Defence Science Journal, Vol. 73, No. 4, July 2023, pp. 445-456, DOI : 10.14429/dsj.73.18626

Detection of Abnormal Vessel Behaviours Based on AIS Data Features Using HDBSCAN+

R. Hari Kumar[#], C.P. Ramanarayanan[#,*] and K.S.R.R.P. Murthy[$]

[#]*Defence Institute of Advanced Technology, Girinagar, Pune - 411 025, India*

[$]*Indian Navy, Delhi - 110 011, India*

[*]*E-mail*: *vcdiat@gmail.com*

ABSTRACT

Achieving maritime security is challenging due to the vastness and complexity of the domain. Monitoring all vessels that use this medium is humanly impossible but is needed for law enforcement. This paper proposes a machine learning solution based on HDBSCAN+ to classify the movements of vessels into 'normal' or 'abnormal'. This classification reduces the number of vessels that have to be monitored by law enforcement agencies to a manageable size. To date, AIS is the primary source of information that can represent vessel movements and enable the detection of maritime anomalies. The proposed model uses latitude, longitude, type of vessel, course and speed as features of the AIS data for analysis. The performance of the proposed model is validated against the marine incidents reported by Information Fusion Centre-Indian Ocean Region (IFC-IOR). The proposed model has successfully detected the incidents reported by IFC-IOR.

Keywords: Maritime anomaly; AIS; Machine learning; Maritime security; HDBSCAN+

NOMENCLATURE

AIS	:	Automatic identification system
COG	:	Course over ground
DBSCAN	:	Speed over ground
HDBSCAN	:	Hierarchical density-based spatial clustering of applications with noise
IFC-IOR	:	Information fusion centre-indian ocean region
MMSI	:	Maritime mobile service identity
SOG	:	Source over ground
SLOC	:	Sea lines of communication
VLCC	:	Very large crude carrier

1. INTRODUCTION

The maritime domain is extraordinarily complex and poses varied challenges due to its vastness, domain peculiarities, and the requirement of a considerable range and scale of resources to achieve domain awareness[1-9].

A constant search is always on for an alternate means (i.e., using lesser resources and modern technologies) to monitor and arrive at an actionable list of vessels suspected of behaving anomalously[10-18]. One such promising means is the application of machine learning and bigdata techniques[5,16] to the AIS data[10-11]. This paper proposes a model based on HDBSCAN+. Furthermore, the model's performance is validated against selected incidents of the real world, i.e., the model's results have been checked to verify if the model could detect the actual events that have occurred in the past[19].

The researchers have addressed anomaly detection broadly into two categories, namely, point-based models and trajectory-based models[10-11,20-34] Point-based models mainly depend on signatures and do not depend on historical data. These algorithms have been devised on rule-based approaches, which look for definite patterns like U-turns, sudden increases or decreases in speed, going through loops, etc[10]. The trajectory-based approaches are mainly data-driven, these approaches detect the abnormality based on the degree of deviation from the learned trajectory paths.

The advantages of rule-based methods are interpretability, and at the same time, it is complex to formalize the exhaustive list of abnormal behaviour of the ships, and it is also difficult to interpret categorical terms like fast, medium, slow, etc, for devising the algorithms. In learning-based approaches, the detection rules will be learnt from the data itself. Since the AIS data does not come with ground truth, the supervised learning algorithms cannot be used for detecting abnormalities[10,11.] Hence unsupervised learning algorithms have been widely adapted for anomaly detection[7,35,] Learning-based anomaly detection has been generally implemented in two stages, in the first stage, a normalcy model has been developed, which will be specific to a particular region of interest[35-40]. Density-based spatial clustering of applications with noise (DBSCAN) algorithm has been widely employed to extract the critical points of waypoints- where the vessels enter the region of interest[26]. The algorithms have been designed to build a graph

Received : 30 November 2022, Revised : 27 April 2023
Accepted : 18 May 2023, Online published : 11 July 2023

being the waypoints as nodes and the edges as the maritime routes. The normalcy model has been employed to fit each node using kernel density estimation[14], Gaussian mixture model[16], and multiple Ornstein-Uhlenbeck processes[27]. After forming the normalcy model, the abnormality is detected based on how likely a new AIS track is aligned with an order to label it as an anomaly. This is typically done by thresholding the distance from the route to the centroid point or by estimating the probability of AIS tracks given the normalcy model. Deep learning algorithms, which have become popular in achieving state-of-the-art in many domains, also have been widely used in maritime traffic analysis[24-25,39,41]. Since the AIS is a time series data, popular architectures like Recurrent Neural Networks, Long short-term memory and probabilistic neural networks have been widely used for path estimation and anomaly detection.

DBSCAN-based approaches have been made in the past by a few researchers[10,11,26]. The approach taken by this paper differs from the earlier papers as our approach takes geographic position (Latitude, Longitude), Type of Vessel (i.e., Tanker, Fishing Vessel, etc.), Course Over Ground (COG), Speed Over Ground (SOG), Moving/ Stopping state of vessels for analysis. Further, these features are applied sequentially, and the logic of arriving at COG-based and SOG-based clusters, normal and abnormal points, is different. This paper also proposes to optimise various parameters used in the proposed model, depending on the specific geographical area.

2. METHODOLOGY

The method followed for building the model and data analysis is elaborated in this section.

2.1 Geographic Area and the AIS Data Selected for the Study

The geographical area bounded by latitudes 04 degrees South to 26 degrees North and longitudes 57 degrees East to 110 degrees East was selected for the study. This area was selected only to limit the data to be processed, as considering the data of the entire world would require greater computing power. Historical AIS data from 01 Aug to 25 Aug 2021 was used for analysis.

2.2 Data Pre-processing & Structuring

From the AIS data collected, eight features were extracted for the study: Latitude, Longitude, SOG, Operating Day, Type of Vessel, MMSI, COG and Timestamp. Each AIS transmission is a data point (or a row in the data frame) used for analysis. The data was further divided into moving data (speed > 1 knot) and static data (speed < 1 knot) based on the vessel's speed (SOG). This was done to get to the places where the ships generally anchor. Further, analysing the course and speed of a static vessel makes little sense. Therefore, while moving vessels were analysed for abnormalities based on the position, type of ship, SOG and COG, the static vessels were analysed for the position and type of ship. The data points plotted for visualisation are shown in Fig. 1. For plotting the data, only the Latitudes and longitudes information is considered.

