Marine Research
of the National Academy of Marine Research
ISSN: 2709-6629 (Vol.4 Iss.1)
GPN: 2011200033

DOI: http://dx.doi.org/10.29677/MR.202406_4(1).0006
2024, 4(1),79-99

Automatic Identification System Data Fusion

Lie Yang [1] Shao-Hua Hsu [2] Zhi-Jie Yu [3] Yung-Cheng Yao [4*]

[1] Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

[2] Department of Marine Leisure Management, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

[3] Auray Technology Corp, Kaohsiung, Taiwan

[4] Department of Microelectronics Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

ABSTRACT

With the rise of intelligent computing applications such as artificial intelligence and big data analytics, the quality of data sources has become crucial for accurate decision-making. This is primarily because data collected from various sources needs to be organized and processed to extract information for analysis and subsequent decision-making. With advancements in communication technology and the growth of maritime transportation, the Automatic Identification System (AIS) has become a fundamental communication device for maritime vessel navigation. Due to the extensive range of data provided by the AIS and its diverse functional requirements, there may be instances of incomplete or duplicated data packets during the transmission process to shore-based stations, caused by signal interference or poor reception. This study, with the assistance of the National Academy of Marine Research (2024) which provided AIS data sources, focuses on establishing a data management mechanism for the AIS database. Additionally, it involves the development of AIS data processing algorithms for tasks such as data format parsing, field interpretation, and cleaning. The aim of the study is to establish an operational and standardized AIS data quality control mechanism that effectively identifies and eliminates abnormal or inconsistent AIS data within the system. By doing so, the study seeks to maintain the data stability and reliability of the AIS database while providing high-quality AIS data.

Keywords: Automatic Identification System (AIS), Data Quality, AIS Database, Data Processing.

* Corresponding author, e-mail: yuchyao@nkust.edu.tw

Received 7 August 2023, Accepted 26 March 2024.

1 INTRODUCTION

Surrounded by the sea, Taiwan is located in the eastern part of Asia, where East Asia and South Asia converge, and occupies a central position in the East Asian archipelago. Therefore, the maritime transportation in its surrounding areas is highly developed. To the west, across the Taiwan Strait, lies China. To the south, through the Bashi Strait, one can reach countries such as the Philippines, Vietnam, Singapore, and Malaysia. To the north, through the East China Sea and the South China Sea, one can reach countries like Japan and South Korea. Taiwan's exceptional geographical location has made it a crucial transportation hub in Asia and the Western Pacific. It plays a significant role in various aspects, including the economy and transportation within the Asian region. The waters near Taiwan are characterized by a high number of vessels, bustling maritime trade activities, and frequent fishing operations. With limited navigational space in the surrounding seas of Taiwan, the increasing variety of vessels poses a significant risk of accidents during navigation.

As maritime safety concerns have gained significant attention from countries worldwide, the International Maritime Organization (IMO) has issued the "International Convention for the Safety of Life at Sea (SOLAS)" (RECOMMENDATION ITU-R M.2092-1, 2022), which mandates the installation of the Automatic Identification System (AIS) on all passenger and cargo ships with a gross tonnage of 300 tons or above for international voyages, as well as cargo ships not engaged in international voyages with a gross tonnage of 500 tons or above. While navigational instruments and equipment such as radar charts and Automatic Radar Plotting Aids (ARPA) have the capability to identify vessels at sea, their primary purpose is to provide relevant navigational information during ship operations. However, they are unable to acquire and relay other dynamic vessel information such as longitude, latitude, heading, and speed of neighboring vessels. This limitation hinders effective measures to avoid and mitigate potential maritime accidents, including collisions that can lead to capsizing and grounding. In response to the increasing volume of various types of vessels entering and leaving ports (Nieh, et al., 2019) and to ensure safe navigation in the waters surrounding Taiwan, the Maritime and Port Bureau, Ministry of Transportation and Communications (MOTC) revised the "Ship Equipment Regulations" (Huang et al., 2019) in 2008 based on the International Convention for the Safety of Life at Sea. According to these regulations, all types of vessels with a gross tonnage of 20 tons or above are required to be equipped with the AIS, and nationwide AIS-related hardware facilities have been established. These measures aim to ensure the safety of vessel navigation and acquire precise dynamic data such as ship speed and heading.

In the research project titled "Application of Ship Monitoring and Early Warning System" published by Su and Xu from the Harbor and Marine Technology Center in 2021, the authors discussed how to utilize historical dynamic data from ships' AIS to analyze vessel trajectories in the waters surrounding Taiwan and compile statistics on various types of vessels entering and leaving major commercial ports. The purpose of this study is to provide a basis for maritime traffic management by the competent authorities and reduce the probability of ship collisions at sea. Therefore, the accuracy and correctness of AIS data significantly impact the accuracy of the analysis. In the 2020 paper "Analysis of Ship AIS Applications in the Waters and Ports Surrounding Taiwan" by Huang and Chen et al., several issues regarding the current operation of the AIS were identified. Specifically, the paper highlights instances of missing or incorrect parameters in the AIS data fields of terrestrial AIS base stations receiving data from ships. To address this concern, we utilize AIS receiving stations to collect ship AIS data and verify whether the received data exhibits issues such as missing or incorrect parameters in the aforementioned AIS data fields. Examination of the system database interface for AIS data decoding, depicted in Figure 1, reveals that certain parameters in the AIS data indeed exhibit abnormal conditions.

Marine Research of the National Academy of Marine Research
DOI: 10.29677/MR.202406_4(1).0006

413389080	0	8.6	0	118.59984
416009194	0	0	0	120.136513
413400630	0	11.3	0	118.837022
416008000	0	0.5	0	121.746733
416488000	0	14.5	1	119.900283
413695810	-128	4.8	0	118.153383
256548000	1	18.5	1	119.569773
413700470	-128	0.1	0	118.065673
414401540	2	8.8	0	120.689855

MMSI	ROT	SOG	Position_Accuracy	Longitude	Latitude	COG	True_Heading	Ship_and_Cargo_Type	Fixing_Device	MAX_Draught	Destination	DTE	Record_Time
416006693	0	0.8	0	120.034742	23.56447	144	169	40	1	2.2	YUNLIN OWF	0	2024-03-25 16:03:05.000
538006052	0	11.3	0	118.585723	23.893123	49.9	46	70	1	12.7	KR PTK	0	2024-03-25 16:03:05.000
563096700	0	15.7	1	119.585767	22.493117	252	254	74	1	12.3	TW KHH > HK HKG	0	2024-03-25 16:03:05.000
414694000	0	11.7	0	119.046027	24.539828	47.3	45	70	1	7	CAO FEI DIAN	0	2024-03-25 16:03:05.000
370341000	0	19.1	0	119.231258	24.340322	235.4	235	72	1	14.7	CNNSA	0	2024-03-25 16:03:05.000
412702910	0	11.8	0	119.899317	26.198783	29.5	27	70	0	3.8	NING DE	0	2024-03-25 16:03:05.000
247009600	0	15.8	1	120.48395	25.54495	220	222	74	1	12.5	CN NGB>SG SIN	0	2024-03-25 16:03:05.000
413389080	0	8.6	0	118.59984	24.373238	49.9	46	70	1	7.4	CHANGJIANGKOU	0	2024-03-25 16:03:05.000
416009194	0	0	0	120.136513	23.382117	12.8	90	70	1	3.2	LONGMEN	0	2024-03-25 16:03:06.000
413400630	0	11.3	0	118.837022	24.678855	44.8	43	80	1	4.5	MEIZHOU WAN	0	2024-03-25 16:03:06.000
416008000	0	0.5	0	121.746733	25.1505	198	301	70	1	9	KEELUNG	0	2024-03-25 16:03:06.000
416488000	0	14.5	1	119.900283	23.34915	2	1	70	1	9.1	TWTXG	0	2024-03-25 16:03:06.000
413695810	-128	4.8	0	118.153383	24.35675	310.2	286	33	1	2.5	XIAMEN	0	2024-03-25 16:03:06.000
256548000	1	18.5	1	119.569773	24.770605	232	233	71	1	10.7	CNSHK	0	2024-03-25 16:03:06.000
413700470	-128	0.1	0	118.065673	24.461172	220.7	175	60	0	2.1	XIA MEN	0	2024-03-25 16:03:06.000
414401540	2	8.8	0	120.689855	26.228212	204.7	204	70	1	9.4	YANG JIANG	0	2024-03-25 16:03:06.000

Figure 1. AIS data field erroneous parameter illustration.

Therefore, for this study, the AIS data provided by the National Academy of Marine Research (2024) was used to analyze the decoded AIS data for both dynamic and static information. In the first phase, based on the international standards set forth in the ITU M.1371-5 protocol for AIS data field parameters, non-compliant and duplicate AIS data was removed. This included parameters such as MMSI, latitude and longitude, ship speed, ship heading, and bow direction. In the second phase, we used MATLAB to analyze the dynamic information of AIS data. Ship speed, heading, bow direction, latitude and longitude, and other dynamic data points were graphically represented. By analyzing a large amount of data, different ship behavior characteristics under various navigation states (e.g., underway, at anchor, berthed) could be understood. Ship behavior feature data was graphically represented to identify ship behavior patterns and determine if they conformed to the continuous changes in AIS dynamic data, including the influence of speed changes on navigation distance, the effect of heading on latitude and longitude, and the relationship between heading and bow position. Data algorithms were used to calculate AIS data that exhibited abnormal ship behavior patterns and to subsequently exclude them. In the third phase, AIS data of good quality, processed and analyzed through data algorithms, was systematically stored in a database. Subsequently, this data can be displayed on the developed interface of the dynamic ship information system.

