Fishing Vessels Activity Detection from Longitudinal AIS Data
(Industrial Paper)

Saeed Arasteh
Engineering Science
Simon Fraser University
British Columbia, Canada
sarasteh@sfu.ca

Mohammad A. Tayebi, Zahra Zohrevand
Uwe Glässer, Amir Yaghoubi Shahir
Computing Science
Simon Fraser University
British Columbia, Canada
{tayebi,zzohreva,glaesser,sayaghou}@sfu.ca

Parvaneh Saeedi
Engineering Science
Simon Fraser University
British Columbia, Canada
psaeedi@sfu.ca

Hans Wehn
MDA Systems Ltd.
Earth Observation and Intelligence
British Columbia, Canada
hans.wehn@mdacorporation.com

ABSTRACT

The impact of marine life on the oceans of our planet is undeniable and overfishing is a serious threat to marine ecosystems worldwide. Maritime domain awareness calls for continuous monitoring and tracking of fisheries using data from maritime intelligence sources to detect illegal fishing activities. Marine traffic data from vessel tracking services is a promising source for identifying, locating, and capturing vessel information. Given the volume of such data, manual processing is impossible, raising an immediate need for autonomous and smart systems to follow the footprints of vessels and detect their activity types in near real-time. To achieve this goal, we propose *FishNET*, a simple yet effective convolutional neural network (CNN) model for vessel trajectory classification. The model is trained using a set of invariant spatiotemporal feature sequences extracted from the behavioral characteristics of vessel movements.

While existing approaches present point-based classification models, in this paper we not only discuss that a segment-based classification model has more realistic real-world applications but also show, by using expert-labelled data, that *FishNET* outperforms state-of-the-art fishing activity detection models. Our method does not require information about the fishing vessels type or type of fishing gear which is deployed.

To show applications in taking action against illegal fishing, we apply the trained model on large real-world but unlabelled fishing vessel data from the U.S. and Denmark gathered over a period of four years. In this analysis, we show how *FishNET* can contribute to managing fisheries by learning more about spatiotemporal fishing effort distribution, and to law enforcement agencies by detecting unreported and underreported fishing effort of individual vessels.

SIGSPATIAL '20, November 3–6, 2020, Seattle, WA, USA

ACM ISBN 978-1-4503-8019-5/20/11…$15.00
https://doi.org/10.1145/3397536.3422267

CCS CONCEPTS

• **H.2.8 [Database Management]: Database Applications—Data Mining**;

KEYWORDS

trajectory classification, convolutional neural network, spatiotemporal data mining, illegal fishing

ACM Reference Format:
Saeed Arasteh, Mohammad A. Tayebi, Zahra Zohrevand, Uwe Glässer, Amir Yaghoubi Shahir, Parvaneh Saeedi, and Hans Wehn. 2020. Fishing Vessels Activity Detection from Longitudinal AIS Data (Industrial Paper). In *28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '20), November 3–6, 2020, Seattle, WA, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3397536.3422267

1 INTRODUCTION

Illegal, unreported and unregulated (IUU) fishing is one of the most serious threats to the sustainability of world fisheries [1–3] and a major threat to marine biodiversity, the sustainability and balance of marine ecosystems, fish populations worldwide, and global food security as 4.3B people depend on fisheries for protein [4]. Entities that engage in IUU fishing circumvent conservation and management measures, avoid the operational costs associated with sustainable fishing practices, and may derive economic benefit from exceeding harvesting limits [5]. It is estimated that IUU fishing accounts for up to 30% of all fishing worldwide with an estimated economic damage of \$23B per year [6–9]. Maritime regulatory authorities struggle to enforce compliance of fisheries in the face of an enormous fleet of ships and boats that operate across coastal waters and open oceans covering about 71% of our planet's surface.

