Ship Type Classification Based on The Ship Navigating Trajectory and Machine Learning

Siyeol Baeg[1], Tracy Hammond[1]

[1]Texas A&M University, USA

Abstract

Accurate identification of ships is vital to ensure safe maritime activities. Research methods for the classification of ship types mostly use traditional radar recognition and optical recognition, but these methods all have their limitations. However, in the case of ship identification based on Automatic Identification System (AIS) data, not only is it less affected by the weather, but also static information and dynamic information can be utilized. So Automatic Identification System (AIS) data applications have been actively researched in ship identification as a more advanced and reliable method than traditional methods. However, incorrect AIS information may be transmitted due to an operator's mistake. In addition, some ships intentionally change AIS data information, such as ship type, to hide abnormal operations or illegal activities. In order to solve this problem, it is necessary to devise a new method to classify ship types correctly. So A ship-type classification scheme based on a ship navigating trajectory with Automatic Identification System (AIS) data is proposed to solve this problem. First, to acquire training data, historical AIS data provided by the Danish Maritime Authority have been converted into ship trajectories based on the Maritime Mobile Service Identities (MMSI), including corresponding ship types. As one of the main challenges in handling raw datasets is cleaning them to ensure the removal of invalid data, pre-processing is applied. Next, we extracted 39 features, including behavioral, geographic properties, and measurement of ship appearance characteristics. We especially proposed new features that could represent the shape of the overall trajectory using ink features designed for sketch recognition. Based on the extracted features, several benchmark classification algorithms (i.e., Decision Tree, Random Forest) are trained to classify four types of ships: Fishing, Passenger, Tanker, and Cargo. Finally, we check which features are valuable for recognizing ship types and which models can implement good performance in ship classification through performance analysis. The results demonstrate that the ink features designed for sketch recognition could express essential characteristics of ship trajectories and could be used for ship classification. Furthermore, Random Forest performs better than other classifiers in the classification of AIS data, and the classification accuracy of the four types of ships could reach 84.05% with a 39-dimensional feature vector.

Keywords

AIS data, Ship classification, Machine Learning, Ship trajectory

1. Introduction

1.1. Motivation

The international shipping industry is responsible for the carriage of around 90% of world trade[1]. Ship as the critical transportation tool on the vast ocean always raises much attention. In early 2022, the total fleet of seagoing merchant vessels amounted to 102,899 ships of 100 gross tons and above, equivalent to 2,199,107 thousand deadweight tons of capacity[2]. Furthermore, the world had an estimated 4.1 million fishing ships in 2020[3]. Along with the rapid increase in vessels, the maritime traffic environment has become more complex, and the possibility of maritime traffic accidents has increased. Maritime traffic accidents are complex and might result in the loss of human and irreversible economic damage[4]. Over the past decade, improved technology, regulation, and risk management systems have contributed to a 70% drop in reported shipping losses[1]. Among them, developing systems capable of monitoring vessel activities is one of the most crucial factors in strengthening navigational safety and security, such as preventing collisions and detecting unreported ships.

Traditional maritime navigation primarily relied on charts, watches, and radars and was judged by the sailor's long sailing experience. However, it is difficult to quickly and accurately identify numerous ships because it is limited in vision and radar coverage and provides only incomplete information, such as the speed and direction of movement of ships. However, the Automatic Identification System (AIS), which appeared with the development of communication technology, played an essential role in navigational safety, such as collision avoidance and navigation assistance. In the case of vessel identification based on AIS, not only is it less affected by the weather but also static information (vessel's name, dimensions, vessel's type) and dynamic information (vessel's position, speed) can be utilized. In particular, since the type of ship is closely related to the vessel's maneuverability, it is crucial information that sailors and Vessel Traffic Ser-

Joint Proceedings of the ACM IUI Workshops 2023, March 2023, Sydney, Australia
baek124@tamu.edu (S. Baeg); hammond@tamu.edu (T. Hammond)
0000-0001-8836-122X (S. Baeg); 0000-0001-7272-0507 (T. Hammond)

vice (VTS) controllers must grasp in advance to predict the following behavior of ships in a limited area. These unique characteristics that can affect their behavior can be represented such as :

- **Speed** : Different ship types may have different propulsion systems and speeds, affecting their ability to navigate in a given area, especially in crowded waterways or areas with shallow water.
- **Turn radius** : Larger ships with a deep draft and slow speeds may have a larger turning radius than smaller, more maneuverable ones.
- **Stoppage time** : Some ship types, such as container ships and bulk carriers, may take longer to stop due to their size and weight compared to smaller ships, such as tugs and fishing ships.

By knowing the type of ship and its characteristics from a distance, people can quickly recognize the situation, take precautionary measures in a limited area, such as a port or shipping lane, and improve the safety and efficiency of maritime operations. This information can also be used to optimize traffic management, design navigational aids, and prepare for emergencies.