The remainder of this study is organized as follows: Section II provides an overview of the key technologies used in this study, including AIS, APIs, data transmission, and MATLAB. It also highlights the current application of AIS technology in Taiwan for various maritime purposes. The primary objective of Section III is to explain the methods used to ensure the accuracy and reliability of the AIS data used in the study by removing erroneous or inconsistent data formats. By organizing the AIS data according to the defined criteria, the study could proceed with reliable and consistent data for further analysis and interpretation. Section IV aims to provide a streamlined and efficient process for handling AIS data, ensuring the accuracy of the data and facilitating its utilization for further analysis and decision-making purposes. The outcomes, emphasizing the significance of the study contributions and suggesting directions for future studies to advance the field of AIS data analysis and its practical applications in maritime operations and safety are summarized in Section V.

Marine Research of the National Academy of Marine Research
DOI: 10.29677/MR.202406_4(1).0006

2 AUTOMATIC IDENTIFICATION SYSTEM

In order to strengthen maritime vessel management, navigation safety, enforcement of maritime regulations, and rapid response to maritime incidents, relevant maritime safety agencies have established ship monitoring systems. The Automatic Identification System (AIS) plays a crucial role in monitoring the dynamic movements of vessels, enabling the tracking of distressed vessels and improving rescue efficiency. This system exchanges electronic data with neighboring vessels, AIS shore stations, and satellites, allowing AIS information to be integrated into maritime radar systems. By prioritizing collision avoidance, the system helps prevent accidents in maritime traffic. Additionally, it can broadcast weather information and danger warning zones to vessels, thereby enhancing navigational safety. Huang et al., (2019) discussed the use of AIS receiving stations established by the Harbor and Marine Technology Center. These stations collect AIS data and utilize ship dynamic information as the basis for analysis and applications. This allows for a clearer understanding of the fluctuation of ship traffic in major ports in Taiwan and its surrounding waters, providing valuable insights for ship channel planning and policy implementation. By the end of 2017, AIS base stations in Taiwan had been installed in various locations, including Keelung Port, Taipei Port, Miaoli Waipu Fishing Port, Taichung Port, Changhua Wangong Fishing Port, Budai Port, Tainan Anping Port, Kaohsiung Port, Hualien Port, Pingtung (Donggang, Cat Nose, Xuhai), Taitung (Fu'ao, Changbin), Yilan (Su'ao Port, Toucheng), and New Taipei (Ruifang, Shimen), totaling 18 locations in Taiwan proper, and including outlying islands such as Penghu (Magong Port, Jibei Island), Orchid Island Kaiyuan Port, Kinmen (Shuitou Port, Wuqiu), and Matsu (Beigan, Dongyin, Dongju, Fuao Port), totaling 9 locations. In total, 27 AIS receiving stations were established (Liu et al., 2019). The AIS data collected from these stations is utilized to construct a real-time ship information system for the waters surrounding Taiwan. This system offers publicly available information on ship navigation, including vessel types, Maritime Mobile Service Identities (MMSI), speed, heading, latitude and longitude, destination ports, and departure locations, which encompasses both dynamic and static ship information, as depicted in Figure 2.

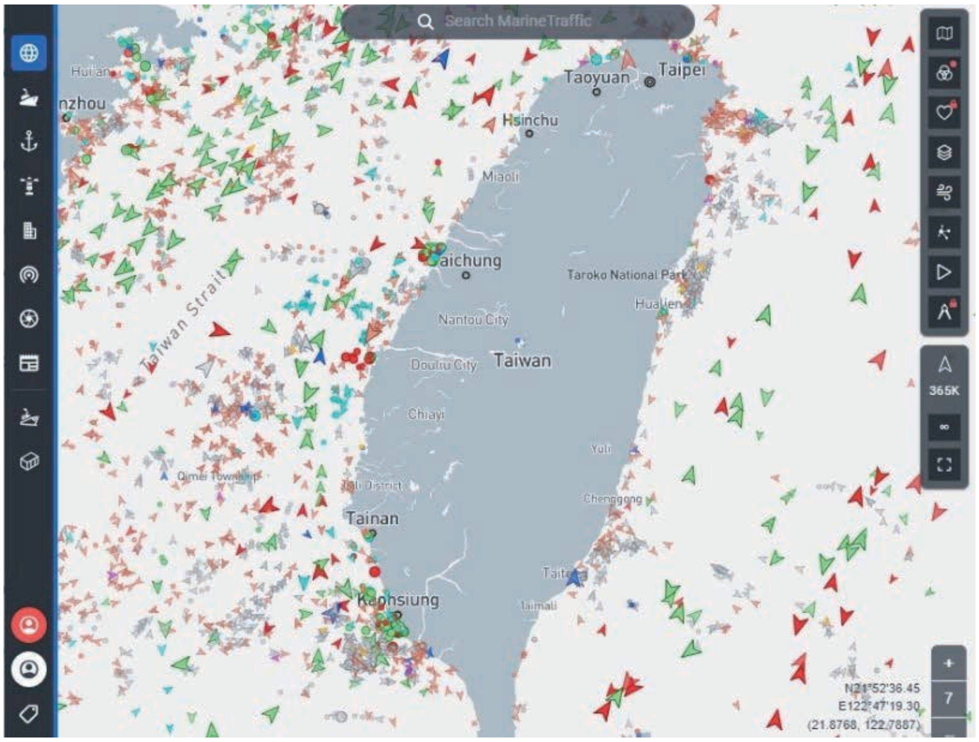

Figure 2. The real-time ship information system. (Marine Traffic, 2024)

Marine Research of the National Academy of Marine Research
DOI: 10.29677/MR.202406_4(1).0006

In cooperation with the MOTC to promote the Smart Navigation and Safety Service Development Plan, since 2018, the Maritime and Port Bureau has been leading the integration of AIS. Based on the national lighthouse locations, 19 beacon stations and 14 base stations have been established. These stations provide services such as electronic navigation, early warning, and monitoring to enhance vessel navigation safety, reduce maritime incidents, strengthen maritime rescue operations, monitor traffic flow in Taiwanese waters, and continuously assess navigation risks for the planning and adjustment of beacon and navigational systems. Additionally, the system has integrated the existing 16 AIS receiving stations to serve as signal reinforcement and backup.

The AIS operates in three modes (Wijaya and Zhang, 2013; Perera et al., 2015; Wei et al., 2020). The first mode is automatic continuous transmission of vessel information to external entities. The second mode involves vessels providing monitoring capabilities to the competent authorities within their jurisdiction. The third mode involves mutual data exchange and response between vessels and the competent authorities. The purposes of the AIS are to provide vessel identification, aid in tracking vessel targets, simplify vessel information exchange, and prevent vessel collisions. AIS messages provide two types of data: static information and dynamic information (Goerlandt and Kujala, 2014; Xiao et al., 2015; Xin et al., 2019). Static information includes vessel details such as vessel name, IMO number, MMSI, vessel type, length, width, and more. Dynamic information includes data like longitude, latitude, heading, speed, and other relevant parameters. The AIS enables accurate and timely reporting of vessel information. AIS electronic signals are transmitted through the air to shore-based receiving stations. These signals are then transferred to the processing center via a wired network for decoding and computation (Dogancay et al., 2021; Qu et al., 2019; Zaman et al., 2015). The processing center utilizes the AIS messages to build a database, which incorporates static and dynamic vessel data, weather data, electronic charts, and other relevant information (Rawson & Brito, 2021). This database is analyzed and processed to provide the latest information services to various government and search and rescue agencies (Mao et al., 2018; Rong et al., 2015; Wang et al., 2013). As mentioned in the preceding section regarding the current development of AIS technology and the goals of this study, our focus is on establishing a quality management mechanism for the AIS database. The aim is to establish an operational and standardized AIS data quality control mechanism that effectively identifies and eliminates abnormal or inconsistent AIS data within the system. In related research on AIS data cleaning and management, a spatial clustering method (SPTCLUST-II) is introduced. This study clusters a substantial amount of ship trajectory data in a short time by extracting historical data from ship AIS, enabling an understanding of maritime vessel navigation behavior. This, in turn, may contribute significantly to the benefits of ship traffic management in intelligent transportation systems. The study also outlines a method for clustering vessel trajectory segments and extracting maritime traffic routes, comprising four main parts. The first part involves the collection and cleaning of AIS data. Initially, the study includes the removal of duplicate data in the collected raw AIS data. Through the use of GIS applications, the original AIS data was verified, encompassing checking its spatial distribution, comparing data discrepancies within the same vessel AIS entry, and removing observations outside the target area and land-based observations.

According to the ITU-R M.1371-5 recommendation (Murray and Perera, 2021), there are 27 defined data types for AIS data, and each data type has its own defined data field format. This study utilizes various AIS static and dynamic data, along with their definitions and data formats, as shown in Table 1 below:

Table 1. AIS Data types, definitions, and formats. (Zhang et al., 2021)

Data Type	Definition	Data Field Format
Type 1	Position Report Class A	Format A
Type 2	Position Report Class A (Assigned)	Format A
Type 3	Position Report Class A (Response)	Format A
Type 4	Base Station Report	Format B
Type 5	Static and Voyage-related Data	Format B
Type 6	Binary Addressed Message	Format B
Type 7	Binary Acknowledge	Format B
Type 8	Binary Broadcast Message	Format B
Type 9	Standard SAR Aircraft Position Report	Format C
Type 10	UTC/Date Inquiry	Format D
Type 11	UTC/Date Response	Format D
Type 12	Addressed Safety-Related Message	Format E
Type 13	Safety-Related Acknowledge	Format E
Type 14	Safety-Related Broadcast Message	Format E
Type 15	Interrogation	Format F
Type 16	Assignment Mode Command	Format F
Type 17	DGNSS Broadcast Binary Message	Format G
Type 18	Standard Class B CS Position Report	Format H
Type 19	Extended Class B Equipment Position Report	Format H
Type 20	Data Link Management	Format I
Type 21	Aid-to-Navigation Report	Format J
Type 22	Channel Management	Format K
Type 23	Group Assignment Command	Format L
Type 24	Static Data Report	Format M
Type 25	Single Slot Binary Message	Format H or I
Type 26	Multiple Slot Binary Message	Format H or I
Type 27	Long Range AIS Broadcast Message	Format H or I

Note: The data field formats mentioned above (A, B, C, D, E, F, G, H, I, J, K, L, M) refer to the specific formatting structures defined in the ITU-R M.1371-5 recommendation.