3. SUITABILITY OF HDBSCAN+ FOR THE PURPOSE

The suitability of using Machine Learning algorithms for detecting anomalies in maritime traffic patterns has been established by researchers in the past.[10-11,29,31,34-35] Due to the characteristics of vessel motion at sea, various patterns are formed. Vessels travelling at sea in a particular geographical area tend to follow a particular pattern. However, most vessels move on the same paths/ positions on the sea surface. This could be due to the requirement of safe navigation, reaching the destination through the shortest possible route and thus saving fuel and time, etc. These patterns are called Sea Lanes of Communication (SLOC) within a geographical area. SLOCs are the primary maritime routes between ports, mainly used

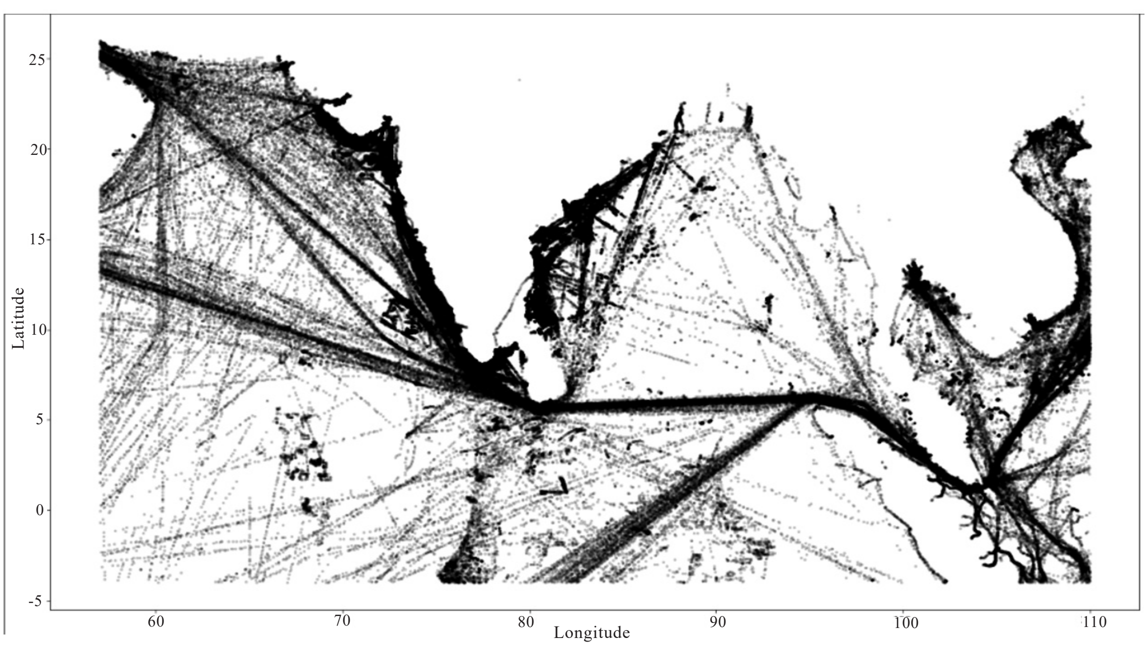

Figure 1. Visualisation of AIS data points (latitude & longitude) from 05 August to 09 August 2021.

for trade and logistics. Hence, AIS transmissions of vessels moving in SLOCs help us to figure out these maritime routes.

This paper's premise in classifying vessels' movements into normal and abnormal is that most vessels follow these set movement patterns/ SLOCs. Different types of vessels may follow slightly different patterns. A VLCC would follow a different path when compared to the path followed by a dhow or a fishing vessel. Hence, it has been assumed that a vessel with clear intentions would generally take the SLOCs for transiting between two ports. However, if the intentions are not transparent, a vessel may deviate from the set patterns. Hence, in this paper, 'Normal Behaviour' and 'Abnormal Behaviour' are defined as follows:

- **Normal Behaviour**: A vessel following the path for transit, which is being followed by the majority of the vessels of that type in a particular geographical area
- **Abnormal Behaviour**: A vessel deviating from the path for transit, which is being followed by the majority of the vessels of that type in a particular geographical area.

Hence, using a suitable algorithm to cluster the AIS transmitted positions (latitude and longitude) based on density would cluster these AIS positions and highlight the paths followed by most vessels. The points designated as the noise could be the points/ parts of a path/ paths, followed by vessels deviating from the paths for transit being taken by the majority. Hence**,** Normal points are those points that are part of a cluster, and abnormal points are points classified as noise.

Although the density-based methods can cluster the trajectory of any shape, the clustering effect of the classic density-based algorithm is poor for detecting clusters of data with different density distributions in the density space. However, HDBSCAN+[42] can overcome this to an extent.

4. WORKING OF HDBSCAN+ ALGORITHM

HDBSCAN+ is an extension of the original HDBSCAN algorithm that uses a hierarchical approach to clustering and density-based clustering techniques to identify the clusters[42]. HDBSCAN+ algorithm improvised over HDBSCAN in two features namely- Multi-metric clustering and Cluster splitting and merging. HDBSCAN+ allows the clustering of data points using multiple distance metrics and allows for more accurate clustering of complex data sets. The HDBSCAN+ algorithm is implemented as follows:

- Transform the space based on the sparsity or density of the feature samples
- The distance-weighted graph for the minimum spanning tree will be built
- Based on connected components cluster hierarchy is constructed
- Condense the cluster hierarchy based on minimum cluster size
- From the condensed tree, stable clusters are extracted

We need to set a few parameters for the HDBSCAN+ algorithm which can have a significant impact on the performance of the clustering, they are- minimum cluster size, minimum samples, cluster selection, leaf clustering and allowing of a single cluster. Minimum cluster size parameter will give a handle to decide the size of the smallest grouping that we wish to consider. Minimum number of samples parameter will provide a measure of how conservative we need the cluster to be. The larger the value of minimum samples, the more conservative the clustering and it will end up declaring more samples as noise and clusters are restricted to more dense areas. Leaf cluster parameter will determine how the algorithm chooses flat clusters from the cluster tree hierarchy. Allowing a single cluster option will enable to return a single cluster, and this parameter will help if we encounter a lot of small clusters, and these small clusters are seen as a part of a large cluster.

The mathematical inference about the selection of the parameters and working of HDBSCAN+ algorithm is as follows:

4.1 Mutual Reachability Distance

The core distance d_{core} is computed as the distance of an object to its minimum point nearest neighbour. For the two objects x_p, x_q, the mutual reachability distance is computed as

$$\max\{d_{core}(x_p), d_{core}(x_q), d(x_p, x_q)\} \tag{1}$$

Where $d(x_p, x_q)$ represents the normal distance computed based on the Euclidean metric. This approach helps in separating the sparse points from the dense points. Based on the clusters formed, a condensed cluster hierarchy is formed. The methodology is applied for different features like latitude, longitude and SOG, COG separately for forming the clusters.