In particular, since 2004, the International Maritime Organization (IMO) has mandated the installation of AIS on international passenger ships and ships of 300 tons or more to strengthen maritime safety and security. In addition, many countries and intergovernmental agencies, such as regional fisheries management organizations, are creating AIS requirements within their waters. However, with the widespread use of AIS, the reliability in accuracy of AIS information has been addressed in recent years. In particular, since some static information provided by AIS is directly entered by the ship owner, incorrect and missing information may be provided intentionally or unintentionally. Furthermore, this may cause a loss of reliability for the provided information. A study by Abbas[5] reveals errors related to the type of ship in the AIS data. According to the "VTS-based AIS study" by Abbas, some ships had no available ship type and were defined as "vessels" rather than a specific ship type. Meanwhile, researchers and VTS operators were unhappy with some of the observed vessel types. Therefore, research on ship-type identification is needed to solve the missing or tampering of ship-type information in AIS information. One of the other approaches to solve this problem is using other types of sensors and technologies, such as satellite imagery and drones, to supplement and improve AIS data for maritime traffic management. There are several advantages to using satellite images for ship classification compared to AIS data. Satellite images can provide information about ships that do not have AIS or have turned off their AIS, which is a common practice for some vessels to avoid detection. Moreover, it can cover a much larger area than AIS data, which is limited to the range of ship-based receivers. However, there are also several disadvantages, such as satellite images having a limited resolution, which can make it challenging to identify smaller vessels, and high-resolution satellite images are not always readily available, and acquiring them can be expensive.

Given these situations, developing a proper ship-type recognizer is vital to solving the missing or tampering of ship-type information in AIS data. Some papers have devised new methods for ship trajectory and type classification based on many AIS dataset. For example, they create a ship's trajectory image based on AIS data and input it into Convolutional Neural Networks (CNN), or they transform the ship's trajectory data into graph data and use Graph Neural Networks (GNN) to classify ship types. However, the following deep learning-based model not only makes it difficult to check what process is taken to identify a ship but also requires a large amount of training data. In contrast, generating meaningful features directly for machine learning algorithm operation involves a lot of effort but is more efficient at classifying objects once they can be found. Some papers obtained significant classification results using kinetic and geographical information obtained from AIS data as features for ship identification. However, it is necessary to create more meaningful features to improve the classification accuracy of ships. Thus, in this paper, we use various ink features used in sketch recognition to create new features practical for vessel identification and compare and analyze the performance of machine learning algorithms based on these.

1.2. Summary of Solution

We can take advantage of the Danish Maritime Authority AIS data to decide the type of a particular ship. We assume that ships cruising for different purposes would have different paths crossing the same area during a similar time. Given AIS data and analyzing such patterns from the data can reveal the type of the ship. Some of the differences can be taken from reasoning. For example, passenger ships, tankers, and cargo ships are likely to cruise along a straight path under good weather since they are moving from one destination to another. However, fishing ships are looking for schools of fish and are less likely to move along a straight line. There can be other differences that could be more obvious. Analyzing the AIS data may help us find such differences. To achieve the goal, we will preprocess the AIS data so that the data falsely collected would not affect our classification. For the data, each CSV file in raw AIS data contains the timestamps and locations of a ship. The Maritime Mobile Service Identities (MMSI) and the type of the ship are also included. The type of ships other than Cargo, Passenger, Tanker, and Fishing will be deleted. Then we will build our features and apply different classifiers to the feature set.

2. Related Work

With the explosion of sensors and GPS-enabled devices, research on trajectory recognition, which is analyzing objects over position and time data, has generated considerable interest. This research is often performed using machine learning algorithms and is vital in various applications, such as object tracking, activity recognition and behavior analysis. As most ships carry an AIS system as a matter of law, the data collected from it can be used in various fields, such as ship navigation route prediction, trajectory classification and anomaly detection in ship behavior[6, 7, 8, 9, 10, 11].

2.1. Ship Type Recognition Approaches

Among many applications, accurately identifying the type of ship is particularly important to ensure the safety and efficiency of maritime operations. This is because knowing the ship's type and characteristics can help predict its behavior in a limited area and can be used to design navigational aids, and prepare for emergencies.

Along with this need, many efforts have been made to identify ships operating at sea. One of the ship-type classification approaches studied so far is image-based ship-type recognition. In recent years, with the advent of Convolutional Neural Networks(CNNs), which do not require human supervision to identify essential features, many attempts have been made to develop a feasible ship classification system using satellite or aerial imagery[12, 13, 14, 15]. However, optical sensors have the disadvantage that they cannot be used at night or in adverse weather conditions. And ship classification in Synthetic Aperture Radar (SAR) images still has challenging to identify detailed sub-lists such as cargo and tankers beyond sorting out general categories such as ships and airplanes from different types of vehicles[15]. Also, the number of labeled samples in the SAR domain is limited.

Unlike image-based systems that attempt ship classification based on images, there are various studies to classify ships using AIS data. Research in this area focuses on developing algorithms and models to classify ships based on their attributes, such as size, speed, using AIS data. Some common approaches in ship classification with AIS data include using machine learning algorithms, such as decision trees, random forests, support vector machines, and neural networks. Although the AIS data already includes information on the type of ship, as mentioned in the introduction, considering human error and the problem of missing ship type information, it is essential to make various efforts to identify the type of ship using other information from AIS data. In particular, as a large amount of trajectory data are generated in real-time based on the AIS system, Ship classification based on trajectory data can compensate for the deficiency of traditional radar and optical identification.