Marine Research of the National Academy of Marine Research
DOI: 10.29677/MR.202406_4(1).0006

3 AIS DATA CLEANING AND DATA ORGANIZATION OF ERRONEOUS FORMATS

AIS data cleaning and data organization involve the process of removing erroneous formats and organizing AIS data to ensure accuracy and consistency. This process can be divided into two main steps: data filtering and handling of erroneous formats. During the data filtering stage, AIS data that meets specific criteria or standards is selected. This may include removing data with empty or default MMSI values, eliminating duplicate data, and discarding data that does not comply with the defined international standards for AIS data as specified in ITU-R M.1371-5 recommendations. In the handling of erroneous formats stage, the focus is on addressing AIS data that does not adhere to the correct format. This could involve handling data with missing essential data fields or erroneous data field formats (Bengio et al., 2014; Karataş et al., 2021). Erroneous format data can be corrected or excluded to ensure the integrity and consistency of the dataset. By implementing AIS data cleaning and data organization of erroneous formats, a dataset that conforms to the standards and contains accurate information can be obtained (Daranda 2016; Eljabu et al., 2022; Nguyen et al., 2018). This provides a reliable foundation for subsequent analysis and applications.

The AIS data used in this study was provided by the National Academy of Marine Research (2024). Through research, it has been discovered that certain AIS data parameters do not comply with international standard specifications. Additionally, some dynamic and static information contains default or empty values. Moreover, due to the multiple reception and decoding of AIS signals by multiple shore-based receiving stations, the same AIS data is stored repeatedly. These errors in AIS data can negatively affect data accuracy, while duplicate AIS data can increase the burden on the database. Therefore, this study first focused on the preliminary exclusion of erroneous format data and the removal of duplicate AIS data from the original AIS data received from the NAMR. This process aimed to provide accurate AIS data for subsequent parameter analysis. The flowchart of the preliminary AIS data organization process is shown in Figure 3.

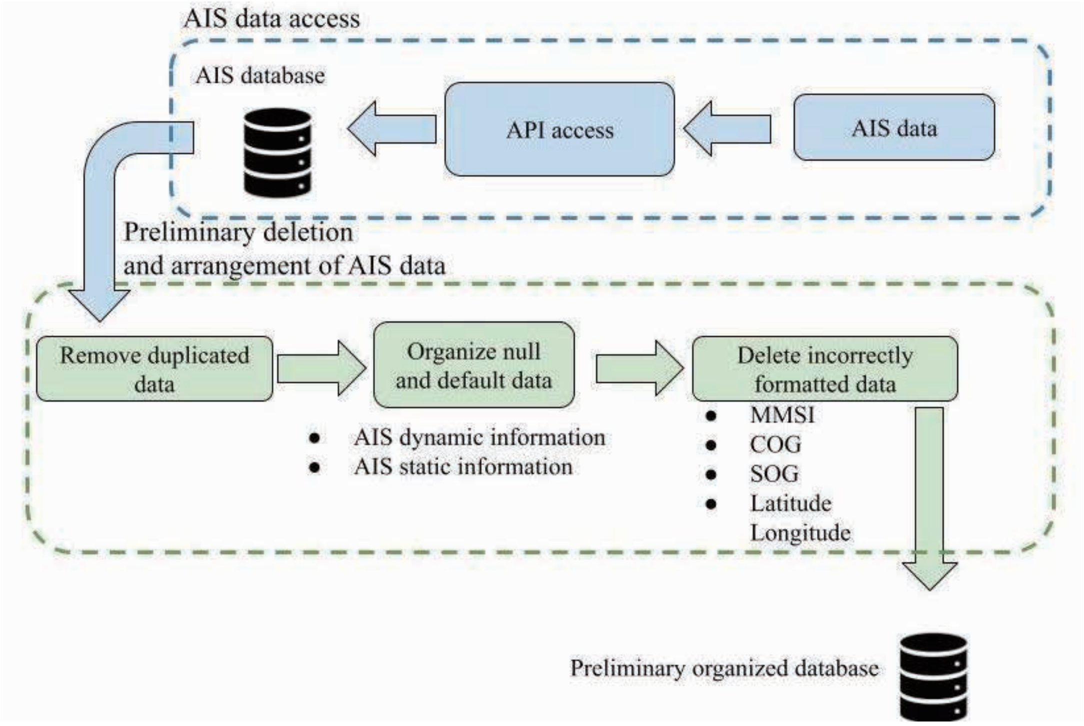

Figure 3. AIS preliminary data collation flowchart.

Marine Research of the National Academy of Marine Research
DOI: 10.29677/MR.202406_4(1).0006

3.1 AIS data access

The data provided by the NAMR was accessed through an HTTP API, and the data format was JSON, as shown in Figure 4. In this study, a web crawler was used to scrape the data from the HTML API webpage. The entire JSON data on the webpage was copied and converted into a data file format that could be read by the database (such as TXT, Excel, CSV, etc.). The data from the file was imported into the designated database location, table, columns, and data types as specified.


```
$PGHP,1,2023,5,4,20,2,59,0,416,,994161810,1,16*22
\s:994161810,c:1683230579*01\!ANVDM,1,1,7,B,EN14bJg64bPVh32@9T4SVV3ph@@D>IQH6mNvH00003v010,4*16
$PGHP,1,2023,5,4,20,2,59,0,416,,994161808,1,77*2C
\s:994161808,c:1683230579*08\!ANVDM,1,1,6,B,169L17502<8Vio<>m>ai30o108Rv,0*77
$PGHP,1,2023,5,4,20,2,58,0,416,,4161814,1,5F*53
\s:004161814,c:1683230578*04\!BSVDM,1,1,,B,169L17502<8Vio<>m>ai30o108Rv,0*5F
$PGHP,1,2023,5,4,20,2,59,0,416,,994161839,1,72*2B
\s:994161839,c:1683230579*0A\!ANVDM,1,1,6,A,16:91b0002`MvGh=nmD@7aCn2<3d,0*72
$PGHP,1,2023,5,4,20,2,59,0,416,,994161839,1,33*2E
\s:994161839,c:1683230579*0A\!ANVDM,1,1,7,B,36:7Sv1P00`NILP=nd5RF?wn2<5b,0*33
$PGHP,1,2023,5,4,20,2,59,0,416,,994161839,1,26*2A
\s:994161839,c:1683230579*0A\!ANVDM,1,1,8,B,177P7S001r`NvTh=l<1`bFqn0HS3,0*26
$PGHP,1,2023,5,4,20,2,58,0,416,,4161807,1,58*2F
\s:004161807,c:1683230578*06\!BSVDM,1,1,,A,B6<fa`00IR9UV7S@5@:cVvu3QP06,0*58
$PGHP,1,2023,5,4,20,2,59,0,416,,994161832,1,5B*52
\s:994161832,c:1683230579*01\!ANVDM,1,1,3,A,16<f`QwP00`U3I8>v3`6f0wn20S9,0*5B
$PGHP,1,2023,5,4,20,2,58,0,416,,4161810,1,4D*54
\s:004161810,c:1683230578*00\!BSVDM,1,1,,B,B6=D8J@01B:r9TSCNRMWh0101P06,0*4D
$PGHP,1,2023,5,4,20,2,59,0,416,,994161837,1,2D*56
\s:994161837,c:1683230579*04\!ANVDM,1,1,8,B,B6:7;Q00>j9<Ht3gjFj:3wv41P06,0*2D
$PGHP,1,2023,5,4,20,2,58,0,416,,4161812,1,1E*52
\s:004161812,c:1683230578*02\!BSVDM,1,1,,B,15SGF>002?8j>n<>JPqH:VR006P<,0*1E
$PGHP,1,2023,5,4,20,2,59,0,416,,994161824,1,44*22
\s:994161824,c:1683230579*06\!ANVDM,1,1,5,B,36<fv<?P00`eOMv>HpDv4?wn25:C,0*44
$PGHP,1,2023,5,4,20,2,59,0,416,,994161836,1,47*22
\s:994161836,c:1683230579*05\!ANVDO,1,1,4,,E>161c0cbaT4T14P00000000000DEqQ16aAjP30HHKuf00,4*47
$PGHP,1,2023,5,4,20,2,59,0,416,,994161836,1,7E*53
\s:994161836,c:1683230579*05\!ANVDM,1,1,8,B,C>10dtP07j;1vkSCJ192kwu06aUKeScWUa11111111bd0000002P,0*7E
$PGHP,1,2023,5,4,20,2,59,0,301,,994161809,1,1A*5C
\s:994161809,c:1683230579*09\!ANVDM,1,1,1,A,16:60;P100`N9t0>37IT1pWn0HS3,0*1A
$PGHP,1,2023,5,4,20,2,59,0,416,,994161806,1,7C*56
\s:994161806,c:1683230579*06\!ANVDM,1,1,1,B,16<fbL@P018V>Gh<hENFkOwd0000,0*7C
$PGHP,1,2023,5,4,20,2,59,0,416,,994161813,1,41*23
\s:994161813,c:1683230579*02\!ANVDM,1,1,5,B,13aQf?P005`UI7L=Pdu0DCUn28Rw,0*41
$PGHP,1,2023,5,4,20,2,59,0,416,,994161808,1,0D*58
\s:994161808,c:1683230579*08\!ANVDM,1,1,7,B,16:?vT001;8U@Np>bsV82bsn08Rw,0*0D
$PGHP,1,2023,5,4,20,2,59,891,416,,994161833,1,3E*52
\s:994161833,c:1683230579891*30\!ANVDM,1,1,8,A,C8N4HP@00R9md<S=c0du@iN04bNjwce;1111111111100000002P,0*3E
$PGHP,1,2023,5,4,20,2,58,0,416,,4161807,1,10*23
\s:004161807,c:1683230578*06\!BSVDM,1,1,,B,16<fv>wP00`Vas@<s85>4?wn2@S5,0*10
$PGHP,1,2023,5,4,20,2,59,920,416,,994161820,1,34*2A
\s:994161820,c:1683230579920*39\!ANVDM,2,1,7,B,56<fet@000008c?GK21<QDF0IF1UF0pv3?J2220N000000@Q108888888888,0*1B
\s:994161820,c:0*34\!ANVDM,2,2,7,B,88888888888,2*2F
$PGHP,1,2023,5,4,20,2,58,0,416,,4161805,1,3E*56
\s:004161805,c:1683230578*04\!BSVDM,1,1,,B,16<fpC0P00`V66h=VPe>4?wn2D0f,0*3E
$PGHP,1,2023,5,4,20,2,58,0,416,,4161805,1,7A*56
\s:004161805,c:1683230578*04\!BSVDM,1,1,,A,16<g9rkP@08Tgof=1jDufGn008S5,0*7A
$PGHP,1,2023,5,4,20,2,58,0,416,,4161805,1,70*27
\s:004161805,c:1683230578*04\!BSVDM,1,1,,A,16<g29@0068V0s8=h?Ekc9aj06P8,0*70
$PGHP,1,2023,5,4,20,2,59,0,416,,4161811,1,19*2C
\s:004161811,c:1683230579*00\!BSVDM,1,1,,B,16<fj0OP00`dk`d=fgtcVOwb0<0M,0*19
$PGHP,1,2023,5,4,20,2,59,0,416,,994161837,1,66*20
\s:994161837,c:1683230579*04\!ANVDM,1,1,9,A,C6904M@0029ooS3h1Eh000N1QkScKKkK1U;11111111003h0PB2P,0*66
$PGHP,1,2023,5,4,20,2,59,0,416,,994161837,1,4E*51
\s:994161837,c:1683230579*04\!ANVDM,1,1,0,B,16;3PR001F``9@P>sjC`hnon2<EG,0*4E
$PGHP,1,2023,5,4,20,2,59,0,416,,4161801,1,2F*51
\s:004161801,c:1683230579*01\!BSVDM,1,1,,B,C695=I@4AJ8T1tSI5h`f0wP0`B:HNL?SgcgKSgKkk:0000000020,0*2F
$PGHP,1,2023,5,4,20,2,59,0,416,,4161801,1,4C*52
\s:004161801,c:1683230579*01\!BSVDM,1,1,,B,B6<fsj@0HR8VVQ3IMB9N9>u3QP06,0*4C
$PGHP,1,2023,5,4,20,2,59,0,416,,994161825,1,3E*55
```


