Defence Science Journal, Vol. 75, No. 4, July 2025, pp. 490-497, DOI : 10.14429/dsj.20660
© 2025, DESIDOC

Leveraging HDBSCAN, LSTM and R-DTW for Proactive Detection and Collision Prediction in Maritime Traffic

Nitish Raj[!,*] and Prabhat Kumar[#]

[!]*Weapons and Electronics System Engineering Establishment, Delhi - 110 066, India*
[#]*National Institute of Technology, Patna - 800 005, India*
[*]*E-mail: Nitish.raj1@navy.gov.in*

ABSTRACT

Detecting anomalies in Automatic Identification System (AIS) data is crucial for marine safety, especially with over 60,000 vessels navigating seaways at any moment. This study proposes an enhanced approach to AIS data analysis for detecting anomalous ship behaviours and predicting collisions in maritime environments. Unlike traditional methods that rely on static threshold-based rules or simpler clustering techniques, our approach integrates advanced machine learning methods like Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) and Long Short-Term Memory (LSTM) networks, along with Rhumb line approach Dynamic Time Warping (R-DTW) for improved trajectory similarity assessment and Closest Point of Approach (CPA) calculations. The study outperforms existing techniques by leveraging HDBSCAN's ability to handle varying-density trajectory clusters, LSTM's temporal sequence learning for more accurate movement predictions, and R-DTW's adaptability in identifying anomalous route deviations. The method includes a robust AIS data preprocessing pipeline, the use of HDBSCAN for dynamically grouping complex maritime trajectories, and LSTM models trained using a sliding window approach to predict future ship movements. CPA computations are employed to assess collision risks based on predicted trajectories. The proposed method significantly enhances anomaly detection accuracy and collision prediction reliability over conventional approaches. This integrated and data-driven approach to anomaly detection and trajectory prediction provides a substantial improvement in maritime traffic management and collision avoidance, contributing to proactive maritime safety measures.

Keywords: Anomaly detection; AIS; HDBSCAN; Long short-term memory; Closest point of approach; Deep learning

1. INTRODUCTION

The International Maritime Organisation (IMO) suggests utilising the Bridge Navigational Watch Alarm System (BNWAS, Electronic Chart Display and Information (ECDIS)) and Automatic Identification System (AIS) to assist the Officers on Watch (OOW).

AIS, which is critical for vessel identification and location, transmits data to satellites, base stations, and ships, aiding collision avoidance by broadcasting navigational data such as position, speed, and course. Mandated for passenger ships and commercial vessels over 300 gross tonnes, AIS enhances situational awareness and marine traffic management, crucial for real-time navigation and post-voyage analysis. Vessel movement, influenced by currents, weather, and traffic, complicates trajectory prediction and anomaly detection. Advanced data analysis, including machine learning and clustering algorithms, is vital for identifying anomalous behaviours and enhancing maritime safety.

Since 1959, substantial volumes of cargo worth billions of dollars have been transported between Canada, the U.S., and other countries. With growing seaside congestion, ensuring maritime safety is becoming increasingly crucial. Collision avoidance depends on AIS data, which includes geographical position (LAT, LON), course over ground, speed over ground, and vessel type. However, vast AIS datasets and irregularities complicate trajectory prediction and risk assessment. This study uses HDBSCAN for anomaly detection and an LSTM model for trajectory prediction to improve safety through proactive collision avoidance. AIS, operating on Very High Frequency (VHF) frequencies (156–162 MHz), enhances navigational safety, collision prevention, and traffic management and is classified into Class A and Class B types.

Rhumb Line - Dynamic Time Warping (R-DTW) is an algorithm measuring similarity between temporal sequences that is useful in time series analysis and other applications. The authors have modified the Dynamic Time Warping (DTW) algorithm, which was previously used in comparable research studies, but with the difference that the distance measurements were based on the Euclidean metric. Nevertheless, the research carried out for this paper employed the Rhumb Line for more accuracy in a longer range.

Research employing Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) demonstrates that LSTM models outperform Exponential Smoothing (ETS), Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), and Recurrent Neural Networks (RNNs) in path

Received : 30 September 2024, Revised : 22 March 2025
Accepted : 27 March 2025, Online published : 26 June 2025

prediction. HDBSCAN also surpasses other density-based clustering algorithms, like DBSCAN. This paper leverages these advancements, using HDBSCAN for anomaly detection and an LSTM model for trajectory prediction.

Recent maritime incidents, such as the South China Sea collision, underscore the limitations of conventional collision avoidance systems, which often rely on rule-based or heuristic approaches. This study integrates an advanced data-driven framework combining HDBSCAN for anomaly detection with LSTM models for trajectory prediction, using Rhumb Line based Dynamic Time Warping (R-DTW) and Closest Point of Approach (CPA) computations to identify hazards proactively, rather than reactively.

The methodology enhances traditional AIS data filtering by incorporating Maritime Mobile Service Identity (MMSI), position, and Course Over Ground (COG) parameters, enabling a more structured preprocessing pipeline. Unlike existing studies that primarily use static trajectory mapping, this work employs dynamic visualization with Folium and applies cubic spline interpolation to smooth movement patterns, improving prediction reliability. HDBSCAN clusters contextually similar movement behaviors, automatically distinguishing dense traffic areas from outliers, unlike conventional DBSCAN or k-means clustering. This reduces noise while preserving navigational trends, allowing for more precise anomaly identification.