Sanchez et al. [16] extracted features from spatiotemporal data that represent the trajectories of ships and used the Support Vector Machine(SVM) and decision tree to classify ship trajectories into either fishing or non-fishing ships. Sheng et al. [17] partitioned each trajectory into three basic movement patterns: anchored-off, turning, and straight-sailing, extracted 17 trajectory features based on it, and classified fishing and cargo ship by using a logistic regression model. Li et al. [18] converted ship trajectory data into graph data and used Graph Neural Network(GNN) to classify four types of ships. They could recognize fishing ships, passenger ships, tankers, and containers with an accuracy of 82.7%. Yang et al. [19] generated ship trajectory images containing operating states such as static, standard navigation, and maneuvering. They used Convolutional Neural Networks(CNNs) to identify eight types of ships from ship trajectory images and achieved an accuracy of 87.5% with the optimal configuration of CNNs. Wang et al. [20] used static information in AIS data such as width, length, and draught and identified five types of ships with an accuracy of 86.14%. Yan et al. [21] extracted some ship appearance and behavior characteristics and classified the five types of ships with an accuracy of 92.7% using the Random Forest model. Machine learning has been applied successfully in ship classification tasks, providing a promising solution to challenges such as missing or tampering with ship-type information in AIS data. However, the accuracy of machine learning algorithms for ship classification can be affected by factors such as the quality and availability of data, the extraction of valuable features, and the choice of algorithms and models. Researchers have also explored integrating AIS data with other sources of information, such as satellite imagery and radar data, to enhance the accuracy and reliability of ship classification[22, 23].

2.2. Features Used For Recognition

Modern sensors and digital devices enable the collection of a large amount of data from moving objects. For example, smartphones, smartwatches, and wildlife with tracking tags generate timestamped position data. Even the handwriting or hand-drawn sketches of users with the digital pen are positional coordinates data created on digital paper over time. Based on the acquired data, we analyze various hidden patterns and use them to solve new problems, such as trajectory and digital ink recognition. Many features have been proposed to solve many different recognition problems. As different problems have their characteristic, it is essential to add valuable features to improve the recognition rate for their specific problem. Jones et al.[24] grouped trajectory features according to their characters to make it easier to select a relevant set for a specific problem. These categories for features can be represented as :

- **Kinematic Features**: Kinematic Features describe an object's motion independent of the forces that cause it moves, such as total distance traveled, average speed, and maximum altitude.
- **Temporal Features**: Temporal Features described when some event of interest took place, such as start and end time, duration, and time when nearest to a fixed point or region.
- **Geospatial Features**: Geospatial Features describe where the trajectory is observed, such as the place where first seen, the destination point, and the nearest distance to a fixed point or region.
- **Shape Features**: Shape Features describe some aspects of the trajectory's geometry, such as convex hull area and divergence from a given shape.

AIS data-based ship type recognition has tried applying various trajectory features. Some researchers have already proposed various features to classify ship types, including kinematic, temporal, geospatial, and geometric features. Kraus et al. [25] extracted geographical characteristics, such as the distance to the nearest coastline, and temporal features, such as if the trajectory starts/ends at night. Yan et al. [21] extracted various ship appearance characteristics, including length, width, and shape complex and Sanchez et al. [16] extracted kinetic features such as average and maximum of course variation from segments of the trajectory.

Although it is a different domain, many features have been proposed for sketch recognition, and they can be divided into two types: gesture-based features and geometric features[26]. The first type describes how a sketch was drawn and used to classify input strokes containing x,y points, and time values into a set of pre-defined gestures[27]. For example, Dean Rubine[28] proposed thirteen features used for basic shape and gesture recognition, and Long et al. [29] added nine new features to Rubine's existing set. Unlike the first one, the second type describes the object's shape and arrangement, focusing on what the sketch looks like[27]. Paulson et al.[30] proposed new features, the Normalized Distance between Direction Extremes (NDDE) and Direction Change Ratio (DCR), which are suitable for classifying polylines and curved strokes. Furthermore, Blagojevie et al.[31] composed a comprehensive library of 114 ink features from previous work in sketch recognition and developed a taxonomy of feature types such as curvature, density, and direction.

Intuitively, we understand that particular features developed for one domain problem can be valuable for some problems in different domains. However, until now, research papers have yet to extract features from the shape of the overall trajectory for ship classification. So our system builds on this kind of work by seeking to identify ship types with features from a different domain, such as sketch recognition. Even if it is a feature for sketch recognition, if it can reflect the characteristics of a general trajectory well, the corresponding feature can be applied to ship type classification.

3. Methodology

The main objective of this work is to develop and evaluate a ship type classifier using machine learning. In other words, the goal of this project is to address the following two questions: 1) What features are valuable for recognizing the ship types? 2) How can we achieve reasonable ship type recognition using different classification algorithms?.