Fishing Activity Detection. Maritime domain awareness calls for continuous monitoring and tracking of fisheries using data and information from maritime intelligence sources to detect illegal fishing activities. Vessel tracking services routinely gather data for identifying, locating, and capturing information about ships and boats operating in coastal waters and across open oceans. Marine traffic data in the form of multi-variate time series is interpreted as *trajectories* of vessel movements over time periods ranging from a few hours to several years. The total data volume renders manual processing impossible, raising an immediate need for autonomous and smart systems to follow the vessels' footprints and detect their basic activity types in near real-time. In this paper, we propose *FishNET*, a simple yet effective deep learning model using a one-dimensional convolutional neural network (1-D CNN) for vessel trajectory classification[1]. The aim is to automatically detect fishing activity based on characteristic movement patterns and independent of the type of ship or boat and fishing gear being deployed (as detailed in Sect. 3.2). Our model is trained based on a set of invariant spatiotemporal feature sequences extracted from the observable behavioral characteristics of vessel movements.

Scientific Contribution. We are tackling here a complex real-world problem with acute ecologic and economic implications as is, that is, avoiding any simplifying but unrealistic assumptions. While existing approaches build on point-based classification models, in this paper, we not only discuss that a segment-based classification model has more realistic real-world applications but also show that *FishNET* outperforms state-of-the-art fishing activity detection methods on expert-labelled data. Our method does not require any input information about the fishing vessel type or type of fishing gear being deployed while fishing.

Experimental Evaluation. We train *FishNET* on marine traffic data labeled by domain experts and made available by Global Fishing Watch (GFW) [10]. Labels refer to fishing and non-fishing activity respectively. In the experimental evaluation, our model achieves 93.63% accuracy as a binary classifier for separating fishing from non-fishing activity (using 10-fold cross-validation). *FishNET* outperforms state-of-the-art fishing activity detection models.

To show applications of *FishNET* in combating illegal fishing and developing fisheries policies, we apply *FishNET* on large real-world but unlabelled fishing vessel data from the U.S. and Denmark gathered over a period of four years. In this analysis, we show how *FishNET* can assist marine authorities in establishing fishing regulations by learning facts extracted from the spatiotemporal fishing effort distribution and how this knowledge can help improve the compliance level by detecting unreported and underreported fishing effort of individual vessels.

Paper Organization. The remainder of this paper is structured as follows. Section 2 discusses related work. Section 3 explains the background concepts used for tracking fishing vessels and their activities. Next, Section 4 introduces the fishing activity detection framework. Section 5 presents the experimental evaluation, while Section 6 illustrates the practical application of *FishNET* in complex real-world scenarios. Section 7 concludes the paper.

2 RELATED WORK

In this section, we review related work in tackling IUU fishing by vessel tracking and deep neural network applications for fishing activity detection.

Tackling IUU Fishing by Vessel Tracking. Monitoring and tracking fishing vessel activity is one of the most critical elements in combating IUU fishing. While vessels in IUU fishing typically try to conceal their illicit activities, machine learning techniques have emerged in recent years as a promising tool to differentiate regular from irregular fishing vessel activity patterns.

Instead of AIS, or in addition to, other data sources are routinely used for monitoring and tracking fishing vessels. Shui-Kai et al. [11] use VMS (Vessel Monitoring System) data to predict fishing status of large-scale tuna longlines based on speed changes at specific times of the day. While VMS is routine in commercial fishing for enforcing compliance with the legal and regulatory framework and offers a more complete and detailed picture of activities than AIS, VMS data is considered highly sensitive and is usually not available for academic research.

Deep Neural Network for Fishing Vessels Activity Detection. Deep learning is being used in the field of maritime awareness and security for two main purposes: detecting the type of vessels and identifying anomalous activities. Detecting the vessel type has important applications as some vessels present themselves using a fake identity [12]. Identifying anomalous movement patterns helps in ensuring regulatory compliance with fisheries management. Venskus et al. [13] use a self-organizing map (SOP) network to identify abnormal movements of vessels in areas of heavy traffic. In another study, Nguyen et al. [14] use a multitask method based on a combination of recurrent neural networks with latent variable modeling to detect vessels' abnormal behavior.

