Classifying Maritime Vessel Behaviour From AIS Using Statistical Features

James Cormack[1] [0000-0001-7103-3043], Matthew Roughan [0000-0002-7882-7329], and Hung Nguyen [0000-0003-1028-920X]

University of Adelaide, Adelaide 5000, Australia
{matthew.roughan,hung.nguyen}@adelaide.edu.au

Abstract. Remotely monitoring maritime vessels is crucial for policing and for ensuring their safety along Australia's extensive coastline. Detecting unusual behaviour is vital for evaluating naval activity, responding to illegal activities, and detecting ships in distress. AIS (Automatic Identification System) data provides a rich source for research in this domain; however, it only provides a limited set of features and the data is noisy and irregular. In this study, we augment sampled trajectory segments of base AIS data with a suite of statistical features and label them with behaviours in order to perform supervised learning. We evaluate these features in classifying maritime vessel behaviour using three different machine learning classifiers - Random Forests, K-Nearest Neighbour and a Multilayer Perceptron. The latter performed best using a set of the twenty most important statistical features, as determined by Gini Impurity. It returned a weighted-average F1 Score of 83.6%. These results demonstrate that a classifier built from a manually labelled AIS dataset can discriminate vessel behaviour using only simple statistical features that are easy to generate and can be applied to a range of classifiers. This finding lays the groundwork for future research aimed at identifying anomalous behaviour in maritime traffic.

Keywords: AIS · maritime behaviour · shipping classification

1 Introduction

The ability to identify a vessel based on its behaviour and detect unusual behaviour is important for several reasons. With such a large coastline, Australia depends heavily upon remote sensing to ensure the safety of vessels at sea. Also, maritime policing has become increasingly important for Australia. Detecting unusual behaviours is needed for evaluating military naval activity and for responding to activities such as illegal fishing, illegal immigration, goods trafficking and piracy. Aside from detecting malicious behaviour, early alerting of search and rescue authorities to unusual behaviour allows for a quicker response time, saving vessels in distress and potentially human lives.

[1] It is with deep sadness that we note that the lead author on this paper – James Cormack – passed away during the final preparation of the paper. He is sorely missed by his family, friends and colleagues.

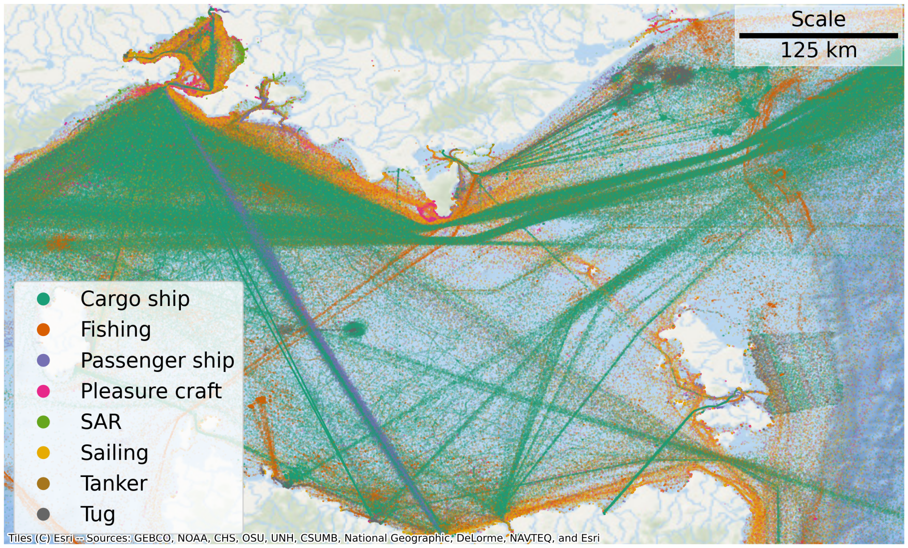

Fig. 1: The maritime traffic in this study in and around Bass Strait (on the Southern coast of Australia). The map shows the recorded vessel trajectories of the top 8 most common vessel types for the period January 2021 to March 2023. Each vessel type is represented by a different colour (SAR refers to search and rescue craft).

The goal of this research is to use AIS maritime vessel data to classify the behaviour of short, fixed-length segments of vessel tracks using simple summary statistical features into one of four states of motion: Stopped, Straight, Turning or Manoeuvring.

Automatic Identification System (AIS) shipping data contains important information about vessels but does not contain labels indicating their type of motion, so we created a labelled dataset using a combination of automatic rule-based and manual labelling using a custom-built labelling tool. After labelling, we resample the segments to rebalance the distribution of segments with each motion type. We augment the base AIS features (Latitude, Longitude, Course, Speed, Timestamp) with over sixty statistical features. Using selected combinations of these features, we train three classifiers using Multi-layer Perception (MLP), Random Forest (RF) and k-Nearest Neighbour (k-NN). We use the Gini Impurity from the RF to order the importance of our statistical features. We evaluate the efficacy of the resulting classifier/feature combinations by looking at F1 scores and generating confusion matrices to visualise the results on a per-category basis.