4.2 Stability-Based Cluster Selection

After forming the condensed cluster, we need to select the leaf nodes. The leaf nodes can be computed based on the excess of mass, which gives the optimal global solution for the problem of finding the clusters with the highest stability. It is defined as follows:

$$S(C_i) = \sum_{x_j \in C_i} (\lambda_{\max}(x_j, C_i) - \lambda_{\min}(C_i))$$

$$= \sum_{x_j \in C_i} \left(\frac{1}{\varepsilon_{\min}(x_j, C_i)} - \frac{1}{\varepsilon_{\max}(C_i)}\right) \tag{2}$$

where, the λ is set to $\frac{1}{\varepsilon}$. The value of λ leaves increases as we move from roots towards leaves, where as the corresponding ε distance values becomes smaller. Subtracting $\lambda_{min}(C_i)$ represents the density level for cluster C_i first appears from the value beyond which object $x_j \in C_i$ no longer belong to C_i, resulting in a lifetime measure for x_j. The sum of all the object lifetimes within C_i leads to overall cluster life time $S(C_i)$, which is called stability. Based on this, we can formalize the sum of the cluster stabilities as an optimization problem which maximises the sum of these cluster stabilities as follows:

$$\max_{\delta_2,\ldots,\delta_k} J = \sum_{i=2}^{k} \delta_i S(C_i) \ \ subject\ to \begin{cases} \delta_i \in \{0,1\}, i = 2,\ldots,k \\ \sum_{j \in I_h} \delta_j = 1, \forall h \in L \end{cases} \tag{3}$$

With L={h|C_h as leaf cluster} as leaves, I_h are the clusters sets on the paths from leaves to excluded root, and δ_i as Boolean expression which tells the whether the respective cluster is selected or not. The same clustering algorithm is applied

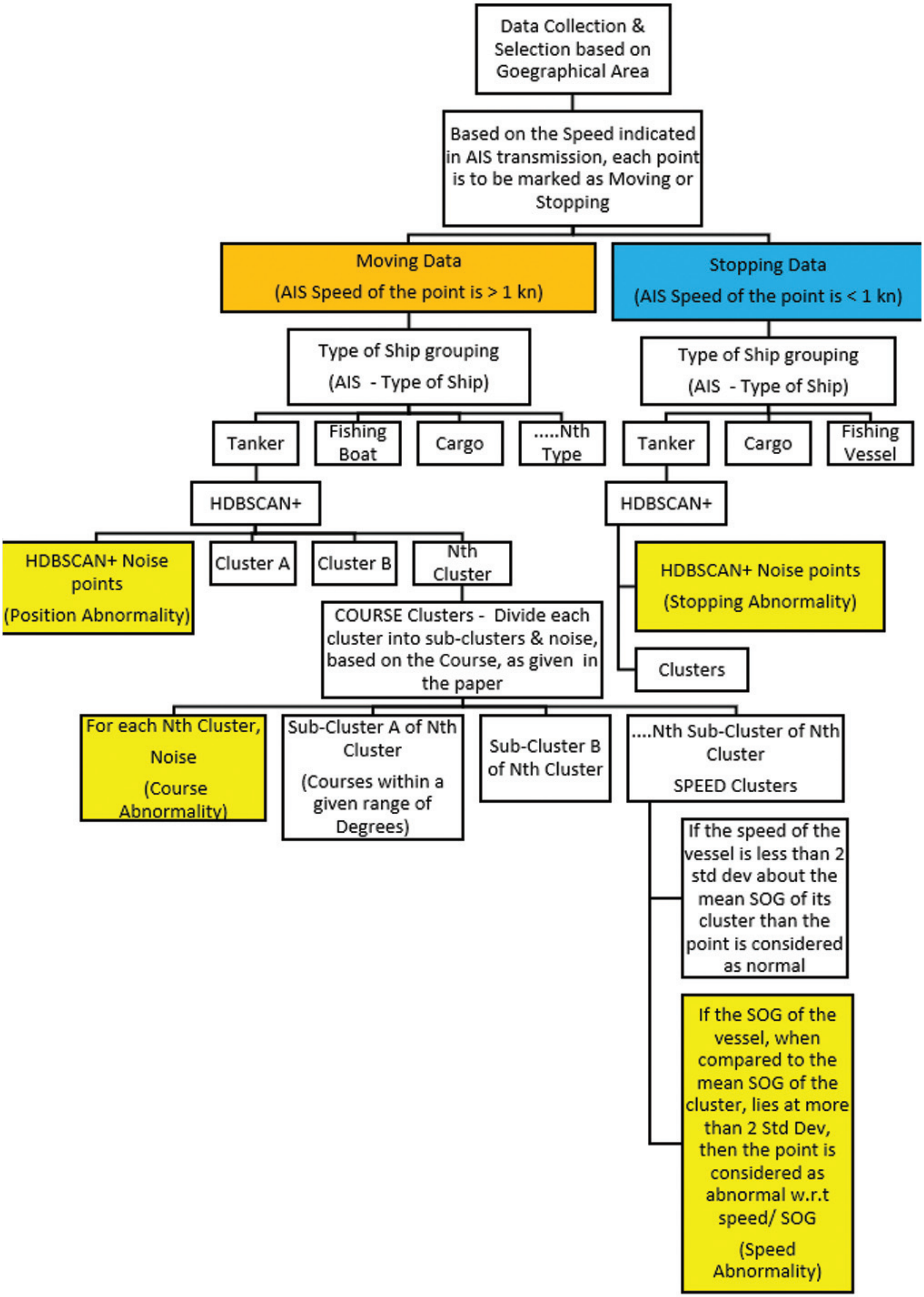

Figure 2. Process for anomaly detection model.

hierarchically to all the features such as latitude and longitude, COG and SOG to form the normal clusters. The mean of the individual clusters is computed, and if the data sample lies within the two standard deviations then it is considered a normal sample, otherwise, it is labelled as an anomaly sample.

5. CLASSIFYING DATA INTO NORMAL AND ABNORMAL POINTS

The selected data were divided into two groups, i.e., 'moving' and 'static'. The moving data was then divided into groups based on the 'AIS type of vessel'. The data of each of these groups was then separately clustered using HDBSCAN+ (based on the latitude and longitude). These clusters were later divided into sub-clusters based on the COG of the vessels inside that cluster. The said sub-clusters were then again divided into sub-sub-clusters based on the SOG of the vessels in the sub-clusters.

It is highlighted that the HDBSCAN+ clustering, clustering into sub-clusters based on COG and sub-sub-clusters based on SOG, is done to label the data as normal and abnormal. The proposed model is an unsupervised machine learning approach, where the normal and abnormal are decided by the model based on the data fed for analysis. The model assigns any data point that falls into any cluster/ sub-cluster/ sub-sub-cluster a positive number, and all the abnormal points that do not fall into any cluster are assigned '-1' as the cluster number. Hence, if a data point has been assigned a positive cluster number by HDBSCAN+ analysis, it means that the point falls in a cluster and hence is a normal point with respect to the position (i.e., latitude and longitude). Similarly, if the same point is assigned a positive number for the 'course sub-cluster, it means that the point is normal w.r.t course, and if it is assigned '-1' as the label, then that point is abnormal w.r.t that analysis. The following flow chart depicts the process undertaken to arrive at various clusters and, finally, the classification of the data into normal and abnormal points as shown in Fig. 2.

The implementation of the HDBSCAN+ algorithm on each group of data, based on the type of vessel, produced various clusters, which can be called the HDBSCAN+ clusters. After this, each cluster was again divided based on the COG of the vessels in that particular cluster. Vessels belonging to a particular HDBSCAN+ cluster with COGs falling within a given range of degrees from each other are clustered into one sub-cluster. We call these sub-clusters HDBSCAN+ Course clusters. Hence, for every HDBSCAN+ cluster, we will have many HDBSCAN+ Course clusters based on the COG of the vessels of that cluster. Data points whose COG does not fall within the ten-degree range of at least five vessels are considered abnormal.