3.1. Data Source

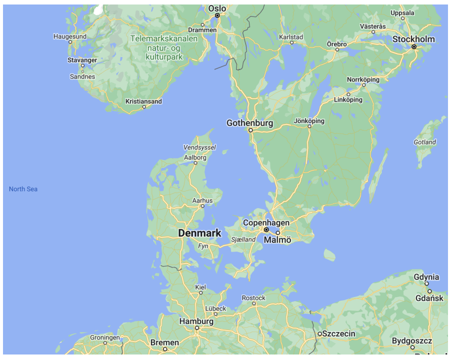
The data used in this work were historical AIS data from the Danish Maritime Authority [33], with a time distribution of 5 days in November 2022. This data includes AIS information obtained from the coast of Denmark since 2006, and the size of each file stored per day reaches approximately 1.5GB. Therefore, it can be used as a sufficient amount of data to create a trajectory for this project. Considering the subsequent trajectory creation process, the information from the AIS data needed includes MMSI, Timestamp, Ship type, Latitude, Longitude, Width, and Length.

3.2. Pre-processing

The quality of AIS data is crucial for trajectory analysis and an essential factor for classification model performance. We used the Pandas library for data preprocessing to remove invalid values in AIS data before converting raw AIS data to trajectories. For example, the data such as the latitude exceeding 90 degrees or the longitude exceeding 180 degrees was cleared because they were out of range.

3.3. Trajectory Creation

After obtaining the pre-processed AIS data, the next step is to generate the trajectory of each ship and label the trajectory according to the type of ship. A trajectory can be defined as consecutive coordinates of the ship, and each trajectory coordinate is a tuple comprising a position(latitude & longitude) and a timestamp. Because the ship



#_Timestamp	MMSI	Latitude	Longitude	SOG	COG	Heading	Ship_type	Width	Length
04/11/2022 23:59:58	230683000	55.806527	10.807122	10.0	182.9	186.0	Cargo	26.0	160.0
04/11/2022 23:59:58	219000431	54.516383	11.238650	13.7	196.9	199.0	Passenger	26.0	141.0
04/11/2022 23:59:58	355254000	57.369693	11.471433	17.9	162.9	163.0	Cargo	32.0	237.0
04/11/2022 23:59:58	219551000	55.344105	7.248375	18.8	262.6	264.0	Cargo	30.0	196.0
04/11/2022 23:59:58	246541000	57.571883	11.646500	5.2	337.8	323.0	Tanker	15.0	116.0

Figure 1: Danish Waters map from Google[32] and examples of AIS raw data

trajectory has various lengths, shorter trajectories of less than 3 hours of operation are excluded to ensure the trajectory carries enough information for feature extraction. Moreover, most of the trajectories consist of more than thousands of coordinates, so the computational cost for feature extraction is expensive. A compression algorithm is applied to reduces the number of coordinates while preserving the character of the shape of the trajectory [34]. We used MovingPandas, a Python library for handling movement data based on Pandas and GeoPandas, to create and generalize the trajectories [35]. Based on this, samples of the trajectory for each ship type is shown in figure 2. The color shown in the legend expresses relative time, and the initially acquired coordinates are set to 0, and the color gradually changes to green as time passes. According to the unique MMSI, CSV files, including the information required for trajectory generation, were created and saved in the folder classified by ship type where the label matches. Finally, we obtained 1,298 trajectories based on unique MMSI from AIS data. Table 1 shows the number of trajectories for each of the four ship types.

Table 1
Number of created trajectories by the four types of ship

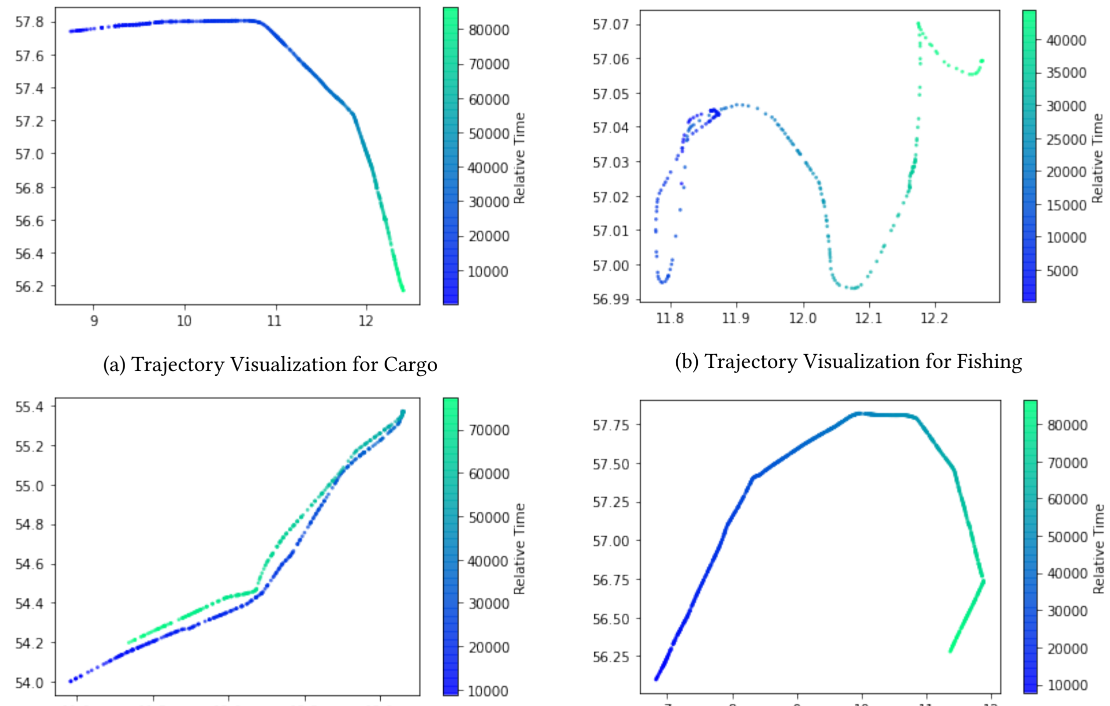
Ship Type	Cargo	Fishing	Passenger	Tanker	Total
Quantity	568	197	289	244	1298