The study found that summary statistics can accurately classify maritime vessel motion in open water. The best results were achieved using an MLP classifier with the top 20 (by Gini Impurity score) statistical features, obtaining an average F1-score of 83.6%. The precision/recall for the 'Stopped' motion type reached as high as 95%.

In summary we make the following contributions:

– We present a study that covers open-water trajectories in Australian waters, detecting behaviour in randomly chosen segments, and showing that behaviour classification of these segments is practical from AIS data.
– We present a mixed-mode method for labelling data with behaviours, partially automated and partially manual, to demonstrate the creation of a labelled dataset with reduced effort.
– We propose a large set of statistical scalar values as features to summarise and represent a vessel's motion and show which features are most useful for behaviour classification.

This paper outlines a method of preparation, segmentation and labelling of AIS for supervised learning. It applies a set of simple statistical features and uses them to classify a vessel's behaviour.

2 Background

The Automatic Identification System (AIS), originally a system to aid vessels in collision avoidance, is now used as a rich source of data for numerous other maritime navigational research activities such as anomaly detection, route planning, trajectory prediction and vessel classification [15, 16, 22, 23].

Numerous publicly available AIS datasets exist [1–4]; however, data quality is often reduced by such factors as noise, errors and missing data [13, 21–24]. To overcome these issues, data cleaning methods like deletion [25], interpolation [8, 17] and resampling [8] are often employed before using the data to train models. In this study, we use deletion and resampling; however, choose not to interpolate new data points due to their tendency to mask underlying patterns of motion.

Many studies decompose long trajectories into segments for analysis. The beginning and end point of a segment can be defined using known origin and destination ports [9], known stopping locations [12, 20] or anchoring/turning points [26]. Some techniques attempt to produce homogeneous segments where each segment exhibits a single behaviour, or characteristic [19]. For further reading on more complicated segmentation techniques, please refer to [7]. All of these techniques produce segments of variable length. We choose to use fixed size randomly selected segments as they are simple to generate, allow us to generate comparable statistics and mimic real-life data streams where the most recent points, often containing mixed behaviours, are the most important.

In this study we use summary features, that is, scalar features that summarise all track points over a segment rather than a vector of feature points. Dominant direction of travel, maximum speed, number of stops per hour and various ratios of speed and course are used in [18]. Max, min, median, mean, std dev, skew, kurtosis, range, interquartile range and a series of quantiles are used in [10]. Summary features are simple to generate and easy to apply to a range of classifiers.

The most similar paper to ours [10] classifies 6 vessel types from AIS kinematic data using simple statistical features (max, min, median, standard deviation, skew, kurtosis, range, quantiles) and an XGBoost classifier. We use a much larger set of statistical features, but also we are focused on classifying behaviour, not vessel type, and are using short segments to do so.

3 Data

3.1 Raw Data

This study makes use of a publicly available AIS data set of maritime traffic provided by the Australian Maritime Safety Authority (AMSA). Files are provided as geospatial vector data files in the ESRI shapefile format and can be downloaded from [6]. AMSA has modified the dataset prior to public release, anonymising the vessel identification and restricting the update interval to a minimum of 15 minutes. This interval restriction reduces the resolution of data points and hides activity that occurs within this 15-minute window. This is a limitation of the data; however, it also presents an opportunity to evaluate how such a limited dataset can be utilised.

Also noteworthy is that the data interval is only approximately uniform, making many naive signal processing approaches to the data inappropriate. Interpolation is often used on such sequences [8, 17] to create uniformly sampled sequences, however, we avoid that approach here to avoid introducing artifacts that skew behaviour classification.

Latitude and Longitude use the WGS1984 coordinate system. We calculate distances by calculating the great circle distance (arc) on an ellipsoid defined by the WGS84 datum. Positional accuracy depends on the device reporting and the environment but the stated horizontal accuracy is no worse than $\pm$10m.

We examine vessel traffic data between January 2021 and March 2023 in the region between mainland Australia and Tasmania, commonly called the Bass Strait, bounded by latitude -41.2 to 37.7 and longitude 143.6 to 149.3. This is a region of interest because (i) it has a high volume of shipping (Melbourne is a major Australian port, for instance), and (ii) the straight is a potentially challenging maritime environment notorious for treacherous weather, strong currents and large waves.

3.2 Data Preparation

The AIS data is imported from multiple shapefiles and concatenated into a single dataframe. We perform cleaning activities to remove incomplete records and manipulate the data into a format suitable for our analysis and classification task. The data is transformed from a single data point per row to a time-ordered sequence of data points per row for each unique vessel ID. Each row then represents a single vessel's movement over the entire study period (vessel track). We later sample from these rows to generate the segments of motion in this study.

