MACHINE LEARNING FOR THE MARITIME INDUSTRY

SAIRAGHUNANDAN PULIBANDLA[1], MOHAMED EL-DOSUKY[2,3], SHERIF KAMEL[2,4]

[1]Next Tech Lab, SRM University - AP, Andhra Pradesh, India.

[2]Computer Science Department, Arab East Colleges, Saudi Arabia

[3]Computer Science Department, Faculty of Computers and Information, Mansoura University, Egypt

[4]Department of Communications and Computer Engineering, October University for Modern Sciences and Arts, Egypt

E-mail: maldosuky@arabeast.edu.sa

ABSTRACT

This article handles two problems in maritime industry. The first is how to track ships and vessels. The second is the fact that numerous maritime trade routes are utilized by ships depending on the nation, topographical elements, and ship characteristics. This article proposes a system for tracking ships and developing maritime traffic routes using statistical density analysis. It uses information from an automatic identification system (AIS) to create quantifiable traffic routes. The approach includes preprocessing, deconstruction, and database management. DBSCAN detects boat waypoints, and kernel density estimation analysis (KDE) assesses the breadth of sea routes. The waypoints along the primary route are assessed while taking into account statistical data on all maritime traffic. The findings can be used to plan paths for autonomous surface ships, ensuring safe routes for ships in designated ocean regions.

Keywords: *Maritime, Ship Trajectory, Ship Maneuvering Instructions, Machine Learning*

1. INTRODUCTION

The maritime sector is a complex and multifaceted industry that facilitates global trade and connects nations and businesses [1]. It involves various participants such as shipping firms, port operators, freight forwarders, shipbuilders, maritime equipment producers, and maritime service providers [2]. Technological advancements like autonomous ships and digitization are transforming the industry, offering opportunities for innovation, sustainability, and expansion.

Machine learning can provide numerous benefits to the maritime sector, including increased productivity, cost savings, and increased safety. For instance, machine learning can forecast machinery breakdowns [3], optimize shipping routes [4], and optimize cargo handling [5]. It can also monitor vessel performance [6], identify safety issues [7], and forecast future trends [8].

However, the maritime industry is complex and requires further research on machine learning and AI applications. As the industry develops and new technologies are developed, researchers will have more opportunities to explore machine learning in areas like vessel optimization, risk management, and environmental sustainability [9].

The rest of the paper is organized as follows. Section 2 is for literature review. Section 3 is for proposed methodology. Section 4 is for describing the dataset and results. Finally, section 5 is for conclusions and future work.

2. LITERATURE REVIEW

Maritime businesses often use slow steaming for fuel efficiency, but this can lead to less effective engine lubrication and structural flaws like crosshead bearings [10]. Traditional monitoring techniques struggle to detect complex problems, and routine inspections are expensive. Predictive maintenance is essential for minimizing failures and avoiding over-maintenance.

The maritime industry has adopted specific anomaly detection methods, such as combining vibration data [11] with performance data to detect failure scenarios in big diesel engines [12]. Self-organizing maps (SOM) have been evaluated for tracking engine state and identifying abnormal data clusters [13]. A one-class technique for forecasting ship system status has been developed, using a one-vs.-one technique [14].

Extreme learning machine to handle 14 defects and correct operating modes are studied [15]. Anomalies that are both point and collective can be found using many approaches [16]. DBSCAN algorithm is utilized to develop efficient waypoints for trip planning [17].

The maritime sector has adapted machine learning techniques to improve delivery and eliminate uncertainty in sea transportation [18]. This allows for clever algorithms for data analysis, directing logic for potential issues, and can be used in maintenance procedures, trip planning, cargo optimization, and maritime network design.

The maritime industry has faced challenges in adopting customer-centric solutions, particularly in the area of the Internet of Vessels [19], where the study identified seven design standards that affect consumer values in liner businesses, revealing that eco-ship and container technology, big data solutions for ship information management, and system automation are the most effective strategies. However, barriers to big data analytics adoption include lack of awareness, executive sponsorship, and competences.

DBSCAN is a density model-based clustering approach used in machine learning, utilizing data location information [20]. It differs from hierarchical and fuzzy clustering, which use distance between points. The algorithm K-means clusters data based on cluster distance. Clustering activities occur within a grid structure, involving multiple probability distributions. DBSCAN involves locating every neighbour point within eps and determining the most important or frequently visited points. It creates a new cluster for each core point, locates all density-connected points, and groups them with the core point in the same cluster.