By dynamically comparing predicted trajectories against both historical movement trends and detected anomalies, the LSTM model offers improved forecasting accuracy compared to traditional autoregressive or Kalman filtering methods. CPA calculations incorporate predicted deviations rather than relying solely on static course projections, ensuring a more accurate assessment of collision risks. The effectiveness of this approach will be tested using real AIS data, model accuracy measured through Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), and overall performance validated through comparative simulations against existing maritime traffic prediction techniques.

This paper introduces a novel, integrated anomaly detection and predictive collision avoidance system, overcoming the constraints of prior work by combining AIS data filtering, adaptive visualization, HDBSCAN clustering, and deep-learning-based LSTM trajectory prediction. The proposed framework demonstrates a significant advancement over traditional methods by integrating both spatial and temporal insights, leading to a proactive and adaptive maritime safety mechanism.

2. LITERATURE REVIEW

AIS data has been pivotal in maritime safety, traffic management, and trajectory pre-diction research. Various methodologies enhance vessel trajectory prediction accuracy and efficiency, Yang[1], *et al.* used Bi-directional LSTM (Bi-LSTM) for trajectory pre-diction, improving accuracy by denoising AIS data through trajectory separation, data cleaning, and standardization. The Bi-LSTM outperformed models like ARIMA, SVR, and LSTM. Mozaffari[2], *et al.* give a detailed assessment of deep learning-based techniques for car behaviour prediction, emphasising their better performance in complicated driving settings. compared to traditional approaches, and outlining major research gaps and prospects. Raj[3], *et al.* reviewed deep learning techniques for vessel trajectory prediction, discussing methodologies, data sources, and challenges in the field. They emphasised the need for robust models to handle vast AIS data and diverse vessel patterns. Lee[4], *et al.* developed an LSTM algorithm to predict maritime traffic conditions, thereby assisting autonomous ships. Their model effectively predicted traffic conditions using features like traffic volume and congestion.

Zhao and Shi[5] combined density-based clustering and RNN for maritime anomaly detection, identifying deviations from standard navigational practices, and enhancing maritime safety. Alam and Torgo[6] introduced a clustering-based framework focussing on vessel type-specific behaviours for location prediction, improving accuracy and computational efficiency. Rong[7], *et al.* used trajectory compression and clustering for probabilistic maritime traffic characterisation and anomaly detection, enhancing traffic management. Lin[8], *et al.* developed a collision risk prediction model using CNN and LSTM, achieving high accuracy in predicting spatial-temporal collision risks. Murray and Perera[9] proposed a deep learning framework using variational recurrent autoencoders and HDBSCAN for high-fidelity regional ship behaviour prediction.

Wu[10], *et al.* integrated ConvLSTM and Sequence-to-Sequence (Seq2Seq) models for ship trajectory prediction showing excellent performance in predicting turning and straight-line trajectories. Olesen[11], *et al.* developed a contextually supported abnormality detector using clustering and deep learning for early risk warnings. Alam[12], *et al.* studied enhanced short-term vessel trajectory prediction by clustering route patterns for each vessel type, thereby improving accuracy and computational efficiency. The study by Raj and Kumar[13] demonstrates the effectiveness of integrating linear regression (LR) and long short-term memory (LSTM) methods to forecast vessel positions using AIS data, to improve maritime operations. These studies demonstrate advancements in vessel trajectory prediction and anomaly detection. These methods improve maritime safety and traffic management by combining deep learning models like LSTM, Bi-LSTM, and ConvLSTM with clustering algorithms like HDBSCAN. They do this by dealing with problems in AIS data and complicated vessel movement patterns.

Building on these insights, this paper combines anomaly detection and an LSTM model for trajectory prediction to improve maritime safety. The method accurately predicts neighbouring vessel trajectories and estimates CPA, providing a robust solution for collision avoidance. Experiential validation with real-world AIS data highlights the method's practical applicability in maritime traffic management and safety enhancement.

3. METHODOLOGY

3.1 Data Acquisition

Selecting the right data source is key for AIS analysis due to quality and availability differences. Researchers found that Marine Cadastre and AIS Explorer offer the highest time