Figure 4. AIS data formats.

DOI: 10.29677/MR.202406_4(1).0006

An application programming interface (API) allows for the analysis and processing of input data and its presentation or storage in a designated interface or environment. By utilizing an API, communication between different hardware devices, computer software, and platform facilities can be achieved. An API serves as a bridging interface between different environments, significantly reducing the complexity and cost of data transmission and computation. With the increasing scale and variety of software and hardware in recent years, communication between different complex systems has become crucial. APIs enable the proper division of responsibilities among software systems in different environments. Well-designed APIs can reduce interdependencies between systems, thereby enhancing system maintainability and scalability, as shown in Figure 5.

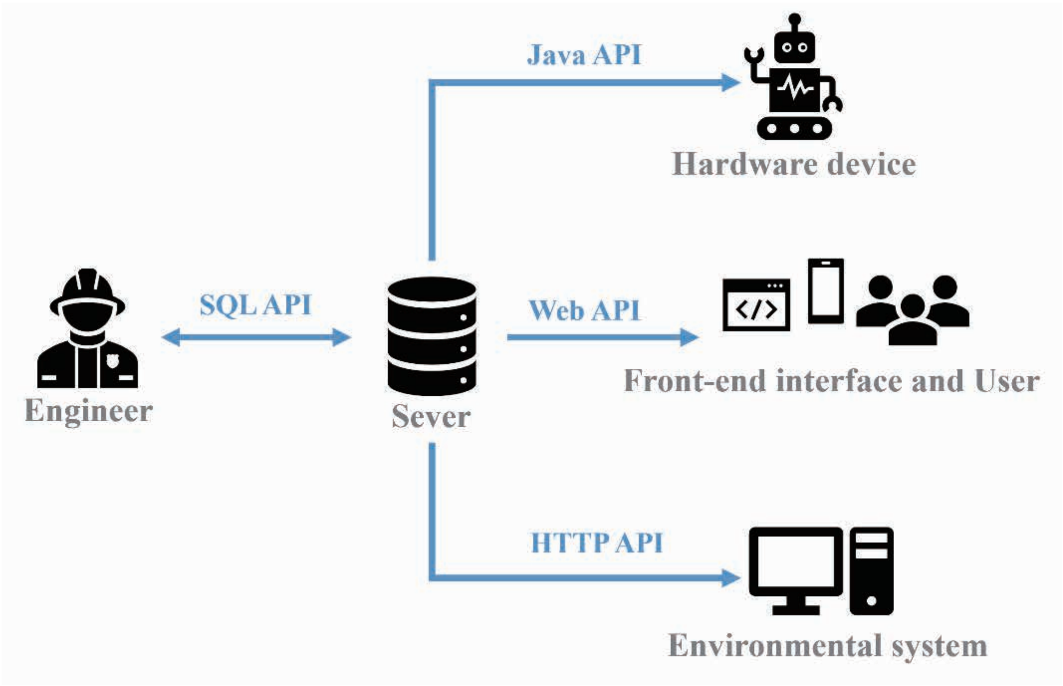

Figure 5. API interface architecture.

In this study, the powerful data interoperability of APIs was leveraged between different software applications. AIS data was transmitted in different data formats among compiled programs, front-end webpages, mobile devices, and backend databases, while preserving the integrity of the original data. APIs are categorized into five main types. This study utilized protocol-based APIs, device APIs, and web APIs to achieve high compatibility across different operating environments. The data format is an important component of APIs, and the designers of these interfaces have the freedom to determine the format and specifications of the data to be returned. As for an HTTP API, it serves as the communication medium for data transmission between the database server and the user interface. HTTP utilizes the HTML structure as the original format for data exchange between web servers and computer system browsers. Nowadays, users can access the internet on various devices such as computers, smartphones, and wearable devices. The availability of different devices allows for diverse approaches to data processing. When using an HTTP API or web API to transmit data, there are commonly two response formats. In this study, the JSON format was used as the primary format for transmitting AIS data.

DOI: 10.29677/MR.202406_4(1).0006

3.2 Duplicated data deletion

Taking the example of accessing AIS data from the NAMR, as shown in Figure 6, the left half of the image shows AIS data for a vessel with MMSI 412508457. At the timestamps of 13:58:15 and 13:40:15, there are 11 and 6 data entries, respectively. The values for ship speed (SOG), course over ground (COG), longitude, and latitude are identical for these entries. In the right half of the image, AIS data for a vessel with MMSI 412950009 is displayed. There are 3 and 9 data entries highlighted with red and blue boxes, respectively. The values for ship speed (SOG), course over ground (COG), longitude, and latitude are the same for these entries, with only the timestamp (Record_Time) being different. The data marked with arrows has the same timestamp (Record_Time) as the AIS data in the blue box, while the other values differ. These data entries are considered abnormal data.

MMSI	ROT	SOG	Longitude	Latitude	COG	True_Heading	Record_Time
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0	118.766382	24.745142	301.1	511	2022-05-16 13:40:15.000
412508457	128	0	118.766382	24.745142	301.1	511	2022-05-16 13:40:15.000
412508457	128	0	118.766382	24.745142	301.1	511	2022-05-16 13:40:15.000
412508457	128	0	118.766382	24.745142	301.1	511	2022-05-16 13:40:15.000
412508457	128	0	118.766382	24.745142	301.1	511	2022-05-16 13:40:15.000
412508457	128	0	118.766382	24.745142	301.1	511	2022-05-16 13:40:15.000

MMSI	ROT	SOG	Longitude	Latitude	COG	True_Heading	Record_Time
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:46.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:46.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:46.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:46.000
412950009	128	1.2	120.159941	26.233512	158	511	2021-11-06 00:10:47.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:47.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:47.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:47.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:47.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:47.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:47.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:47.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:47.000

Figure 6. Duplicated AIS data.

After removing duplicate entries for the vessel with MMSI 412508457, only 2 data entries with the timestamps of 13:58:15 and 13:40:15 remained. For the vessel with MMSI 412950009, only 2 data entries with the timestamps of 0:10:46 and 0:10:47 were retained, as shown in Figure 7.

MMSI	ROT	SOG	Longitude	Latitude	COG	True_Heading	Record_Time
412508457	128	0.2	118.766345	24.745233	299.3	511	2022-05-16 13:58:15.000
412508457	128	0	118.766382	24.745142	301.1	511	2022-05-16 13:40:15.000

MMSI	ROT	SOG	Longitude	Latitude	COG	True_Heading	Record_Time
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:46.000
412950009	128	0.8	120.159893	26.233472	176	511	2021-11-06 00:10:47.000

Figure 7. None duplicate AIS data.

3.3 Erroneous format data cancelation

Some of the AIS dynamic data fields, including SOG, COG, true heading, ship rate of turn (ROT), latitude, and longitude, contained default values or were empty, including AIS data with ship speed (SOG) equal to 102.3 knots, ground heading (COG) equal to 360 degrees, true heading (True Heading) equal to 511 degrees, ship rate of turn (ROT) equal to ±128°/min, longitude (Longitude) equal to 181, and latitude (Latitude) equal to 91. Since the content of these data entries could not be determined as erroneous, this study deleted the aforementioned data. The AIS data used in this study is the raw data provided by the NAMR, which contains many erroneous format entries. These erroneous format AIS data entries can impact subsequent data analysis. For this study, data entries with MMSI codes less than 9 digits or greater than 10 digits, ship speeds (SOG) exceeding 70 knots, ship headings (COG) greater than 360 degrees or less than 0 degrees, true headings exceeding 360 degrees or less than 0 degrees, longitudes exceeding 180 degrees or less than -180 degrees, and latitudes exceeding 90 degrees or less than -90 degrees were deleted.

