

DETECTING VESSELS CARRYING MIGRANTS USING MACHINE LEARNING

A. Sfyridis^a, T. Cheng^{a*}, M. Vespe^b

^a SpaceTime Lab, University College London, Gower Street, London WC1E 6BT UK –
{alexandros.sfyridis.15, tao.cheng}@ucl.ac.uk

^b European Commission, Joint Research Centre (JRC), Directorate for Space, Security and Migration, Demography, Migration and Governance Unit, Via E. Fermi 2749, I-21027 Ispra (VA), Italy, michele.vespe@ec.europa.eu

KEY WORDS: Machine Learning, SVM, Anomaly Detection, AIS, GPS, Data Mining, Pattern Recognition

ABSTRACT:

Political instability, conflicts and inequalities result into significant flows of people worldwide, moving to different countries in search of a better life, safety or to be reunited with their families. Irregular crossings into Europe via sea routes, despite not being new, have recently increased together with the loss of lives of people in the attempt to reach EU shores. This highlights the need to find ways to improve the understanding of what is happening at sea. This paper, intends to expand the knowledge available on practices among smugglers and contribute to early warning and maritime situational awareness. By identifying smuggling techniques and based on anomaly detection methods, behaviours of interest are modelled and one class support vector machines are used to classify unlabelled data and detect potential smuggling vessels. Nine vessels are identified as potentially carrying irregular migrants and refugees. Though, further inspection of the results highlights possible misclassifications caused by data gaps and limited knowledge on smuggling tactics. Accepted classifications are considered subject to further investigation by the authorities.

1. INTRODUCTION

In recent years, the geopolitical situation in EU neighborhood countries and the beginning of the civil war in Syria in 2011, led to rapid increase in the figures of both international migrants and refugees. Smugglers often take advantage of desperate migrants and refugees and provide assistance and equipment to reach Europe. The irregular crossing of the Mediterranean Sea towards Europe has been undertaken using different types of vessels and is continuously changing forms (EPSC 2017).

The busiest routes to date are the Eastern and Central Mediterranean. In this work, we examine an additional route from the East to the Central Mediterranean Sea, followed by large cargo "ghost" ships carrying hundreds of irregular migrants and refugees in 2015. The route expands from the coasts of Israel, Lebanon and Egypt in the east and south east, to the Sea of Marmara in the north and the eastern and southern coasts of Sicily and the Italian province of Calabria respectively (Figure 1). Based on the concept of "Anomaly Detection" and utilising vessel signals, abnormal behaviours are modelled and detected. In particular, this project attempts to detect vessels that behave irregularly with respect to a predefined definition of abnormality in a spatio – temporal context within the dataset. However, this project does not assume that migrant vessels are necessarily anomalies. Vessels may behave in anomalous ways even without carrying migrants or refugees as well as smuggling vessels may behave normally. In order to shape a sensible approach to the ghost ship identification problem, understanding of data structure, migration flows and smuggler tactics as well as anomaly detection techniques are considered as the key areas and foundations of this project.

1.1 The Automatic Identification System (AIS)

AIS is an automatic tracking system for vessels, used to provide its position and other characteristics by exchanging information with other vessels, onshore stations - including Vessel Traffic Service (VTS) centres - and satellites (Transportation Research

Board 2003). The International Maritime Organisation (IMO), under the International Convention for the Safety of Life at Sea (SOLAS), has established as mandatory that all vessels of 300 gross tonnage (GT) or more in international traffic; all cargo vessels of 500 GT or more not operating in international voyages and all passenger vessels are required to be fitted with AIS (Bošnjak et al. 2012). However, not all vessels are covered by SOLAS and thus not all vessels are fitted with AIS. AIS operates in two Very High Frequency (VHF) channels and is able to detect ships within its range, which is typically expected to be around 20 to 30 nautical miles, depending on the height of antenna (IMO 2016). AIS data transmitted by any ship, can be classified in three different types. The Static data are entered into the AIS system on installation and include the Maritime Mobile Service Identity (MMSI), call sign and name, IMO number, ship's length and beam, type of ship and location of electronic position fixing system (EPFS) antenna. Dynamic data, are automatically updated from the sensors and include ship's position (latitude and longitude) in the WGS84 geodetic coordinate system, position time stamp, course over ground (COG), speed over ground (SOG), heading, navigational status and rate of turn (ROT). Voyage related data which are usually manually entered and updated during the voyage, include ship's draught, hazardous cargo, route plan, as well as destination and estimated time of arrival (ETA). Data transmission is automatic and its frequency depends on the type of information transmitted. Static and voyage – related data is transmitted every 6 minutes or upon request, while dynamic data transmission depends on the speed and course of the vessel. (IMO, 2016). Despite being conceived for collision avoidance, AIS is nowadays a cornerstone of maritime surveillance, and is currently used by the scientific community to extract patterns, predict routes and detect anomalies (e.g Pallotta et al. 2013)

1.2 Migration Routes and Smuggling Tactics

As far as the smuggling networks, vessel routes and trajectories are concerned little is known about. To build knowledge in this area we summarise all available information with regards to migrant vessel behaviours (Table 1).