We filter the raw dataset to include only the top eight most common ship types: cargo ship, tanker, passenger ship, pleasure craft, fishing, sailing, tug, and SAR (search and rescue). Vessel types that are sub-classes of these majority types are grouped into the same generic vessel type, and all but the top eight most common types are dropped from the data set.

From the AIS data source, we use latitude, longitude, course, speed, type and timestamp (see Table 1). All other fields are discarded from the dataset. We choose not to interpolate missing values; instead, we drop any vessel tracks that contain N/A values.

Table 1: AIS fields used here.

Field	Range	Units
LATITUDE	-180.0 - +180.0	Decimal degrees
LONGITUDE	-90.0 - +90.0	Decimal degrees
COURSE	0 - 360	Decimal degrees
SPEED	0 - inf	Knots
TIMESTAMP	0 - inf	UTC timestamp
TYPE	Tug, Cargo, Tanker, Passenger, Sailing, Pleasure craft, Fishing or SAR (Search And Rescue)	

3.3 Trajectory Segmentation

We generate short segments of vessel journeys by random sampling (with replacement). Each sequence consists of four consecutive track points. Four points (equating to 45 to 90 minutes of activity) were chosen as the shortest reasonable window for classification. Four data points before and after the sample were also collected to provide context for the manual labelling activity. Each segment has, therefore, a total size of 12 track points, but only the middle four are used in our analysis.

We made no attempt to detect the logical boundary of the segments or ensure that the vessel's motion remains homogeneous within the segment. Samples may encapsulate more than one type of behaviour, and, as a consequence, some mixed-mode segments appear too complex and were therefore discarded as indeterminate during manual labelling.

Each data point is at least 15 minutes apart so four sampled track points represent a minimum time window of 45 minutes. To ensure a relatively regular interval between track points, any sampled segment with an average interval greater than 30 minutes or a standard deviation greater than 10 minutes is excluded. This removes a small number of segments with large irregularities in the time between track points.

Once the vessel trajectory segments have been sampled, we generate a visual image of each vessel's movement overlaid onto a map as shown in Figure 2. The map displays the four sampled track points of the segment and the four surrounding track points on either side of the segment (12 total). Each segment's linearly interpolated path is plotted along with an arrow indicating the vessel's reported course and coloured to indicate its reported speed. These images are used to aid the manual labelling process in 4.4.

4 Methodology

4.1 Feature Generation

We choose to generate statistical features with the goal of finding a set of features that perform well in classifying the motion of the trajectory segments. These are derived from the four middle points, ignoring the context data points surrounding them.

Intermediate Features A set of intermediate vector features are created to aid the creation of our final statistical features. These vector features are values calculated point-by-point based on the points in the base AIS features, that is, features with values that change for each data point of a vessel's journey. They have the same dimension as the source features, in this case, 12 points (4 main + 8 contextual).

We introduce the concept of Direction of Travel (DOT) and Velocity, which are calculated from the interpolated straight line between track points in the segment. Although Speed and Heading values are already reported in AIS data, Velocity and DOT allow us to compare reported values with calculated ones. Moreover, our features are interval averages whereas AIS metrics are point values of the time of measurement.

The following are our intermediate features:

- **Interval**: Difference (in seconds) between the current Timestamp and the previous Timestamp.
- **DOT_Dist**: Distance along the interpolated DOT line between the current Lat/Lon track point and the previous Lat/Lon track point.
- **Velocity**: DOT_Dist divided by Interval.
- **DOT_Difference**: Difference in the direction of travel (DOT) between the current track point and the previous track point.
- **Course_Difference**: Difference between the current reported course and the previously reported course.
- **DOT_Course_Differrence**: Difference in the current direction of travel (DOT) and the current Course.
- **Vel_Speed_Diff**: Difference between Speed and Velocity.
- **DOT_Accel**: Difference of Velocity between the current track point and the previous track point.

We also add a contextual feature called **Dist2Coast**, being the distance to the nearest coastline, using a lookup table provided by NASA [5].

Statistical Summary Features Although AIS data may seem large, modern machine-learning techniques require very, very large datasets. It is plausible that we can alleviate this need by augmenting the data with features that are known to help in classification. Towards determining this we created over 60 statistical summary features, representing each trajectory segment. Some of the features are discussed below to provide illustrative examples, though space prohibits a complete listing.

We create both piece-wise and straight-line summary features. Piece-wise takes each segment point into account, while straight-line features calculate summaries between the first and last data point of a segment. As an example, the feature Piecewise-Distance is the sum of all the distances between successive track points, whereas the Straightline-Distance feature is the distance between the first and the last track point of the segment. Both types of features are useful, in particular, for comparison purposes. For example, differencing the two features above gives a measure of the amount of curvature in the segment.

