Research Article

Zahra Sadeghi* and Stan Matwin

Anomaly detection for maritime navigation based on probability density function of error of reconstruction

https://doi.org/10.1515/jisys-2022-0270
received November 21, 2022; accepted July 14, 2023

Abstract: Anomaly detection is a fundamental problem in data science and is one of the highly studied topics in machine learning. This problem has been addressed in different contexts and domains. This article investigates anomalous data within time series data in the maritime sector. Since there is no annotated dataset for this purpose, in this study, we apply an unsupervised approach. Our method benefits from the unsupervised learning feature of autoencoders. We utilize the reconstruction error as a signal for anomaly detection. For this purpose, we estimate the probability density function of the reconstruction error and find different levels of abnormality based on statistical attributes of the density of error. Our results demonstrate the effectiveness of this approach for localizing irregular patterns in the trajectory of vessel movements.

Keywords: anomaly detection, time series trajectories, deep learning, autoencoder, probability density function

1 Introduction

Anomaly detection is one of the classical problems in machine learning. This problem has been defined and explored in a multitude of different contexts, such as visual data [1], audio/video files [2,3], and EEG signals [4]. The goal is to detect patterns of data that do not conform to the general trend in the data. This problem is closely related to outlier detection. Outlier removal is considered as one of the key components of data cleaning and data preprocessing [5]. One historic approach for unsupervised detection of anomalies in time series data is based on clustering methods in which anomalies are considered as samples that do not fit in the formed clusters. In this regard, there is a large pool of related literature that employ clustering algorithms such as k-means [6], density-based spatial clustering of applications with noise (DBSCAN) [7], and hierarchical clustering methods [8,9]. A multi-scale learning approach based on the clustering trajectory of data is proposed in ref. [10]. One obvious disadvantage of clustering-based approach is that, it is not directly developed and optimized for anomaly detection and depends on hyperparameter tuning. In these methods, cluster shapes and the number of clusters can directly influence the detected anomalies. Another highly applied method for unsupervised identification of anomaly samples in time series data is based on forecasting algorithms using autoregression methods such as autoregressive integrated moving average (ARIMA) [11] and seasonal autoregressive integrated moving average [12]. These approaches pursue the idea that the data points that cannot be estimated accurately are more likely to be anomalous. One potential drawback of these methods is that they perform poorly for forecasting long trajectories and require parameter selection. There are also approaches

*** Corresponding author: Zahra Sadeghi,** Faculty of Computer Science, Institute for Big Data Analytics, Dalhousie University, Halifax, B3H 1W5, Canada, e-mail: zahras@dal.ca
Stan Matwin: Faculty of Computer Science, Institute for Big Data Analytics, Dalhousie University, Halifax, B3H 1W5, Canada, e-mail: stan@cs.dal.ca

based on linear regression modeling in which outliers are detected based on standardized residuals [13]. Bianco et al. studied outlier detection in regression models by proposing a robust estimation for ARIMA parameters [14]. In the study by Yuen and Mu [15], a robust outlier detection method is proposed based on considering the optimal regression parameters, residuals, and outlier probability calculation. This method is later extended by the same authors, where the method is improved by considering the correlation between residuals [16]. A stable real-time cleansing method was proposed by Mu et al. [17] based on Extended Kalman Filter (EKF), which can be used for abnormal data removal in real time.

In this work, we are concerned with analyzing the behavioral patterns of vessels' trajectories in order to detect anomalous data points. Anomaly detection of maritime data is important to support safe travel and manage sea traffic [18]. More specifically, we develop a method for anomaly detection in the maritime domain using a public Automatic Identification System (AIS) dataset. AIS is a system that is developed for automatically tracking vessels and monitoring traffic over oceans. Each vessel is equipped with a transmitter that must be turned on all the time in order to broadcast AIS signals about their status to other vessels to ease navigation and prevent collisions. The AIS messages include static and dynamic information about vessel identity, position, speed, and course [19]. However, this dataset is not annotated, and there is no supervised information about the anomalous patterns. Therefore, an additional problem is the unknowability of the training and test data. Hence, in our approach, instead of categorizing data into anomalous and nonanomalous patterns, we propose a method for creating different Confidence Levels (CLs) of irregularity and introduce a formula for measuring the anomaly score. Our anomaly detection method is based on Autoencoders (AEs). A notable attribute of AEs is that they can be trained without requiring labeled data. However, one common approach to applying AEs for anomaly detection is a supervised learning strategy. In a supervised-learning approach, an AE is trained on normal data and then tested with test data. This approach is also known as a one-class classification problem [20]. The reconstruction loss is then leveraged to detect anomalous patterns. If the error of reconstruction is high for a sample, then the sample will be considered as an anomalous data. Xia et al. [21] followed this idea but only trained an AE on positive anomalous data and then analyzed the reconstruction error using a clustering mechanism. Zhou and Paffenroth [22] in their study have proposed a supervised algorithm for outlier and anomaly detection by developing a robust AE-based technique that computationally relies on Robust Principal Component Analysis (RPCA). However, all of these methods require annotated anomalous data. A number of researchers have made an attempt at softly labeling data before applying AE. For instance, Li et al. [23] first applied clustering to the data in order to estimate and collect normal data and then trained an AE on the normal data. The second approach for anomaly detection using AE benefits from the nonlinear feature space resulting from training an AE model. More specifically, AEs are capable of projecting data into a low-dimensional space with better separability features and thus provide the possibility of applying different clustering techniques to the projected data for outlier identification [24]. Here, we provide a fully unsupervised method for detecting and localizing abnormal behavior in serial data. In general, the unsupervised approach to anomaly detection relies mostly on thresholding the error. However, there is no efficient method for finding a proper cutoff value for a threshold. Often, the mean error value is leveraged as a fair threshold value. But depending on the distribution of data and the task in demand, the mean value could give rise to misleading results. To alleviate this issue, we propose a method for finding thresholds at certain CLs by estimating the distribution of the reconstruction error obtained from training an AE. The encoder–decoder architecture of this network yields an efficient data representation. The reconstruction error encompasses information about nonconformities in a series of data. Hence, the probability density of error can provide insight about the rare events in a trajectory. In this study, we identify the irregular points by computing CLs of anomaly. The rest of this article is organized as follows: Section 2 explores prior art and related work. Section 3 introduces the proposed method. In Section 4, we present the experimental results. Finally, we conclude in Section 5.