MMSI	BaseDateTime	LAT	LON	SOG	COG	VesselType	Date	Time	time_diff_minutes
209729000	01-05-2023 14:53	26.73888	-87.5327	14.1	33.1	70	01-05-2023	14:53:41	
209729000	01-05-2023 15:04	26.77623	-87.50645	14.2	32.6	70	01-05-2023	15:04:53	11.2
209729000	01-05-2023 15:08	26.78922	-87.49723	14.3	33.5	70	01-05-2023	15:08:47	3.9
209729000	01-05-2023 15:09	26.79294	-87.49456	14.3	33.4	70	01-05-2023	15:09:53	1.1
209729000	01-05-2023 15:11	26.79996	-87.48955	14.3	33.3	70	01-05-2023	15:11:59	2.1
209729000	01-05-2023 15:14	26.80963	-87.4826	14.3	33.3	70	01-05-2023	15:14:53	2.9
209729000	01-05-2023 15:18	26.82225	-87.47338	14.3	33.9	70	01-05-2023	15:18:41	3.8
209729000	01-05-2023 15:26	26.84874	-87.45369	14.3	33.8	70	01-05-2023	15:26:41	8
209729000	01-05-2023 15:30	26.85998	-87.44512	14.3	33.6	70	01-05-2023	15:30:06	3.416666667
209729000	01-05-2023 15:31	26.86487	-87.44142	14.3	34.5	70	01-05-2023	15:31:35	1.483333333
209729000	01-05-2023 15:32	26.86922	-87.4381	14.3	34.2	70	01-05-2023	15:32:54	1.316666667
209729000	01-05-2023 15:35	26.87775	-87.4316	14.4	34.6	70	01-05-2023	15:35:29	2.583333333
209729000	01-05-2023 15:38	26.88765	-87.42405	14.4	35.3	70	01-05-2023	15:38:29	3
209729000	01-05-2023 15:40	26.89329	-87.4197	14.4	34.9	70	01-05-2023	15:40:12	1.716666667
209729000	01-05-2023 16:23	27.03484	-87.30628	14.4	37	70	01-05-2023	16:23:41	43.48333333
209729000	01-05-2023 16:30	27.05866	-87.28998	14.1	30.4	70	01-05-2023	16:30:47	7.1
209729000	01-05-2023 16:32	27.06472	-87.28592	14.1	30.5	70	01-05-2023	16:32:35	1.8
209729000	01-05-2023 16:33	27.0691	-87.28296	14.1	30.1	70	01-05-2023	16:33:53	1.3
209729000	01-05-2023 16:38	27.08393	-87.27297	14.1	30.6	70	01-05-2023	16:38:17	4.4
209729000	01-05-2023 16:42	27.09748	-87.26378	14.1	31.6	70	01-05-2023	16:42:18	4.016666667
209729000	01-05-2023 16:43	27.10112	-87.2613	14	30.8	70	01-05-2023	16:43:23	1.083333333

Figure 1. Filtered data.

resolution among non-commercial sources. Marine Cadastre provides comprehensive, free data but is limited to U.S. waters and does not include current-year data, only up to the previous year.

Marine Cadastre was chosen for its daily AIS data in ZIP format. Initially, the entire U.S. coastline was analysed, but the focus shifted to a region below Panama (longitude -88.2729 to -83.9557, latitude 26.6614 to 29.8356) for the period from December 2, 2022, to June 30, 2023. This area, with its significant maritime traffic and key ports, improved accuracy and relevance.

3.2 Data Filtering Methodology

This dataset initially contains 4,220,319 rows with the following columns: MMSI, Latitude LAT, LON, SOG, COG, Vessel Type, and BaseDateTime. Meticulous filtering steps were performed to ensure the dataset's integrity and suitability for predictive modelling.

3.2.1 Exploratory Data Analysis (EDA)

- **Handling Missing Values:** The dataset was examined for any missing values (NAs). Any rows containing missing values were dropped to ensure data completeness and integrity
- **Removing Duplicate Values:** Duplicate rows, where all values were identical, were removed to ensure data unique-ness and prevent redundancy. After this process, the dataset was reduced to 72,253 rows
- **Initial Duplicate Values**
    - LON: Values were constrained to the range [-180, 180] degrees. No rows were excluded
    - LAT: Values were constrained to the range [-90, 90] degrees. No rows were excluded
    - COG: Values were constrained to the range [0, 359.9] degrees. No rows were excluded.

3.2.2 Feature Engineering

- **Splitting BaseDateTime:** The BaseDateTime column was split into separate Date and Time columns for granular analysis, focusing on May 1-10, 2023, reducing the dataset to 317,815 rows and increasing the columns to 9.
- **Adding time_difference_minutes Feature:** A time_difference_minutes feature was added to calculate the time difference between consecutive data points. Differences exceeding 300 minutes indicate new trajectories, enhancing temporal resolution.
- **SOG Grouping:** SOG values were grouped by MMSI to calculate the average SOG. Vessels with an average SOG greater than 5 knots were retained to exclude stationary vessels, reducing the dataset to 158,449 rows.

3.2.3 Filtering Process

- **MMSI Filtering:** To ensure valid MMSI, the dataset was filtered to retain only 9-digit MMSI numbers. This step reduced the dataset to 4,219,742 rows, removing 577 rows.
- **Vessel Type Filtering:** Certain vessel types, such as fishing boats or those operating near ports, exhibit erratic movement patterns that complicate clustering and prediction processes. To retain specific vessel types, the data was filtered using AIS type codes: 80: Tanker Type Ships, 35: Military Ops, 60: Passenger Ships and 70: Cargo Ships. This filter further reduced the dataset to 72,254 rows.
- **Sorting by BaseDateTime:** The dataset was sorted by BaseDateTime to ensure proper sequencing of vessel trajectories, which is crucial for accurate trajectory prediction.
- **Length of the Trajectory:** MMSIs with fewer than 1,000 data points were filtered out to ensure substantial trajectory information. This step reduced the dataset to 35,624 rows.