While previous research has made significant strides in predictive maintenance and route optimization within the maritime industry, several limitations remain that justify the need for this study. Traditional monitoring methods often focus narrowly on engine performance and struggle with noise sensitivity and real-world variability. Machine learning approaches like SOM, and one-class classifiers have shown promise but often lack scalability, interpretability, and adaptability to dynamic maritime conditions. Moreover, route clustering methods using DBSCAN typically emphasize spatial data while overlooking contextual and temporal aspects, limiting their effectiveness for real-time autonomous navigation. Additionally, most existing studies treat maintenance and route planning as separate domains, missing the opportunity to integrate insights for more holistic decision-making. Given the slow adoption of big data analytics due to organizational and technical barriers, there is a clear need for a unified, intelligent framework that combines object detection and adaptive route analysis. This research addresses that gap by proposing a comprehensive system designed to enhance operational efficiency and safety for autonomous surface vessels.

3. PROPOSED METHODOLOGY

As shown in Figure 1, the process involves loading images from the dataset directory, preprocessing images, loading the pre-trained model, detecting objects, and post-processing results. The images are read, stored, and normalized before being preprocessed. The pre-trained model is loaded using a deep learning framework and initialized. Objects are detected by performing a forward pass through the model, obtaining bounding box predictions and class probabilities, applying confidence thresholds, and removing duplicate bounding boxes. Post-process results are drawn and labeled with class names and confidence scores.

Figure 2 provides the proposed methodology for maritime traffic routes. The framework involves loading AIS data, setting up an environment, preprocessing data, deconstructing data, organizing data in a database, detecting waypoints using DBSCAN clustering, assessing route density using Kernel Density Estimation (KDE), analyzing primary routes, and incorporating statistical data. It then plans navigation paths for autonomous surface ships, testing their effectiveness in different scenarios. The findings are then output into reports or dashboards, highlighting identified routes and safety recommendations. The database is then updated to store the final route data and related analysis results for future reference. The framework is designed to meet safety standards and efficiency for autonomous surface ships.

4. DATASETS AND RESULTS

For detecting maritime objects, the Singapore Maritime Dataset (SMD) is utilized which includes high-definition on-shore and on-board videos, as well as Near Infrared (NIR) videos captured using a Canon 70D camera with a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

removed hot mirror and Mid-Opt BP800 Near-IR Bandpass filter [21].

It contains 10 different classes of objects that are Ferry, Buoy, Vessel, Speed boat, Boat, Kayak, Sail boat, Swimming person, Flying object, and Other.

Figure 3 displays the distribution of object motion in the SMD dataset, categorized as "Moving," "Stationary," and "Other" across three scenarios: "on-shore," "on-board," and "near infra-red." Figure 4 represents the distribution of the distance of objects in the SMD dataset, categorized by "Near," "Far," and "Other" across the three scenarios: "on-shore," "on-board," and "near infra-red."

The ground truth in maritime settings refers to accurate, manually labeled data used to evaluate object detection models. This includes bounding boxes (as shown in Figure 5), object classes, attributes, and temporal consistency, which provide precise coordinates, labels, and additional details about detected objects.

Samples of the SMD dataset are shown in Figure 6 (Visible on-Shore sample), Figure 7 (Visible on-board sample), and Figure 8 (NIR on-Shore sample).

To detect objects in an image from the Singapore Maritime Dataset using Python, you can utilize a pre-trained deep learning model like YOLO (You Only Look Once [24]) or SSD (Single Shot MultiBox Detector [25]).

Example results on SMD dataset are shown in Figure 9 (for Faster R-CNN with Inception v2 deep learning model) and Figure 10 (for SSD with Resnet 50 deep learning model). Fast object detection is shown in Figure 11.

Table 1 summarizes various sources to download AIS data. However, let us generate synthetic AIS data and then running the DBSCAN and kernel density estimation analysis (KDE) on that data. Figure 12 shows KDE of maritime routes, while Figure 13 shows the primary route assessment.

5. CONCLUSION AND FUTURE WORK

The proposed system generates safe and quantifiable maritime routes using statistical density analysis, suitable for autonomous ship navigation and route planning.

Unlike prior research that typically focuses on either predictive maintenance or route optimization in isolation, this study introduces a unified framework that combines image-based object detection with data-driven maritime route analysis. While earlier works rely heavily on single-source data (e.g., vibration signals or AIS logs), this research leverages multi-modal inputs and incorporates both visual recognition and spatial-temporal clustering techniques such as DBSCAN and KDE. Furthermore, previous clustering applications often ignore contextual factors like vessel type or navigation environment, whereas this study aims to integrate such dimensions for smarter, scenario-aware route planning. Additionally, most traditional methods are static or rule-based, whereas the proposed framework is designed to be adaptive, scalable, and suitable for real-time decision-making in autonomous maritime systems. This positions the current work as a more comprehensive and forward-looking solution aligned with the future of smart, autonomous shipping.

Despite its comprehensive design, the proposed methodology has certain limitations. First, the accuracy of the object detection component heavily depends on the quality and diversity of the training dataset; limited or biased image data can lead to poor generalization in real-world maritime conditions. Second, the effectiveness of DBSCAN clustering is sensitive to the choice of parameters, which may not be optimal across different regions or traffic densities without adaptive tuning. Third, while the framework processes AIS data for route planning, it may not fully capture real-time environmental variables like weather conditions, sea state, or dynamic obstacles, which are crucial for safe navigation. Moreover, integrating multiple data sources and running deep learning models in real-time can be computationally intensive, potentially requiring significant onboard processing power or reliable network infrastructure.