*Corresponding author

Table 1. Summary of Migration Routes and Smuggling Tactics

Ports of Interest (POIs)	
Turkey	Mersin (Channel 4 2015; Correct!V 2015), Istanbul and Izmir (BBC 2014; Vice News 2015)
Syria	Tartus (Correct!V 2015; Arab Reporters for Investigative Journalism 2016)
Libya	Misrata, Zuwara (Lewis & Elumami 2016) and Sabratha (Xchange 2016)
Greece	Islands of the Aegean Sea (Lister (CNN) 2015)
Italy	The port of Gallipoli (Lister (CNN) 2015; International Business Times 2015) the islands of Lampedusa and Pentellera. Ports of Sicily and Calabria.
Other	Malta (UNODC 2011), Egypt (Fenwick 2016) and Slovenia (Washington Post 2016)
Types of Vessels	
Turkish Coast to Greek Islands	small wooden boats and inflatable dinghies (UNODC 2011)
Libya to Italy	Fishing trawlers (BBC 2016; CBCnews 2015), commercial boats, ferries and ocean – going pleasure yachts (Frontex 2016), since 2016 Smugglers shift ing from vessels to inflatable dinghies (EPSC 2017)
Other	old cargo and livestock carriers, the so called “ghost ships”, usually departing from Mersin, Turkey (Frontex 2016; Arab Reporters for Investigative Journalism 2016). They have also been spotted in Korfez, Turkey (news.com.au 2015), Tartus, Syria and Famagusta in the Turkish controlled Cyprus (The Telegraph 2015)
Behaviour at Sea	
Mainly followed by smugglers from Egypt	Use of two different vessels. The first vessel transfers the migrants to the open sea and the second vessel undertakes the second and usually longest part of the journey, while the first returns to the point of origin (Frontex 2016)
In the case of old cargo and livestock carriers departing from Turkey	Smugglers abandon vessels and set to autopilot towards Italy (Daily News 2016; The Guardian 2015). Vessels sailing in unusual ways, zig-zagging in open waters between Cyprus and Turkey (The Guardian 2015) as well as spending days in open waters prior commencing a regular course (Correct!V 2015), have been observed.

1.3 Anomaly Detection

Anomalies are considered as parts or patterns in a dataset, “that do not conform to a well-defined notion of normal behaviour” (Chandola et al. 2009). In other words anomalies are the outliers in a dataset and consequently anomaly detection is the process of identifying these outliers. Chandola et al. (2009) defines anomaly detection as “the problem of finding patterns in data that do not conform to expected behaviour” and it has been used in a variety of domains, such as intrusion detection in computer related

systems, credit card and mobile phone fraud detection, industrial damage detection, text data, as well as healthcare and disease detection amongst others.

Though, anomalies must be distinguished from noise. Noise can be caused by unwanted variables in the dataset which can impact the analysis and by considered as not necessary, they can be removed. Thus, noise is not necessarily an anomaly. On the other hand, anomalies occur when a variable of interest deviates from what is defined as normality within the dataset. However, both anomalies and noise are considered as barriers in data analysis.

The specific type of anomaly, data labels as well as anomaly detection methods, techniques and outputs should be considered in an anomaly detection problem. Anomalies are categorised in three main types; Point, Contextual and Collective. Data labels are the ones that classify each data instance as being normal or anomalous and are usually defined by human experts. Getting labels for anomalous data, is typically more difficult than getting labels for normal behaviour and acquisition of data labels for anomalous data instances can be rare, since anomalies tend to be less frequent than normalities (Chandola et al. 2009). Data labels are also important due to their significance in the selection of anomaly detection mode, classified as Supervised, Semi – Supervised and Unsupervised. Regardless the technique used to identify anomalies within a dataset, the output of the process can be generally reported in two ways; Scores and Labels. According to the type of anomaly, a relevant method should be selected in order to have reliable results out of the process. The generalized methods commonly used are: Classification, Nearest neighbour, Clustering, Statistical, Information Theoretic and Spectral.

2. DATA

2.1 Dataset Description

The main dataset includes one month’s vessel AIS traces in the east Mediterranean, expanding from the coast of Lebanon and the Suez Canal in the south east, to the south coast of Italy including Sicily. It covers all the Aegean, Marmara and Ionian seas and the raw dataset consists of over 6.7 million points, representing approximately 6000 vessels (Figure 1). Out of these vessels, three are labelled as “migrant” (Figure 2). For each vessel only dynamic and static attributes are provided (excluding voyage related) as shown in Table 2. For the purposes of this research two additional datasets indicating landmasses and ports in the area of interest (AOI), have been extracted from open sources.