2 Related work

Anomaly detection and outlier identification are two closely related problems. In general, outlier detection is considered a preprocessing stage for data cleansing and noise removal [25], whereas anomaly detection is

regarded as a technique for finding unexpected and abnormal behavior. Anomaly detection is correlated with salient pattern detection, i.e., localizing and identifying the sources of nonconformity and conspicuous patterns. The primary goal is to detect rare and unexpected patterns. Anomaly detection has a broad range of applications in various domains. These applications include fraud detection [26], fault detection [27], theft detection [28], intrusion detection [29], and abuse detection [30]. In this work, our focus is on detecting anomalous behavior from maritime vessel navigation. However, there are various anomalous situations that can occur in the maritime environment. Port arrival time [31], entering restricted zones [32], close proximity between vessels [33], route deviation [34], AIS off-on switching [35], and silent period in message transmission [35] are the most studied problems. Our work can be categorized as a route deviation method since we look at the behavior of vessels of the same type. Maritime anomaly detection has been investigated as maritime situational awareness using machine learning models based on feature engineering for a long time. To this end, various classification and clustering methods are proposed. Zhen et al. first clustered the normal trajectories and then measured the posterior probability of assigning a new trajectory into one of the clusters using the Bayesian rule [36]. Another Bayesian-based approach is proposed by Mascaro et al. using Bayesian network learners [37]. Support Vector Machine (SVM) has also been applied for transforming data into a high-dimensional feature space and grouping data behavior into normal and abnormal classes [38]. In the study by De Vries and Van Someren [39], trajectories are first compressed, then a one-class SVM model is applied. Unsupervised learning based on grouping- or clustering-based anomaly detection is another category of traditional approaches that have been widely investigated. Murray and Perera, in their study [40], applied a Gaussian Mixture Model (GMM) to cluster data and exploited the trace of covariance matrices to threshold them in order to identify anomalous clusters. The Self-Organizing Map (SOM) is another classical technique that has been applied to model clusters of normal and anomalous points [41]. A spline-based trajectory clustering approach is presented for detecting anomalous events in a coastal surveillance scenario [42]. The Gaussian Process (GP) has been exploited for anomaly detection by predicting vessel positions [43]. As a matter of fact, finding abnormal patterns in time series data is challenging because of the conditional dependency between subsequent time points. In addition, due to various weather conditions and unpredictability and complexity of maritime navigation, marine vessels do not follow a predefined road map, and the pattern of movements of similar vessel types varies greatly. Thus, it is very difficult to apply traditional similarity-based and regular data mining algorithms [44] to vessel movements to analyze their activities, and thus, traditional machine learning algorithms are not capable of finding efficient answers. Therefore, many researchers have applied Deep Neural Networks (DNNs) to vessel trajectories for handling different problems, such as vessel trajectory prediction [45,46], collision avoidance [47], movement or trajectory classification [48,49] and traffic prediction [50]. In the context of anomaly detection, DNNs such as Recurrent Neural Network (RNN) [51], Long-Short-Term Memory (LSTM) [52], and AEs [53] have been adopted and applied by several research groups. There are also a few attempts for combining different methods. For instance, in the study by Zhao and Shi [51], an RNN with an LSTM unit is trained on the clustered trajectories using the DBSCAN method. An anomaly detector is proposed in the study by Nguyen et al. [54], which exploits variational RNN architecture to build the normal representation of movement patterns and then applies the geo-spatial information of AIS messages to find abnormal data. A behavior-based anomaly detection approach is also proposed in the study by Blauwkamp et al. [55], where images of the trajectory of vessel movements are created and then used for training a DNN. Nevertheless, there are still limited research on applying deep learning methods to a time series of AIS messages. Essentially, two types of anomaly detection can be considered: (1) contextual anomaly detection and (2) behavioral anomaly detection. Contextual anomaly detection is about anomalous data that depend on time [56]. In this regard, the same pattern can be determined as anomalous at one particular time and nonanomalous at another time based on the context, which is usually provided by meta data. In contrast, behavioral anomaly detection is totally dependent on the geometric and kinematic behavior of trajectories, regardless of the time of occurrence. This problem can be assumed as temporal anomaly detection as the objective is to find the time that the typical pattern of the sequence changes. In contrast to some of the domains in which the time series data can be categorized into known stages [57], the trajectory patterns of movement of vessels in the maritime environment are largely unpredictable, which leads to a lack of annotated data in this domain. In essence, the anomalous patterns of vessel movements are diverse and variable and do not generally represent

a well-defined signature. Hence, in this domain, the anomalous behavior can be grounded on changes in the movement patterns of each vessel, or they can be decided in a collective manner depending on the trajectories taken by other similar samples from the same vessel-type category [58]. As aforementioned, another challenge in regard to the task of anomaly detection in the maritime domain is related to the lack of labeled public datasets. While it is very intuitive for humans to identify anomalies, solving this problem with machine learning techniques relies on a trained model with an effective feature representation that can distinguish anomalous data points. In order to accomplish this purpose, a common approach is to train a model on an annotated dataset for learning the distinction between groups of regular and irregular data, and therefore, typically, anomaly detection problem has been addressed in a supervised setting. However, labeling sequence data for the task of anomaly detection is an expensive manual task. Not only does it take a huge human effort to annotate each time step in a long data sequence, but it is also not feasible due to different anomalous scenarios in the maritime environment. One issue with anomaly annotation deals with the abundance of unknown and undefined anomalous events. Therefore, there is no annotated AIS dataset for anomaly identification that is publicly available for research. Consequently, machine learning researchers have endeavored to attack AIS anomaly detection in a semi-supervised and unsupervised fashion. The semi-supervised approach rests on the assumption that normal data are known and available [59]. Hence, even though it requires less annotated data, it still relies on labeled normal data. Rhodes et al. proposed a method to measure vessel route deviation according to Hebbian learning [60]. This technique requires a large history of trajectories to accurately predict anomalous data. The aforementioned methods suffer from requiring labeled data or using a feature engineering mechanism for modeling the data. This article uses an unsupervised technique based on deep learning to overcome these issues.

3 Method

As mentioned earlier, our method is based on an unsupervised feature learning procedure by training AE models. In this procedure, unlabeled data x is first compressed into a low-dimensional feature representation h. Then, the input data are reconstructed into x' by decompressing the hidden representation and mapping it to its original dimensional space. The reconstruction error $|x - x'|$ is the loss value for training the network, which needs to be minimized. Here, we train an AE on the AIS dataset and exploit the Speed Over Ground (SOG) feature. This feature consists of sequential information about the speed of vessels at each time step. The sequence of SOG is then divided uniformly into fixed-sized segments and treated as a sample. In mathematical terms, having a sequence data $s = x_1, \ldots, x_n$, we divide it into equal sequences of size m and treat each segment as a sample. We then train a convolutional AE in a fully unsupervised setting. Following the study by Xia et al. [21], we treat the reconstruction error as a signal for anomaly detection. We expect to obtain a lower error for the sequences that are less likely to comply with the frequent patterns. In contrast to previous methods, we use the Probability Density Function (PDF) of error for localizing the anomalous behavior in the time series of movement data.