A selection of the summary features are shown below:

- **Avg_Interval**: Average interval between all track points.
- **Pcw_Std_Course**: Standard deviation of Course over all track points.
- **Pcw_Avg_Accel**: Average acceleration over all track points.
- **Pcw_Max_Speed_Change**: Maximum change in Speed over all track points.
- **Pcw_Avg_Dist2Coast**: Average distance to coast over all track points.
- **Str_Change_In_DOT**: Difference in DOT between the first track point and the last track point.
- **Pcw_vs_Str_Speed_Diff**: Difference between the piecewise average speed and the straightline average speed.
- **Pcw_Avg_Course_vs_DOT**: Difference between the piecewise average course and the piecewise average DOT.

We min-max normalise the values of each feature so that they range between 0 and 1.

4.2 Labelling

Based on the type of motion the vessel is exhibiting, one of four labels is applied to each trajectory segment:

Stopped Trajectory segments are labelled as Stopped if the vessel is not moving. However, GPS noise and minor variations in position suggest that insisting on 0.0 knots is impractical. In many cases, even if the vessel is not actively moving, there is still a small amount of movement due to currents, as shown in Figure 2-a. Following [12], we label a speed less than 0.5 knots as stopped. This includes trajectories where the vessel is moving for some of the time but stationary for at least half of the time.

Straight Trajectory segments are labelled as Straight if the vessel is travelling on a constant trajectory. This can be a straight direction of travel (interpolated) but can also be a curved trajectory, including small turns and gradual curves in motion. There is an allowance for noisy trajectories and trajectories with varying courses provided the vessel is overall travelling in a single general direction.

Turning Turns are a single and abrupt change in direction. Segments are labelled as Turning if the vessel changes from one constant trajectory to another.

Manoeuvring Trajectory segments are labelled as Manoeuvring if the vessel is undergoing multiple changes in trajectory. This occurs when a vessel is negotiating a port or avoiding other traffic but can also occur outside of ports, *e.g.,* in fishing activities.

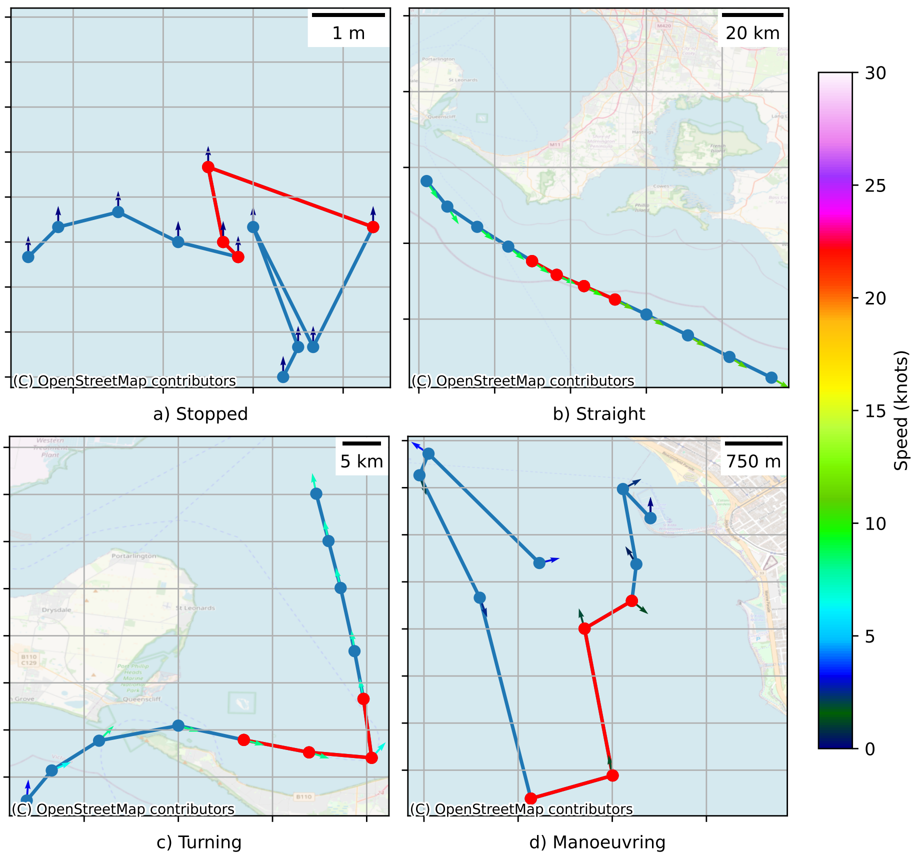

Fig. 2: Examples of the four types of motion. a) Stopped trajectories include vessels that are partially moving and also stopped vessels that are moving slightly due to currents b) Straight trajectories can also include small turns and gradual curves in motion c) Turns are a single and abrupt change in direction d) Manoeuvring segments show multiple abrupt changes in direction.