To classify the vessel's behaviour based on speed, each of these 'HDBSCAN+ Course clusters' is again divided into normal and abnormal based on the SOG/ Speed of the vessels in that cluster. The average/ mean and standard deviation w.r.t SOG for each cluster is calculated separately, and the vessel positions of each cluster whose SOG is beyond two standard deviations are marked as abnormal.

6. ESTIMATING PARAMETERS FOR FINETUNING THE MODEL AND OBTAINING OPTIMAL RESULTS

The proposed model of determining normal and abnormal positions requires estimating a few parameters. Like the epsilon, the minimum number of samples, the minimum cluster size for HDBSCAN+, a COG range for the HDBSCAN+ Course

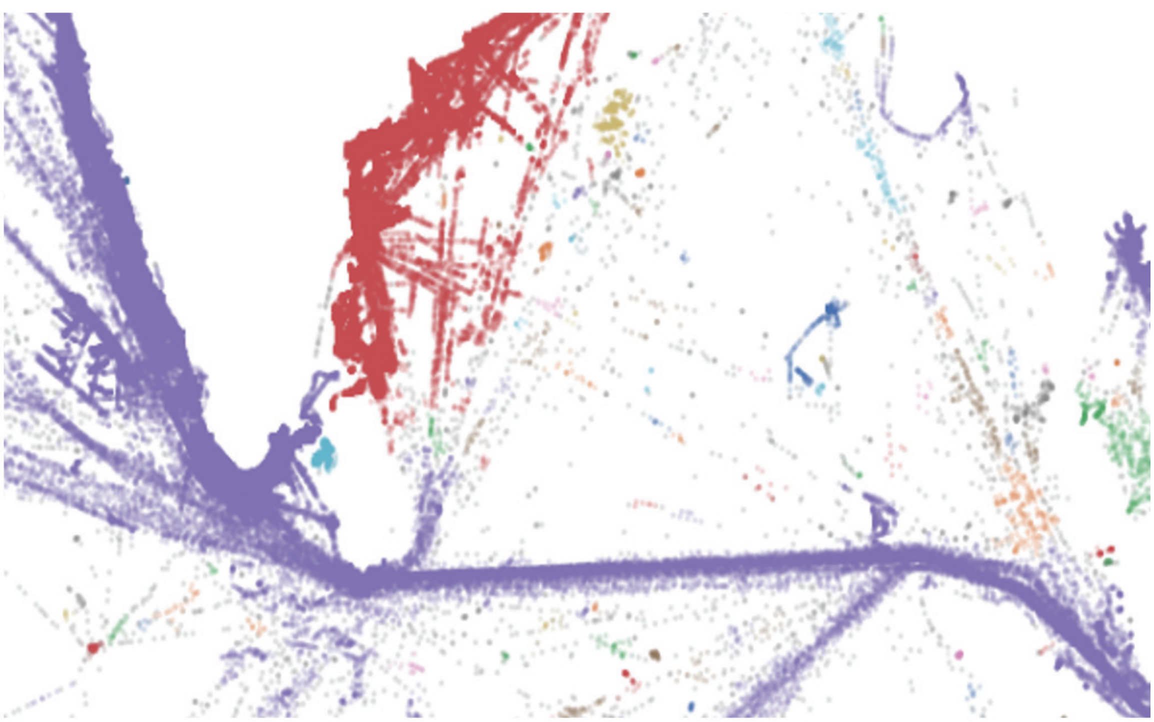

Figure 3. Visualisation of the HDBSCAN+ Clusters, based on AIS Position of Vessels (After grouping the data based on type of vessel).

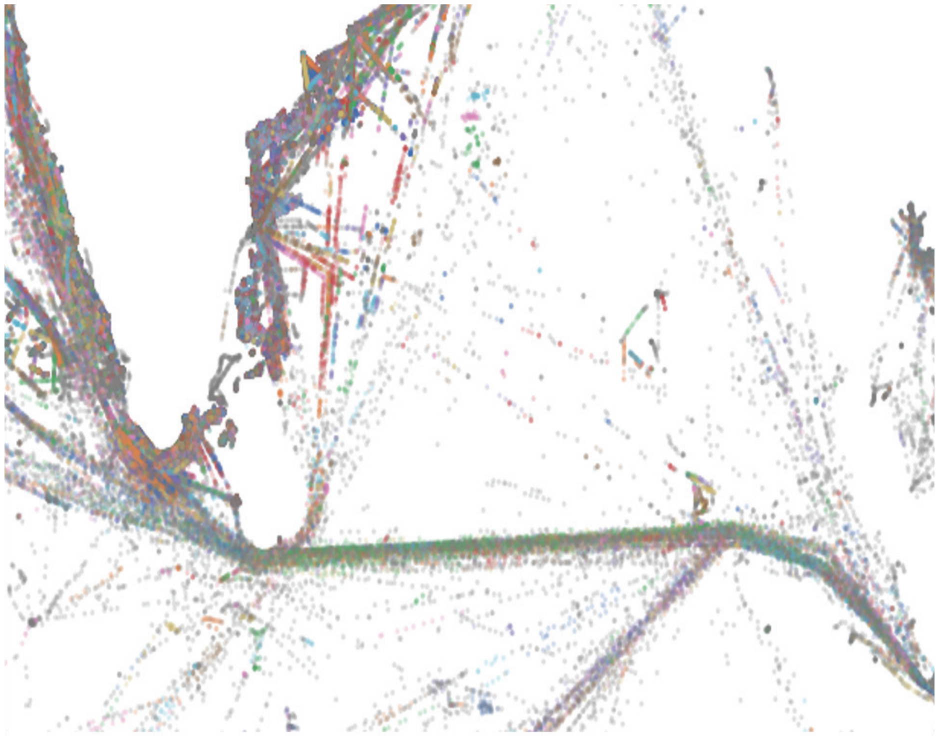

Figure 4. Visualisation of the 'Normal' and 'Abnormal' Points after Classifying them with HDBSCAN + based on the 'Type of vessel', AIS position, COG and SOG (The points shown in black are noise or abnormal, and the points shown in other colours are normal points).

clusters and the number of standard deviations to declare a position normal when dividing the data based on SOG. These parameters must be optimised depending on the data available and the geographic location to which it pertains. The process of optimisation is essential to finetune the model as shown in Fig. 3 and Fig. 4.

The paper looks at detecting anomalies in each position instead of combining the positions into tracks and then trying to detect the anomaly in the tracks. This approach has been adopted as our interest is in detecting the anomaly even in a single vessel position (i.e., the smallest part of a track).

7. RESULTS

Results from the proposed model, provide a data frame that holds the following details:

- All the features of the initial AIS data used for analysis.
- Position Abnormality of the Data, i.e., Normal and Abnormal labels for each point w.r.t HDBSCAN+ based clustering of geographical positions among a particular type of vessels.
- Course Abnormality, i.e., Normal and Abnormal labels for each point w.r.t COG.
- Speed Abnormality, i.e., Normal and Abnormal labels for each point w.r.t SOG.
- Stopping Abnormal, i.e., Normal and Abnormal labels for each point w.r.t its stopping position.