3.1 PDF of error

In order to analyze patterns of vessel movement and detect anomalies, we estimate the PDF of error. The density estimation can be conducted in a parametric or nonparametric manner. The fundamental assumption of parametric approaches is that the underlying distribution is known, and so the emphasis is on finding the proper parameters of the distribution function. For instance, one of the popular methods for density estimation is a mixture of Gaussian functions known as GMM [61]. This method is developed based on the assumption that the underlying data are generated from Gaussian functions with different parameters. Obviously, this method and other parametric algorithms cannot be appropriate for cases – such as ours – in which there is no information about the source of data. Therefore, in our approach, we apply a nonparametric method for

modeling the underling distribution of error. Nonparametric approaches are free of parameters in the sense that they attempt to fit a model to data without any strict assumption about the specific distribution functions. Histograms are known as the simplest method in this category. Each partition of histogram indicates the number of samples that fall within a certain interval. However, histograms are unable to provide a smooth and continuous density estimation. For this purpose, we leverage Kernel Density Estimation (KDE) at each data point x, as shown by equation (1):

$$p_n(x) = \frac{1}{n\delta}\sum_{i=1}^{n} K(x, X_i;\ \delta). \tag{1}$$

In contrast to parametric methods, KDE deals with density estimation without making any particular assumptions about the probability distribution from which the data is sampled. The smoothness of this method is adjusted by bandwidth size δ. The lower values of this parameter result in a smoother estimation. In our implementation, we always fix it to be one-fortieth of the length of the error signal. We have tested the results with three different kernels, namely, Gaussian, Epanechnikov, and Tophat kernels, as shown in equations (2)–(4). By estimating the probability distribution of the error signal, we can elicit statistical information about how the density of error behaves and decide about the likelihood degree of each data point, which can lead to anomaly CLs which is explained in the next Section 3.2.

$$K(x, z;\ \delta) \propto \exp(-||x - z||^2)/2\delta^2 \tag{2}$$

$$K(x, z;\ \delta) \propto 1 - ||x - z||^2/\delta^2 \tag{3}$$

$$K(x, z;\ \delta) \propto 1 \quad \text{if } ||x - z|| \leq \delta/2 \text{ else } 0. \tag{4}$$

3.2 Anomaly Confidence Level

Anomaly detection methods generally depend on the selection of a threshold value. In other words, an anomaly threshold specification is required to discriminate between anomalous and non-anomalous patterns. The threshold is considered as a cutoff point for the binary categorization of data. In fact, the primary challenge of unsupervised anomaly detection is threshold selection. In practice, a threshold is chosen based on the desired performance and domain knowledge. There are several thresholding techniques that are based on measurements of standard deviation [62], median absolute deviation [63], and clever standard deviation [64]. However, these methods are biased and produce imprecise results. In our approach, for detecting anomalous patterns in an unsupervised manner, a threshold needs to be applied to the reconstruction error. To this end, we propose a density-based approach for threshold specification. The simplest method for density estimation is a histogram. In a histogram-based approach, the partitions at higher error intervals can be attributed to irregular data. Here, we propose a thresholding approach based on the calculation of a continuous estimation of the distribution of error. Having estimated the PDF of error of reconstruction with KDE (as explained in Section 3.1), we set the threshold based on a specified anomaly Confidence Level (CL). We define CLs by calculating the percentile on the PDF of error. The nth percentile indicates a threshold below which n percent of error of the whole population lie. Hence, by n% anomaly CL, we refer to the nth percentile in the sorted distribution of reconstruction error, which specifies a certainty degree for capturing anomalies. This CL refers to the degree of uncertainty of anomalies' severity and is a user-defined or task-based parameter. This degree can be specified based on the purpose and/or context and can vary in an interval of (0, 100), which controls the sensitivity for capturing abnormal patterns. The higher the CL, the lower the number of anomalous samples that the model can capture. This feature provides high flexibility over the number of anomalies that can be caught by the algorithm. An advantage of this method over standard deviation-based techniques is the fact that this calculation is more robust and not based on the assumption of a normal distribution of data. With this approach, we can adjust the irregularity level of anomalies that we are interested to detect. Hence, we can find an appropriate threshold by controlling the CL of anomalies that we aim to detect.

3.3 Anomaly score

In Section 3.2, we focused on a threshold value that can result in a crisp labeling of data into anomalous and non-anomalous patterns. Here, we address the question of how we can associate a continuous value to each data, which demonstrates the deviation from normality. More specifically, instead of seeking a threshold value to categorize data into two distinct groups of anomalous and nonanomalous patterns, we propose a method for measuring the anomaly score. This way, we can provide a score to each data point, which shows the anomalous degree or the likelihood of a sample to be categorized as an anomaly. In this article, we propose to use the idea of the cumulative distribution function as a way to compute this score by using the equation (5), where e is the PDF of error, which is estimated by applying KDE according to equation (1). $e_{\min}$ and $e_{\max}$ are the minimum and maximum errors associated with a data point in the error signal of each trajectory sample corresponding to one vessel movement. The anomaly score $A(x)$ ranges between 0 and 1 and specifies the probability that the error is less than or equal to x. With this equation, we link between error of reconstruction and anomality. In other words, by specifying the error of each data point, which is resulted from the trained AE on the vessel movements information, we can predict the level of irregularity in data. Hence, for a given data point and its associated error value, we can compute the probability of finding lower error values. When error is the minimum, the anomaly score is close to 0, and when it increases, the anomaly score monotonically goes up until it converges to 1. It can also be implied that by taking the derivative of anomaly score function, we can obtain the PDF of error, hence the changes in anomaly score are according to the PDF of error function values.

$$\begin{aligned} A(x) &= \int_{e_{\min}}^{x} e(u)\mathrm{d}u \\ 0 \le A(x) &\leqslant 1 \\ \lim_{x \to e_{\min}} A(x) &= 0 \\ \lim_{x \to e_{\max}} A(x) &= 1. \end{aligned} \tag{5}$$

4 Experimental results

For all the experiments, we used the AIS dataset for 2020, which is publicly available for download[1]. This dataset contains both static and dynamic attributes about vessels' identity, location, and movement [65]. The dynamic attributes include latitude, longitude, SOG and course over ground features. For the purpose of this research, we focus on the SOG dimension. This feature contains the footprint about speed of vessels at each time-step and therefore encompasses movement patterns. We rely on this feature for the anomaly analysis of vessels' movement. We collect maritime trajectories of all tankers corresponding to vessel IDs 80–89. We first normalize the dataset and process it to eliminate unknown values. The size of trajectory for each vessel is not equal, and therefore, in order to feed the data into the network, we segment each trajectory into segments of equal length. Data are then divided into training and test sets with a split ratio of 80:20, and all trajectories with length less than 10,000 are removed. We developed an AE with four hidden layers. For the architecture of the AE model, we used a simple structure consisting of two convolutional layers in the encoder and two transposed convolutional layers in the decoder. The number of nodes in each layer is 32, 16, 32, and 16, respectively. Kernel sizes are 7 and ReLU activation functions are applied to the output of convolution layers. DNN training includes several hyperparameters that control the learning and training process. Adjusting the values of these hyperparameters can lead to an increases in the performance of the trained models. However, there is no analytical method for hyperparameter configuration. Hyperparameter optimization is an approach

1 https://coast.noaa.gov/htdata/CMSP/AISDataHandler/2020/index.html.