Following these thorough filtering steps, the dataset (Fig. 1) underwent significant refinement, retaining only 26 unique MMSIs from an initial 3,132, thereby providing a robust foundation for accurate trajectory prediction modelling.

3.3 Trajectory Extraction and Interpolation

After initial filtering, vessel trajectories from AIS data were processed by dividing each vessel's journey into multiple segments using MMSI and a unique trajectory ID (Fig. 2). Gaps

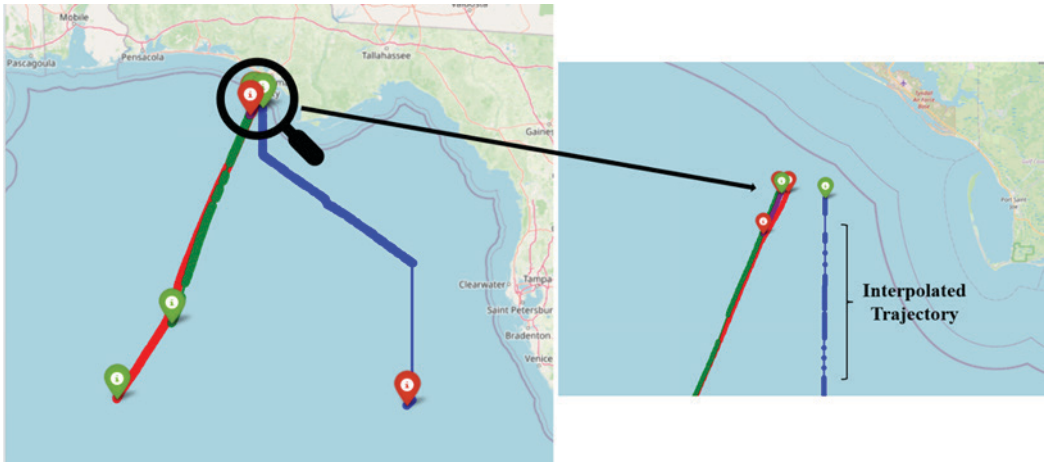

Figure 2. Interpolation of trajectories.

in AIS messages were handled by interpolating only where messages were not broadcast between 3 and 300 minutes, as longer gaps typically indicate extended halts. Cubic spline interpolation was used to estimate positions during these gaps, based on the vessel's average broadcast frequency in that trajectory. This method offers smoothness, flexibility, minimal oscillation, and effective trend preservation, making it ideal for interpolating vessel trajectories in AIS data.

3.4 R-DTW + HDBSCAN

R-DTW and HDBSCAN were employed to analyse and cluster vessel trajectories. This combination allowed us to account for temporal and spatial variations in the trajectories, providing more accurate and meaningful clustering results (Fig. 3).

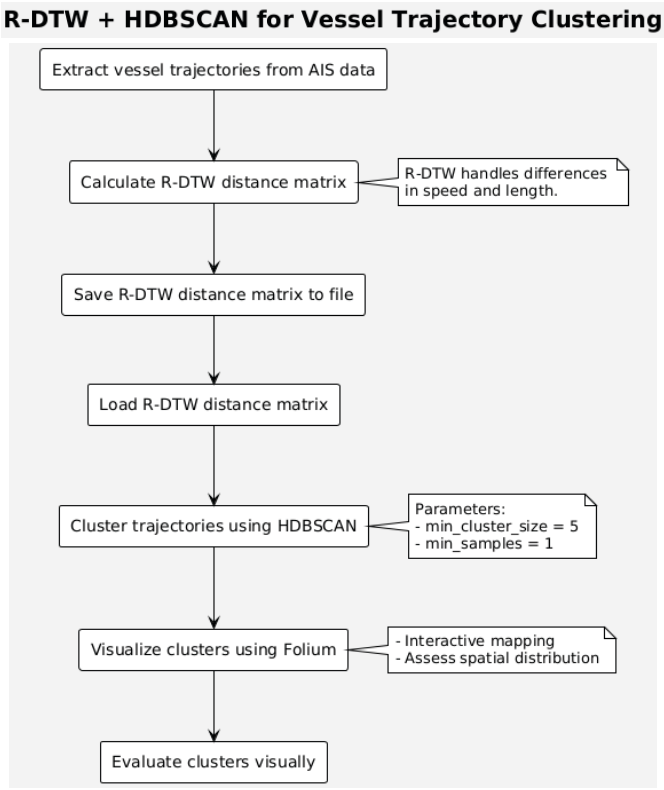

Figure 3. Workflow for detecting vessel behaviour patterns with R-DTW & HDBSCAN.

3.4.1 R-DTW

Before clustering, R-DTW was used to measure trajectory similarity, handling speed and length differences better than traditional Rhumb Line distances. R-DTW aligns sequences based on shape rather than point-by-point, making it ideal for comparing maritime paths with varying temporal dynamics.

3.4.2 Clustering with HDBSCAN

LAT and LON were used for clustering, focusing on these for a generalized model. HDBSCAN was chosen over DBSCAN for its density handling and automatic epsilon determination. Key parameters are "min_cluster_size" (cluster size) and "min_samples" (core point classification). A grid search (Table 1) was performed, and cluster quality was assessed with silhouette scores. Due to path overlaps, scores were sometimes misleading, so results were visually inspected. "min_samples" was set from 1 to 11 and "min_cluster_size" from 2 to 21, with several parameter combinations yielding silhouette scores above 0.2.