4 AIS DATA AUTOMATIC CALCULATIONS AND DEBUGGING

After the initial sorting process to eliminate AIS data that did not comply with international standards, duplicate data, and dynamic data with default values or empty fields, there may still be errors in the remaining AIS data. It is not possible to determine the accuracy of individual AIS data entries based on their values alone. Therefore, the ship's heading, speed, and distance traveled based on two consecutive AIS data entries for each vessel were calculated for this study. The calculated values were compared and analyzed against the actual AIS data values, and any abnormal AIS data was removed. The cleaned AIS data was imported into the database, effectively reducing the system and database load by removing vessel data with unclear behavior trajectories.

4.1 AIS parameters calculation

In this study, the AIS data was classified based on the same MMSI. AIS data with the same MMSI was grouped into the same category, and then further divided into two categories based on the ship's speed (SOG): stationary vessel data and moving vessel data. To avoid misidentifying vessels that were undergoing deceleration or acceleration as stationary vessels, a condition had to be met. The difference in ship's speed (SOG_1-SOG_2) between two consecutive AIS data entries (D_1 and D_2) had to be 0 to be classified as stationary vessel data. If the difference in ship's speed (SOG_1-SOG_2) between D_1 and D_2 was not 0, then the ship's speed ($SOG_{1/2}$) of D_1 and D_2 was individually evaluated. If the ship's speed ($SOG_{1/2}$) is 0, it was classified as stationary vessel data. Otherwise, it was classified as moving vessel data. The relevant process flowchart is shown in Figure 8.

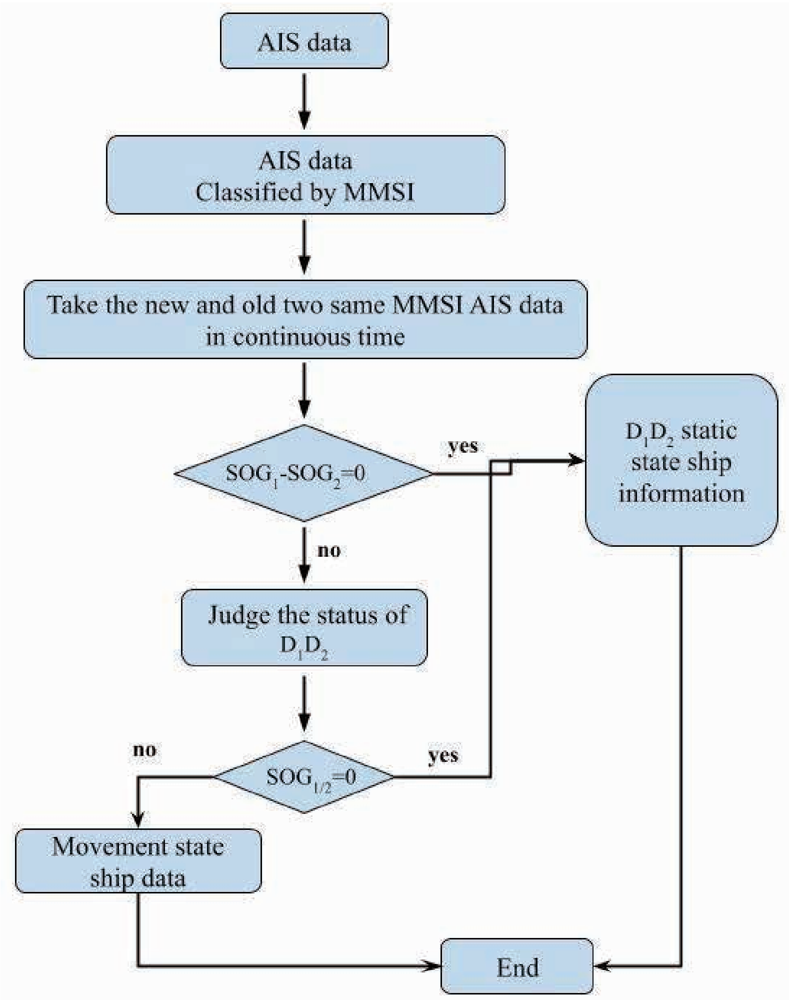

Figure 8. AIS data classification flowchart.

4.2 Calculation and deletion of the latitude and longitude of ships in abnormal static state

To determine if the stationary position of the same vessel was abnormal, we selected the current AIS data and the previously received AIS data, defining them as D_1 and D_2, respectively. The great circle distance, denoted as d, between D_1 and D_2 was then calculated. The purpose was to assess whether the linear distance change in latitude and longitude between D_1 and D_2 was normal for a stationary vessel without any acceleration. Upon examination, it was observed that the GPS devices used in AIS data may have introduced fluctuations in the fourth decimal place of longitude and latitude readings, even when there was no actual change in position. These fluctuations were considered as positioning errors. Considering that the length of the fourth decimal place in longitude and latitude is approximately 10 meters of great circle distance, and based on the international standard specified in ITU-R M.1371-5, the positioning accuracy error of GPS devices should not exceed 10 meters. Therefore, if the great circle distance d between D_1 and D_2 exceeded 10 meters, it was classified as abnormal data. The process for determining abnormal ship latitude and longitude positions is illustrated in the right flowchart of Figure 9. Since the Earth is a sphere, the calculation of latitude and longitude requires considering the curvature of the sphere. The Haversine formula was used in this study to calculate the great circle distance (1). λ_1, ϕ_1, λ_2, and ϕ_2 represent the longitude and latitude of two points, and $hav(\theta)$ represents the half of the chord length between two points on the surface of the sphere for the central angle θ. The great circle distance d is measured in kilometers, and r is the average radius of the Earth, which is approximately 6,371 kilometers.

$$hav(\theta) = hav(\phi_2 - \phi_1) + \big(1 - hav(\phi_1 - \phi_2) - hav(\phi_1 + \phi_2)\big) * hav(\lambda_2 - \lambda_1)$$

$$d = 2r\ arcsin\left(\sqrt{hav(\theta)}\right) = 2r\ arcsin\left(\sqrt{sin^2\left(\frac{\phi_2-\phi_1}{2}\right) + cos\phi_1 * cos\phi_2 * sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right) \quad (1)$$

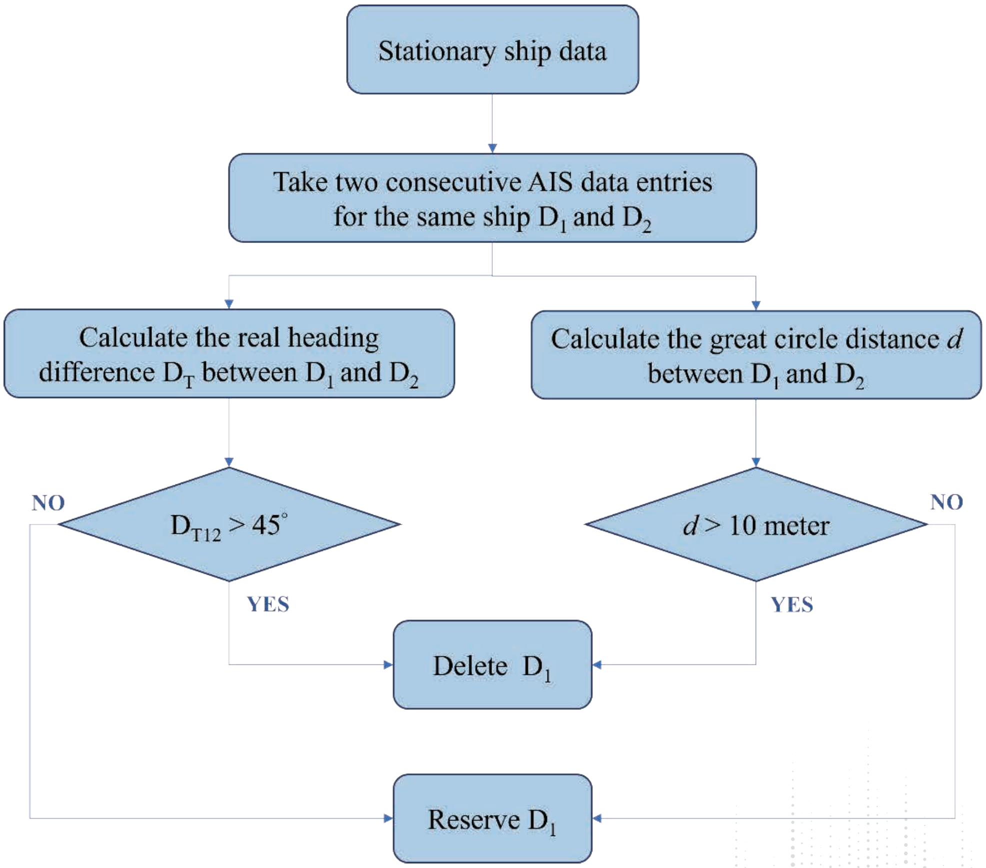

Figure 9. The process for determining abnormal latitude and longitude positions along with abnormal heading.

DOI: 10.29677/MR.202406_4(1).0006

Based on the AIS data from the NAMR, for the vessel with MMSI 440687000, all of its AIS entries had a ship speed (SOG) value of 0, indicating a stationary state. In Figure 10, this vessel was classified as a stationary vessel. However, when examining the great circle distance between this AIS entry (marked in the red box) and the consecutive AIS entry, it exceeded the average great circle distance for the vessel's stationary period as shown in the red box in Figure 11. Therefore, this AIS entry was considered abnormal and should be deleted. The abnormal location of this vessel is indicated by the red box in Figure 12, while the vessel's actual berth is represented by the yellow box. The distance between these two locations is approximately 611.883 meters according to the great circle distance calculation.