Machine learning's potential in the maritime industry is significant, particularly in sustainable transportation. However, its application in the digital world is limited.

Future applications should focus on algorithms for voyage optimization, sustainable transportation, maintenance forecasting, cargo rate control, energy efficiency management, and maritime security enhancement.

CODE AVAILABILITY:

The code that supports the findings of this paper is available from author Mohamed Eldosuky, upon request.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

REFERENCES:

[1] Makridis, G., Kyriazis, D. and Plitsos, S., 2020, September. Predictive maintenance leveraging machine learning for time-series forecasting in the maritime industry. In 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC) (pp. 1-8). IEEE.

[2] Rawson, Andrew, and Mario Brito. "A survey of the opportunities and challenges of supervised machine learning in maritime risk analysis." Transport Reviews 43, no. 1 (2023): 108-130.

[3] Akyuz, E., Cicek, K. and Celik, M., 2019. A Comparative research of machine learning impact to future of maritime transportation. Procedia Computer Science, 158, pp.275-280.

[4] Munim ZH, Dushenko M, Jimenez VJ, Shakil MH, Imset M. Big data and artificial intelligence in the maritime industry: a bibliometric review and future research directions. Maritime Policy & Management. 2020 Jul 3;47(5):577-97.

[5] Du Y, Chen G, Wang S. Sustainable Operations in Maritime Industry. Journal of Marine Science and Engineering. 2023 Apr 26;11(5):922.

[6] Jung, H.R.; Chang, H.J.; Song, Y.E. Trend of Autonomous Navigation Technology for Unmanned Ship. J. Inst. Control Robot. Syst. 2019, 25, 76–87.

[7] IMO. Regulatory Scoping Exercise for the Use of Maritime Autonomous Surface Ships (MASS) MSC 99/WP.9; IMO: London, UK, 2018.

[8] Campbell, S.; Naeem, W.; Irwin, G.W. A review on improving the autonomy of unmanned surface vehicles through intelligent collision avoidance manoeuvres. Annu. Rev. Control 2012, 36, 267–283.

[9] Geng, X.; Wang, Y.; Wang, P.; Zhang, B. Motion Plan of Maritime Autonomous Surface Ships by Dynamic Programming for Collision Avoidance and Speed Optimization. Sensors 2019, 19, 434.

[10] Chae, C.-J.; Kim, M.; Kim, H.-J. A Study on Identification of Development Status of MASS Technologies and Directions of Improvement. Appl. Sci. 2020, 10, 4564.

[11] Jo, S.; D'agostini, E. Disrupting technologies in the shipping industry: How will MASS development affect the maritime workforce in Korea. Mar. Policy 2020, 120, 104139.

[12] Hexeberg, S.; Flåten, A.L.; Brekke, E.F. AIS-based vessel trajectory prediction. In Proceedings of the 20th International Conference on Information Fusion (Fusion), Xi'an, China, 10–13 July 2017.

[13] Son, W.-J.; Lee, J.-S.; Lee, H.-T.; Cho, I.-S. An Investigation of the Ship Safety Distance for Bridge across Waterways Based on Traffic Distribution. J. Mar. Sci. Eng. 2020, 8, 331.

[14] Lee, H.T. Analysis of Factors Influencing Berthing Velocity of Ship Using Machine Learning Prediction Algorithm. Master's Thesis, Korea Maritime and Ocean University, Busan, Korea, 2019.

[15] Lee, J.-S.; Son, W.-J.; Lee, H.-T.; Cho, I.-S. Verification of Novel Maritime Route Extraction Using Kernel Density Estimation Analysis with Automatic Identification System Data. J. Mar. Sci. Eng. 2020, 8, 375.

[16] Rong, H.; Teixeira, A.P.; Soares, C.G. Data mining approach to shipping route characterization and anomaly detection based on AIS data. Ocean Eng. 2020, 198, 106936.

[17] Lee, S.-W.; Cho, J.-W.; Cho, I.-S. Estimation of Berthing velocity using probability distribution characteristics in tanker terminal. Korean Inst. Navig. Port. Res. 2019, 25, 809–818.

[18] Lee, D.J. Real-time monitoring of ship's dynamic behavior characteristics by AIS. J. Korean Soc. Fish. Ocean Technol. 2008, 44, 218–228.

[19] B. Murray and L. P. Perera, "An AIS-based deep learning framework for regional ship behavior prediction", Rel. Eng. Syst. Saf., vol. 215, Nov. 2021.