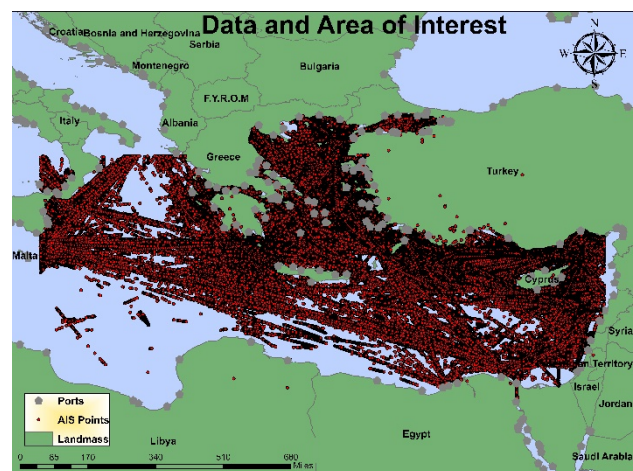


Figure 1. Data Area of Interest

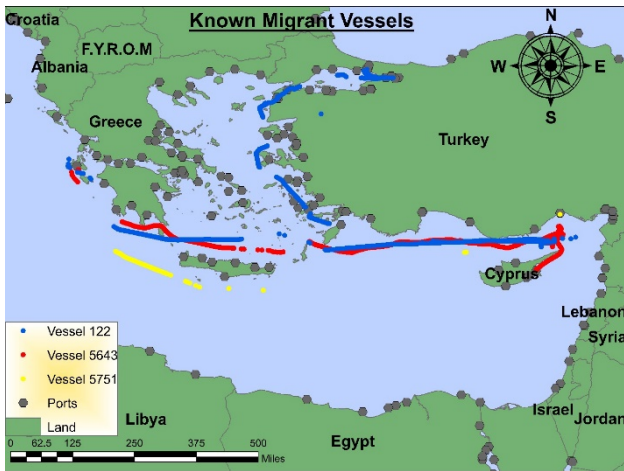


Figure 2. Known Migrant Vessels

Table 2. Vessel Data Attributes

Information Type	Attribute	Unit	Notes
Static	Latitude	Decimal Degrees	In WGS84 Coordinate System
	Longitude	Decimal Degrees	In WGS84 Coordinate System
	Course Over Ground (COG)	Degrees	
	Speed Over Ground (SOG)	Knots	
	Rate of Turn (ROT)	Degrees per Minute	
	Timestamp	Seconds	progressive time elapsed from the oldest messages in the dataset
Dynamic	Identification Number (ID)	Numeric	Unique identification number for each vessel, replacing MMSI
	Ship Type	Numeric code	
	Mid_Code	Numeric code	Representing country code (flag).

2.2 Data Filtering

Based on the smuggling tactics identified (Table 1) and attributes exploration, data clearance and manipulation are considered.

2.2.1 Ship Type: Inflatable boats and dinghies used mainly for the crossing from Turkey to the Greek islands as well as along other route (e.g. Central Mediterranean route), do not have AIS system fitted and thus, they are not expected to be found in this dataset. Moreover, it is assumed that passenger vessels, ferries and large yachts that have occasionally been used for smuggling

purposes are unlikely to exhibit suspicious behaviour. This types of vessels hide migrants under their decks (Frontex 2016) and thus it is not expected to behave abnormally while sailing. Consequently, two main types of vessels used directly for smuggling are included in the dataset; cargo and fishing. In addition, the ship type numerical code; reveals the three known migrant vessels' type, being either cargo or "other". Thus, based on the ship type attribute for each vessel, initial data filtering considering both the vessels most likely to carry migrants (cargo and fishing) as well as the three known migrant vessel types in the dataset is taking place. The filtering results to 2972 vessels and over 3 million points.

2.2.2 Other Attributes: In addition, the ROT attribute is completely removed from the dataset due to large number of zero values (over 4.5 million null values) and mid_code is not considered any further, due to reliability concerns because of manual input. The ID field is only used to represent each vessel in the migrant vessel identification process.

Thus, longitude, latitude, SOG, COG and timestamp are the five attributes to be further processed and utilized for anomaly detection.

3. METHODOLOGY

The process of identifying anomalies is conducted in three stages. First, we define the concepts of normality and abnormality. Similar to the work of Teng et al. (1990), abnormality is defined by setting rules, describing the migrant vessel behaviours identified. That is, a vessel's behaviour that does not match the predefined abnormality can be considered as normal. Second, we design the variables capturing the defined anomalous behaviour. Finally, we select the most information rich designed features and suitable algorithm for training.

3.1 Rule Definition

Following data filtering based on vessel types as presented in Table 1, a vessel's "behaviour at sea" can now be used to define abnormality.

In the first case; where two vessels are involved, it can be assumed that the first vessel would have a regular behaviour during the first part of each journey, exhibiting relatively constant speed and course until it reaches the second vessel. During the second stage, where people are transferred from one vessel to the other, both vessels would potentially exhibit zero or very low speed in the open water, with the possibility of either constant or no change on their course depending on water currents and weather conditions. The third stage - where the first vessel starts its journey back to the point of origin - a potentially large change in course and speed is expected. Finally, the vessels would possibly exhibit regular speed and course until the first one returns to its point of origin and the second reaches its destination.

For the second case, where the vessel stops or exhibits random movement in open waters, two major inferences can be drawn. While the vessel spends days in open waters, it is expected that zero or very low speed and significant change in course will be evident; again, depending on currents and weather conditions (e.g. wind). In addition, while the vessel sails towards random directions in open waters, a constant variation in its speed and course is expected to be evident. Both behaviours are observed in open waters; hence, in a relatively long distance from the coast.

Considering these behaviours, it can now be argued that a vessel exhibits anomalous behaviour in the event¹:

- i. It has zero or very low speed in a particular geographic area that is not a port, harbour, oil platform, or any other manmade infrastructure designed to serve the process of anchoring, loading and/or unloading goods or passengers from and to a vessel.
- ii. It exhibits a constant change in its speed over a period of time in a similar geographic area.
- iii. It exhibits constant change in its course over a period of time in a similar geographic area.

3.2 Feature Engineering

Feature engineering is the process where empirical knowledge is used to create features that describe specific conditions of the data. Hence, the purpose of this section is to design the variables illustrating each vessel's behaviour and reveal potential anomalies as defined by the rules.

In our AIS dataset, each vessel is represented as a temporal sequence of events having spatial characteristics, expressed as longitudinal and latitudinal attributes. Thus, considering the sequential spatio – temporal structure of the data, an anomaly is expected to be evident in a series of consecutive data instances (subsequence). That is, a vessel, may exhibit anomalous as well as normal behaviour during different parts of its trajectory.

Consequently, the designed features should be such, that will capture the spatio – temporal data structure and behaviour of each vessel as well as identify subsequences in the vessel trajectories that could be considered as anomalous.