4.3 Auto-Labelling

To reduce manual effort, we use automatic labelling to generate a large number of labelled samples and decrease the pool of unlabelled samples requiring manual labelling. The auto-labelling is deliberately conservative, *i.e.,* it classifies cases that definitely lie in one category, using an auto-labelling script based on a set of prescribed rules of motion:

- **Straight**: Vessel is travelling within 5 degrees of a straight line (Course and DOT)
- **Stopped**: Speed $<$ 0.5 knots for first 2 data points or Speed $<$ 0.5 knots for the last two data points or Speed $<$ 0.5 knots for all points.
- **Manoeuvring** or **Turning** segments are not automatically labelled.

The labelling rules are intentionally very tight, leaving a large margin on each of the decision boundaries. Auto-labelling only classifies samples that are clearly one category of motion, and the remainder were labelled "Unclassified" for the moment. The goal is to label easily classified segments and leave the samples that remain outside the margins for manual labelling.

4.4 Manual Labelling

Manual labelling was performed on the segment samples that remain outside the margins for auto-labelling. A purpose-built graphical tool was used to present each segment, and the label was applied by a person (a single person performed all manual classification to improve consistency).

The human labeller identified the type of motion the vessel exhibited. A vessel may exhibit multiple types of motion in a single segment sample, and in these cases, the labeller was tasked to identify the most prominent motion. However, if the type of motion is unclear, then the sample is left undefined and excluded from our study. In rare cases, artifacts in the data lead to impossible action, such as a ship travelling over land, and in these cases the segment was marked as abnormal.

4.5 Training Sample Preparation

Many learning techniques, including deep learning, perform better when there is an even distribution of class labels [14]. However, ships spend different proportions of time exhibiting each behaviour category, so our samples were unevenly distributed across the four segment motion types. We randomly undersample the classes that have automatically generated samples (Stopped and Straight) to have the same number as the largest manually labelled class (Turning). Manually labelled samples are taken first with the automatically labelled examples sampled afterwards to top up the number in each class to the required size. Samples in the Manoeuvring class are unchanged. Table 2 shows the final distribution of classes after resampling.

Label	Samples
Turning	1151
Stopped	1151
Straight	1151
Manoeuvring	746

Table 2: Distribution of class labels after resampling. Note that Manoeuvring is underrepresented as these can only be labelled manually.

Data was randomly partitioned into train, test and validation sets in a 64/20/16 split to make the best use of the data for training.

4.6 Feature Selection

The Random Forest classifier provides an ordered list of feature importance based on the Gini Index [11, p. 338]. We utilise this list to select the most significant features as inputs for the neural network classifier. Additionally, we choose a number of additional feature combinations as shown in Table 3. These were aimed at providing a more exhaustive understanding of the value of these features than the Gini Index alone.

Feature Name	Description
All Features	All scalar features
Top 5	Top 5 most important features from the RF Gini Index
Top 10	Top 10 most important features from the RF Gini Index
Top 15	Top 15 most important features from the RF Gini Index
Top 20	Top 20 most important features from the RF Gini Index
Top 25	Top 25 most important features from the RF Gini Index
Top 30	Top 30 most important features from the RF Gini Index
Basic	Simple feature set derived from average/standard deviation of SPEED, COURSE and DISTANCE
Distance	Features derived only from Distance (5 features)
LatLon	Features derived only from AIS Lat and Lon (24 features)
Speed-only	Features derived only from AIS Speed
Speed-Course	Features only derived from AIS Speed and Course
Speed-Vel-Course-DOT	Features derived from Speed, Velocity, Course, DOT and Distance
Comparison Features	Features comparing Straightline vs Piecewise, Speed vs interpolated Velocity and Course vs interpolated Heading

Table 3: Feature sets used as input to the classifiers.

4.7 Classifier

We train three types of classifiers: a simple multi-layer perception (MLP), a Random Forest (RF) classifier, and a k-Nearest Neighbour (k-NN) classifier.

Our neural network is a simple feed-forward classifier consisting of only 3 layers. The first and second layers consist of nodes of size 140 and 300 using the ReLU activation function. The last layer is a softmax to generate a final classification output. The Batch Size hyper-parameter is varied to determine a combination that performs well using the validation dataset. The training epoch is set to 300 and we employ early stopping and 20% dropout for regularisation.

The Random Forest Classifier varies the number of estimators (4, 10, 50, 100, 250, 500, 1000) and max depth (1, 5, 10, 20, 40, 80, 150) and uses randomised cross-validation search to find the best hyper-parameters. Finally, the k-Nearest Neighbours Classifier varies the number of neighbours (1, 4, 8, 20, 50, 100, 150) and also uses randomised cross-validation to find the best hyper-parameters.