7.1 A Summary of Abnormalities Detected by the Proposed Model

The abnormalities detected by the proposed model have been summarised based on the 'type of vessel' and 'type of abnormality. A few crucial details are given in Table 1.

7.2 Number of AIS transmitted Positions Vs Number of Vessels

The proposed model attempts to detect abnormality for every AIS position used in the analysis. However, the path taken by one vessel to transit between two ports will consist of many AIS positions. Hence, while the number of positions analysed is huge, these positions belong to a much lesser number of vessels.

7.3 Ground Truth

Many researchers have brought out that the lack of ground truth is a major challenge in this area of research[10,11]. Many researchers struggle to find a valid ground truth when using real-world data and resort to the artificial generation of AIS anomalies[4]. As per a review study, out of the 32 papers studied, there were only eight publications with a decent ground truth derived, e.g., from reported incidents in Europe[22] or suspected illegal fishing rendezvous, tracks labelled by domain experts, or situations with severe weather conditions, there are also 11 publications with makeshift ground truth. These include drawing anomalies by hand, introducing random data, or adding synthetic anomalies, among others.

7.4 Ground Truth for the Paper

In this paper, the ground truth has been obtained from the real incidents reported by IFC-IOR. However, it is essential to note that the existing means and resources cannot detect most of the anomalies. Hence, what is detected is a fraction of the actual occurrences, and what is reported is a fraction of what is

Table 1. A summary of the results obtained from the proposed model

Type of vessel	Total number of positions (latitude, longitude) for the vessel type -- (a)	Number of positions - marked abnormal based only on latitude & longitude -- (b)	Position abnormality % (b/a *100)	Number of positions - marked abnormal based on position latitude & longitude, cog -- (c)	Course abnormality % (c/a*100)	Number of positions - marked abnormal based on latitude & longitude, cog, sog -- (d)	Speed abnormality % (d/a*100)
Fishing vessel	14,36,203	4,411	0	16,951	1	61,969	4
Undefined	8,63,538	9,597	1	25,715	3	53,309	6
Cargo ship	6,45,277	58,586	9	72,727	11	98,391	15
Towing vessel	2,10,219	7,242	3	36,836	18	40,878	19
Tanker	3,78,093	46,786	12	62,495	17	82,698	22
Ship according to rr resolution no. 18	99,164	6,256	6	30,443	31	33,896	34
Vessel engaged in dredging or underwater operations	25,232	254	1	9,326	37	9,644	38
Passenger ship 69	3,962	275	7	1,512	38	1,646	42
Tanker category D	70,898	17,747	25	32,324	46	33,872	48
Other vessel	5,685	227	4	3,643	64	3,703	65
Loran C	4,675	29	1	2,999	64	3,120	67
Passenger ship	27,008	8,161	30	20,771	77	21,167	78
Cargo ship 79	1,348	521	39	1,165	86	1,169	87
Tanker category A	318	28	9	281	88	281	88
Tug	14,140	196	1	12,434	88	12,502	88
Tanker category B	812	50	6	754	93	754	93
Tanker 89	1,466	107	7	1,428	97	1,428	97
Local vessel type 56	3,947	687	17	3,917	99	3,917	99
Pilot vessel	6,174	25	0	6,174	100	6,174	100
Hsc	1,705	44	3	1,705	100	1,705	100
Pleasure craft	1,586	243	15	1,586	100	1,586	100
Vessel engaged in military operations	2,848	146	5	2,848	100	2,848	100
Other vessel 95	189	143	76	189	100	189	100
Search and rescue vessel	707	93	13	707	100	707	100
Passenger ship category A	329	0	0	329	100	329	100
Sailing vessel	918	5	1	918	100	918	100
Law enforcement vessel	176	106	60	176	100	176	100
Tanker 85	387	0	0	387	100	387	100
Cargo ship category A	666	130	20	666	100	666	100
Hsc category C	434	0	0	434	100	434	100

detected. Thus, it has been seen that the anomalies detected by the model are large when compared to the occurrences reported. However, the model has detected all the reported anomalies in all those cases where the data w.r.t to the time and position of occurrence of the anomaly were available for analysis.

7.5 Efficacy of the Proposed Model

The proposed model's efficacy can be established by verifying if the model can detect either the Position abnormality, COG-based abnormality or SOG-based abnormality in the AIS positions that correspond to the time/ date and geographical area of occurrence of the past incidents. For this purpose, historical AIS data from 01 August 21 to 25 August 21 was selected, and this data was analysed using the model developed. The details of the maritime incidents that have been reported (which is only a fraction of the actual occurrences) by IFC-IOR for the period 01 August 21 to 25 August 21 have been used for the analysis as ground truth[19]. The results from the model were analysed to check if they had indicated any abnormality at the location and the date, corresponding to the incidents reported by IFC-IOR/ ground truth.

The results of the model, w.r.t the vessels involved have been analysed around the position and time of occurrence of the incident. A square of approximately two degrees by two degrees (i.e., approx. 14,400 sq nm) was made with the location of the incident at the centre of this square. After the initial analysis, the square was expanded as required to analyse more data. The model had indicated at least one kind of abnormality, i.e., the position abnormality, the COG abnormality, the SOG abnormality, or the stopping abnormality for all the incidents analysed.

Five cases have been analysed in detail. The succeeding paragraphs elaborate on the analysis undertaken and bring out the inference that can be gained from the proposed model.

7.5.1 Case 1: Collision of 'Green Pacific' with a Fishing Boat.

7.5.1.1 Incident Reported

On 14 August 2021, a Vietnam-flagged container ship named 'Green Pacific' collided a fishing vessel 'TTH 92206TS', at 30 nm Northeast of Con Co Island, Quang Tri province, Gulf of Tonkin, Vietnam. It is reported that after collision the fishing vessel sank. Then the container ship Green Pacific rescued seven fishermen and reported that other two fisher men were missing.

7.5.1.2 Detection by the Model

The MMSI number of the vessel Green Pacific is 574003640. The model could detect various abnormalities w.r.t this vessel. The results from the model showing abnormality are given at Table 2.

7.5.1.3 Inference From the Model

For 14 Aug 21, the database used for analysis has five AIS transmissions of the said vessel. As per the results of the model, the vessel was moving in a normal manner till 07:00

Table 2. Abnormalities detected by the model, w.r.t the vessel Green Pacific.