Table 1. Silhouette score in different situations

Min_samples	Min_cluster_size	Silhouette score
1	3	0.26725
1	5	0.26970
1	6	0.23613
1	7	0.23613
1	8	0.23613
1	9	0.23613
1	10	0.23613
1	11	0.23613
1	12	0.32652
1	13	0.32652
1	14	0.32652
2	2	0.25493
2	3	0.25493
2	4	0.25016
2	5	0.25016
2	6	0.25453
2	11	0.29606
2	12	0.29606
2	13	0.29606
3	10	0.27605
3	11	0.27605
3	12	0.27605

The best configuration was "min_samples=1" and "min_cluster_size=5", which resulted in five distinct clusters, including one for noise. Although "min_cluster_size=12" with a silhouette score of 0.3265 produced only two clusters, they were unsatisfactory. The final setup effectively captured the data's patterns, with 17 noise points identified (Fig. 3).

3.4.3 Evaluation and Visualization

Due to overlapping trajectories, the silhouette score was less effective for evaluating clustering quality. Instead, visual inspection with Folium, an interactive mapping library, was used to assess the spatial distribution and coherence of clusters (Fig. 4).

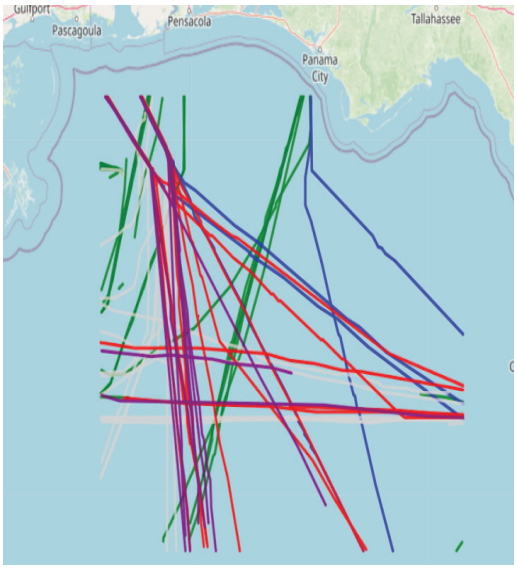

Figure 4. Clustered trajectories.

3.5 LSTM Model

LSTM networks, a type of RNN, overcome vanishing and exploding gradient problems using a unique cell structure with Memory Cells, Input Gate, Forget Gate, and Output Gate. These gates allow LSTMs to selectively remember and forget information, capturing long-term dependencies effectively.

3.5.1 Sliding Window Concept

The LSTM model captures temporal dependencies by using a sliding window of 50 data points to predict the next point.

3.5.2 LSTM Model Deployment

The hyperparameters for training the model (Table 2) are chosen based on empirical evaluations to optimize performance while preventing overfitting:

- **Epochs (20):** The model undergoes training for 20 cycles across the complete dataset. This value was determined by monitoring validation loss, where additional epochs provided diminishing returns and signs of overfitting
- **Batch Size (32):** The model updates its weights after every 32 data points. This batch size balances computational efficiency and gradient stability, preventing excessive noise while maintaining reasonable convergence speed
- **Validation Split (0.2):** Twenty percent of the training data is reserved for validation. This ratio was selected after testing different splits (10 %, 30 %), with 20 % providing the best trade-off between training data availability and model generalization.

Table 2. Hyperparameter settings for model

Hyper_Parameter	Values
Epochs	20
Batch_size	32
Validation_split	0.2

3.5.3 Model Training and Validation

Training and validation loss values were monitored, and found 0.0001& 0.0006 respectively. These low values indicate effective learning and good generalization (Fig. 5).

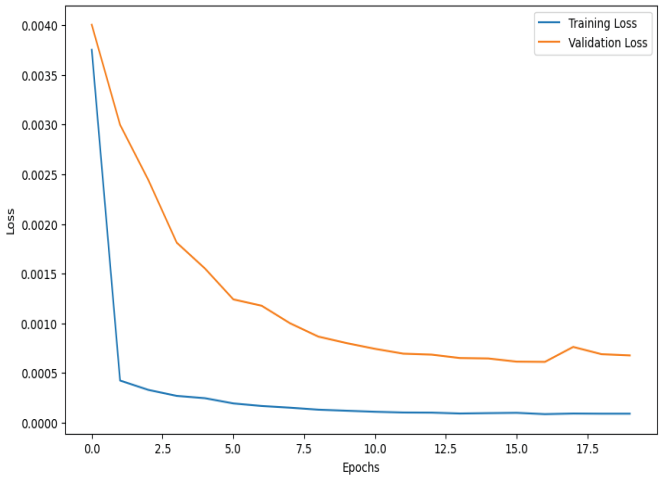

Figure 5. Training and validation loss curves over epochs.

3.5.4 Trajectory Prediction and Anomaly Detection

3.5.4.1 Prediction Process

The last 50 points of the vessel's trajectory are removed to evaluate the model's prediction accuracy. The actual and predicted trajectories are compared after the model predicts these 50 points. (Table 3) Anomalies are identified when the predicted trajectory significantly deviates from the actual trajectory.