[20] Han, J.; Pei, J.; Kamber, M. Cluster Analysis: Basic Concepts and Methods. In Data Mining: Concepts and Techniques, 3rd ed.; Elsevier: Waltham, MA, USA, 2011; pp. 443–494.

[21] Prasad DK, Rajan D, Rachmawati L, Rajabally E, Quek C. Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey. IEEE Transactions on Intelligent Transportation Systems. 2017 Jan 10;18(8):1993-2016.

https://www.kaggle.com/datasets/mmichelli/singapore-maritime-dataset/data

[22] Tilemachos Bontzorlos, "Singapore Maritime Dataset frames ground truth generation and statistics", GitHub repository, Feb. 2019. https://github.com/tilemmpon/Singapore-Maritime-Dataset-Frames-Ground-Truth-Generation-and-Statistics.

[23] Tilemachos Bontzorlos, "Singapore Maritime Dataset trained Deep Learning models", GitHub repository, Feb. 2019. https://github.com/tilemmpon/Singapore-Maritime-Dataset-Trained-Deep-Learning-Models.

[24] Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: Unified, real-time object detection. InProceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 779-788).

[25] Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. Ssd: Single shot multibox detector. InComputer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 2016 (pp. 21-37). Springer International Publishing.

```
1. Start

2. Load Images
  - Define the path to the dataset
  - Load images from the Singapore Maritime Dataset (SMD) directory
  - For each image in the dataset:
    - Read the image
    - Store the image in a list or array

3. Preprocess Images
  - For each image in the list:
    - Resize the image to the required input size for YOLO (e.g., 416x416 pixels)
    - Normalize pixel values (e.g., scale pixel values to the range [0, 1])
    - Convert the image to the appropriate format (e.g., BGR to RGB if needed)
    - Optionally, apply data augmentation techniques (e.g., rotation, flipping)

4. Load YOLO Model
  - Define the paths to YOLO configuration file, weights file, and class names file
  - Load the YOLO model using a deep learning framework (e.g., TensorFlow, PyTorch, OpenCV)
    - Initialize YOLO with the configuration file and weights
    - Load the class names

5. Detect Objects
  - For each preprocessed image:
    - Perform forward pass through the YOLO model
    - Get bounding box predictions and class probabilities
    - Apply confidence threshold to filter out low-confidence predictions
    - Apply Non-Maximum Suppression (NMS) to remove duplicate bounding boxes
    - Extract final bounding boxes and class labels

6. Post-process Results
  - For each detected object:
    - Draw bounding boxes on the image
    - Label the bounding boxes with class names and confidence scores
    - Optionally, save or display the results

7. End
```

Figure 1: Proposed Methodology for detecting maritime objects

```
Step 1: Initialize Framework
        Load AIS Data:
                Import AIS data containing ship positions (latitude, longitude), speed, and timestamp.
        Set Up Environment:
                Initialize necessary libraries for data processing, clustering, and density estimation.
Step 2: Data Preprocessing
        Filter Data:
                Remove duplicate records and filter out entries with invalid or missing values.
                Select relevant fields like latitude, longitude, speed, and time.
        Clean Data:
                Filter based on criteria such as speed thresholds (e.g., exclude ships moving below a
                certain speed).
        Normalize Data:
```

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

If needed, normalize geographical coordinates for consistent processing.
Step 3: Deconstruction
Identify key waypoints where ships frequently change direction or speed.
Segment the data by geographic region or ship type for more targeted analysis.
Step 4: Database Management
Organize the cleaned and decomposed data in a database or data structure.
Ensure the data is indexed by geographic location and time for efficient retrieval during analysis.
Step 5: Waypoint Detection using DBSCAN Clustering
Use DBSCAN to detect dense clusters of ship waypoints in the AIS data.
Set the eps (epsilon) and min_samples parameters based on data characteristics.
Exclude points identified as noise by DBSCAN (points that don't belong to any cluster).
Step 6: Route Density Assessment using Kernel Density Estimation (KDE):
Apply KDE to estimate the density of maritime routes based on the identified waypoints.
Generate density plots for various confidence intervals (e.g., 75%, 90%).
Create visualizations (e.g., heatmaps) to show the density and spread of the maritime routes.
Step 7: Primary Route Assessment
Analyze Density Levels:
Identify the primary maritime routes based on KDE results at 75% and 90% confidence intervals.
Incorporate Statistical Data:
Use additional statistical data on maritime traffic to refine the assessment of the primary routes.
Adjust the primary routes if necessary to ensure they are representative and safe.
Step 8: Application for Autonomous Surface Ships
Plan Routes:
Use the primary routes to plan navigation paths for autonomous surface ships.
Ensure these routes meet safety standards and are efficient for the designated ocean regions.
Simulate and Test:
Run simulations to test the effectiveness of the planned routes in different scenarios.
Make adjustments based on simulation results and real-world testing.
Step 9: Output and Reporting
Generate Reports:
Compile the findings into a report or dashboard, highlighting the identified routes and any safety recommendations.
Update Database:
Store the final route data and related analysis results in the database for future reference.