4.8 Training

We train each of the three algorithms using the various subsets of feature. Due to the stochastic nature of training, we opt to run each configuration multiple times and calculate results as an average over five test runs.

The training history of the best-performing classifier (MLP), shown in Figure 3, indicates a typical learning curve. The classifier trains until the validation

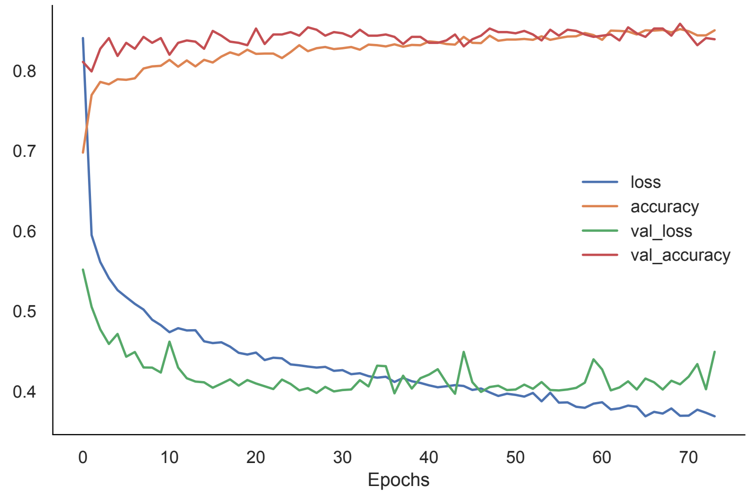

Fig. 3: Training history of the best-performing classifier - A MLP using the top 20 Gini features with a batch size of 32. Early stopping stops training at 43 epochs.

loss (green) stops improving, using early stopping to prevent overfitting. Both the validation (red) and training (orange) accuracy approaches a limit confirming that further training will not greatly improve performance.

4.9 Metrics

We choose to use a weighted-average F1-Score to measure the performance of the classifiers. The F1-score is the harmonic mean of precision and recall. F1-score reaches its best value at 1 (perfect precision and recall) and worst at 0.

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

We use a normalised confusion matrix (Figure 4) to provide an understanding of how class affects the results. In our confusion matrix, each cell contains three values. The middle value represents the count of accurate predictions for a specific class. The top-left value shows the percentage of accurate predictions within that column, offering insight into precision ($TP/(TP+FP)$). The bottom-right value displays the percentage of accurate predictions within that row, providing information about recall ($TP/(TP+FN)$). This enables us to visualise precision and recall on a per-class basis.

5 Results

5.1 Features

All features were evaluated using the Gini Impurity score [11, p. 338] calculated from the best random forest classifier. The Gini score indicates the order of importance of each feature, and we used it to find the most important features for our neural network. Table 4 shows the ten most important features.

Gini Importance	Description
0.0792	Distance travelled - distance between first and last track points
0.0580	The average distance travelled between segments
0.0562	The total distance travelled over all segments
0.0480	The average change in direction over all segments
0.0439	The cumulative change in direction over all segments
0.0364	The average velocity over all segments
0.0348	The number of stops $velocity < 0.5knots$
0.0336	The standard deviation of the direction of travel over all segments
0.0326	The average velocity between the first track point and the last
0.0325	The standard deviation of the course over all segments

Table 4: Top 10 most important features as ordered by the Gini impurity. Detailed description of these features can be found in the technical paper *(ref redacted for review).*

The most important feature captures the difference in distances between the sum of the linear interpolated distances between each point in the segment and the straight line ‘as the crow flies’ distance between the first point and the last point. This intuitively captures the curvature of the trajectory, which explains why it is considered the most important.

The second and third most important features are the average and total distance travelled between segments, both indications of the speed of the vessel. The fourth and fifth most important features measure the amount of direction change (in either direction), indicating how much turning is occurring. The seventh is a count of the number of stops, which indicates how often the vessel is stationary. It can be seen that all features add a different statistical perspective on the vessel’s motion, aiding in the differentiation of the four motion types.

5.2 Evaluation

The MLP performed slightly better than the RF and much better than the k-NN classifier. The best-performing classifier was an MLP using the top 20 Gini features with a batch size of 32, which early stopped at 43 epochs (Table 5). It returned a weighted-average F1-Score of 83.6%. The best RF classifier achieved an average F1 score of 82.9%, and the best-performing K-NN classifier returned a weighted average F1 score of 80.4%.

Examining the confusion matrix in Figure 4 shows that the MLP performs well in classifying stopped vessels. 95% of stopped vessels are correctly predicted, and 94% of the predicted stopped vessels are actually stopped. Straight and Turning vessels are classified reasonably well; however, Manoeuvring vessels are not classified as well as the other types of motion. Random Forest and K-NN classifiers offer similarly distributed but slightly worse results.