to search for the best values in the entire space of feasible values. In this research, we do not focus on optimizing these values, and hence, hyperparameter tuning for training deep network is performed in a linear search fashion. For learning rate, we tried the values in the range of $[0.1, 10^{-6}]$ with a step size of 0.1 and chose 0.001. For batch size, the values within the set 64,128,256 are tried, and finally, batch size of 128 is used for training all the networks. We also checked Adam and RMSProp and opted for Adam. We trained three AEs that are similar in structure but different in segment sizes. Mean absolute error (MAE) is used as the loss function for training the deep network. The idea of anomaly detection based on the error of AE is depicted in Figure 1. As can be understood, the anomalous patterns generate higher error values. The learning curves that include training and validation loss values over the course of learning are shown in Figure 2. We then estimate the density of error. As previously mentioned, a histogram cannot provide a smooth and accurate estimation. Hence, KDE with three different kernels, i.e., Gaussian, Epanechnikov, and Tophat kernels, is modeled and fitted to the error as shown in Figure 3. After applying KDE, the threshold values are calculated for four different CLs. The CLs for the KDEs are obtained based on the idea of percentile and are exactly computed based on the distribution of data. Four levels of thresholds are marked in each distribution. For histogram, threshold levels can be selected based on bin size, which need to be manually decided, whereas in our approach for KDE distributions, the thresholds are computed based on their CLs. Figure 4(a) and (b) presents the estimated distribution of error by histogram and KDE methods based on the idea of anomaly CLs. The dotted line specifies the cutoff value. All samples on the right side of this line correspond to anomaly, whereas the samples on the left side are normal samples. The impact of each CL on the original error signal is indicated in Figure 5. As can be observed, the higher the level of confidence, the lower the number of anomalous patterns that can be captured. The anomaly score that links the error value with an anomaly degree in a continuous manner is presented in Figure 6. These scores are specified with dashed lines and compared for different PDF estimations, as explained in previous sections. As mentioned before, AIS dataset does not include labels for anomaly detection problem. Therefore, in order to assess the performance of our algorithm quantitatively, we create pseudo labels for the dataset. For this purpose, we make time series stationary by utilizing the difference mechanism [66]. Then, we convert each time series into a binary vector by employing a cutoff value. This value serves as an approximate estimation of aberrant fluctuations in data and is arbitrarily chosen as 0.1 in our experiments. We have also tried two other values of 0.1 and 0.2 in our implementation and did not find a significant effect on the results. Our approach is not dependent on this value, and it is only used for the sake of numerical evaluation. Figure 7 demonstrates the predicted labels at different levels of confidence according to the pseudo labels, which are shown in Figure 8. As evident from the results, the number of correct anomalous points increases when we try a lower CL.

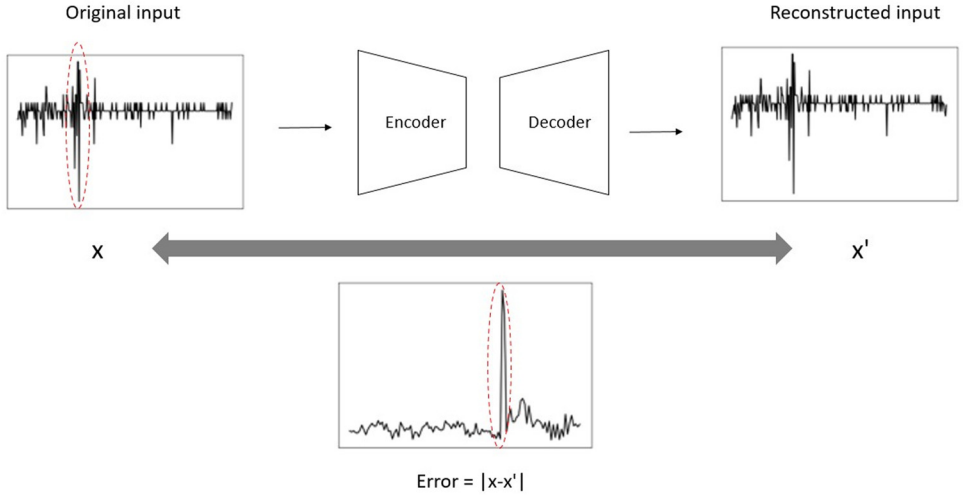


Figure 1: Anomaly detection with AEs.

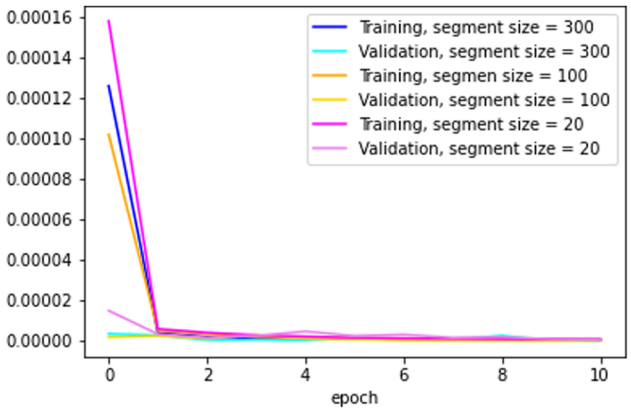
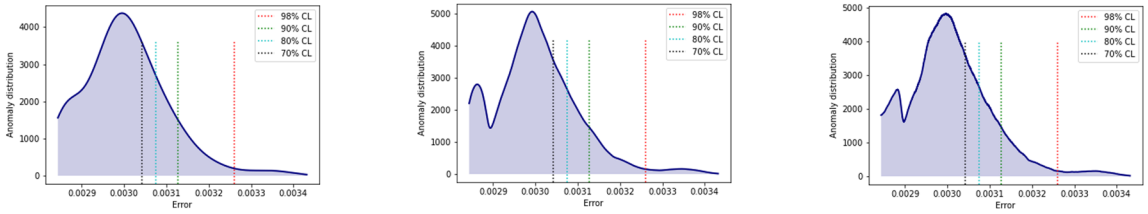
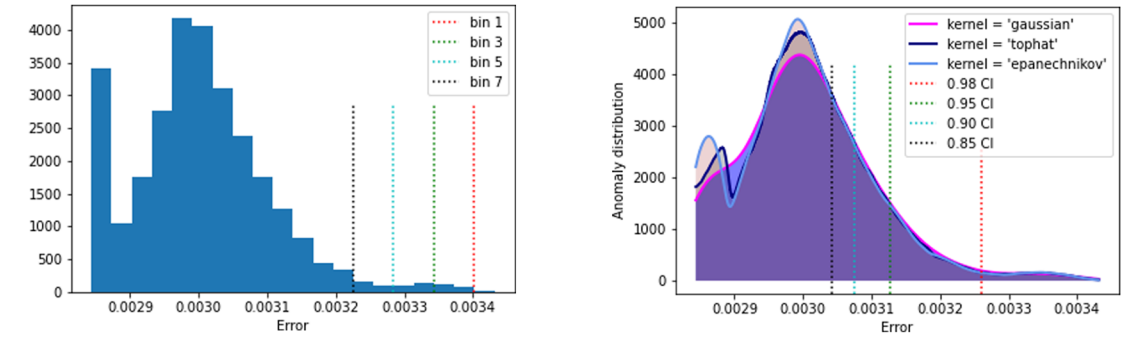


Figure 2: Learning curves for three trained AEs.



(a) (b) (c)

Figure 3: Kernels for density estimation: (a) Epanechnikov kernel, (b) Gaussian kernel, and (c) Tophat kernel.



(a) (b)

Figure 4: Anomaly CLs computation on distribution of error. The dashed lines indicate threshold levels. Each line divides samples into anomalous (on the right-hand side) and nonanomalous (on the left-hand side) samples. (a) Histogram and (b) KDE.