Table 3. Predicted vessel trajectory data

MMSI	LAT	LON	Time
309761000	27.953092	-87.937632	03:26:35
309761000	27.946034	-87.940097	03:26:42
309761000	27.938481	-87.942439	03:26:48
309761000	27.930610	-87.944713	03:26:55
309761000	27.922408	-87.946932	03:27:02
-	-	-	-
-	-	-	-
309761000	27.529361	-87.998084	03:31:32
309761000	27.519787	-87.998984	03:31:39
309761000	27.510228	-87.999878	03:31:45
309761000	27.500685	-88.000766	03:31:52
309761000	27.491158	-88.001649	03:31:59

3.5.4.2 Intersection and Collision Detection

Using interpolated data and predicted trajectories, vessels that intersect the target vessel's path are identified. MMSI 309761000 intersects with the following MMSIs: 367633000, 636014278, 311001144, 303520000, 311018700, 258288000, and 477050700. Further analysis (Table 4) confirms that no vessels will disrupt the predicted path within 10 minutes, ensuring safe navigation.

To validate the effectiveness of our proposed method, we compared its performance with a baseline linear regression model and a Kalman filter-based trajectory prediction approach. The results demonstrate that our model achieves significantly lower error rates (MSE, RMSE, MAE, and MAPE) and a higher R-squared value, indicating better predictive accuracy and robustness.

Table 4. Model Evaluation Metrics

Performance matrix	Proposed model (RDTW+LSTM +HDBSCAN)	Kalman filter	Linear regression
MSE	0.0001	0.0023	0.0041
RMSE	0.0105	0.0480	0.0642
MAE	0.0037	0.0201	0.0314
MAPE	0.0001	0.0028	0.0039
R-squared	0.9996	0.9342	0.8793

The comparison highlights that our model outperforms traditional approaches by reducing prediction errors while maintaining high accuracy. The LSTM model effectively captures non-linear motion patterns, whereas the Kalman filter and linear regression struggle with complex trajectory variations. These findings reinforce the robustness and reliability of our proposed approach for proactive maritime navigation.

3.6 CPA

In maritime navigation, the Closest Point of Approach (CPA) assesses collision risk by identifying the closest meeting point of two converging vessels. This metric helps mariners and automated systems avoid collisions. The research includes CPA, which is crucial for evaluating the safety of predicted vessel trajectories.

3.6.1 Time to CPA (TCPA)

TCPA measures the time until vessels reach their closest point of approach. It helps mariners assess collision imminence (Fig. 6) and make timely decisions to alter course or speed to avoid accidents.

The formula for TCPA:

$$TCPA = \frac{DCPA}{v_{relative}} \quad (1)$$

In Eqn. 1, DCPA is the distance at the closest point of approach and $v_{relative}$ is the relative velocity between the two vessels.

3.6.2 DCPA

DCPA measures the minimum distance between two vessels if they maintain their current courses and speeds. A smaller DCPA indicates a higher collision risk, while a larger DCPA suggests a safer distance (Fig. 6). It's a key spatial measure for assessing the closest approach of vessels.

The formula for DCPA:

$$DCPA = \frac{|v_1 \times v_2|}{\sqrt{v_1^2 + v_2^2}} \quad (2)$$

In Eqn. 2, V_1 and V_2 are the velocity vectors of the two vessels.

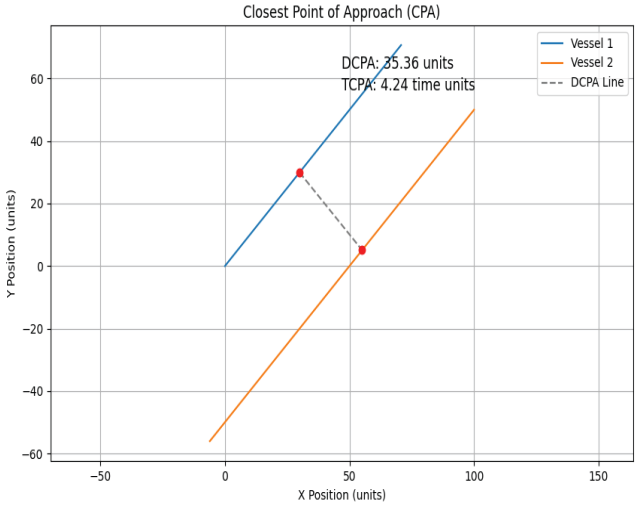

Figure 6. Vessel trajectories with CPA, DCPA, and TCPA.

3.6.3 CPA's Importance in Detecting Anomalous Behaviour and Collision Risk

3.6.3.1 Collision Hazards

CPA calculations with AIS data and LSTM models predict future vessel positions, enabling proactive collision prevention by altering course or speed.

3.6.3.2 Anomalous Behaviours

- Direct Detection: Monitoring CPA and low DCPA values reveals anomalies like sudden speed or course changes.
- Real-Time Monitoring: Continuous CPA calculations trigger alerts for deviations and prompt preventive actions.
- Historical Analysis: Analysing past AIS data with CPA uncovers trends of anomalous behaviour.

