J.H., Liu, Z.J., 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), 682.
Singh, S.K., Heymann, F., 2020. Machine Learning-Assisted Anomaly Detection in Maritime Navigation using AIS Data, 2020. IEEE/ION Position. Location and Navigation Symposium, Portland, OR, USA, pp. 832–838.
Svanberg, M., Santén, V., Hörteborn, A., et al., 2019. AIS in maritime research. Marine Policy 106, 103520.
Tang, H., Yin, Y., Shen, H., 2019. A model for vessel trajectory prediction based on long short-term memory neural network. Journal of Marine Engineering & Technology 21 (3), 136–145.
Tedjopurnomo, D.A., Bao, Z., Zheng, B., et al., 2022. A survey on modern deep neural network for traffic prediction: Trends, methods and challenges. IEEE Transactions on Knowledge and Data Engineering 4 (4), 1544–1561.
Tritsarolis, A., Chondrodima, E., Pelekis, N., et al., 2022. Vessel Collision Risk Assessment using AIS Data: A Machine Learning Approach, 2022. 23rd IEEE International Conference on Mobile Data Management (MDM). Paphos, Cyprus, pp. 425–430.
Tu, E., Zhang, G., Mao, S., et al., 2020. Modeling historical AIS data for vessel path prediction: A comprehensive treatment, arXiv:2001.01592.
Tu, E., Zhang, G., Rachmawati, L., et al., 2018. Exploiting AIS Data for Intelligent Maritime Navigation: A Comprehensive Survey From Data to Methodology. IEEE Transactions on Intelligent Transportation Systems 19 (5), 1559–1582.
Unctad, 2022. Review of Maritime Transport 2022. United Nations publications.
Venskus, J., Treigys, P., Markevičiūtė, J., 2021. Unsupervised marine vessel trajectory prediction using LSTM network and wild bootstrapping techniques. Nonlinear Analysis: Modelling and Control 26 (4), 718–737.
Vespe, M., Gibin, M., Alessandrini, A., et al., 2016. Mapping EU fishing activities using ship tracking data. Journal of Maps 12 (Suppl. 1), 520–525.
Virjonen, P., Nevalainen, P., Pahikkala, T., et al., 2018. Ship movement prediction using k-NN method. Baltic Geodetic Congress 2018, 304–309.
Volkova, T.A., Balykina, Y.E., Bespalov, A., 2021. Predicting ship trajectory based on neural networks using AIS data. Journal of Marine Science and Engineering 9 (3), 254.
Wang, X., Liu, X., Liu, B., et al., 2014. Vessel route anomaly detection with Hadoop MapReduce. 2014 IEEE International Conference on Big Data, Washington, DC, USA, 2014, 25–30.
Wang, S., He, Z., 2021. A prediction model of vessel trajectory based on generative adversarial network. The Journal of Navigation 74 (5), 1161–1171.
Wang, C., Ren, H., Li, 2020. Vessel trajectory prediction based on AIS data and bidirectional GRU. In: 2020 International Conference on Computer Vision, Image and Deep Learning, pp. 260–264.
Wang, K., Xiao, Y., He, Y., 2023. Charting the future: Intelligent and connected vehicles reshaping the bus system. Journal of Intelligent and Connected Vehicles 6 (3), 113–115.
Watson, R. T., Holm, H., & Lind, M., 2015. Green steaming: A methodology for estimating carbon emissions avoided. Thirty Sixth International Conference on Information Systems, Fort Worth, the U.S.A.
Weng, J., Liao, S., Yang, D., 2020a. Methodology for Estimating Waterway Traffic Capacity at Shanghai Estuary of the Yangtze River. The Journal of Navigation 73 (1), 75–91.
Weng, J., Shi, K., Gan, X., et al., 2020b. Ship emission estimation with high spatial-temporal resolution in the Yangtze River estuary using AIS data. Journal of Cleaner Production 248, 119297.
Wolsing, K., Roepert, L., Bauer, J., et al., 2022. Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches. Journal of Marine Science and Engineering 10, 112.
Wu, P., Bucknall, R., 2020. Hybrid fuel cell and battery propulsion system modelling and multi-objective optimisation for a coastal ferry. International Journal of Hydrogen Energy 45, 3193–3208.
Wu, P., Partridge, J., Bucknall, R., 2020. Cost-effective reinforcement learning energy management for plug-in hybrid fuel cell and battery ships. Applied Energy 275, 115258.
Wu, P., Partridge, J., Anderlini, E., et al., 2021. Near-optimal energy management for plug-in hybrid fuel cell and battery propulsion using deep reinforcement learning. International Journal of Hydrogen Energy 46, 40022–40040.
Xu, J., Zhang, Y., Dai, Z., Li, F., Chen, J., 2022. Vehicle trajectory oscillation characteristics and lane width control under human natural driving conditions. Journal of Automotive Safety and Energy 13 (4), 718–728. https://doi.org/10.3969/j.issn.1674-8484.2022.04.013.
Yan, R., Wang, S., 2019. Study of data-driven methods for vessel anomaly detection based on AIS data. In: Smart Transportation Systems. Springer, Singapore, pp. 29–37.
Yan, R., Wang, S., Psaraftis, H.N., 2021a. Data analytics for fuel consumption management in maritime transportation: Status and perspectives. Transportation Research Part e: Logistics and Transportation Review 155, 102489.
Yan, R., Wang, S., Zhen, L., et al., 2021b. Emerging approaches applied to maritime transport research: Past and future. Communications in Transportation Research 1, 100011.
Yang, D., Wu, L., Wang, S., et al., 2019. How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications. Transport Reviews 39 (6), 755–773.
Yuan, Z., Liu, J., Zhang, Q., et al., 2021. Prediction and optimisation of fuel consumption for inland ships considering real-time status and environmental factors. Ocean Engineering 221, 108530.
Zeng, Z., Wang, S., Qu, X., 2023. Consolidating Bus Charger Deployment and Fleet Management for Public Transit Electrification: A Life-Cycle Cost Analysis Framework. Engineering 21, 45–60.
Zhang, X., Liu, G., Hu, C., 2019. Wavelet analysis based hidden Markov model for large ship trajectory prediction, et al., 2019. Chinese Control Conference. Guangzhou, China, pp. 2913–2918.
Zhang, X., Fu, X., Xiao, Z., et al., 2022a. Vessel Trajectory Prediction in Maritime Transportation: Current Approaches and Beyond. IEEE Transactions on Intelligent Transportation Systems 23 (11), 19980–19998.
Zhang, M., Kujala, P., Hirdaris, S., 2022b. A machine learning method for the evaluation of ship grounding risk in real operational conditions. Reliability Engineering and System Safety 226, 108697.
Zhang, J., Liu, J., Hirdaris, S., et al., 2023. An interpretable knowledge-based decision support method for ship collision avoidance using AIS data. Reliability Engineering and System Safety 230, 108919.
Zhang, Z., Ni, G., Xu, Y., 2020. Trajectory prediction based on AIS and BP neural network. 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference. Chongqing, China 9, 601–605.
Zhao, L., Roh, M.I., 2019. COLREGs-compliant multiship collision avoidance based on deep reinforcement learning. Ocean Engineering 191, 106436.
Zheng, C., Yan, Y., Liu, Y., 2023. Prospects of eVTOL and modular flying cars in China urban settings. Journal of Intelligent and Connected Vehicles. https://doi.org/10.26599/JICV.2023.9210029.
Zhou, H., Chen, Y., Zhang, S., 2019. Ship trajectory prediction based on BP neural network. Journal of Artificial Intelligence 1 (1), 29.