MMSI	SOG	Longitude	Latitude	Record_Time
440687000	0	119.263217	25.4267	2022-05-29 01:48:00.000
440687000	0	119.2632	25.426683	2022-05-29 02:10:00.000
440687000	0	119.263217	25.426683	2022-05-29 03:11:00.000
440687000	0	119.2632	25.42675	2022-05-29 03:49:00.000
440687000	0	119.263217	25.426717	2022-05-29 04:42:00.000
440687000	0	119.263217	25.426717	2022-05-29 04:57:00.000
440687000	0	119.264236	25.43212	2022-05-29 05:13:00.000
440687000	0	119.263217	25.4267	2022-05-29 05:14:00.000
440687000	0	119.263233	25.426667	2022-05-29 06:15:00.000

Figure 10. Ship information with MMSI of 440687000.

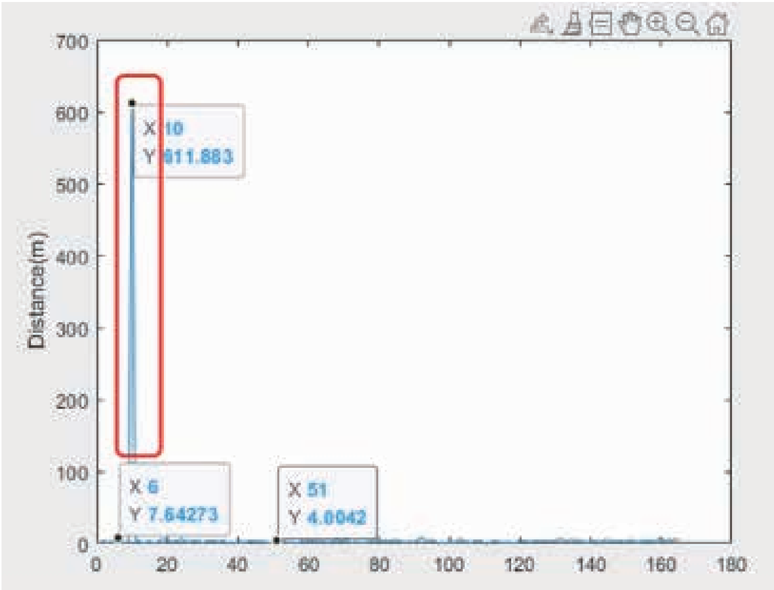

Figure 11. MMSI 440687000 mean great circle distance during ship stay.

DOI: 10.29677/MR.202406_4(1).0006

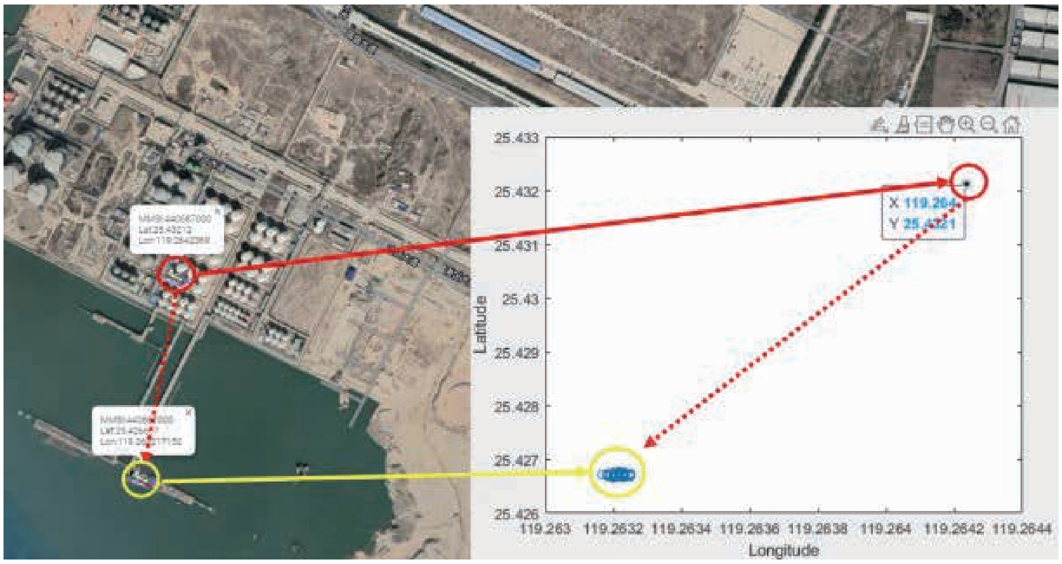

Figure 12. MMSI 440687000 unusual position of ship.

4.3 Calculation and deletion of ship course data in abnormal static state

In AIS data, there are two types of heading information: course over ground (COG) and true heading. True heading is defined as the direction of the vessel's bow, represented by the red arrow in Figure 13, while COG represents the vessel's direction of travel, indicated by the yellow dashed line in Figure 13. During vessel navigation at sea, the vessel's heading should align with its bow direction. However, when the vessel is turning, the inertia of motion causes it to continue moving in the original direction, resulting in a drift. This can introduce discrepancies between the COG and true heading. If the difference between these two values exceeded 45 degrees, the AIS data entry was considered abnormal.

During periods when the vessel is berthed or anchored, the vessel's position remains relatively unchanged. However, due to GPS positioning errors, there may be slight positional fluctuations within a radius of 0 to 10 meters around the vessel's location, as shown in Figure 14. These positioning errors can mistakenly indicate vessel displacement and introduce inaccuracies in the parameters, including COG. Therefore, when the vessel is in a stationary state, COG is not considered in the anomaly detection criteria.

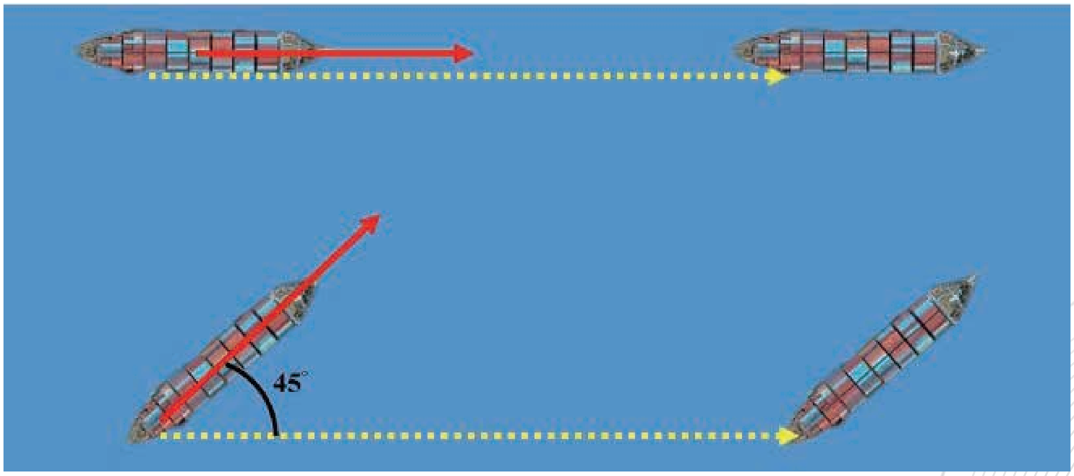

Figure 13. Schematic diagram of true heading and heading over ground.

DOI: 10.29677/MR.202406_4(1).0006

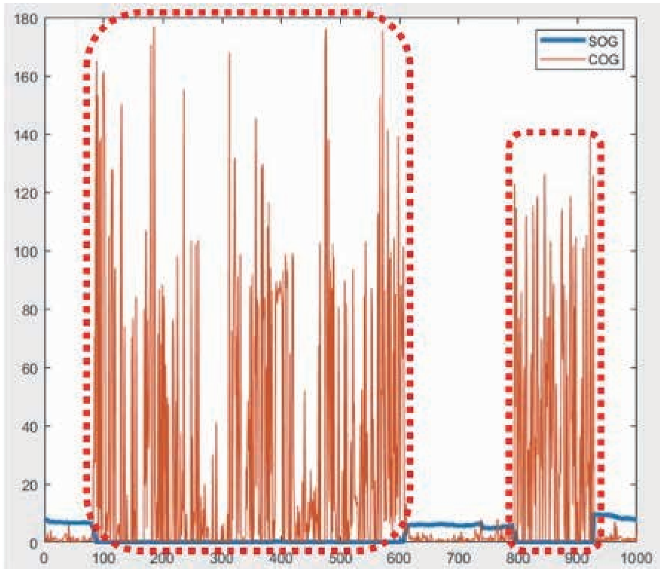

Figure 14. Schematic diagram of abnormal course over ground.

To assess whether there were anomalies in the heading data of a vessel in a stationary state, we started by obtaining two consecutive AIS data entries, D_1 and D_2, for the same vessel. Subsequently, we calculated the difference in true heading (D_T) between D_1 and D_2 to evaluate whether the true heading data was normal for a vessel without a significant rate of turn. The purpose of considering the true heading difference was to account for potential errors in sensor measurements. If the difference in vessel heading exceeded 45 degrees, indicating a significant change in the vessel's bow direction, the AIS data entry was classified as abnormal. The abnormal judgment process for ship's course data is illustrated in the left flowchart of Figure 9.

Based on the AIS data from the NAMR, the AIS data entries for the vessel with MMSI 413212680 indicate that the vessel was in a stationary state, as the vessel's speed over ground (SOG) was consistently 0, as shown in Figure 15. However, upon examining the true heading values for this vessel, there was a significant and rapid change within a minute, as shown in the highlighted region in Figure 16. The true heading difference exceeded the threshold of 45 degrees, indicating an abnormal change in the vessel's heading. Therefore, this AIS data entry was classified as abnormal and was removed from the dataset.

MMSI	SOG	True_Heading	Record_Time
413212680	0	83	2020-05-01 02:59:31.000
413212680	0	83	2020-05-01 03:00:52.000
413212680	0	84	2020-05-01 03:01:50.000
413212680	0	283	2020-05-01 03:11:52.000
413212680	0	83	2020-05-01 03:12:39.000
413212680	0	86	2020-05-01 03:15:22.000
413212680	0	86	2020-05-01 03:16:53.000
413212680	0	84	2020-05-01 03:18:42.000
413212680	0	84	2020-05-01 03:19:32.000
413212680	0	84	2020-05-01 03:20:32.000
413212680	0	83	2020-05-01 03:21:32.000
413212680	0	86	2020-05-01 03:23:02.000
413212680	0	82	2020-05-01 03:25:52.000

Figure 15. Schematic diagram of ship information with MMSI 413212680.