3.4. Trajectory Data Exploration

In order to create meaningful features for ship type classification, it is crucial to identify the corresponding trajectory pattern. Trajectory coordinates are plotted using a different color to indicate ship type using the Kelper.Gl, designed for geospatial data analysis. Figure 3 shows some trajectory patterns, such as passenger ships using common routes and fishing ships tending to cluster many points within a limited area intensively.

3.5. Feature Extraction

One of the goals of this work is to find features for recognizing the ship types based on trajectory. Some papers have already proposed various features to classify ship type, including geographical, behavioral, and Geometric features. The author of [25] extracted geographical characteristics, such as the distance to the coast, to classify the type of ship. In [21], the author extracted various ship appearance characteristics, including length, width, and shape complex. Moreover, the author of [17] extracted trajectory features based on the fundamental movement patterns and divided them into three categories such as global, straight-sailing, and turning features. However, the most challenging thing is to discover additional practical features that can describe the overall behavior of a ship along with the previously used features. However, until now, research papers have yet to extract features from the shape of the overall trajectory for ship classification. For this purpose, we propose new features that could measure an essential characteristic of ship trajectories using ink features designed for sketch recognition. In [31], the author composed a comprehensive library of 114 ink features from previous work in sketch recognition and developed a taxonomy of feature types. We have selected several features from these lists that will be useful for the ship type classification problem. Finally, feature extraction was performed by dividing it into three categories. The first was the trajectory shape feature, which generated 21 features, including Rubine Features, some Long Features, and DCR(Direction Change Ratio). The second category used latitude and longitude, such as mean, max, and standard deviation, as a geographic features. Lastly, we extracted some features from the measurement of ship appearance characteristics to improve the classification performance[36]. Finally, we created a 39-dimensional feature vector for each trajectory. The detailed feature list is shown in Table 2.



(a) Trajectory Visualization for Cargo

(b) Trajectory Visualization for Fishing

(c) Trajectory Visualization for Passenger

(d) Trajectory Visualization for Tanker

Figure 2: Subset of ship trajectory images made from AIS data

Table 2
Extracted features : shape features(F1-F21), geographic features(F22-F33), ship appearance features(F34-F39)

Feature	Description	Feature	Description	Feature	Description
F1	Cosine of initial angle	F2	Sine of initial angle	F3	Length of diagonal bounding box
F4	Angle of bounding box	F5	Distance between endpoints	F6	Cosine of endpoints angle
F7	Sine of endpoints angle	F8	Trajectory length	F9	Total angle traversed
F10	Absolute sum of angle traversed	F11	Squared sum of angle traversed	F12	Maximum speed of trajectory
F13	Total duration of trajectory	F14	Curviness	F15	Total angle traverse / stroke length
F16	Total angle traverse / absolute total angle	F17	Trajectory length / distance between endpoints	F18	Trajectory length / bounding box diagonal length
F19	Distance between endpoints / bounding box diagonal length	F20	Log of bounding box area	F21	DCR
F22	Max Latitude	F23	Max Longitude	F24	Min Latitude
F25	Min Longitude	F26	Mean Latitude	F27	Mean Longitude
F28	VAR Latitude	F29	VAR Longitude	F30	STD Latitude
F31	STD Longitude	F32	Latitude Span	F33	Longitude Span
F34	Shape Complex	F35	Naive Perimeter	F36	Naive Area
F37	Length of ship	F38	Width of ship	F39	Length of ship / Width of ship

4. Results and Discussion

Extracted features are tested on classification algorithms with 10-fold cross-validation provided by the WEKA [37]. We experimented with five different classifiers for our evaluation: J48, Random Forest, Random Tree, Logistic, and Multilayer Perceptron. We used standard classifier evaluation metrics such as Accuracy, Precision, Recall, and F-measures to measure the model's effectiveness in classifying the four ship types. Accuracy, average precision, average recall, and average F1 score are calculated for each classifier, and the results are shown in Table 3.