Lat	Lng	SOG (kn)	Date	COG (deg)	Time	Pos abnorm (Cluster number)	COG abnorm (Cluster number)	SOG abnorm (Cluster number)
19.95632	107.08	15	14-08-2021	166	00:00:00	108	5254	86872
19.37502	107.25	15	14-08-2021	164	00:54:38	108	5255	86878
17.19963	107.6	15	14-08-2021	232	17:00:03	-1	-1	-1
17.12667	107.5	15	14-08-2021	240	17:48:46	-1	-1	-1
17.00786	107.38	12	14-08-2021	112	20:44:17	-1	-1	-1

Table 3. Abnormalities detected by the said model w.r.t Navios Amaryllis

Lat	Lng	SOG (kn)	Date	COG (deg)	Time	Pos abnorm (Cluster number)	COG abnorm (Cluster number)	SOG abnorm (Cluster number)
5.67929	75.49787	11	18-08-2021	248	00:48:49	-1	-1	-1
5.38249	74.7261	12	18-08-2021	250	05:03:38	46	-1	-1
5.4513	74.90289	12	18-08-2021	248	04:05:01	46	-1	-1
5.50404	75.03991	11	18-08-2021	248	03:22:00	46	-1	-1
5.08984	73.96315	12	18-08-2021	250	09:13:27	-1	-1	-1
5.01139	73.75581	12	18-08-2021	249	10:19:14	-1	-1	-1
4.40485	73.34965	0	19-08-2021	0	08:05.6	50	NA	NA
4.40493	73.34962	0	19-08-2021	204	24:46.2	50	NA	NA
4.40486	73.34964	0	19-08-2021	0	58:33.5	50	NA	NA

h on 14 Aug 21. However, the AIS transmissions at 17:00h, 17:48h and 20:44h show abnormality. The vessel has position, COG, and SOG abnormalities. As per the proposed model, the vessel has moved to an abnormal position when compared to the movement pattern of vessels of its type.

7.5.2 Case 2: Grounding of 'Navios Amaryllis'

7.5.2.1 Incident Reported

On 19 August 2021, it was reported that a Panama-flagged carrier named 'Navios Amaryllis' with nineteen crew members onboard ran aground off Kaafu Atoll Reefs, Maldives. The event happened due to engine failure. No environmental damage or injuries were reported in the incident. The tug from Sri Lanka arrived on 23 Aug to assist in refloating the vessel.

7.5.2.2 Detection by the Model

The MMSI number of the vessel Navios Amaryllis is 370317000. The model could detect various abnormalities. The results from the model showing abnormality are given at Table 3.

7.5.2.3 Inference From the Model

The said event of grounding was reported on 19 Aug 21. However, the results of the model indicate that the vessel was moving abnormally from 00:48:49 on 18 Aug 21. It was moving in a geographical path/ position that is not normal for ships of that 'particular type' to move. The vessel then ran aground on 19 Aug 21. The stopping data from the corresponding AIS transmission at 12:08 PM indicates that the vessel has stopped/ run aground at the position indicated. However, vessels generally anchor around this position. Hence, the stopping area is not abnormal.

7.5.3 Case 3: Mechanical Failure Onboard a Panama-Flagged Ship Tan Binh 127

7.5.3.1 Incident Reported

On 01 August 2021, it was reported that a Panama-flagged ship Tan Binh one hundred and twenty-seven crew onboard with eighteen crew onboard encountered mechanical failure about 51 nm off Kyunsu, Myanmar. The vessel was experienced flooding. Later, a Hong Kong-flagged container ship Chittagong rescued all the crew members and vessel was towed to shore (Incident Location is Lat -11.83N, Lon -98.52E)

7.5.3.2 Detection by the Model

The MMSI number of vessel Tan Binh 127 is 354984000. The model could detect various abnormalities. The results from the model showing abnormality are tabulated at Table 4.

7.5.3.3 Inference From the Model

On 01 August 2021, at about 15:13h, the vessel moved abnormally and stopped in a position (10.05417 N, 96.817 E) where generally, vessels don't stop. Hence, a stopping position abnormality was detected. However, the model could detect the abnormality of this vessel's movement from 0845h onwards. It can be seen from the Table 4, that this vessel was moving with an abnormal course and speed before stopping. (In the Table 4, '-1' indicates noise or abnormality)

7.5.4 Case 4: Sinking of a Fishing Vessel off the Coast of Thailand

7.5.4.1 Incident Reported

On 02 August 2021, it was reported a fishing vessel carrying thirty-one crew members that including twenty fishermen onboard, sank about 70 nm NW of Koh Surin, Thailand. The fishing vessel encountered adverse weather prevailing and eventually sank. Entire crew members went

Table 4. Relevant results from the model w.r.t vessel Tan Binh

Lat	Lon	SOG (kn)	Date	MMSI	COG (deg)	Time	Pos Abnorm (Cluster number)	COG Abnorm (Cluster number)	SOG Abnorm (Cluster number)
9.6410	96.570	18	01-08-2021	354984000	343	08:45.4	285	-1	-1
9.4986	96.603	18	01-08-2021	354984000	346	12:55.5	285	-1	-1
10.054	96.817	1	01-08-2021	354984000	69	15:13.0	285	-1	-1

Table 5. Relevant results from the model w.r.t various vessels that were operating in the said area

Lat	Lon	SOG (kn)	Date	AIS type	MMSI	COG (deg)	Time	Pos Abnorm (Cluster number)	COG Abnorm (Cluster number)	SOG Abnorm (Cluster number)
9.40862	97.88543	1	02-08-21	Undefined	111111111	260.9	07:57.9	4	-1	-1
9.72609	98.43486	1	02-08-21	Undefined	111111111	314.5	04:17.5	4	-1	-1
9.72611	98.4348	1	02-08-21	Undefined	111111111	315	21:23.9	4	-1	-1
9.72515	98.4354	2	02-08-21	Undefined	111111111	339.1	01:45.3	4	-1	-1
9.72558	98.43543	1	02-08-21	Undefined	111111111	179	08:28.5	4	-1	-1

missing. (LAT -09.46N, LON - 97.79E)

7.5.4.2 Detection by the Model

The IFC -IOR report has not given the vessel's name or MMSI number. Hence, we have to correlate the vessels present in the area. We have an undefined vessel with MMSI 111111111 in the results produced by the model. This vessel was detected as abnormal by the model. This could have been the vessel. The results from the model are enumerated at Table 5.

7.5.4.3 Inference From the Model

The model had detected abnormal behaviour in a few other vessels too, but as they had MMSI numbers and names, they were not considered. The model could detect various abnormalities in the said area even on the previous day by undefined Vessels. The model detected the abnormal course and speed undertaken by undefined/ suspected vessels. Further, it has also been pointed out that this vessel has stopped at an abnormal position.

7.5.5 Case 5: Seizure of Contraband by Sri Lankan Navy

7.5.5.1 Incident Reported

On 07 August 2021, the Sri Lankan Navy arrested five suspects who are carrying 5372 kilogram of turmeric from a boat off Chilaw, Sri Lanka. The contraband was packed inside 143 sacks and was believed to be smuggled from a foreign country via sea route and the seizure was taken into custody.

7.5.5.2 Detection by the Model

We have several vessels showing abnormality in the area on that particular day. The IFC -IOR report has not given the vessel's name or MMSI number. Hence, we have to correlate the vessels present in the area.