We exploited six metrics, i.e., accuracy, sensitivity, specificity, precision, recall, and f1-measure, for evaluating the results. In fact, sensitivity and recall are the same criterion. Since we did not observe a significant difference with different kernel methods and the results for all three kernels are very similar, we only provide the performance obtained from the Gaussian kernel. Figure 9 outlines the evaluated results on test trajectories in bar charts. We show the results obtained from four CLs evaluated by pseudo labels with a threshold of 0.1 for three segment sizes of 20, 100, and 300. For the sake of comparison, Tables 1 and 2 provide the numerical values organized in two measurement sets. All the measurements are compared with baseline in which the differentiation threshold between anomalous and normal points is considered according to mean error. In addition, we provide the results for three different pseudo labels at threshold values of 0.1, 0.2, and 0.3. As depicted in Figure 10, the parameter of pseudo label creation does not affect the performance of our method

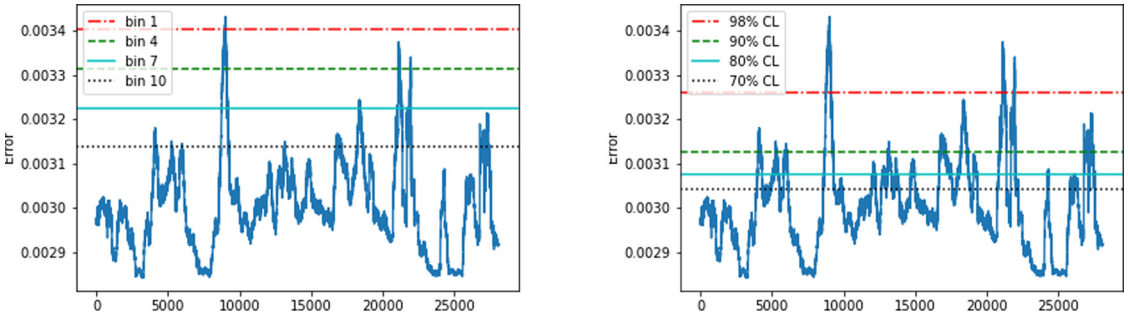


Figure 5: Impact of anomaly CLs on a sample trajectory. (a) Histogram and (b) KDE.

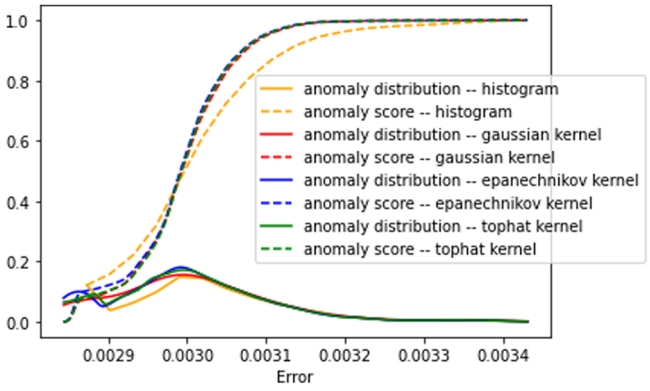


Figure 6: Anomaly score. The dashed graph are corresponding to the anomaly score functions.

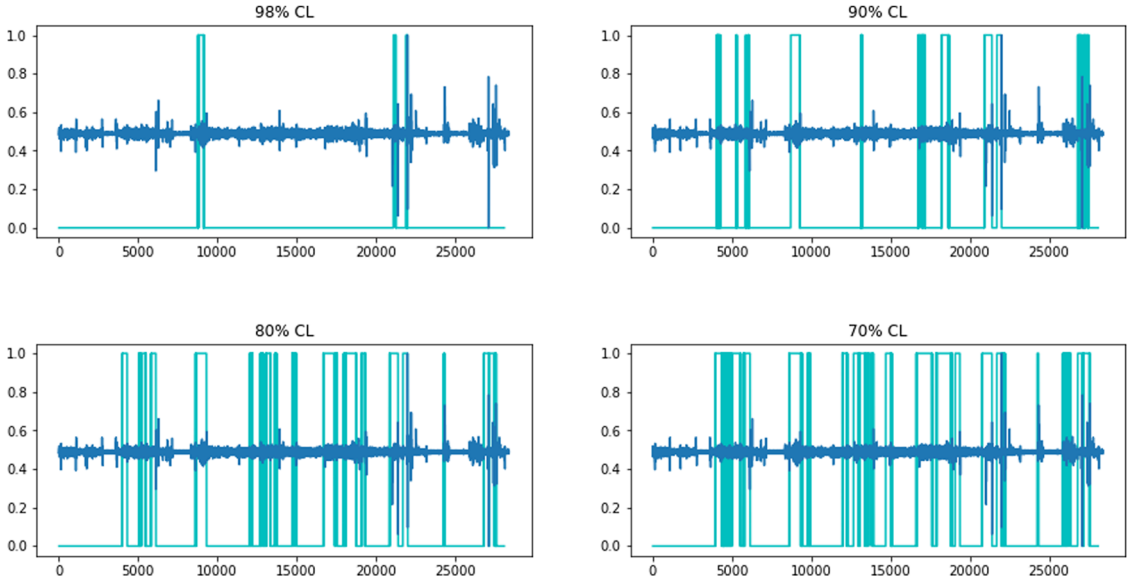


Figure 7: Predicted labels. The signal is shown in dark blue and the pseudo labels are specified with light blue.

significantly. A minor impact is observable on specificity metric, which is related to true negative rate. Nevertheless, true positive samples or true anomalies have been detected at high rates in all three cases. As expected, it is remarkable that the detection performance is slightly better when using pseudo labels with

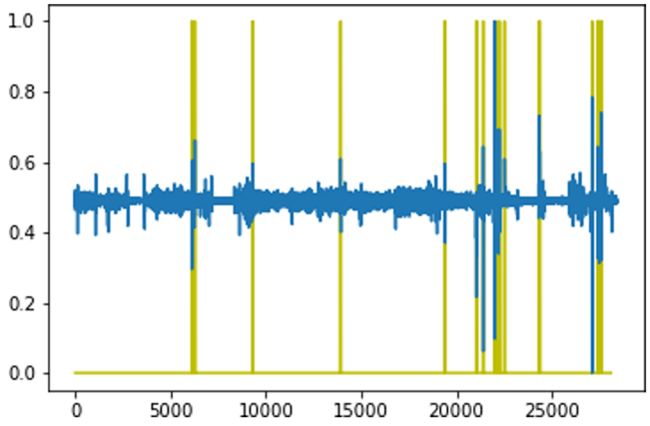


Figure 8: Pseudo labels. The signal is shown in blue and the pseudo labels created with cutoff value of 0.1 are specified with yellow.