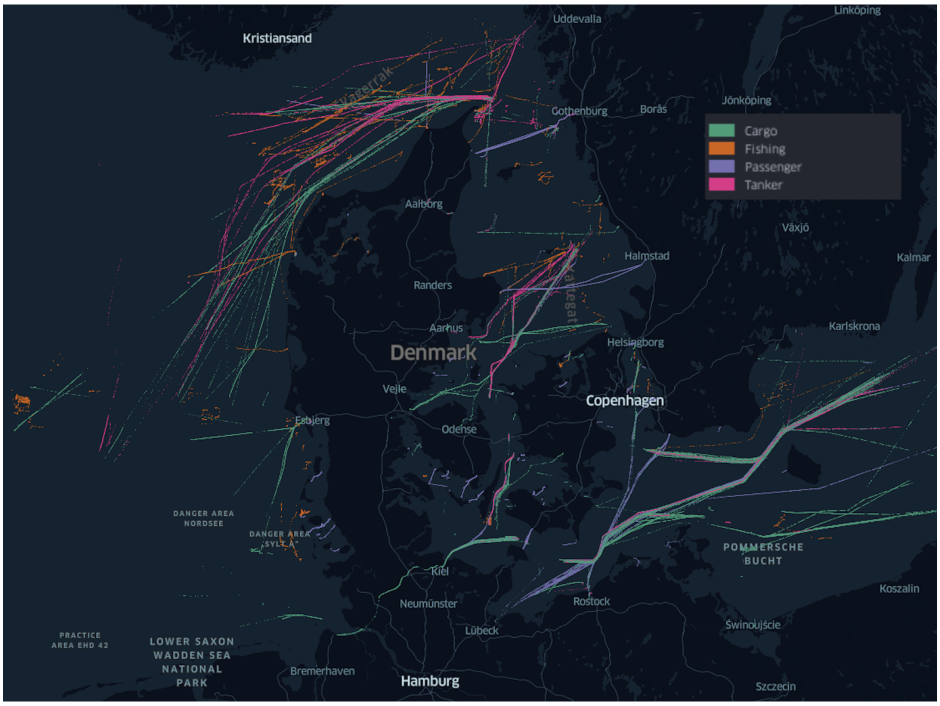


Figure 3: Cargo, fishing, passenger, tanker trajectories, generated using Kepler.Gl

Table 3
Comparison with different classification algorithms

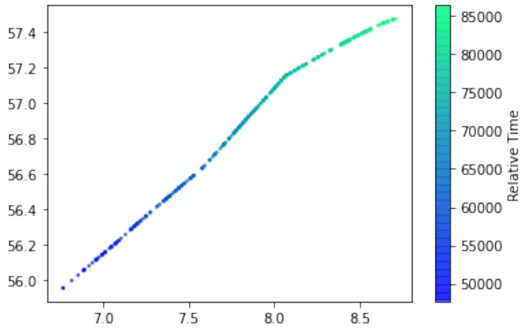
Models	Accuracy(%)	Average Precision(%)	Average Recall(%)	Average F1(%)
Random Forest	84.05	84.2	84.1	83
J48	80.43	79.8	80.4	80
Logistic	76.65	76.5	76.7	75.9
Random Tree	75.42	75.5	75.4	75.4
Multilayer Perceptron	74.88	74.3	74.9	74.5

In weighted average, random forest outperformed other classifiers by nearly 4%.

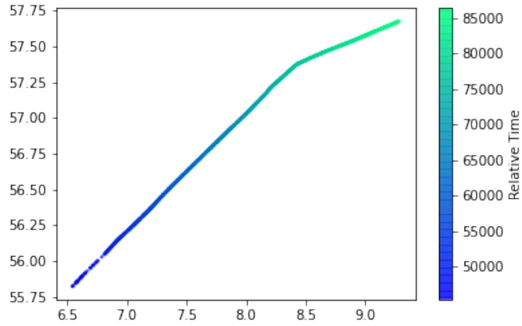
Table 4
Result of the four types of ships with Random Forest

Ship Types	Precision(%)	Recall(%)	F1-Score(%)
Cargo	74.4	91.7	84
Passenger	94.6	91.7	93.1
Fishing	95.1	99	97
Tanker	78.6	45.1	57.3

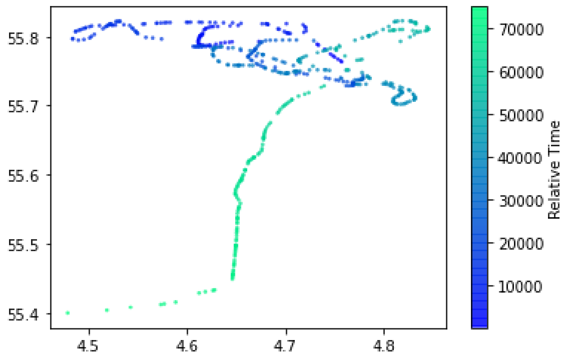
According to Table 3, our model has 84% accuracy for the four-ship types altogether and outperforms the GNN model[18]. Although the dataset differs, it suggests our model successfully matches ships to the cargo, tanker, fishing, and passenger category based on ship trajectory. Regarding each ship type, the precision, recall, and F-score is 94.6%, 91.7%, and 93.1% respectively for passenger ships; 95.1%, 99%, and 97% respective for fishing ships; 78.6%, 45.1%, and 57.3% respective for tankers and 74.4%, 91.7% and 84% respectively for cargo ships. Though all indicators for fishing and passenger ships exceed 90%, it is also alarming that the precision for cargo type and all indicators for tanker type are significantly worse. According to Table 4, the major problem is that tankers are labeled cargo type. Reviewing the trajectory in figure 2a and figure 2d, the trajectory of cargo and tanker are too similar for our model, which is based on shape features, to recognize their difference. Figure 4a and Figure 4b provide a more intuitive view. Those two trajectories are not a simple straight line or polyline. They make