Figure 2: Proposed Methodology for maritime traffic routes

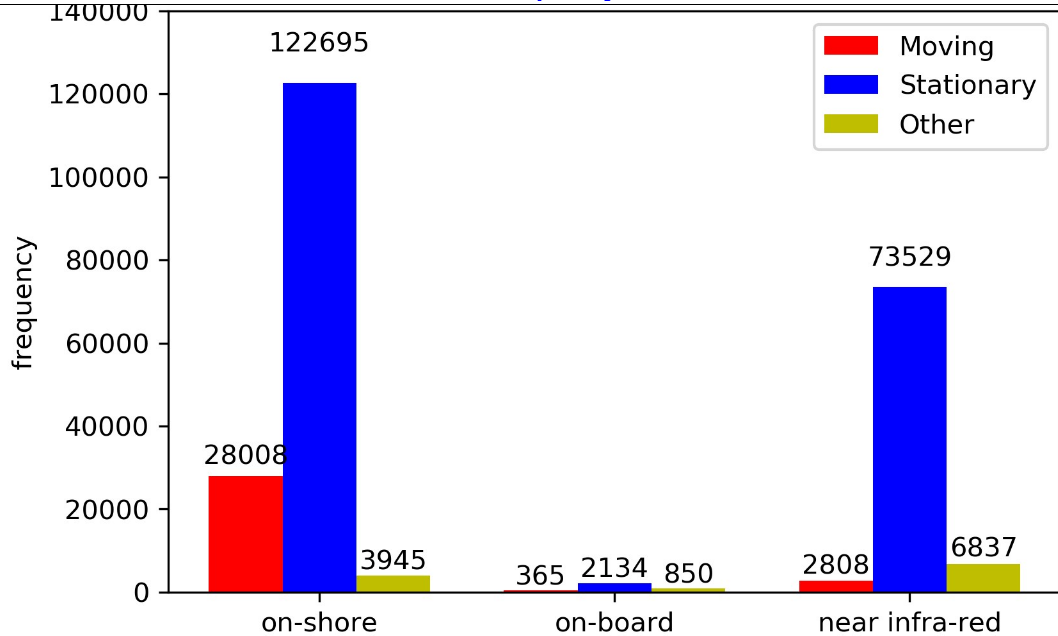

Figure 3: Motion of objects in SMD [22]

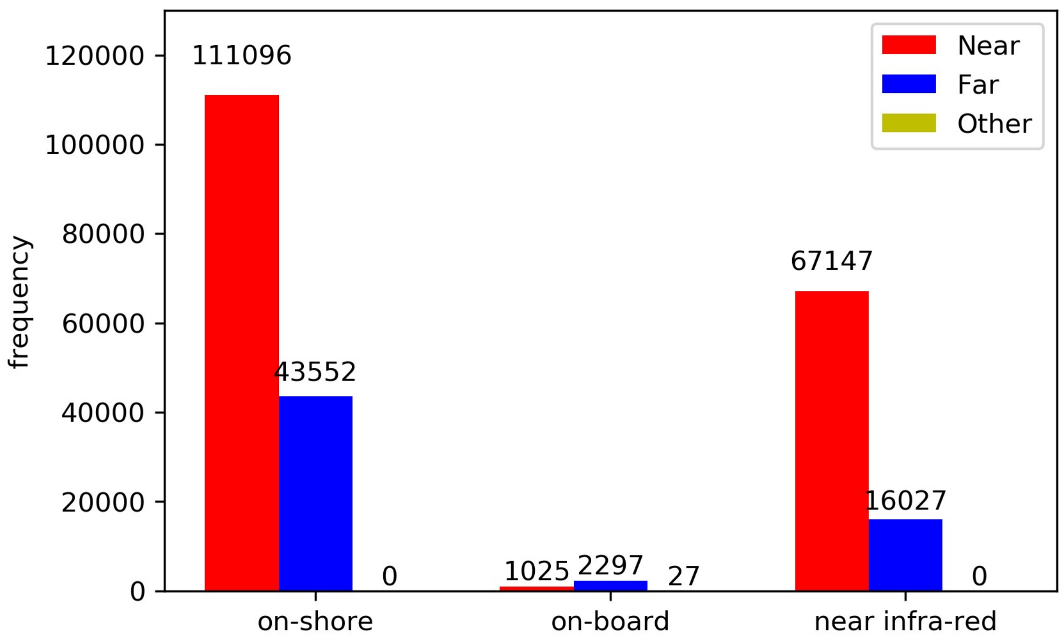

Figure 4: Distance of objects in SMD [22]