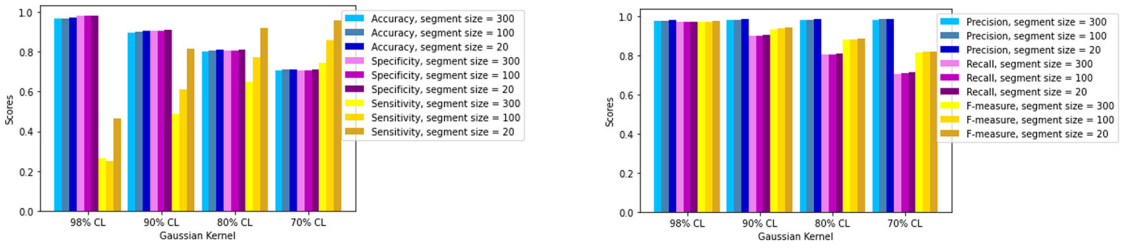


Figure 9: Performance measurements.

Table 1: Performance measurements set 1

	Measurement set 1								
Method	**Accuracy**			**Specificity**			**Sensitivity**		
Segment size	***20***	***100***	***300***	***20***	***100***	***300***	***20***	***100***	***300***
98% CL	0.97	0.97	0.97	0.46	0.25	0.26	0.98	0.98	0.98
90% CL	0.90	0.90	0.90	0.81	0.61	0.49	0.91	0.90	0.90
80% CL	0.81	0.80	0.80	0.92	0.77	0.65	0.81	0.81	0.80
70% CL	0.71	0.71	0.70	0.95	0.86	0.76	0.71	0.71	0.70
Baseline	0.77	0.69	0.66	0.85	0.78	0.76	0.77	0.69	0.66

Table 2: Performance measurement set 2

	Measurement set 2								
Method	**Precision**			**Recall**			**f-Measure**		
Segment size	***20***	***100***	***300***	***20***	***100***	***300***	***20***	***100***	***300***
98% CL	0.99	0.98	0.98	0.98	0.98	0.98	0.97	0.97	0.97
90% CL	0.99	0.98	0.98	0.91	0.90	0.90	0.94	0.94	0.93
80% CL	0.98	0.98	0.98	0.81	0.80	0.80	0.88	0.88	0.88
70% CL	0.98	0.98	0.98	0.71	0.71	0.70	0.82	0.81	0.82
baseline	0.88	0.88	0.88	0.77	0.69	0.66	0.86	0.80	0.77

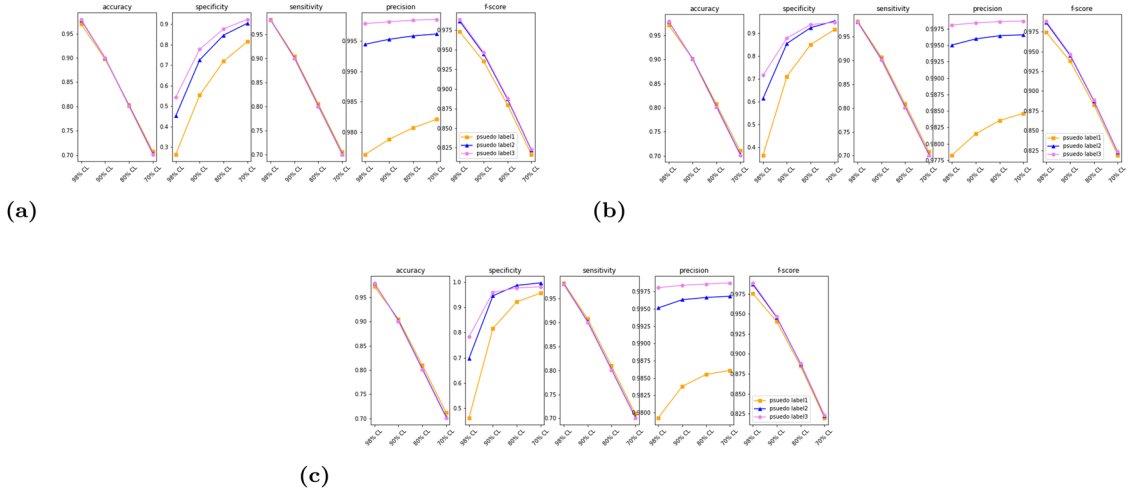


Figure 10: The effect of three different thresholds for creating pseudo labels. The pseudo label 1, pseudo label 2, and pseudo label 3 are corresponding to thresholds of 0.1, 0.2, and 0.3, respectively. Performance measurements for segment sizes of (a) 300, (b) 100, and (c) 20.

higher thresholds, which correspond to easier tasks. However, as we increase the threshold gap for creating pseudo labels, the anomaly detection task becomes easier and the precision improves. The results indicate that the highest sensitivity (or recall) and accuracy are achieved at the highest anomaly CL. Sensitivity is representative of true detected anomalies, and it drops as we decrease the anomaly CL. This is due to the fact that with a lower CL, the number of correctly classified samples decreases. Likewise, accuracy and precision have the top values within the highest amount of CL. On the other hand, we observe that, under the same condition, specificity, which represents the true negative rate, goes up. This suggests that the likelihood of correctly identifying normal patterns increases when a higher degree of freedom for irregularity is applied. More importantly, precision is not affected by the level of confidence, which suggests that the number of correctly detected anomalies against wrongly detected anomalies remains high at all CLs. In anomaly detection methods, two issues of *masking and swamping* can arise [67]. The masking effect occurs when an outlier prevents the identification of other outliers and produces false negative (FN), whereas swamping refers to the situation in which an outlier makes nonoutliers or normal data to be categorized as outliers and hence increases false positive (FP) cases [68]. In order to measure the effect of these situations, we have measured precision and recall. Lower values of FP and FN result in higher precision and recall. Figure 11 depicts the precision-recall curve for four CLs and four segment sizes. As can be observed, both precision and recall values are reasonably high for different segment sizes and CLs. This result suggests that the proposed method is not suffered from masking and swamping issues. Moreover, it can be inferred that lower segment sizes provide superior results. This can be due to the more accurate models that are trained on low-length sequences. For a

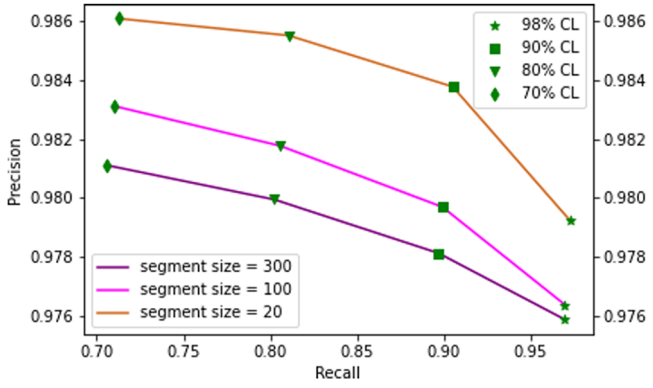


Figure 11: Precision-recall curve.

reliable anomaly detection system, it is crucial to have high sensitivity and precision. In other words, these measurements reflect the correctly identified true positives, which are true anomalies in our case. Therefore, in order to guarantee the best identification of true anomalies, we can focus on the results from higher CLs (i.e., the ones above 90% CL). In contrast, if the requirements tend towards thorough detection of all possible anomalies, lower levels of confidence such as 80% CL should be considered.