Marine Research of the National Academy of Marine Research
DOI: 10.29677/MR.202406_4(1).0006

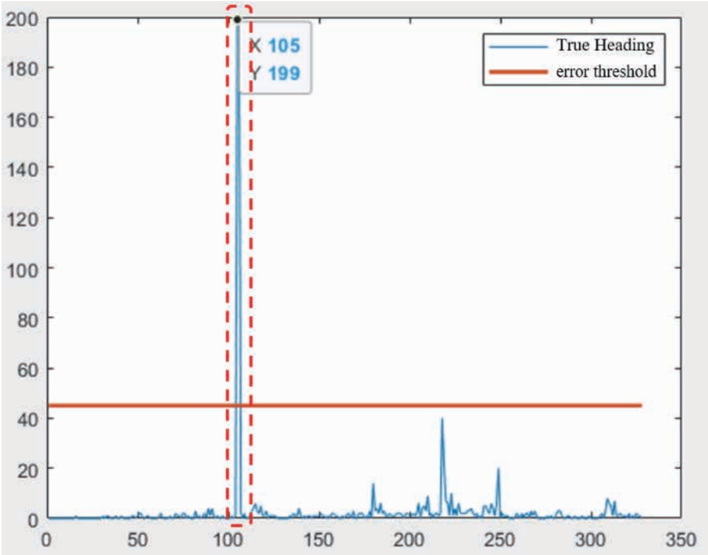

Figure 16. Schematic diagram of real heading anomalies.

4.4 Calculation and deletion of ship heading and sailing distance data in abnormal motion state

To extract the motion state data for vessels with the same MMSI, we can consider two consecutive AIS data pointsD_1 and D_2, with a time difference of $T_{12} = T_1 – T_2$. The vessel's speeds for these data points are SOG_1 and SOG_2, respectively. First, we calculated the differences between the D_1 and D_2's own ship's course over ground (DCT_1 and DCT_2) and true heading (DT_{12}). If the differences in course over ground or true heading exceeded 45 degrees, the data was classified as abnormal and removed. Next, we computed the differences in course over ground (DC_{12}) and true heading (DT_{12}) between D_1 and D_2, and analyzed the changes in heading and ship's bow angle to identify any abnormal turns exceeding 45 degrees and removed those data points. After excluding abnormal heading data, we calculated the great circle distance (d_{12}) between D_1 and D_2. In addition, we determined the vessel's distance traveled based on the speed SOG_2 and time difference T_{12}, represented as $DST_2 = SOG_2 * T_{12}$. If d_{12} was greater than DST_2, it meant that the vessel's actual traveled distance exceeded the maximum great circle distance achievable at the vessel's speed SOG_2. This suggested that the vessel was not maintaining a constant speed or decelerating but rather accelerating. To account for this, we calculated the distance traveled by D_1 using the speed SOG_1 and time difference T_{12}, represented as $DST_1 = SOG_1 * T_{12}$. Assuming an imaginary theoretical calculation point D_X at DST_1, we added the distance from D_2 to D_X ($DST_1 + DST_2 = DST_{12}$). If d_{12} was greater than DST_{12}, it meant that the vessel's traveled distance, even under acceleration, exceeded the maximum great circle distance achievable at the vessel's final speed SOG_1.

AIS devices, GPS positioning errors, ocean currents, and weather conditions can cause vessel position drift and distance measurement errors. To avoid misjudgments due to these factors, we calculated the difference between the great circle distance DST_{12} from D_2 to D_X and the great circle distance d_{12} between D_1 and D_2 (DST_d). We calculated the time required for D_1 to travel DST_d at the final speed SOG_1, represented as TST_d. If TST_d was greater than T_{12}, it meant that the time required to travel that distance was much higher than the time difference between D_1 and D_2. Consequently, such data was classified as abnormal and removed as shown in Figure 17.

Marine Research of the National Academy of Marine Research
DOI: 10.29677/MR.202406_4(1).0006

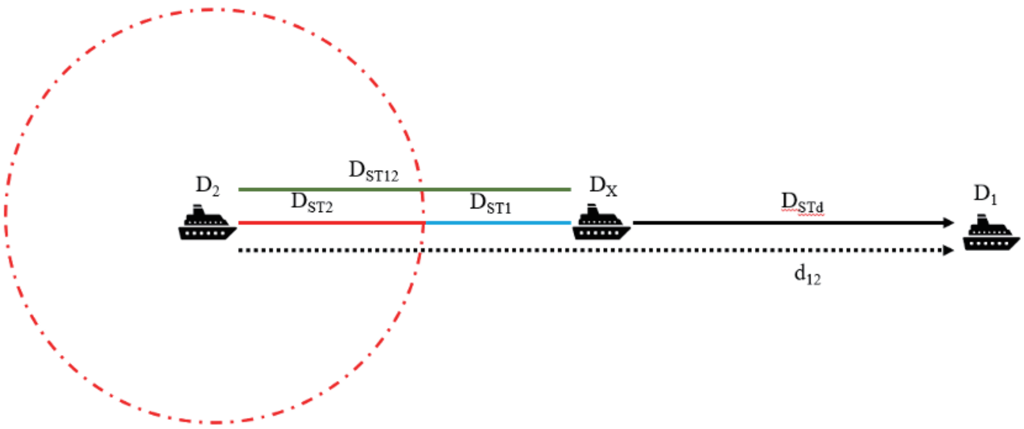

Figure 17. Diagram of abnormal sailing distance.

In the case of the vessel with MMSI 413706040 in the AIS data from the NAMR, the vessel's course over ground and true heading showed a difference of more than 140 degrees. Additionally, there was a clear contradiction between the vessel's bow direction and its actual navigation direction, as shown in Figure 19. Therefore, the AIS data highlighted in the red box in Figure 18 was classified as abnormal data.

MMSI	SOG	Longitude	Latitude	COG	True_Heading	Record_Time
413706040	9.4	117.629948333333	23.73278	34.5	32	2020-05-01 03:13:55.000
413706040	9.2	117.632881666667	23.7373783333333	28	27	2020-05-01 03:15:58.000
413706040	8.9	117.634198333333	23.7396733333333	29	25	2020-05-01 03:16:59.000
413706040	9	117.634813333333	23.7407366666667	26.7	26	2020-05-01 03:17:29.000
413706040	9	117.638451666667	23.7462083333333	36.9	34	2020-05-01 03:20:00.000
413706040	9	117.64208	23.7517216666667	28.6	27	2020-05-01 03:22:37.000
413706040	8.9	117.643266666667	23.75392	27.1	167	2020-05-01 03:23:37.000
413706040	8.7	117.643866666667	23.7550633333333	23.5	25	2020-05-01 03:24:07.000
413706040	8.9	117.64505	23.757325	29.4	27	2020-05-01 03:25:09.000
413706040	8.7	117.646963333333	23.7605433333333	27.4	27	2020-05-01 03:26:39.000

Figure 18. Schematic diagram of ship information with MMSI 413706040.

Figure 19. Schematic diagram of the abnormality of the direction of the ship's bow and the direction of the ship's navigation.

DOI: 10.29677/MR.202406_4(1).0006

In the case of a vessel with MMSI 413358520 in the AIS data from the NAMR, we considered the AIS data labeled as D_2 at 2020-05-01 01:22:54 and the AIS data labeled as D_1 at 2020-05-01 01:23:54, as shown in Figure 20. The great circle distance between D_1 and D_2 was calculated to be 825 meters. However, the calculation of the vessel's travel radius based on D_2's speed was approximately 256 meters, which was significantly less than the actual distance traveled. This indicated that the vessel was accelerating. The calculation of the vessel's travel radius based on D_1's speed was approximately 260 meters. The addition of both travel radii resulted in an estimated total of 518 meters, which was still 307 meters short of the D_1-D_2 great circle distance. According to D_1's speed of 8.5 knots, it would take 1 minute and 10 seconds to travel 307 meters, exceeding the time difference of 1 minute between D_1 and D_2. Therefore, D_1 was classified as abnormal data, as shown in Figure 21.

	MMSI	SOG	Longitude	Latitude	COG	True_Heading	Record_Time	
	413358520	9	119.648101666667	25.319265	33.2	37	2020-05-01 01:08:54.000	
	413358520	9.1	119.650983333333	25.32361	30.5	34	2020-05-01 01:10:54.000	
	413358520	8.6	119.653636666667	25.3275633333333	32.5	34	2020-05-01 01:12:45.000	
	413358520	8.1	119.658941666667	25.335845	27.1	30	2020-05-01 01:16:54.000	
	413358520	8	119.660133333333	25.3378233333333	28.7	31	2020-05-01 01:17:53.000	
	413358520	7.8	119.663588333333	25.3436616666667	27	31	2020-05-01 01:20:54.000	
	413358520	8.2	119.665981666667	25.3475866666667	30.9	35	2020-05-01 01:22:54.000	
	413358520	8.3	119.667406666667	25.3495266666667	36	37	2020-05-01 01:23:54.000	D_2
D_1	413358520	8.5	119.67924	25.3421242	30	32	2020-05-01 01:24:34.000	
	413358520	8.7	119.669701666667	25.3529666666667	29	28	2020-05-01 01:25:34.000	
	413358520	8.6	119.671033333333	25.3555683333333	22.1	24	2020-05-01 01:26:44.000	
	413358520	8.3	119.67193	25.357335	29.6	30	2020-05-01 01:27:34.000	
	413358520	8.1	119.674071666667	25.3606016666667	31.8	31	2020-05-01 01:29:13.000	

Figure 20. Schematic diagram of ship information with MMSI 413358520.