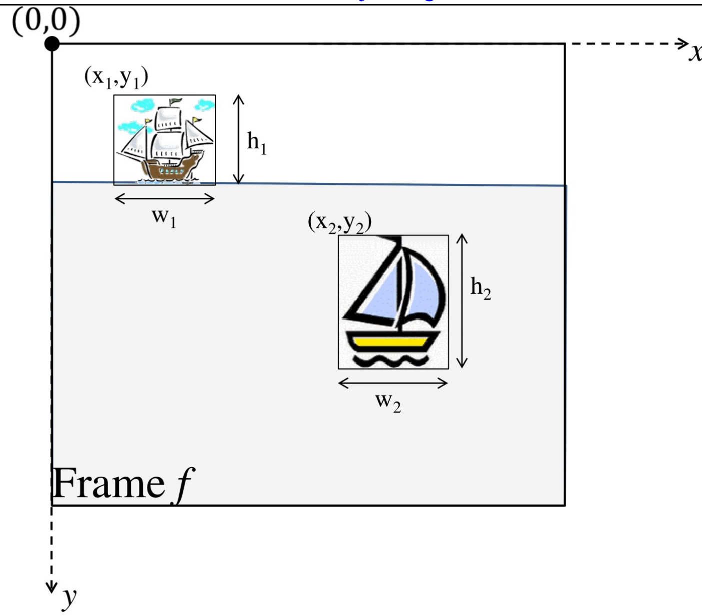

Figure 5: Ground truth for object detection

Figure 6: Sample Visible on-Shore snapshot [21]

Figure 7: Sample Visible on-board snapshot [21]

Figure 8: Sample NIR on-Shore snapshot [21]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

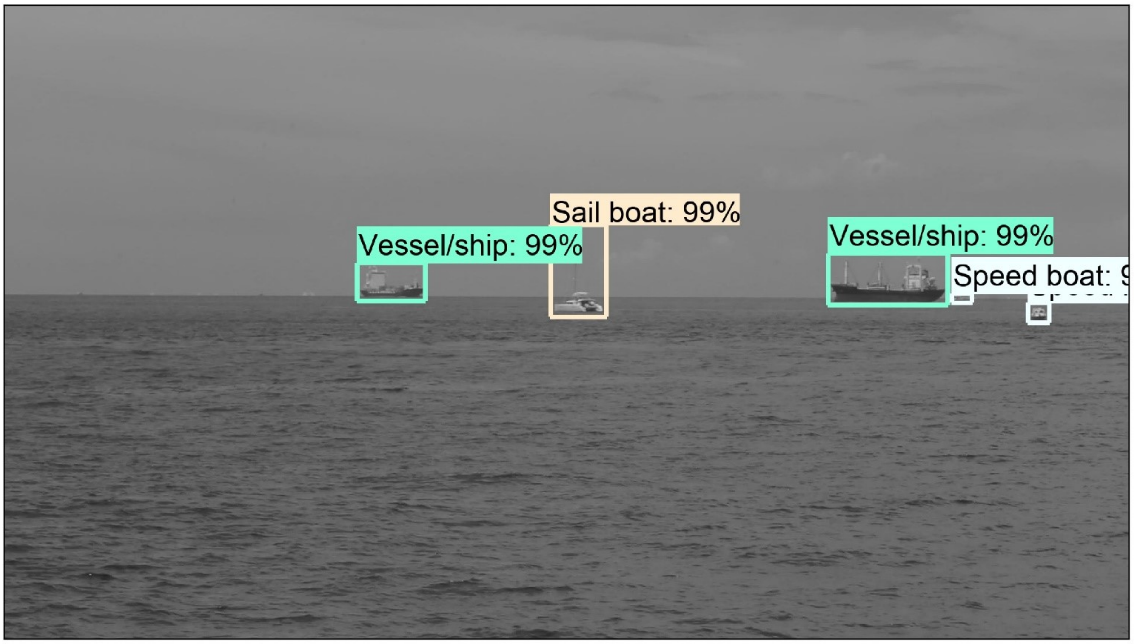

Figure 9: Result for Faster R-CNN with Inception-v2 deep learning model [23]