5 Conclusion

Anomaly and outlier detection is one of the challenging problems in machine learning and pattern recognition due to the lack of clarity of anomaly definition in various real-world applications and the lack of sufficient data for training efficient models. In this article, we proposed an autonomous methodology for anomaly detection for time series data in a fully unsupervised fashion. Since we do not have access to the labels for the time steps at which an anomaly has taken place or information about abnormal trajectories, we detect anomalies at different levels of confidence. We proposed a thresholding mechanism by applying KDE in order to find different levels of confidence for anomaly detection based on computation of percentiles of the PDF of error. In addition, we formulated anomaly score calculation based on cumulative distribution function of error. Our method provides flexibility for users in making decisions about anomalous patterns by specifying a certain CL and can assign a particular anomaly score to each sample. This provides a degree of freedom to decision-makers to identify anomalies with different severity levels of aberration. Especially, this feature is very important in the maritime domain where, due to the complexity and unpredictability of vessels' motions, the suspicious behavior of vessels is not easily noticeable. Hence, setting different levels of confidence and comparing the number of anomalies in each level can facilitate decisions about the alert amount that should be given to each anomalous pattern. One possible avenue of future work is concerned with multi-level thresholding, hence defining multi-level anomaly score functions that can divide a trajectory into multiple categories simultaneously. Each category can then identify a semantic segment of a trajectory that reveals a particular behavior of vessels' movements.

Acknowledgment: The authors would like to thank Martha Dais Ferreira for introducing the topic and the dataset.

Funding information: This work was supported by the Natural Sciences and Engineering Research Council of Canada.

Author contributions: Idea: Z. S., method: Z. S., Investigation: Z.S., Implementation and results: Z. S., Analysis: Z.S., Writing: Z. S., Review: Z.S., S.M., Supervision: S. M., Funding: S. M.

Conflict of interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability statement: The dataset is publicly available for download from https://coast.noaa.gov/htdata/CMSP/AISDataHandler/2020/index.html.

References

[1] Santhosh KK, Dogra DP, Roy PP. Anomaly detection in road traffic using visual surveillance: a survey. ACM Comput Surveys (CSUR). 2020;53(6):1–26.

[2] Zhou JT, Du J, Zhu H, Peng X, Liu Y, Goh RSM. AnomalyNet: an anomaly detection network for video surveillance. IEEE Trans Inform Forensics Security. 2019;14(10):2537–50.

[3] Zhou Y, Yan S, Huang TS. Detecting anomaly in videos from trajectory similarity analysis. In: 2007 IEEE International Conference on Multimedia and Expo. IEEE; 2007. p. 1087–90.
[4] Chen G, Lu G, Xie Z, Shang W. Anomaly detection in EEG signals: a case study on similarity measure. Comput Intell Neurosci. 2020;2020:6925107.
[5] Loureiro A, Torgo L, Soares C. Outlier detection using clustering methods: a data cleaning application. In: Proceedings of KDNet Symposium on Knowledge-based systems for the Public Sector. Bonn: Springer; 2004.
[6] Münz G, Li S, Carle G. Traffic anomaly detection using k-means clustering. In: GI/ITG Workshop MMBnet. vol. 7; 2007. p. 9.
[7] Chesnokov MY. Time series anomaly searching based on DBSCAN ensembles. Scientific Tech Inform Process. 2019;46(5):299–305.
[8] Liang B. A hierarchical clustering based global outlier detection method. In: 2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA). IEEE; 2010. p. 1213–5.
[9] Krleža D, Vrdoljak B, Brčić M. Statistical hierarchical clustering algorithm for outlier detection in evolving data streams. Machine Learn. 2021;110:139–84.
[10] Bomberger NA, Rhodes BJ, Garagic D, Dankert JR, Zandipour M, Stolzar LH, et al. Adaptive spatial scale for cognitively-inspired motion pattern learning & analysis algorithms for higher-level fusion and automated scene understanding. In: MILCOM 2008-2008 IEEE Military Communications Conference. IEEE; 2008. p. 1–7.
[11] Kozitsin V, Katser I, Lakontsev D. Online forecasting and anomaly detection based on the ARIMA model. Appl Sci. 2021;11(7):3194.
[12] Ma S, Liu Q, Zhang Y. A prediction method of fire frequency: Based on the optimization of SARIMA model. PLoS One. 2021;16(8):e0255857.
[13] Arimie CO, Harcourt P, Harcourt P, Harcourt P. Outlier detection and effects on modeling. Open Access Library J. 2020;7(09):1.
[14] Bianco AM, Garcia Ben M, Martinez E, Yohai VJ. Outlier detection in regression models with ARIMA errors using robust estimates. J Forecast. 2001;20(8):565–79.
[15] Yuen KV, Mu HQ. A novel probabilistic method for robust parametric identification and outlier detection. Probabilistic Eng Mech. 2012;30:48–59.
[16] Yuen KV, Ortiz GA. Outlier detection and robust regression for correlated data. Comput Meth Appl Mech Eng. 2017;313:632–46.
[17] Mu HQ, Kuok SC, Yuen KV. Stable robust extended Kalman filter. J Aerospace Eng. 2017;30(2):B4016010.
[18] Lind M, Hägg M, Siwe U, Haraldson S. Sea traffic management-beneficial for all maritime stakeholders. Transport Res Procedia. 2016;14:183–92.
[19] Harati-Mokhtari A, Wall A, Brooks P, Wang J. Automatic Identification System (AIS): data reliability and human error implications. J Navigat. 2007;60(3):373–89.
[20] Ruff L, Vandermeulen R, Goernitz N, Deecke L, Siddiqui SA, Binder A, et al. Deep one-class classification. In: International Conference on Machine Learning. PMLR; 2018. p. 4393–402.
[21] Xia Y, Cao X, Wen F, Hua G, Sun J. Learning discriminative reconstructions for unsupervised outlier removal. In: Proceedings of the IEEE International Conference on Computer Vision; 2015. p. 1511–9.
[22] Zhou C, Paffenroth RC. Anomaly detection with robust deep autoencoders. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2017. p. 665–74.
[23] Li T, Wang Z, Liu S, Lin WY. Deep unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision; 2021. p. 3636–45.
[24] Zhang C, Liu J, Chen W, Shi J, Yao M, Yan X, et al. Unsupervised anomaly detection based on deep autoencoding and clustering. Security Commun Netw. 2021;2021:1–8.
[25] Xiong H, Pandey G, Steinbach M, Kumar V. Enhancing data analysis with noise removal. IEEE Trans Knowledge Data Eng. 2006;18(3):304–19.
[26] Shukur HA, Kurnaz S. Credit card fraud detection using machine learning methodology. Int J Comput Sci Mobile Comput. 2019;8(3):257–60.
[27] Kadam V, Kumar S, Bongale A, Wazarkar S, Kamat P, Patil S. Enhancing surface fault detection using machine learning for 3D printed products. Appl Syst Innovat. 2021;4(2):34.
[28] Zhang Q, Zhang M, Chen T, Fan J, Yang Z, Li G. Electricity theft detection using generative models. In: 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE; 2018. p. 270–4.
[29] Nagaraja A, Aljawarneh S. PAREEKSHA: a machine learning approach for intrusion and anomaly detection. In: Proceedings of the First International Conference on Data Science, E-learning and Information Systems; 2018. p. 1–6.
[30] Zhang C, Xiao X, Wu C. Medical fraud and abuse detection system based on machine learning. Int J Environ Res Public Health. 2020;17(19):7265.
[31] Lane RO, Nevell DA, Hayward SD, Beaney TW. Maritime anomaly detection and threat assessment. In: 2010 13th International Conference on Information Fusion. IEEE; 2010. p. 1–8.
[32] Natale F, Gibin M, Alessandrini A, Vespe M, Paulrud A. Mapping fishing effort through AIS data. PLoS One. 2015;10(6):e0130746.
[33] Shahir HY, Glässer U, Nalbandyan N, Wehn H. Maritime situation analysis: A multi-vessel interaction and anomaly detection framework. In: 2014 IEEE Joint Intelligence and Security Informatics Conference. IEEE; 2014. p. 192–9.
[34] Rong H, Teixeira A, Soares CG. Data mining approach to shipping route characterization and anomaly detection based on AIS data. Ocean Eng. 2020;198:106936.
[35] Singh SK, Heymann F. Machine learning-assisted anomaly detection in maritime navigation using AIS data. In: 2020 IEEE/ION Position, Location and Navigation Symposium (PLANS). IEEE; 2020. p. 832–8.