Hence, by utilizing the data attributes, we design the features as follows:

- i. From the SOG variable, the percentage of time each vessel exhibited low speed with respect to the total of its trajectory is calculated. Low speed threshold has been assumed below 3 knots according to AIS broadcast specifications
- ii. To capture the change of SOG over time we first calculate Acceleration, which is the rate of change of speed per time unit. SOG (in knots) is transformed to m/s, knowing that 1knot = 0.514444m/s. Then:

$$\alpha = \frac{\Delta V}{\Delta t} = \frac{V_f - V_i}{t_f - t_i} \quad (1)$$

Where α = acceleration (in m/s²)
 V_f, V_i = final and initial velocities² (in m/s)
 t_f, t_i = ending and starting time² (in seconds)

Furthermore, four additional features have been calculated:

- a. Mean acceleration
The mean value of acceleration during the vessel's trajectory
- b. Standard Deviation (SD) of Acceleration
The SD of acceleration during each vessel's trajectory.
- c. SD threshold

Same as the SOG variable (i), this feature represents the percentage of time a vessel's acceleration has exceeded a specific threshold. In this case the threshold set is the SD of acceleration (Figure 3)

- d. SD * 3 threshold

Similarly, this feature represents the percentage of time, a vessel's acceleration has exceeded the SD threshold multiplied by three (Figure 4). This threshold has been additionally set, to overcome occasions where a vessel would exceed lower thresholds due to conditions not considered in this project as described in 3.1. It is assumed, that exceeding this threshold is a significant indication of abnormality.

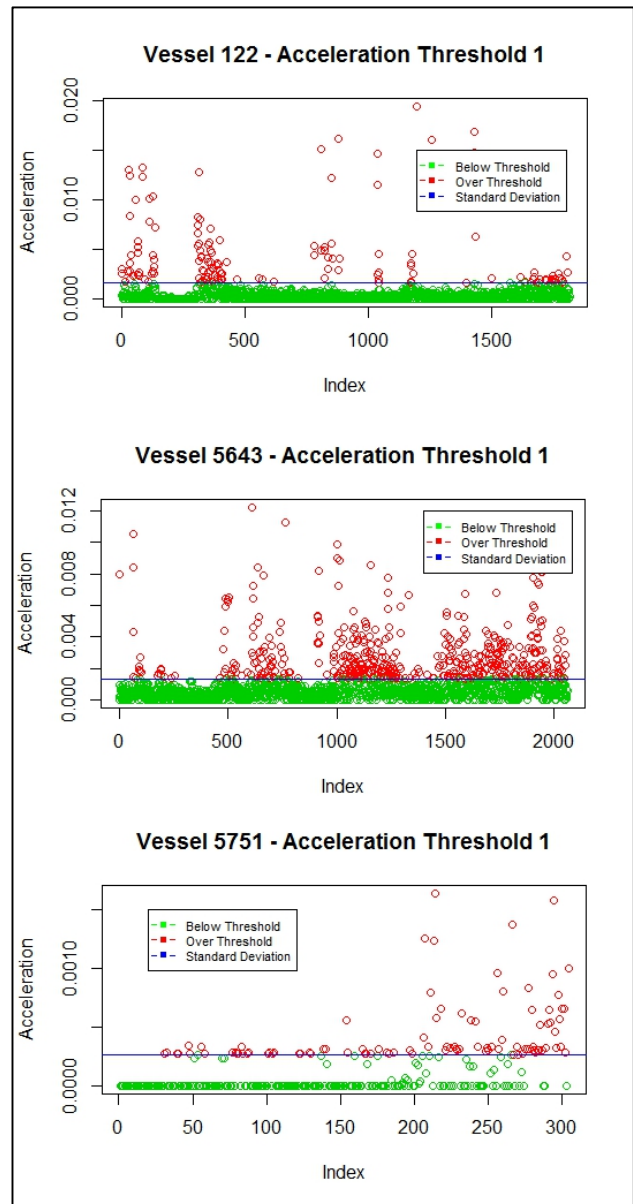


Figure 3. Acceleration Threshold (SD)

¹ These rules are not considering other factors affecting a vessel's behaviour, such as weather conditions or tidal status.

² Starting and ending values, are the values of two consecutive data instances.