Type	Feature Set	F1-Score
1st MLP	TOP 20 (batch size 32)	83.6%
2nd MLP	TOP 15	83.6%
3rd MLP	TOP 20 (batch size 128)	83.5%
Best RF	All Features	82.9%
Best K-NN	TOP 10	80.4%

Table 5: Classifier Results showing the top 3 best-performing classifiers and the best-performing Random Forest and k-Nearest Neighbour classifiers.

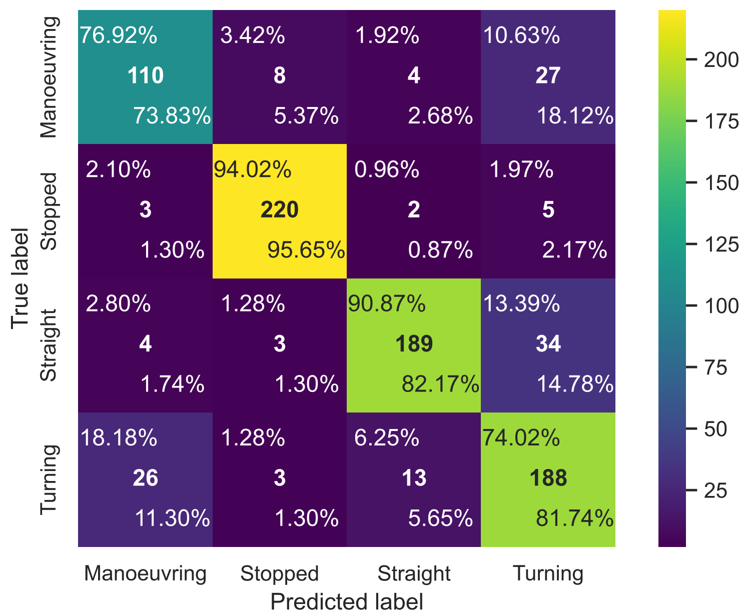

Fig. 4: Confusion matrix for the best-performing classifier (MLP). Centre values are a count of categorised instances. The top left value in each cell displays the percentage of instances normalised along the vertical predicted value axis, and the value in the bottom right cell displays the percentage of instances normalised along the horizontal true value axis.

6 Conclusion

In this work, we classify the motion of maritime vessels in short segments from AIS data. We avoid the common rule-based and unsupervised techniques and attempt to capture the more subtle human interpretation of each vessel's motion by manually labelling the AIS dataset. The results show that simple summary statistics can be used as features to classify the type of motion of a maritime vessel. The best results were obtained with an MLP classifier using the top 20 most important statistical features, returning an F1-score averaging 83.6%. The confusion matrix indicates how well each of the motion types are independently classified. For some motion types (Stopped), the precision/recall is as high as 95%.

Future research should focus on enhancing the quality of the data and/or the classifier. It would be beneficial to conduct further studies using a higher-quality, larger, evenly distributed dataset with smaller intervals between data points. To improve the results, it could be useful to train multiple classifiers, such as one for each ship type or motion type, and then combine them using ensemble methods.

Acknowledgements

This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

References

1. Access to AIS data - Norwegian Coastal Administration. `https://www.kystverket.no/en/navigation-and-monitoring/ais/access-to-ais-data/`
2. Danish Maritime Authority - AIS data. `https://www.dma.dk/safety-at-sea/navigational-information/ais-data`
3. Marine traffic - Open data from Finnish waterways. `https://www.digitraffic.fi/en/marine-traffic/`
4. MarineCadastre.gov - Vessel Traffic Data. `https://marinecadastre.gov/ais/`
5. NASA Ocean Color. `https://oceancolor.gsfc.nasa.gov/resources/docs/distfromcoast/`
6. Spatial@AMSA - Vessel Tracking Data. `https://www.operations.amsa.gov.au/Spatial/DataServices/DigitalData`
7. Amigo, D., Pedroche, D.S., García, J., Molina, J.M.: Segmentation optimization in trajectory-based ship classification. Journal of Computational Science **59**, 101568 (Mar 2022). https://doi.org/10.1016/j.jocs.2022.101568
8. Capobianco, S., Millefiori, L.M., Forti, N., Braca, P., Willett, P.: Deep Learning Methods for Vessel Trajectory Prediction based on Recurrent Neural Networks. IEEE Transactions on Aerospace and Electronic Systems **57**(6), 4329–4346 (Dec 2021). https://doi.org/10.1109/TAES.2021.3096873
9. Eljabu, L., Etemad, M.: Anomaly Detection in Maritime Domain Based on Spatio-Temporal Analysis of AIS Data Using Graph Neural Networks (Dec 2021). https://doi.org/10.1109/ICVISP54630.2021.00033
10. Ginoulhac, R., Barbaresco, F., Schneider, J.Y., Pannier, J.M., Savary, S.: Target Classification Based On Kinematic Data From AIS/ADS-B, Using Statistical Features Extraction and Boosting. 2019 20th International Radar Symposium (IRS) pp. 1–10 (Jun 2019). https://doi.org/10.23919/IRS.2019.8768094
11. James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J.: An Introduction to Statistical Learning: With Applications in Python. Springer Texts in Statistics, Springer International Publishing, Cham (2023). https://doi.org/10.1007/978-3-031-38747-0
12. Kraus, P., Mohrdieck, C., Schwenker, F.: Ship classification based on trajectory data with machine-learning methods. In: 2018 19th International Radar Symposium (IRS). pp. 1–10 (Jun 2018). https://doi.org/10.23919/IRS.2018.8448028