[36] Zhen R, Jin Y, Hu Q, Shao Z, Nikitakos N. Maritime anomaly detection within coastal waters based on vessel trajectory clustering and Naive Bayes classifier. J Navigat. 2017;70(3):648–70.
[37] Mascaro S, Nicholso AE, Korb KB. Anomaly detection in vessel tracks using Bayesian networks. Int J Approx Reason. 2014;55(1):84–98.
[38] Handayani DOD, Sediono W, Shah A. Anomaly detection in vessel tracking using support vector machines (SVMs). In: 2013 International Conference on Advanced Computer Science Applications and Technologies. IEEE; 2013. p. 213–7.
[39] De Vries GKD, Van Someren M. Machine learning for vessel trajectories using compression, alignments and domain knowledge. Expert Syst Appl. 2012;39(18):13426–39.
[40] Murray B, Perera LP. Unsupervised trajectory anomaly detection for situation awareness in maritime navigation. In: International Conference on Offshore Mechanics and Arctic Engineering. Vol. 84379. American Society of Mechanical Engineers; 2020. p. V06AT06A024.
[41] Riveiro M, Johansson F, Falkman G, Ziemke T. Supporting maritime situation awareness using self organizing maps and Gaussian mixture models. Front Artif Intell Appl. 2008;173:84.
[42] Dahlbom A, Niklasson L. Trajectory clustering for coastal surveillance. In: 2007 10th International Conference on Information Fusion. IEEE; 2007. p. 1–8.
[43] Kowalska K, Peel L. Maritime anomaly detection using Gaussian process active learning. In: 2012 15th International Conference on Information Fusion. IEEE; 2012. p. 1164–71.
[44] Fu P, Wang H, Liu K, Hu X, Zhang H. Finding abnormal vessel trajectories using feature learning. IEEE Access. 2017;5:7898–909.
[45] Capobianco S, Millefiori LM, Forti N, Braca P, Willett P. Deep learning methods for vessel trajectory prediction based on recurrent neural networks. IEEE Trans Aerospace Electron Syst. 2021;57(6):4329–46.
[46] Yang CH, Wu CH, Shao JC, Wang YC, Hsieh CM. AIS-based intelligent vessel trajectory prediction using bi-LSTM. IEEE Access. 2022;10:24302–15.
[47] Mou JM, Van Der Tak C, Ligteringen H. Study on collision avoidance in busy waterways by using AIS data. Ocean Eng. 2010;37(5–6):483–90.
[48] Jiang X, Liu X, de Souza EN, Hu B, Silver DL, Matwin S. Improving point-based AIS trajectory classification with partition-wise gated recurrent units. In: 2017 International Joint Conference on Neural Networks (IJCNN). IEEE; 2017. p. 4044–51.
[49] Chen X, Liu Y, Achuthan K, Zhang X. A ship movement classification based on automatic identification system (AIS) data using convolutional neural network. Ocean Eng. 2020;218:108182.
[50] Kim KI, Lee KM. Deep learning-based caution area traffic prediction with automatic identification system sensor data. Sensors. 2018;18(9):3172.
[51] Zhao L, Shi G. Maritime anomaly detection using density-based clustering and recurrent neural network. J Navigat. 2019;72(4):894–916.
[52] Karataş GB, Karagoz P, Ayran O. Trajectory pattern extraction and anomaly detection for maritime vessels. Internet Things. 2021;16:100436.
[53] Sakurada M, Yairi T. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis; 2014. p. 4–11.
[54] 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 Trans Intell Transport Syst. 2021;23(6):5655–67.
[55] Blauwkamp D, Nguyen TD, Xie GG. Toward a deep learning approach to behavior-based AIS traffic anomaly detection. In: Dynamic and Novel Advances in Machine Learning and Intelligent Cyber Security (DYNAMICS) Workshop, San Juan, PR; 2018.
[56] Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Comput Surveys (CSUR). 2009;41(3):1–58.
[57] Zhou B, Liu S, Hooi B, Cheng X, Ye J. BeatGAN: anomalous rhythm detection using adversarially generated time series. In: IJCAI; 2019. p. 4433–9.
[58] Jiang Y, Zeng C, Xu J, Li T. Real time contextual collective anomaly detection over multiple data streams. Proceedings of the ODD. 2014; p. 14.
[59] LeGuillarme N, Lerouvreur X. Unsupervised extraction of knowledge from S-AIS data for maritime situational awareness. In: Proceedings of the 16th International Conference on Information Fusion. IEEE; 2013. p. 2025–32.
[60] Rhodes BJ, Bomberger NA, Zandipour M. Probabilistic associative learning of vessel motion patterns at multiple spatial scales for maritime situation awareness. In: 2007 10th International Conference on Information Fusion. IEEE;2007. p. 1–8.
[61] Reynolds DA. Gaussian mixture models. Encyclopedia Biometrics. 2009;741:659–63.
[62] Miller J. Reaction time analysis with outlier exclusion: Bias varies with sample size. Quarter J Experiment Psychol. 1991;43(4):907–12.
[63] Hampel FR. The influence curve and its role in robust estimation. J Amer Stat Assoc. 1974;69(346):383–93.
[64] Buzzi-Ferraris G, Manenti F. Outlier detection in large data sets. Comput Chem Eng. 2011;35(2):388–90.
[65] Perez HM, Chang R, Billings R, Kosub TL. Automatic identification systems (AIS) data use in marine vessel emission estimation. In: 18th Annual International Emission Inventory Conference. vol. 14; 2009. p. e17.
[66] Engle R, Granger C. Long-run economic relationships: readings in cointegration. UK: Oxford University Press; 1991.
[67] Acuna E, Rodriguez C. A meta analysis study of outlier detection methods in classification. Technical paper, Department of Mathematics, University of Puerto Rico at Mayaguez. 2004; vol. 1. p. 25.
[68] Mishra A, Müller CL. Robust regression with compositional covariates. Comput Stat Data Anal. 2022;165:107315.