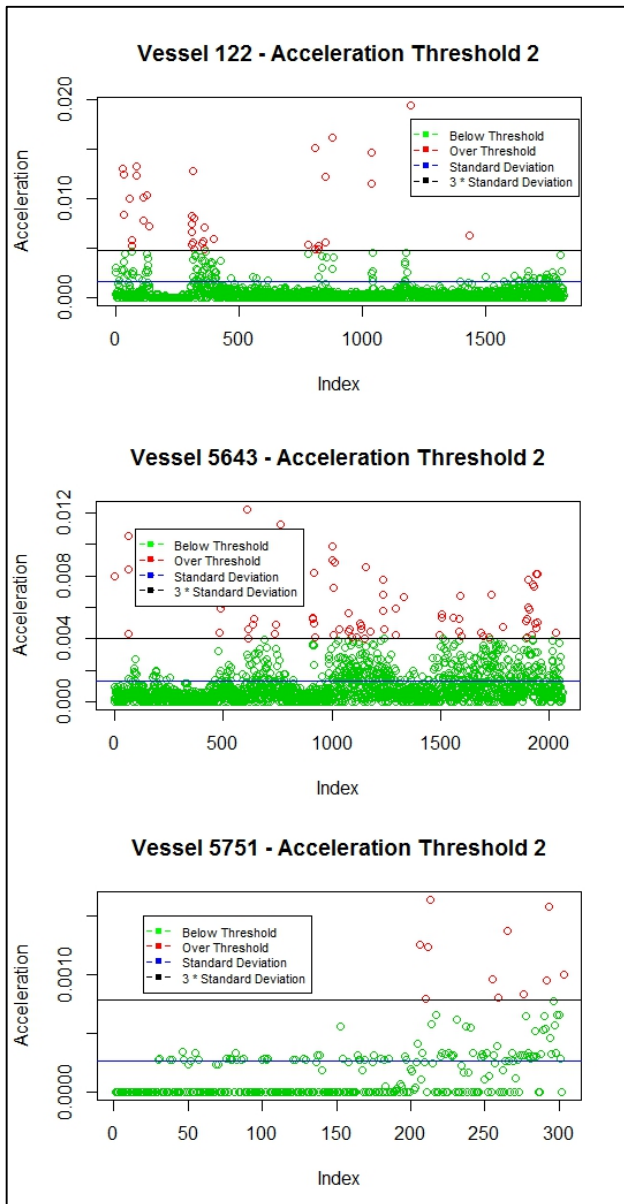


Figure 4. Acceleration Threshold (SD*3)

- iii. To capture the COG change over time we first calculate the Angular Velocity, which represents the change of course per time unit:

$$\omega = \frac{d\theta}{dt} = \frac{\theta_f - \theta_i}{t_f - t_i} \quad (2)$$

Where ω = angular velocity (in degrees/sec)
 θ_f, θ_i = final and initial angles² (in degrees)
 t_f, t_i = ending and starting times² (in sec)

Similar to Acceleration, four additional features have been calculated:

- a. Mean angular velocity
- b. SD of angular velocity
- c. SD threshold
- d. SD *3 threshold

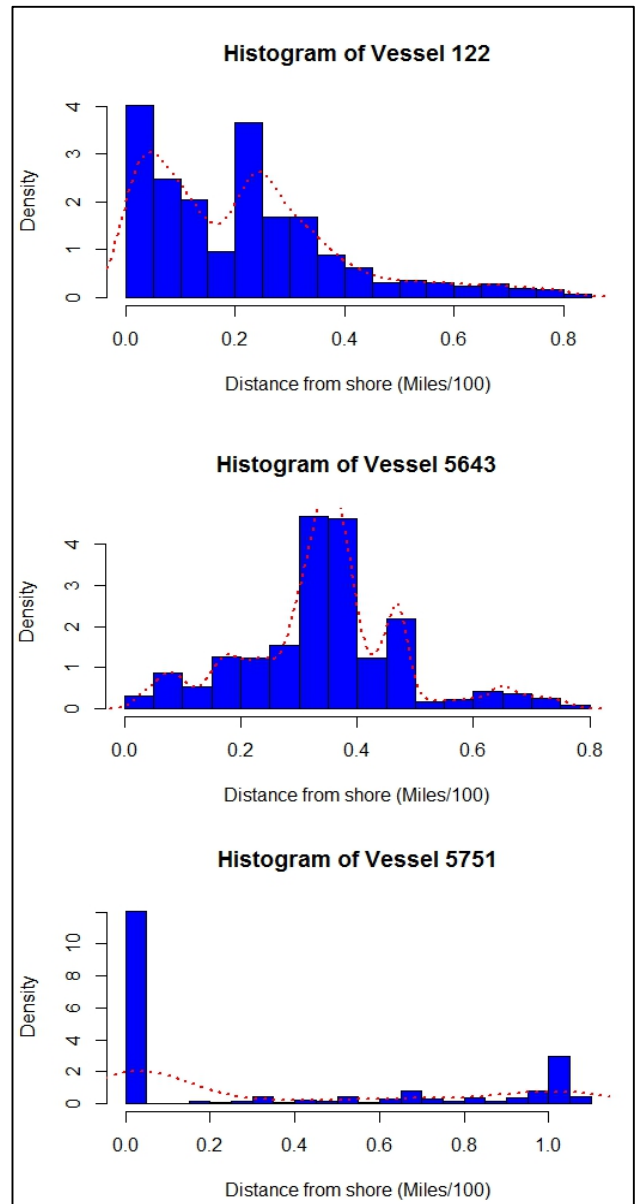


Figure 5. Distance from Shore Histograms

- iv. To include spatial information the distance from shore (Dfs) has been calculated from the relative position of each point, given in longitude and latitude. Dfs has been used for two main reasons. First, it is indicative of standard route lanes used by vessels. Second, using coordinates, would bias the classification towards specific geographic areas. Thereinafter, from the Dfs distribution (Figure 5) skewness has been calculated. Skewness values are expected to be an indicative attribute capturing the spatial behaviour of each vessel, particularly in the event where a vessel is showing random movement in open waters.

For all designed features the values have been calculated where the vessel is in the open sea. That is, data instances indicating that a vessel is moored in a port or anchorage point have been excluded from the calculations. Ports have been identified from the available dataset and anchorage points have been assumed to occur in a maximum distance of 1 nautical mile (1.8km, 1.15miles). These data instances have not been included in our calculations, assuming that when a vessel is stationed in a port or

mooring point cannot exhibit anomalous behaviour. A stationed vessel is expected to exhibit zero or very low speed and randomly change its course according to wind and water currents. Inclusion of these points in the feature engineering process, would bias the variables towards larger non-indicative values of angular velocity and potentially lower values for acceleration.

However, the excluded data instances are used to construct an additional feature, designated as “Port” and added as a categorical feature (dummy variable). Vessels where their trajectory starts, stops or passes for a long or short period of time through a POI, have given different value in the “port” feature than others.

Consequently, eleven features have been designed and will be further processed for anomaly detection. However, even though normality has not been defined, it is expected that the three labelled vessels will exhibit abnormal behaviour at some part of their trajectory.