(a) A trajectory of Cargo ship



(b) A trajectory of Tanker



(c) A trajectory of Fishing

Figure 4: Examples of ship trajectory by ship type

similar turns, and the terminals of their trajectories are similar but different. They are even traveling at a similar speed. The color of their trajectories changes similarly, indicating the relative time of their trips. Going over all trajectories of cargo ships and tankers, it appears that the difference between a cargo ship and another particular tanker can have a relatively high chance of being smaller than its difference with another cargo ship. The similarity between the trajectories of cargo ships and tankers is reasonable since, in some sense, tanker ships are also cargo ships, except that their cargo is oil. To recognize them, either the training data is expanded so that extra shape features can be observed to distinguish cargo ships and tankers, or more features need to be applied, and that solution falls out of the scope of this study.

By the same reasoning, the exceptional outcome in fishing ships (99% recall and 97% F1-score) suggests that fishing ships cruise very differently from the other three ship types. The most noticeable difference could be that the goal of fishing is not traveling from one place to another, which is the objective for all three other types. Figure 4c above is a typical case.

5. Conclusion and Future Work

This project proposed a ship classification method with Danish water's AIS data considering the overall ship's behavior characteristics to solve the missing and tampering of ship type information in AIS data. Firstly, trajectory shape features and geographic features are extracted from many AIS data from different types of ships. After that, this project used various classification algorithms such as Random Forest and Decision Tree to compare the performance. Results and Discussion showed that the Random Forest performs better than other classifiers in the classification of AIS data. The classification accuracy of the four types of ships could reach 84.05% with 39-dimensional feature vectors. In particular, in the case of fishing and passenger ships, the precision was 0.951 and 0.946, respectively, and very high results were confirmed. This confirmed that the ink features designed for sketch recognition could express essential characteristics of ship trajectories and be used for ship classification. In the future, research can be carried out considering the following directions : (1) focusing on extracting additional features and applying feature selection to improve the performance of AIS data ship classification. In particular, finding practical features distinguishing cargo and tanker ships will significantly improve the overall classification model performance. (2) testing the classifier with much larger data volumes to check the classifier's scalability. (3) Expand the AIS data from Danish waters to different regions worldwide to validate its potential in dealing with maritime traffic situations.

References

[1] A. G. C. S. AGCS, Safety and shipping review 2021, 2021.

[2] U. Nations, The review of maritime transport 2022, 2022.

[3] Food, A. Organization(FAO), The state of world fisheries and aquaculture 2022, 2022.

[4] V. L. N. R. C.-E. I. T. R. Dominguez-Péry, C., Reducing maritime accidents in ships by tackling human error: a bibliometric review and research agenda, Journal of Shipping and Trade (2021). doi:https://doi.org/10.1186/s41072-021-00098-y.

[5] A. Harati Mokhtari, A. Wall, P. Brooks, J. Wang, Automatic identification system (ais): Data reliability and human error implications, Journal of Navigation 60 (2007) 373 – 389. doi:10.1017/S0373463307004298.

[6] J.-Y. Oh, H.-J. Kim, S.-K. Park, Detection of ship movement anomaly using ais data: a study, Journal of Navigation and port research 42 (2018) 277–282.

[7] T. Guo, L. Xie, Research on ship trajectory classification based on a deep convolutional neural network, Journal of Marine Science and Engineering 10 (2022) 568.

[8] X. Chen, Y. Liu, K. Achuthan, X. Zhang, A ship movement classification based on automatic identification system (ais) data using convolutional neural network, Ocean Engineering 218 (2020) 108182.

[9] Y. Li, H. Ren, Visual analysis of vessel behaviour based on trajectory data: A case study of the yangtze river estuary, ISPRS International Journal of Geo-Information 11 (2022) 244.

[10] P. Sampath, Trajectory analysis using automatic identification system (ais) in new zealand waters, 2012.

[11] J.-G. Lee, J. Han, X. Li, H. Gonzalez, Traclass: trajectory classification using hierarchical region-based and trajectory-based clustering, Proceedings of the VLDB Endowment 1 (2008) 1081–1094.

[12] Y. Wang, C. Wang, H. Zhang, Ship classification in high-resolution sar images using deep learning of small datasets, Sensors 18 (2018) 2929.