7.5.5.3 Inference From the Model

An interesting part to note here is that the model has the capability to earmark certain vessels that are to be monitored based on their abnormal behaviour. In this incident, the area selected has about 1842 data points. Out of these 1842 data points, the model has labelled only 139 points as abnormal. Further, these 139 points belong to about 30 vessels (unique MMSI numbers); out of these 30 vessels, 09 vessels have changed their AIS-based data about the 'Type of Ship' feature from 'fishing vessel' to 'undefined' or vice versa while in the area. Hence, it is likely that one of these 09 vessels could be the suspected vessel.

8. CONCLUSION

Maritime security is essential for any nation. Close monitoring of the vessels in the maritime domain is a prerequisite for ensuring maritime security. However, the number of physical assets required to monitor the domain could be extremely large. Hence, there is a need to explore alternate measures like AI/ Machine learning for monitoring the vessels in the domain. The paper explores one such means by using HDBSCAN+ based machine learning model and reducing the number of vessels that are to be monitored by physical assets to a manageable number.

Historical AIS data pertaining to the geographical area bounded by 4° South to 26° North and 57° East to 110° East for a few days in the month 21 Aug was analysed. The model for detecting abnormal vessel behaviour/ anomalies has been built using Python. The abnormality is detected based on the various features of AIS data, i.e., Latitude, Longitude, SOG, COG and Type of ship. The results of the model have been validated against the actual incidents that have occurred during the same period. The model could detect and indicate the abnormal behaviour of various vessels involved/ present in the vicinity of the incident.

REFERENCES

1. Lützhöft, M.; Grech, M. & Porathe, T. Information environment, fatigue, and culture in the maritime domain. *Rev. Human Factors and Ergonomics*, 2011, **7**(1), 280-322.
   doi:1177/1557234X11410391.
2. Zhang, Y.L.; Peng, J.; Liu, J.S. & Liu, S.K. AIS data oriented ships- trajectory mining and forecasting based on trajectory delimiter, *In* Proceedings of 10th International Conference on Intelligent Human-Machine Systems and Cybernetics, Hangzhou, China, 2018, 269-273.
   doi: 10.1109/IHMSC.2018.00069.
3. Tang, C.; Chen, M.; Zhao, J.; Liu, T.; Liu, K.; Yan, H. & Xiao, Y. A novel ship trajectory clustering method for finding overall and local features of ship trajectories. *Ocean Engineering*, 2021, **24**(1), 110-108.
   doi:10.1016/j.oceaneng.2021.110108.
4. Kazemi, S.; Abghari, S.; Lavesson, N.; Johnson, H. & Ryman, P. Open data for anomaly detection in maritime surveillance. *Expert Syst. Appl.*, 2013, **40**(14), 5719-29.
   doi: 10.1016/j.eswa.2013.04.029
5. Liu, C.; Musharraf, M.; Li, F. & Kujala, P. A data mining method for automatic identification and analysis of icebreaker assistance operation in ice-covered waters. *Ocean Engineering*, 2022, **266**(2), 112914.
   doi:10.1016/j.oceaneng.2022.112914
6. Song, L.; Wang, R.; Xiao, D.; Han, X.; Cai, Y. & Shi, C. Anomalous trajectory detection using recurrent neural network. *In* Proceedings of 14th International Conference on Advanced Data Mining and Applications, Springer International Publishing, Nanjing, China, 2018, 263–277.
   doi: 10.1007/978-3-030-05090-0_23
7. Arguedas, V.F.; Pallotta, G. & Vespe, M. Maritime traffic networks: From historical positioning data to unsupervised maritime traffic monitoring. *IEEE Transact. Intelligent Transport. Syst.,* 2017, **19**(3),722-732.
   doi: 10.1109/TITS.2017.2699635
8. Mazzarella, F.; Vespe, M.; Alessandrini, A.; Tarchi, D.; Aulicino, G. & Vollero, A. A novel anomaly detection approach to identify intentional AIS on-off switching. *Expert Syst. Appl.*, 2017, **78**, 110-123.
9. Zhang, Y. & Weifeng, Li. Dynamic maritime traffic pattern recognition with online cleaning, compression, partition,

and clustering of AIS Data. *Sensors*, 2022, **22**(16), 6307. doi:10.3390/s22166307.