3.3 Anomaly Detection

Integrating all the 11 features into an algorithm may result into high computational requirement and model overfitting. Thus, feature selection prior training will allow us to improve the prediction performance and reduce the effects from noise of irrelevant features (Chandrashekar & Sahin 2014).

3.3.1 Feature Selection: For the purposes of this study, two different feature selection methods are used. Pearson’s Correlation is applied to rank and remove the highly correlated features and a Random Forest (RF) applies a feature importance measure on correlation outcomes³.

From the correlation results (Figure 6), we select Acceleration SD and Angular Velocity Threshold 2 as not highly positive correlated, speed (as the only feature representing the actual vessel speed in the open sea), Angular Velocity SD and Acceleration Mean (as less correlated for SOG and COG).

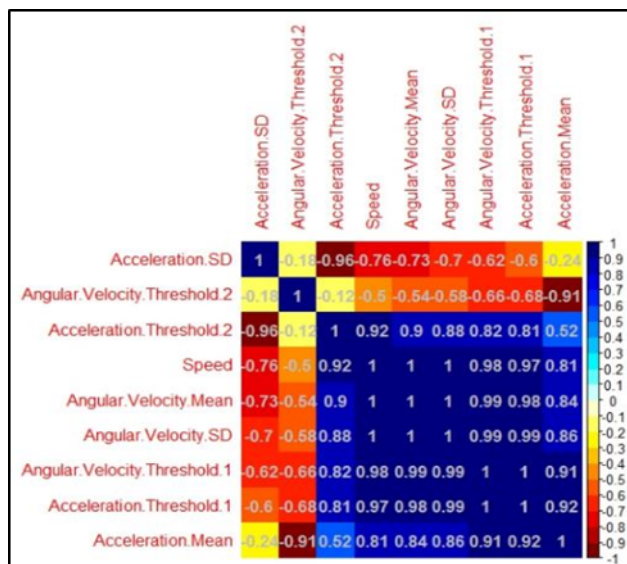


Figure 6. Features' Correlation

³ The port and skewness variables have been omitted from feature selection. The port feature, is the only variable representing a vessel’s origin and skewness is the only spatial variable. In the event these variables had to be excluded, the classification would

The RF feature selection (Figure 7) shows that Acceleration SD and Angular Velocity SD are the two most important attributes where Acceleration Mean and Angular Velocity Threshold 2 are the less important variables in the dataset. We proceed with the selection of the three most important variables, since again; speed is the only variable that can reveal the actual speed of a vessel in each position.

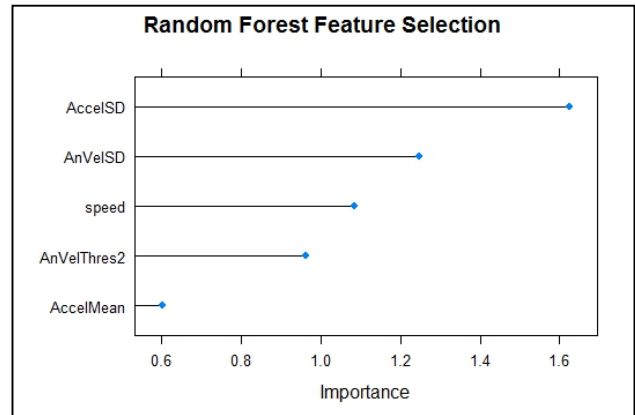


Figure 7. Random Forest Feature Selection

The feature selection process, resulted into two out of the nine initially considered variables to be ranked as information reach. However, the inclusion of the Speed feature is considered as a necessity and the two additional features (ports and skewness) are to be used in the training process as well. Consequently, these five features are further used for training the algorithm.

3.3.2 Algorithm Selection: Since labelled data instances are known only for migrant vessels, the potential algorithm will have to learn from one class (what is considered as anomaly), as far as there is no class for normality. The algorithm will have to classify the new data as anomalous if similar patterns are identified; and normal, if identified patterns deviate significantly from the training data. Moya & Hush (1996) introduced and defined this process as one class classification (OCC) or unary classification, where “a classifier that can recognize new examples of target patterns and distinguish those from non-target patterns is called a *one-class classifier*”.

To classify the data, we apply One – Class Support Vector Machine (OCSVM). Support Vector Machines (SVMs) are applicable to linear and non – linear classification problems. In non - linear classification, SVMs classify the data by applying the Kernel Trick, where Kernel Functions take a low dimensional input space and map the data by transforming it into a higher dimensional feature space (Boser et al. 1992). The OCSVM developed by Scholkopf et al. (1999), sets an upper bound on the fraction of outliers and it is a lower bound on the number of training examples used as Support Vectors (Vlasveld 2013).

4. RESULTS

Training the OCSVM with the five selected variables, resulted into detecting nine vessels of different types, originating from various ports in the study area (Figure 8).

violate the spatio – temporal structure of the data and the key identifications of a migrant vessel’s overall behaviour as presented in Table 1.