3.6.3.3 Enhancing Marine Safety and Security

Integrating CPA with AIS data predicts collisions, provides real-time alerts, maintains safe distances, and enhances maritime security.

3.6.4 Mathematics

3.6.4.1 Definitions and Notations

- Position Vectors: Let P_1 (t), and P_2 (t), be the position vector of vessel 1 and vessel 2 at time t
- Velocity Vectors: Let V_1, and V_2, be the constant velocity vector of vessel 1 and vessel 2
- Initial Positions: Let $P_{1,0}$ and $P_{2,0}$ be the initial position vector of vessel 1 and vessel 2 at time t = 0.

3.6.4.2 Equations of Motion

The position of each vessel at any time t can be described as:

$$P_1(t) = P_{1,0} + V_1 * t \quad (3)$$

$$P_2(t) = P_{2,0} + V_2 * t \quad (4)$$

Relative Position and Velocity:

Define the relative position and velocity vectors:

$$P_{rel}(t) = P_1(t) - P_2(t) = (P_{1,0} - P_{2,0}) + (V_1 - V_2) * t \quad (5)$$

$$V_{rel} = (V_1 - V_2) \quad (6)$$

Let $P_{rel,0} = P_{1,0} - P_{2,0}$ be the initial relative position.

Time to Closest Point of Approach (TCPA) : To find the time t_{CPA} at which the distance between the two vessels is minimized, the time when the derivative of the squared distance concerning time is zero is solved for.

$$\left(\frac{\partial(P_{rel}(t))}{dt}\right) * P_{rel}(t) = 0 \quad (7)$$

This yields:

$$P_{rel}(t) * V_{rel} = 0 \quad (8)$$

Solving for t, the result is

$$t_{CPA} = -\left(\frac{P_{rel,0} * V_{rel}}{V_{rel} * V_{rel}}\right) \quad (9)$$

Range at Closest Point of Approach (RCPA): Substitute t_{CPA} back into the relative position equation to find the relative position at CPA:

$$P_{CPA} = P_{rel,0} + V_{rel} * t_{CPA} \quad (10)$$

The RCPA is the magnitude of P_{CPA}:

$$RCPA = \| P_{CPA} \| = \| P_{rel,0} + V_{rel} * t_{CPA} \| \quad (11)$$

If RCPA is less than a safe distance (D_{safe}) and TCPA is within a critical time frame, a collision risk exists, requiring evasive action. This helps predict collisions and enhance navigation safety. CPA is calculated by selecting a point within a vessel's predicted trajectory and setting a threshold distance to identify nearby vessels. The closest distance between the predicted trajectory and any passing vessel is then determined, aiding in informed route decisions to avoid collisions (Fig. 7).

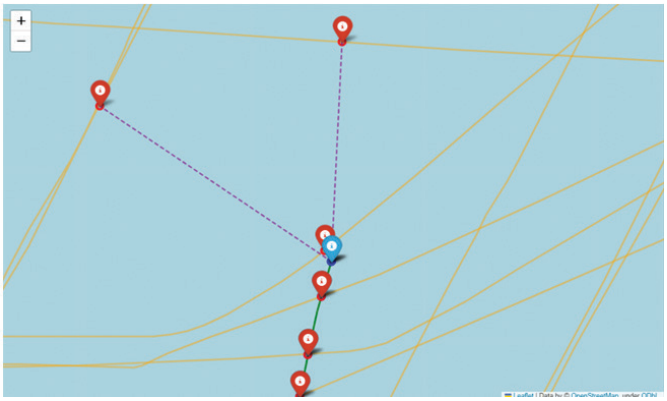

Figure 7. Closest point of approach.

The methods were validated using real-world AIS data, confirming their practical applicability in maritime operations. In the visualization (Fig. 8), the **Blue Line** represents the actual vessel trajectory, while the **Green Line** depicts the predicted path generated by the model. **Red Points** indicate intersection points with other vessels. This visual representation effectively highlights deviations from the actual path and potential points of intersection, providing clear insights into the model's performance and the safety of the predicted trajectory.

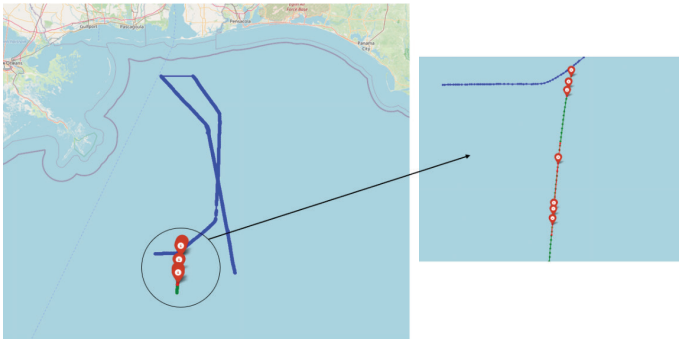

Figure 8. Anomaly detection.