10. Tu, E.; Zhang, G.; Rachmawati, L.; Rajabally, E. & Huang, G.B. Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology. *IEEE Transact. Intelligent Transport.Syst.,* 2017, **19**(5),1559-82. doi: 10.1109/TITS.2017.2724551.
11. Riveiro, M.; Pallotta, G. & Vespe, M. Maritime anomaly detection: A review. *Wiley Interdisciplinary Rev.: Data Mining and Knowl. Discovery*, 2018, **8**(5), 1-19. doi:10.1002/widm.1266
12. Pallotta, G.; Vespe, M. & Bryan, K. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy*, 2013, **15**(6), 2218-45. doi: 10.3390/e15062218
13. Liu, B.; Souza, E. & Hilliard, C. Ship movement anomaly detection using specialised distance measures. *In* Proceedings of 18th International Conference on Information Fusion, Washington, DC, USA, 2015, 1113-20.
14. Laxhammar, R. Anomaly detection for sea surveillance. *In* proceedings of 11th international conference on information fusion, IEEE, Cologne, Germany, 2013, 1-8.
15. Wolsing, K.; Roepert, L.; Bauer, J. & Wehrle, K. Anomaly detection in maritime AIS Tracks: A review of recent approaches. *J. Marine Sci. Eng.*, 2022, **10**(1), 112. doi:10.3390/jmse10010112.
16. Zor, C. & Kittler, J. Maritime anomaly detection in ferry tracks. *In* Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, USA, 2017, 2647-2651. doi:10.1109/ICASSP.2017.7952636
17. Yang, D.; Lingxiao, W.; Shuaian, W.; Haiying, J. & Kevin X.L. How big data enriches maritime research–a critical review of Automatic Identification System (AIS) data applications*. Transport Rev.,* 2019*,* **39**(6), 755-773. doi: 10.1080/01441647.2019.1649315
18. Sidibé, A. & Gao, S. Study of automatic anomalous behaviour detection techniques for maritime vessels. *J. Navigation*, 2017, **70**(4), 847–858. doi:10.1017/S0373463317000066.
19. Soumyajit, M. Information fusion centre-indian ocean region, Indian Navy Annual Report 21, New Delhi, 2021. 40-47
20. Hari Kumar, R.; Ramanarayanan, C.P. & Bharath, R. Machine learning-based automated process for predicting the anomaly in AIS Data. *In* Proceedings of International Conference on Data Management, Analytics and Innovation, Springer, Virtual, India, 2022, 303-314. doi: 10.1007/978-981-19-2600-6_21
21. Yan, Z.; Cheng, L.; He, R. & Yang, H. Extracting ship stopping information from AIS data. *Ocean Eng.*, 2022, **250**, 111004 doi:10.1016/j.oceaneng.2022.111004.
22. Liu, B. Maritime traffic anomaly detection from AIS satellite data in near port regions. Dalhousie University, Nova Scotia, Canada, 2015, PhD Thesis.
23. Ma, C.; Miao, Z.; Li, M.; Song, S. & Yang, M.H. Detecting anomalous trajectories via recurrent neural networks. *In* proceedings of 14th Asian Conference on Computer Vision, Springer, Perth, Australia*,* 2018, 370-82. doi: 10.1007/978-3-030-20870-7_23
24. Bouritsas, G.; Daveas, S.; Danelakis, A. & Thomopoulos, S.C. Automated real-time anomaly detection in human trajectories using sequence-to-sequence networks. *In* Proceedings of International Conference on Advanced Video and Signal Based Surveillance IEEE, Taipei, Taiwan, 2019, 1-8. doi: 10.1109/AVSS.2019.8909844
25. Nguyen, D.; Vadaine, R.; Hajduch, G.; Garello, R. & Fablet, R. GeoTrackNet: A maritime anomaly detector using probabilistic neural network representation of AIS tracks and a contrario detection, *IEEE Transact. Intelligent Transport. Syst.*, 2021, **23**(6), 5655-67. doi: 10.1109/TITS.2021.3055614
26. Varlamis, I.; Tserpes, K.; Etemad, M.; Junior, A.S. & Matwin, S.A. Network abstraction of multi-vessel trajectory data for detecting anomalies. *In* proceedings of the EDBT/ICDT Workshops*,* Lisbon, Portugal*,* 2019.
27. Coscia, P.; Braca, P.; Millefiori, L.M; Palmieri, F.A. & Willett, P. Multiple Ornstein–Uhlenbeck processes for maritime traffic graph representation. *IEEE Transact. Aerospace Electron. Syst.,* 2018*,* **54**(5), 2158-70. doi: 10.1109/TAES.2018.2808098
28. Forti, N.; Millefiori, L.M; Braca, P. & Willett, P. Anomaly detection and tracking based on mean–reverting processes with unknown parameters. *In* International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Brighton, UK, 2019, 8449-53. doi: 10.1109/ICASSP.2019.8683428
29. Zhao, L. & Shi, G. Maritime anomaly detection using density-based clustering and RNN. *J. Navigation*, 2019, **72**(4), 894 – 916. doi: 10.1017/S0373463319000031
30. McInnes, L. & Healy, J. Accelerated hierarchical density-based clustering. *In* Proceedings of IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, USA, 2017, 33-42. doi: 10.1109/ICDMW.2017.12.
31. Zissis, D.; Chatzikokolakis, K.; Spiliopoulos, G. & Vodas, M. A distributed spatial method for modeling maritime routes. *IEEE Access*, 2020, **8**, 47556-68. doi 10.1109/ACCESS.2020.2979612
32. d'Afflisio, E.; Braca, P.; Millefiori, L.M. & Willett, P. Maritime anomaly detection based on mean-reverting stochastic processes applied to a real-world scenario. *In* Proceedings of 21st International Conference on Information Fusion, IEEE, Cambridge, UK, 2018. doi: 10.23919/ICIF.2018.8455854
33. McAbee, A.; Scrofani, J.;Tummala, M.; Garren, D. & McEachen, J. Traffic pattern detection using the Hough transformation for anomaly detection to improve maritime domain awareness. *In* Proceedings of 17th International Conference on Information Fusion, Salamanca, Spain, 2017, 1-6.

34. Mascaro, S.; Kevin, B.K. & Ann, E.N. Learning abnormal vessel behaviour from ais data with bayesian networks at two-time scales. *Tracks a J. Artists Writings*, 2010, 1-34.
35. Vespe, M.; Ingrid, V.; Karna, B. & Paolo, B. Unsupervised learning of maritime traffic patterns for anomaly detection, *In* Proceedings of 9th IET Data Fusion & Target Tracking Conference (DF&TT): Algorithms & Applications, London, UK, 2012, 14-19. doi: 10.1049/cp.2012.0414.
36. Fang, Z.; Hongchu, Y.; Ranxuan, K.; Shih-Lung, S. & Guojun, P. Automatic identification system-based approach for assessing the near-miss collision risk dynamics of ships in ports. *IEEE Transact. Intelligent Transport. Syst,* 2018, **20**(2), 534-43. doi: 10.1109/TITS.2018.2816122
37. Xianbin, L.; Kai, W.; Min, T.; Jiangyi, Q.; Peng, W.; Tingting, Y. & Haichao, Z. Marine drifting trajectory prediction based on LSTM-DNN. *Wireless Commun. Mobile Comput.*, 2022, **2022**, 7099494. doi:10.1155/2022/7099494
38. Singh, S.K. & Frank, H. Machine learning-assisted anomaly detection in maritime navigation using AIS data. *In* proceedings of IEEE/ION Position, Location and Navigation Symposium PLANS, Portland, OR, USA, 2020, 832–38. doi: 10.1109/PLANS46316.2020.9109806.
39. Balduzzi, M.; Alessandro, P. & Kyle, W. A security evaluation of AIS automated identification system. *In* Proceedings of the 30th annual computer security applications conference, New Orleans, USA, 2014, 436–45. doi: 10.1145/2664243.2664257
40. Gao, M.; Shi, G. & Li, S. Online prediction of ship behaviour with automatic identification system sensor data using bidirectional long short-term memory recurrent neural network, *Sensors*, 2018, **18**(12), 4211. doi: 10.3390/s18124211.
41. Ford, J.H.; Peel, D.; Kroodsma, D.; Hardesty, B.D; Rosebrock, U. & Wilcox, C. Detecting suspicious activities at sea based on anomalies in Automatic Identification Systems transmissions. *PLoS One,* 2018, **13**(8), e020164 0. doi: 10.1371/journal.pone.0201640.
42. Malzer, C. & Marcus, B. A hybrid approach to hierarchical density-based cluster selection. *In* IEEE international conference on multisensor fusion and integration for intelligent systems (MFI), Karlsruhe, Germany, 2020. doi: 10.1109/MFI49285.2020.9235263

CONTRIBUTORS

Mr R. Hari Kumar is an alumnus of the National Defence Academy, Pune. He was commissioned into the Indian Navy on 01 Jan 1983.
He contributed towards the conceptualisation work on abnormality detection using AIS-speed, AIS-course and the geographic area of operations of a vessel. He also contributed towards checking the HDBSCAN+ algorithms and has provided overall guidance to the entire code and simulations, and to overall text refinement and modifications.

Dr C.P. Ramanarayanan obtained his Ph.D from JNTU Hyderabad, and working as a Vice chancellor for DIAT, Pune.
He contributed towards the conceptualisation work on abnormality detection using the AIS-Type of Ship and has validated the analytical methods and the model prepared for detecting anomalies. Further, he has contributed towards the validation of the analytical results and the parameter estimation results.

Mr K.S.R.R.P. Murthy is an alumnus of the National Defence Academy, Khadakwasla. He was commissioned into the Indian Navy in 2004. The Officer holds MBA in Logistics Management and MSc. in Defence & Strategic Studies.
He developed the framework and the model used for anomaly detection in Python. In addition, he contributed in drafting and structuring the paper.