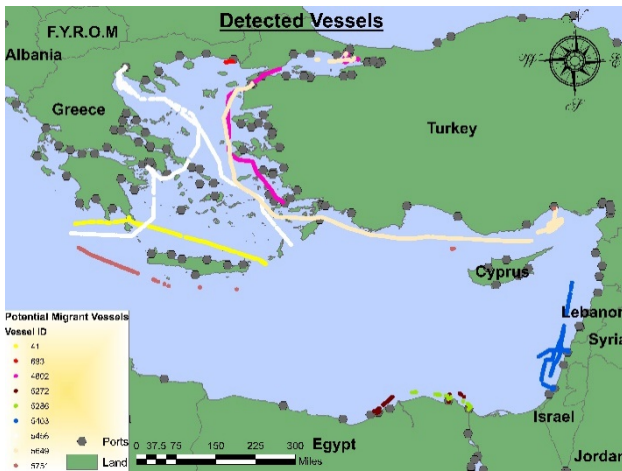


Figure 8. Detected Vessels

The identified vessels, are spread across the East Mediterranean covering every corner of the AOI. From these vessels, one is a fishing boat and the rest are cargo ships. Five vessels appear to have departed or at some point stopped or passed through one of the POIs, including Istanbul, Mersin and three Egyptian ports. The rest four vessels have not been in the vicinity of any of the POIs. In addition, three of the vessels are presented with a sparse number of points across space and time in their trajectory, which appears to be incomplete.

However, an interesting finding is vessel 5649 whose trajectory exhibits very similar pattern with the labelled vessels, particularly in the area off the Turkish port of Mersin (Figure 9).

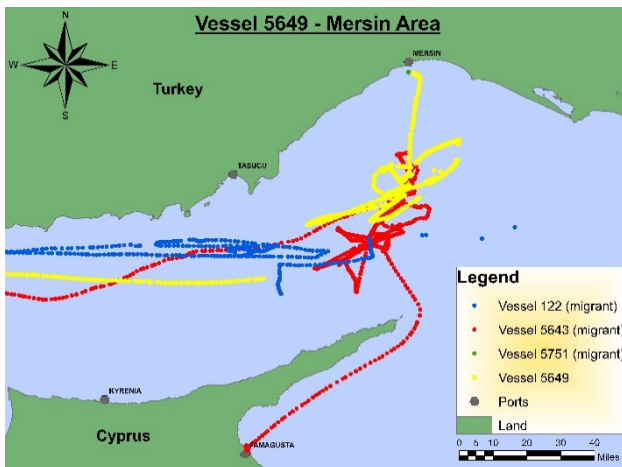


Figure 9. Vessel 5649 - Mersin Area

In addition, another vessel demonstrates similar behaviour in part of its journey. Though, its trajectory ends before reaching the south east coast of Turkey and its behaviour in that area cannot be mapped. Moreover, further exploration of the detected vessels reveals stopping and departing areas such as shipyards and small marinas where scrap vessels can be acquired and smuggling operations are easier to apply.

5. CONCLUSIONS

The importance of anomaly detection in maritime domain awareness can be highlighted by the occurrence of incidents, associated with impacts on safety and security as well as human

lives. This project highlighted the limited documentation on migration flows, smuggling tactics and in particular; vessel behaviours as well as the significance of fact - based feature design, subject to data availability and reliability.

Few abnormal behaviour examples resulted to limited behaviour modelling and consequently, the designed features were not expected to depict all possible abnormal behaviours. Vagueness in identified information was also evident. For example, there was no clear evidence on major ports of ghost ships departure from Egypt, forcing the modelling process to mark all Egyptian ports across the Mediterranean coast as POIs. Moreover, data gaps appear to be a significant barrier. Bearing in mind that the dataset covers one month's period, vessel's trajectories depicted with very few points cannot be considered adequate and reliable. In addition, AIS broadcasting frequencies appear rather random since most vessels in the dataset broadcast intermittently across the timeframe under examination. Spatial issues are also evident. Assuming a vessel is moored when within 1 nautical mile off the coast resulted to possible misclassifications. The projection of combined datasets (landmasses, ports, vessels), acquired from different sources, led to misplacement of data instances in relation to the base map (landmasses). For example, a vessel may appear to be stagnant in the open seas at a distance larger than 1 nautical mile from the coast, where it may be closer, or even inside the port. Though, apart from projection, AIS positioning information is subject to accuracy as well. Nevertheless, knowing that a vessel has stopped within 1 nautical mile or further away from the shore, would not actually affect the smuggling process, but the impacts on the model are of high significance, since all feature design has been based upon this assumption. In particular, considering misplaced data instances, highly affects the skewness and port features, which are based on spatial attributes. In addition, all other features have been designed excluding data instances in range of one nautical mile from the shore, considering that these instances would significantly affect the calculated values. Hence, the spatial attribute has significantly affected the designed features and consequently the results. Thus, it can be argued that the one nautical mile assumption may have been inadequate for the natural characteristics of the study area, where many islands and narrow straits prevail and vessels are likely to sail close to the shore. Furthermore, very few labels as well as the training process with only one class, has been another major barrier affecting the algorithm selection process and potentially the results.

However, taking into account the aforementioned limitations, the identification of nine vessels as potentially being operated by smugglers, indicated that the OCSVM classifier have equally considered all predefined variables to conduct classification and provided a sensible and acceptable outcome.

Though, assuming that all detected vessels are migrant within the anomaly framework, is arbitrary. An anomaly is defined as the demonstration of uncommon values within a sequence of events. This study has used the anomalies of smuggling vessels. However, it is not a necessity that vessels operated by smugglers will exhibit this behaviour. They may behave normally whilst a vessel may be classified as abnormal due to several factors affecting its course, such as involvement in some other kind of illegal operation (e.g. drugs, weapons, etc.), engine failure, weather conditions imposing change of direction or a vessel's diversion to participate in a rescue operation.