13. Liang, M., Su, J., Liu, R.W., Lam, J.S.L.: AISClean: AIS data-driven vessel trajectory reconstruction under uncertain conditions. Ocean Engineering **306**, 117987 (Aug 2024). https://doi.org/10.1016/j.oceaneng.2024.117987
14. Rezvani, S., Wang, X.: A broad review on class imbalance learning techniques. Applied Soft Computing **143**, 110415 (Aug 2023). https://doi.org/10.1016/j.asoc.2023.110415
15. Ribeiro, C.V., Paes, A., de Oliveira, D.: AIS-based maritime anomaly traffic detection: A review. Expert Systems with Applications **231**, 120561 (Nov 2023). https://doi.org/10.1016/j.eswa.2023.120561
16. Riveiro, M., Pallotta, G., Vespe, M.: Maritime anomaly detection: A review. WIREs Data Mining and Knowledge Discovery **8**(5), e1266 (2018). https://doi.org/10.1002/widm.1266
17. Sang, L.z., Wall, A., Mao, Z., Yan, X.p., Wang, J.: A novel method for restoring the trajectory of the inland waterway ship by using AIS data. Ocean Engineering **110**, 183–194 (Dec 2015). https://doi.org/10.1016/j.oceaneng.2015.10.021
18. Sheng, K., Liu, Z., Zhou, D., He, A., Feng, C.: Research on Ship Classification Based on Trajectory Features. The Journal of Navigation **71**(1), 100–116 (Jan 2018). https://doi.org/10.1017/S0373463317000546
19. Soares Junior, A., Cesario Times, V., Renso, C., Matwin, S., Cabral, L.A.: A Semi-Supervised Approach for the Semantic Segmentation of Trajectories. In: 2018 19th IEEE International Conference on Mobile Data Management (MDM). pp. 145–154. IEEE, Aalborg, Denmark (Jun 2018). https://doi.org/10.1109/MDM.2018.00031
20. Spaccapietra, S., Parent, C., Damiani, M.L., De Macedo, J.A., Porto, F., Vangenot, C.: A conceptual view on trajectories. Data & Knowledge Engineering **65**(1), 126–146 (Apr 2008). https://doi.org/10.1016/j.datak.2007.10.008
21. 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 Transactions on Intelligent Transportation Systems **19**(5), 1559–1582 (May 2018). https://doi.org/10.1109/TITS.2017.2724551
22. Wolsing, K., Roepert, L., Bauer, J., Wehrle, K.: Anomaly Detection in Maritime AIS Tracks: A Review of Recent Approaches. Journal of Marine Science and Engineering **10**(1), 112 (Jan 2022). https://doi.org/10.3390/jmse10010112
23. Yang, Y., Liu, Y., Li, G., Zhang, Z., Liu, Y.: Harnessing the power of Machine learning for AIS Data-Driven maritime Research: A comprehensive review. Transportation Research Part E: Logistics and Transportation Review **183**, 103426 (Mar 2024). https://doi.org/10.1016/j.tre.2024.103426
24. Zhang, L., Meng, Q., Xiao, Z., Fu, X.: A novel ship trajectory reconstruction approach using AIS data. Ocean Engineering **159**, 165–174 (Jul 2018). https://doi.org/10.1016/j.oceaneng.2018.03.085
25. Zhang, Y., Wen, Y., Tu, H.: A Method for Ship Route Planning Fusing the Ant Colony Algorithm and the A* Search Algorithm. IEEE Access **PP**, 1–1 (Jan 2023). https://doi.org/10.1109/ACCESS.2023.3243810
26. Zhang, Y., Ma, Y., Liu, J.: Ship trajectory segmentation and semisupervised clustering via geospatial background knowledge. Ocean Engineering **304**, 117872 (Jul 2024). https://doi.org/10.1016/j.oceaneng.2024.117872