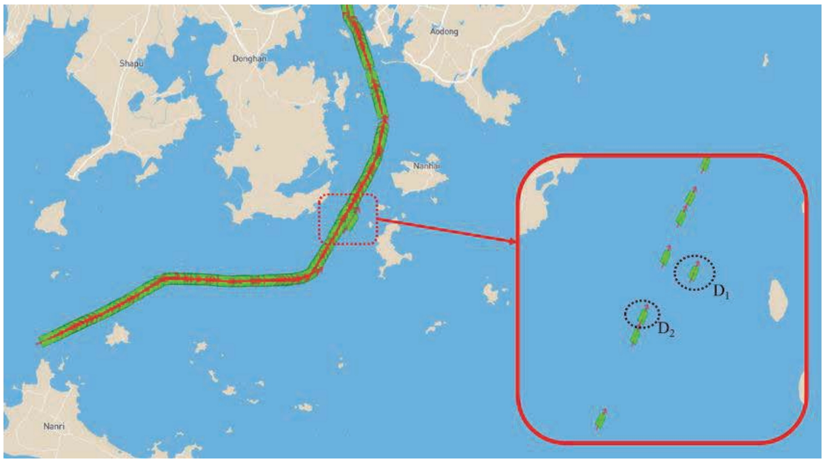

Figure 21. Schematic diagram of the abnormal sailing distance of the ship.

DOI: 10.29677/MR.202406_4(1).0006

5 CONCLUSIONS

This study presents the development of an operational system for data quality management in the Automatic Identification System (AIS). The system focuses on analyzing and debugging decoded AIS static and dynamic data. It systematically analyzes different ship behaviors under various navigation states and infers data and behavior patterns of ship AIS data based on these characteristics. Through the analysis of AIS data, the system automatically identifies abnormal ship data and removes these abnormal data points. The filtered data is then stored in a database for use in a dynamic ship information system, providing real-time and historical AIS data.

It is noteworthy that the performance and real-time nature of AIS data largely depend on the stability and update speed provided by the data source. The AIS data used in this study was provided by the NAMR. During the process of AIS data quality management and abnormal data identification, a considerable amount of abnormal data was discovered to be linked to repetitive AIS data. This result is inferred to be due to the reception of AIS data from the same vessel by adjacent AIS base stations, resulting in a large amount of repetitive AIS data. Incomplete codes or incorrect ship MMSI codes also account for a larger percentage of abnormal data errors. As mentioned earlier, we have explained that the correct format for MMSI codes should be 9 digits, and according to specifications of ITU-R M.585-8 and ITU-T E.217 recommendations, different purposes have different MMSI assignment categories. Therefore, our assessment of abnormal data related to MMSI may be due to AIS base stations receiving AIS data from maritime electronic buoys, handheld VHF transceivers, and other radio beacons. The main reason is that, currently, we cannot verify whether the received AIS data is transmitted by onboard AIS equipment. For future studies, the ability to independently receive and decode AIS signals would be beneficial. This would significantly enhance the accuracy and breadth of ship behavior analysis in this system, enabling more comprehensive and precise identification of abnormal AIS data.

ACKNOWLEDGEMENTS

This article is part of the research project supported by the National Academy of Marine Research, No. NAMR111062. The author thanks the National Academy of Marine Research for funding support.

REFERENCES

Bengio, Y., Courville, A., & Vincent, P. (2014). Representation Learning: A Review and New Perspectives, (1206.5538). *arXiv*. https://doi.org/10.48550/arXiv.1206.5538

Daranda, A. (2016). Neural Network Approach to Predict Marine Traffic. *Baltic Journal of Modern Computing, 4(3)*, 483-495. https://tinyurl.com/3kavvxnr

Dogancay, K., Tu, Z., & Ibal, G. (2021). Research into vessel behaviour pattern recognition in the maritime domain: Past, present and future. *Digital Signal Processing, 119*, 103191. https://doi.org/10.1016/j.dsp.2021.103191

Eljabu, L., Etemad, M., & Matwin, S. (2022). Spatial Clustering Method of Historical AIS Data for Maritime Traffic Routes Extraction. *2022 IEEE International Conference on Big Data (Big Data)*, 893-902. https://doi.org/10.1109/BigData55660.2022.10020646

Goerlandt, F., & Kujala, P. (2014). On the reliability and validity of ship–ship collision risk analysis in light of different perspectives on risk. *Safety Science, 62*, 348-365. https://doi.org/10.1016/j.ssci.2013.09.010

Huang, J. C., Nieh, C. Y., & Kuo, H. C. (2019). Risk assessment of ships maneuvering in an approaching channel based on AIS data. *Ocean Engineering*, *173*, 399-414. https://doi.org/10.1016/j.oceaneng.2018.12.058

Karataş, G. B., Karagoz, P., & Ayran, O. (2021). Trajectory pattern extraction and anomaly detection for maritime vessels. *Internet of Things, 16*, 100436. https://doi.org/10.1016/j.iot.2021.100436

Liu, T. K., Chen, Y. S., & Chen, Y. T. (2019). Utilization of Vessel Automatic Identification System (AIS) to Estimate the Emission of Air Pollutant from Merchant Vessels in the Port of Kaohsiung. *Aerosol and Air Quality Research, 19*(10), 2341-2351. https://doi.org/10.4209/aaqr.2019.07.0355

Mao, S., Tu, E., Zhang, G., Rachmawati, L., Rajabally, E., & Huang, G. B. (2018). An Automatic Identification System (AIS) Database for Maritime Trajectory Prediction and Data Mining. In J. Cao, E. Cambria, A. Lendasse, Y. Miche, & C. M. Vong (Eds.). *Proceedings of ELM-2016, 9*, 241-257. Springer International Publishing. https://doi.org/10.1007/978-3-319-57421-9_20

Murray, B., & Perera, L. P. (2021). An AIS-based deep learning framework for regional ship behavior prediction. *Reliability Engineering & System Safety, 215*, 107819. https://doi.org/10.1016/j.ress.2021.107819

Nieh, C. Y., Lee, M. C., Huang, J. C., & Kuo, H. C. (2019). RISK ASSESSMENT AND TRAFFIC BEHAVIOUR EVALUATION OF INBOUND SHIPS IN KEELUNG HARBOUR BASED ON AIS DATA. *Journal of Marine Science and Technology, 27*(4). https://doi.org/10.6119/JMST.201908_27(4).0002

Nguyen, D., Vadaine, R., Hajduch, G., Garello, R., & Fablet, R. (2018). A Multi-Task Deep Learning Architecture for Maritime Surveillance Using AIS Data Streams. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA),* 331-340. https://doi.org/10.1109/DSAA.2018.00044

National Academy of Marine Research. (2024). *National Ocean Database and Sharing System.* https://nodass.namr.gov.tw/dataInfo?metadataid=1

Perera, L. P., & Guedes Soares, C. (2015). Collision risk detection and quantification in ship navigation with integrated bridge systems. *Ocean Engineering*, *109*, 344-354. https://doi.org/10.1016/j.oceaneng.2015.08.016

Qu, X., Zhen, L., Howlett, R. J., & Jain, L. C. (Eds.). (2019). *Smart Transportation Systems 2019* (Vol. 149). Springer Singapore. https://doi.org/10.1007/978-981-13-8683-1

RECOMMENDATION ITU-R M.2092-1—Technical characteristics for a VHF data exchange system in the VHF maritime mobile band. (2022).

Rawson, A., & Brito, M. (2021). A critique of the use of domain analysis for spatial collision risk assessment. *Ocean Engineering*, *219*, 108259. https://doi.org/10.1016/j.oceaneng.2020.108259

Rong, H., Teixeira, A., & Soares, C. G. (2015). Evaluation of near-collisions in the Tagus River Estuary using a marine traffic simulation model. *Zeszyty Naukowe Akademii Morskiej w Szczecinie*, *43 (115)*, 68-78. https://repository.am.szczecin.pl/handle/123456789/775

Wang, Y., Zhang, J., Chen, X., Chu, X., & Yan, X. (2013). A spatial-temporal forensic analysis for inland-water ship collisions using AIS data. *Safety Science*, *57*, 187-202. https://doi.org/10.1016/j.ssci.2013.02.006

Wei, Z., Xie, X., & Zhang, X. (2020). AIS trajectory simplification algorithm considering ship behaviours. *Ocean Engineering*, *216*, 108086. https://doi.org/10.1016/j.oceaneng.2020.108086

Wijaya, W. M., & Nakamura, Y. (2013). Predicting Ship Behavior Navigating through Heavily Trafficked Fairways by Analyzing AIS Data on Apache HBase. *2013 First International Symposium on Computing and Networking*, 220-226. https://doi.org/10.1109/CANDAR.2013.39

Xiao, F., Ligteringen, H., Van Gulijk, C., & Ale, B. (2015). Comparison study on AIS data of ship traffic behavior. *Ocean Engineering*, *95*, 84-93. https://doi.org/10.1016/j.oceaneng.2014.11.020

Xin, X., Liu, K., Yang, X., Yuan, Z., & Zhang, J. (2019). A simulation model for ship navigation in the "*Xiazhimen*" waterway based on statistical analysis of AIS data. *Ocean Engineering*, *180*, 279-289. https://doi.org/10.1016/j.oceaneng.2019.03.052

Zaman, M. B., Kobayashi, E., Wakabayashi, N., & Maimun, A. (2015). Development of Risk Based Collision (RBC) Model for Tanker Ship Using AIS Data in the Malacca Straits. *Procedia Earth and Planetary Science*, *14*, 128-135. https://doi.org/10.1016/j.proeps.2015.07.093

Zhang, D., Zhang, Y., & Zhang, C. (2021). Data mining approach for automatic ship-route design for coastal seas using AIS trajectory clustering analysis. *Ocean Engineering*, *236*,109535. https://doi.org/10.1016/j.oceaneng.2021.109535

Marine Traffic. *AIS ship monitoring*. https://www.marinetraffic.com/en/ais/home

DOI: 10.29677/MR.202406_4(1).0006