In any event, even though the OCSVM classification can be considered as adequate and thought of as a novel approach in maritime anomaly detection; not much is yet known about

smuggling tactics and vessels used for these purposes. A constant monitoring is required to record and understand these tactics and vessel behaviours when evident. This would potentially lead to additional patterns to be identified and consequently modelled. A dataset with mixed labelled instances (normal and anomalous), focus on smaller specific geographic areas and trajectory analysis of specific types of vessels could highly improve the results and provide less false positives and negatives.

REFERENCES

- Arab Reporters for Investigative Journalism, 2016. Pirates of the Mediterranean Sea. Available at: <http://en.arij.net/report/pirates-of-the-mediterranean-sea/>.
- BBC, 2014. Mapping Mediterranean Migration. Available at: <http://www.bbc.co.uk/news/world-europe-24521614>.
- BBC, 2016. Migrant tragedy: Anatomy of a shipwreck. Available at: <http://www.bbc.co.uk/news/world-europe-36278529>.
- Boser, B.E., Guyon, I.M. & Vapnik, V.N., 1992. A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory*, pp.144–152. Available at: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.3818>.
- Bošnjak, R., Šimunović, L. & Kavran, Z., 2012. Automatic Identification System in Maritime Traffic and Error Analysis.
- CBCNews, 2015. Migrant ship sinking: Collision with cargo ship, movement of migrants to blame. Available at: <http://www.cbc.ca/news/world/migrant-ship-sinking-collision-with-cargo-ship-movement-of-migrants-to-blame-1.3041785>.
- Chandola, V., Banerjee, A. & Kumar, V., 2009. Anomaly Detection : A Survey. , (September), pp.1–72.
- Chandrashekar, G. & Sahin, F., 2014. A survey on feature selection methods. *Computers and Electrical Engineering*, 40(1), pp.16–28. Available at: <http://dx.doi.org/10.1016/j.compeleceng.2013.11.024>.
- Channel 4, 2015. The Mediterranean’s deadliest migration sea routes. Available at: <http://www.channel4.com/news/lampedusa-mediterranean-deadliest-migration-routes-libya>.
- Correct!V, 2015. The Hidden Fleet. Available at: <https://correctiv.org/en/investigations/stories/2015/12/11/hidden-fleet/>.
- Daily News, 2016. Turkish crackdown on “ghost ships” sparked migrant drive to Greece: IOM. Available at: <http://www.hurriyetdailynews.com/turkish-crackdown-on-ghost-ships-sparked-migrant-drive-to-greece-iom-.aspx?pageID=238&nid=86940>.
- EPSC, 2017. Irregular Migration via the Central Mediterranean.
- Frontex, 2016. Apulia and Calabria Route. Available at: <http://frontex.europa.eu/trends-and-routes/apulia-and-calabria-route/>.
- International Business Times, 2015. Turkey opens fire on migrant ship to stop it entering European Union waters. Available at: <http://www.ibtimes.co.uk/turkey-opens-fire-migrant-ship-stop-it-entering-european-union-waters-1491800>.
- International Maritime Organisation (IMO), 2016. Revised Guidelines for Onboard Operational use of Ship Automatic Identification Systems (AIS) . , 917(22).
- Lewis, A. & Elumami, A., 2016. Dangerous Migrant Smuggling Routes Flourish in Lawless Lybia. Available at: <http://www.reuters.com/article/us-europe-migrants-libya-idUSKCN0YO1GZ>.
- Lister (CNN), T., 2015. Desperation, dire conditions for migrants fleeing to Europe. Available at: <http://edition.cnn.com/2015/01/03/world/migrants-europe/>.
- Moya, M. & Hush, D., 1996. Network Constraints and Multi-objective Optimization for One-class Classification.
- news.com.au, 2015. Why people smugglers have jumped to dangerous “ghost ships.” Available at: <http://www.news.com.au/world/why-people-smugglers-have-jumped-to-dangerous-ghost-ships/news-story/65386d4d928746ae67e7b32cd2215c4a>.
- Pallotta, G., Vespe, M. & Bryan, K., 2013. Vessel pattern knowledge discovery from AIS data: A framework for anomaly detection and route prediction. *Entropy*, 15(6), pp.2218–2245.
- Scholkopf, B. et al., 1999. Support Vector Method for Novelty Detection. , pp.1–7.
- Teng, H.S., Chen, K. & Lu, S., 1990. Adaptive Real-time Anomaly Detection Using Inductively Generated Sequential Patterns.
- The Guardian, 2015. Smugglers abandon migrant ship off Italy in new tactic to force rescue. Available at: <https://www.theguardian.com/world/2015/jan/02/smugglers-abandon-migrant-ship-italy-ezadeen>.
- The Telegraph, 2015. Abandoned migrant ship reaches Italy. Available at: <http://www.telegraph.co.uk/news/worldnews/europe/italy/11322954/Abandoned-migrant-ship-reaches-Italy.html>.
- Transportation Research Board, 2003. *SHIPBOARD AUTOMATIC IDENTIFICATION SYSTEM DISPLAYS: Meeting the Needs of Mariners*
- UNODC, 2011. Smuggling of Migrants by Sea.
- Vice News, 2015. Death Boats to Greece (Part 1): Europe Or Die (Episode 2). Available at: <https://news.vice.com/video/death-boats-to-greece-part-1-europe-or-die-episode-2>.
- Vlasveld, R., 2013. Introduction to One-class Support Vector Machines. Available at: <http://rvlasveld.github.io/blog/2013/07/12/introduction-to-one-class-support-vector-machines/>.
- Washington Post, 2016. Slovenia says 12 migrants found on cargo ship from Turkey.
- Xchange, 2016. More than 1,000 Feared Death in Updated Death Toll – IOM. Available at: <http://migrantreport.org/700-feared-dead-worst-week-2016/>.