4. RESULT & DISCUSSION

The study showcased notable advancements in the detection of anomalous ship behaviors and the prediction of potential collisions using AIS data, contributing significantly to maritime safety and operational efficiency. Through the application of HDBSCAN clustering, the research achieved high precision in identifying irregular patterns, such as abrupt course changes and unauthorized breaches of restricted zones, which are critical indicators of anomalous activity. Collision prediction was enhanced by integrating Long Short-Term Memory (LSTM) networks with Closest Point of Approach (CPA) calculations, providing a reliable framework for anticipating and mitigating potential collision scenarios. Furthermore, advanced preprocessing techniques, including noise removal and trajectory interpolation, were instrumental in refining AIS data, ensuring greater clarity and accuracy for downstream analysis. These methodologies were rigorously validated with real-world maritime datasets, demonstrating their effectiveness and practical applicability in dynamic maritime environments.

5. CONCLUSION

This study presents a novel approach combining HDBSCAN clustering and LSTM networks to enhance maritime safety by detecting anomalous ship behaviours and predicting collisions. The results demonstrate the method's accuracy and reliability, highlighting its potential for real-world applications in maritime traffic management. Future research should focus on refining these techniques, integrating additional data sources, and exploring emerging technologies to further improve detection and prediction capabilities. This approach provides a valuable tool for proactive risk management and collision avoidance in congested seaways.

REFERENCES

1. Yang CH, Wu CH, Shao JC, Wang YC, Hsieh CM. AIS-based intelligent vessel trajectory prediction using bi-LSTM. Ieee Access. 2022 Feb 25;10:24302-15. doi: 10.1109/ACCESS.2022.3154812
2. Mozaffari S, Al-Jarrah OY, Dianati M, Jennings P,

Mouzakitis A. Deep learning-based vehicle behavior prediction for autonomous driving applications: A review. IEEE Transactions on Intelligent Transportation Systems. 2020 Aug 4;23(1):33-47.
doi: 10.1109/TITS.2020.3012034.

3. Raj N, Kumar P. Navigating the Future: A Comprehensive Review of Vessel Trajectory Prediction Techniques. Defence Science Journal. 2025 Jan;75(1):129-38.
doi : 10.14429/dsj.20287
4. Lee E, Khan J, Son WJ, Kim K. An efficient feature augmentation and LSTM-based method to predict maritime traffic conditions. Applied Sciences. 2023 Feb 16;13(4):2556.
doi: 10.3390/app13042556
5. Zhao L, Shi G. Maritime anomaly detection using density-based clustering and recurrent neural network. The Journal of Navigation. 2019 Jul;72(4):894-916.
doi: 10.1017/S0373463319000031
6. Alam MM, Torgo L. A clustering-based approach for predicting the future location of a vessel. InCanadian AI 2022 May 27.
doi: 10.1016/j.knosys.2021.107561.
7. Rong H, Teixeira AP, Soares CG. Data mining approach to shipping route characterization and anomaly detection based on AIS data. Ocean Engineering. 2020 Feb 15;198:106936.
doi: 10.1016/j.trc.2020.01.011.
8. Lin C, Zhen R, Tong Y, Yang S, Chen S. Regional ship collision risk prediction: An approach based on encoder-decoder LSTM neural network model. Ocean Engineering. 2024 Mar 15;296:117019.
doi: 10.1109/TITS.2024.3115102.
9. Murray B, Perera LP. An AIS-based deep learning framework for regional ship behavior prediction. Reliability Engineering & System Safety. 2021 Nov 1;215:107819.
doi: 10.1109/ACCESS.2021.3088154.
10. Wu W, Chen P, Chen L, Mou J. Ship trajectory prediction: An integrated approach using ConvLSTM-based sequence-to-sequence model. Journal of Marine Science and Engineering. 2023 Jul 25;11(8):1484.
doi: 10.1016/j.oceaneng.2023.111147.
11. Olesen KV, Boubekki A, Kampffmeyer MC, Jenssen R, Christensen AN, Hørlück S, Clemmensen LH. A Contextually Supported Abnormality Detector for Maritime Trajectories. Journal of Marine Science and Engineering. 2023 Oct 31;11(11):2085.
doi: 10.1016/j.eswa.2022.118270.
12. Alam MM, Spadon G, Etemad M, Torgo L, Milios E. Enhancing short-term vessel trajectory prediction with clustering for heterogeneous and multi-modal movement patterns. Ocean Engineering. 2024 Sep 15;308:118303.
doi: 10.1109/TITS.2023.3157812.
13. Raj N, Kumar P. A Novel and Efficient LR-LSTM AIS Route Data Prediction for Longer Range. Defence Science Journal. 2024 Jul 1;74(4):583-91
doi: 10.14429/dsj.74.19336

CONTRIBUTORS

Dr Nitish Raj obtained PhD from NIT Patna and working as a Scientist at DRDO, posted at the Weapons and Electronics Systems Engineer Establishment, Ministry of Defence in New Delhi. His research interests encompass: System design & development, systems integration, and Machine Learning.
He contributed to the current work by coming up with the idea and designing the experiment, optimising the deep learning techniques used in the experiment, creating the programme, analysing the data, and finalising the manuscript.

Dr Prabhat Kumar holds a PhD in Computer Science and working as a Professor in the Computer Science and Engineering Department at NIT Patna, India. His research focuses on Wireless sensor networks, internet of things, cyber security, data science, software engineering, and e-Governance.
He made contributions to the current study by assisting in the conceptualization of the review, helping in the identification and contributing to the analysis and synthesis of findings.