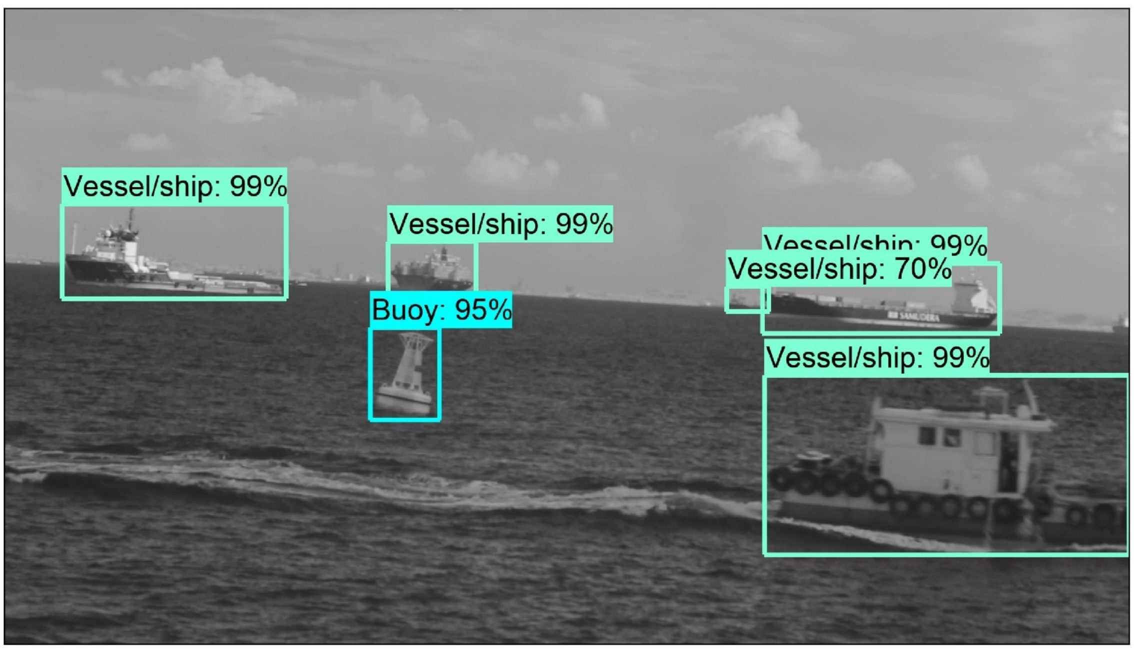

Figure 10: Result for SSD with Resnet-50 deep learning model [23]

Journal of Theoretical and Applied Information Technology
15th June 2025. Vol.103. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

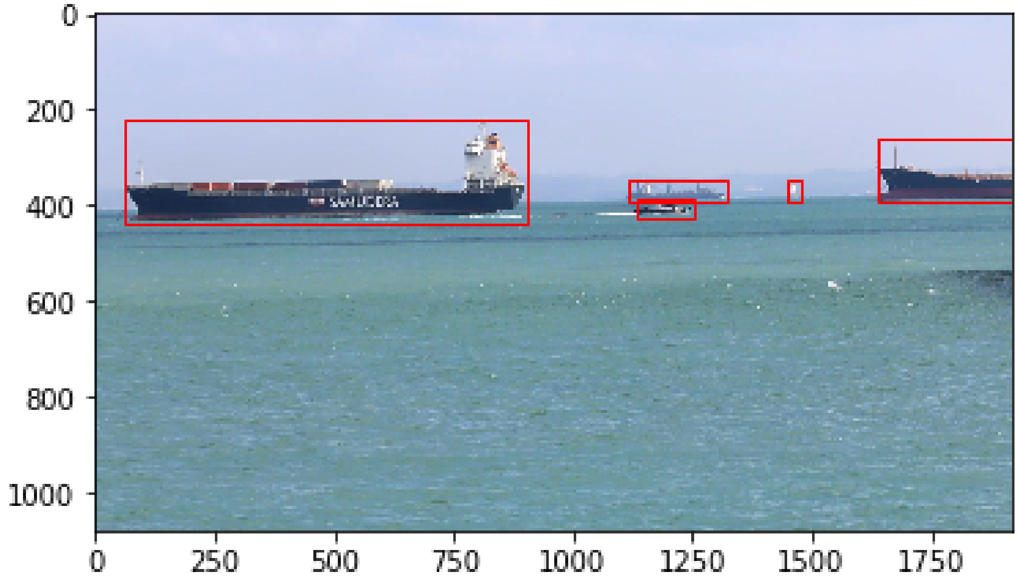

Figure 11: Fast object detection

Table 1: Various sources to download AIS data

Source	***Website***	***Details***	***Type of Data***	***Cost***
MarineTraffic	*MarineTraffic*	*Provides real-time and historical AIS data for a fee.*	*Real-time, Historical*	*Paid*
Spire	*Spire*	*Offers AIS data with global coverage, useful for detailed maritime analytics.*	*Real-time, Historical*	*Paid*
ExactEarth	*ExactEarth*	*Satellite-based AIS data with global coverage and high-quality data.*	*Real-time, Historical*	*Paid*
AISHub (AIS Open Data)	*AISHub*	*Offers a free AIS data stream for non-commercial use, with some historical datasets available.*	*Real-time, Historical (limited)*	*Free (Non-Commercial)*
Global Fishing Watch	*Global Fishing Watch*	*Provides AIS data focused on fishing vessels, useful for marine conservation studies.*	*Real-time, Historical*	*Free*

Journal of Theoretical and Applied Information Technology
15th June 2025. Vol.103. No.11