[13] Z. Shao, J. Wang, L. Deng, X. Huang, T. Lu, R. Zhang, X. Lv, Q. Ding, Z. Wang, Glsd: The global large-scale ship database and baseline evaluations, arXiv preprint arXiv:2106.02773 (2021).

[14] J. Li, C. Qu, J. Shao, Ship detection in sar images based on an improved faster r-cnn, in: 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), IEEE, 2017, pp. 1–6.

[15] Y. Xu, H. Lang, Ship classification in sar images with geometric transfer metric learning, IEEE Transactions on Geoscience and Remote Sensing 59 (2020) 6799–6813.

[16] D. Sánchez Pedroche, D. Amigo, J. García, J. M. Molina, Architecture for trajectory-based fishing ship classification with ais data, Sensors 20 (2020) 3782.

[17] K. Sheng, Z. Liu, D. Zhou, A. He, C. Feng, Research on ship classification based on trajectory features, The Journal of Navigation 71 (2018) 100–116.

[18] T. Li, H. Xu, W. Zeng, Ship classification method for massive ais trajectories based on gnn, in: Journal of Physics: Conference Series, volume 2025, IOP Publishing, 2021, p. 012024.

[19] T. Yang, X. Wang, Z. Liu, Ship type recognition based on ship navigating trajectory and convolutional neural network, Journal of Marine Science and Engineering 10 (2022) 84.

[20] Y. Wang, L. Yang, X. Song, X. Li, Ship classification based on random forest using static information from ais data, in: Journal of Physics: Conference Series, volume 2113, IOP Publishing, 2021, p. 012072.

[21] Z. Yan, X. Song, H. Zhong, L. Yang, Y. Wang, Ship classification and anomaly detection based on spaceborne ais data considering behavior characteristics, Sensors 22 (2022) 7713.

[22] Z. Yan, X. Song, L. Yang, Y. Wang, Ship classification in synthetic aperture radar images based on multiple classifiers ensemble learning and automatic identification system data transfer learning, Remote Sensing 14 (2022) 5288.

[23] H. Lang, S. Wu, Y. Xu, Ship classification in sar images improved by ais knowledge transfer, IEEE Geoscience and Remote Sensing Letters 15 (2018) 439–443.

[24] J. L. Jones, B. D. Newton, K. L. Wisniewski, A. T. Wilson, M. J. Ginaldi, C. A. Waddell, K. Goss, K. J. Ward, M. D. Rintoul, Large-Scale Trajectory Analysis via Feature Vectors, Technical Report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2021.

[25] P. Kraus, C. Mohrdieck, F. Schwenker, Ship classification based on trajectory data with machine-learning methods, in: 2018 19th International Radar Symposium (IRS), IEEE, 2018, pp. 1–10.

[26] P. Corey, T. Hammond, Gladder: Combining gesture and geometric sketch recognition., in: AAAI, volume 8, 2008, pp. 1788–1789.

[27] B. Paulson, P. Rajan, P. Davalos, R. Gutierrez-Osuna, T. Hammond, What!?! no rubine features?: Using geometric-based features to produce normalized confidence values for sketch recognition, in: HCC workshop: sketch tools for diagramming, 2008, pp. 57–63.
[28] D. Rubine, Specifying gestures by example, ACM SIGGRAPH computer graphics 25 (1991) 329–337.
[29] A. C. Long Jr, J. A. Landay, L. A. Rowe, J. Michiels, Visual similarity of pen gestures, in: Proceedings of the SIGCHI conference on Human Factors in Computing Systems, 2000, pp. 360–367.
[30] B. Paulson, T. Hammond, Paleosketch: accurate primitive sketch recognition and beautification, in: Proceedings of the 13th international conference on Intelligent user interfaces, 2008, pp. 1–10.
[31] R. Blagojevic, S. H.-H. Chang, B. Plimmer, The power of automatic feature selection: Rubine on steroids., SBIM 10 (2010) 79–86.
[32] Google, Google map of danish waters, (n.d.). URL: https://www.google.com/maps/@57.1858914,10.0813199,6z, retrieved February 22, 2023.
[33] D. M. Authority, Ais data, (n.d.). URL: https://dma.dk/safety-at-sea/navigational-information/ais-data, accessed November 3, 2022.
[34] N. Meratnia, R. A. de By, Spatiotemporal compression techniques for moving point objects, in: Advances in Database Technology-EDBT 2004: 9th International Conference on Extending Database Technology, Heraklion, Crete, Greece, March 14-18, 2004 9, Springer, 2004, pp. 765–782.
[35] A. Graser, Movingpandas: efficient structures for movement data in python, GIForum 1 (2019) 54–68.
[36] H. Lang, S. Wu, Y. Xu, Ship classification in sar images improved by ais knowledge transfer, IEEE Geoscience and Remote Sensing Letters 15 (2018) 439–443. doi:10.1109/LGRS.2018.2792683.
[37] M. A. H. Eibe Frank, Ian H. Witten, The WEKA Workbench. Online Appendix for ”Data Mining: Practical Machine Learning Tools and Techniques”, fourth ed., Morgan Kaufmann, 2016.