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

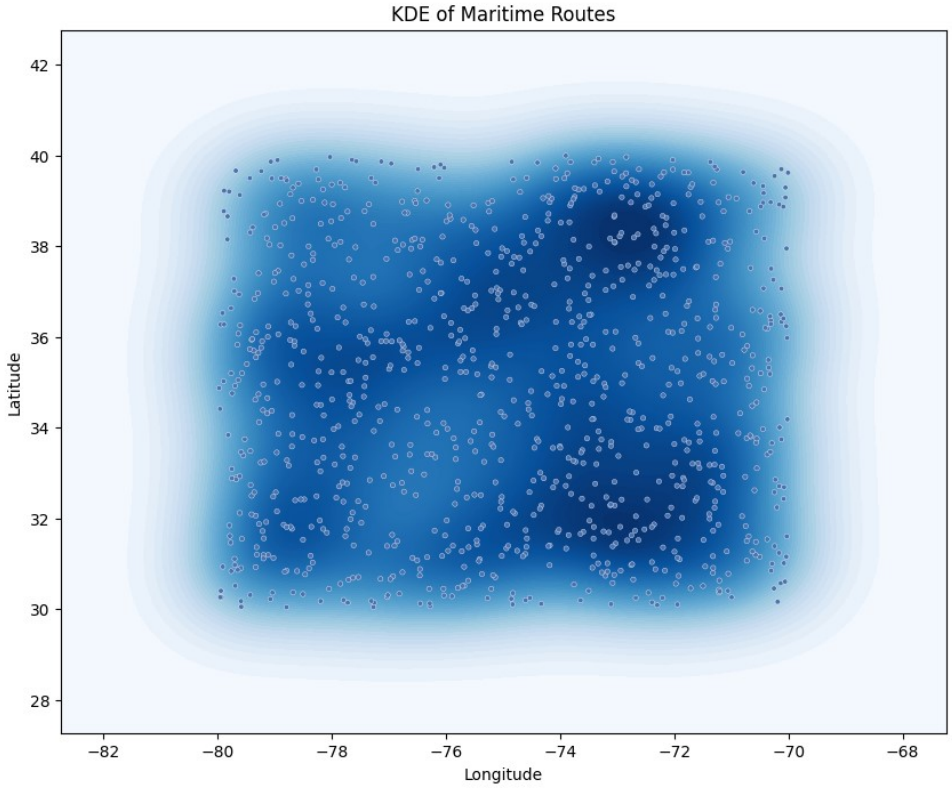

Figure 12: KDE of Maritime Routes

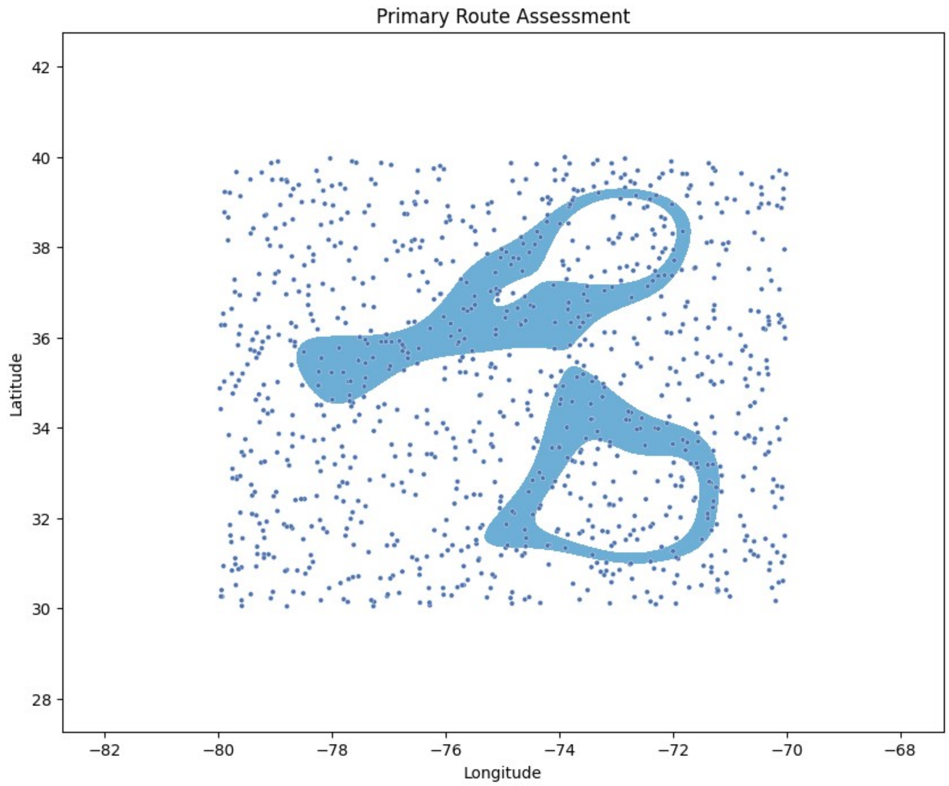

Figure 13: Primary Route Assessment