AIS data is a valuable source to track vessel movements, despite all its shortcomings. X. Jiang et al. [15] propose a model to detect the fishing activity of longlines using sequence labelling and autoencoders. Using a sliding window, they obtain segments of labelled sequences based on latitude and longitude of AIS reports. In another study, they used partition-wise recurrent neural network (p-RNN) for point-based AIS trajectory classification to detect fishing activity in longlines[16]. In another work, De Souza et al. developed three separate models, Hidden Markov models (HMM) for trawlers, Data Mining (DM) for longlines and a multi-layered filtering strategy based on the detection of fishing in purseseines[17].

In a prominent work GFW presents a CNN-based approach for fishing activity detection (henceforth refered to as *GFW*) [18]. They extract vessel type information from the registration history of fishing and non-fishing vessels and trained their model using labelled data. To the best of our knowledge [18], this work is the most similar one to ours. The accuracy of *FishNET* obtained in our experimental evaluation exceeds the results published in [18].

3 FISHING FOOTPRINTS OF VESSELS

The Automatic Identification System (AIS) consists of an external antenna connected to a shipboard AIS transceiver, which regularly sends an updated AIS report several times per minute, effectively rendering a trajectory allowing to track vessel movements over time. Marine traffic information comes in various forms and formats from sources including marine radar, base stations and satellites using different technologies for monitoring and tracking a wide range of ships and boats. Most prominently, the Automatic Identification System (AIS) [19], an automated, autonomous tracking system is

[1]This research is funded by the Natural Sciences and Engineering Research Council of Canada and the Technology Demonstration Program under Project STAR (Satellite Technology and Advanced Research).

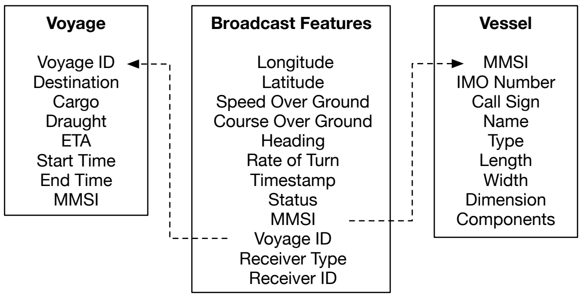


Figure 1: AIS report structure

widely used across the maritime world to periodically transmit data and information about a vessel and its voyage [20].

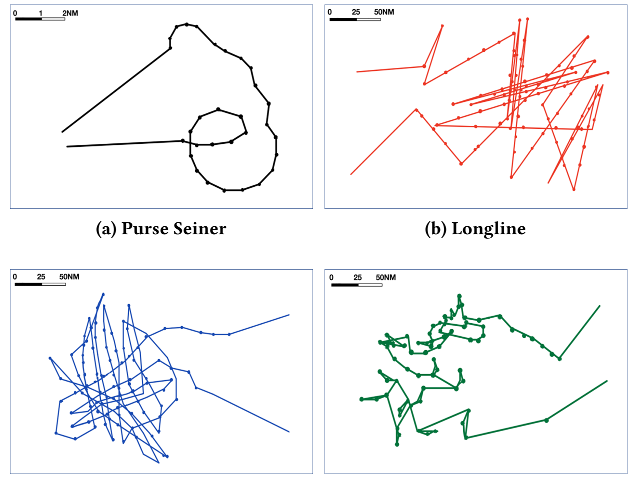
3.1 Tracking Fishing Vessels

AIS reports can be received through other vessels operating nearby, vessel tracking services from ashore locations and satellites. Despite scattering data points, satellite AIS provides extensive coverage for large regions out on the open ocean. In addition to a vessel's unique Maritime Mobile Service Identity (MMSI), course and speed over ground, and GPS coordinates, AIS reports include various other static and dynamic information about the vessel and its voyage as listed in Fig. 1. A status flag manually set by the crew indicates the current navigational status of the vessel, where status value "7" means "engaged in fishing".

International maritime regulations require all fishing vessels over 15 meters to be fitted with AIS and to maintain AIS in operation at all times, except where international agreements, rules or standards provide for the protection of navigational information [19]. AIS information is self-reported and prone to manipulation by fisheries who tend to conceal their operations; they can easily misreport their navigational status and location to conceal illegal fishing. This kind of reporting makes AIS unreliable to account for the detection of fishing activity across vast coastal waters and open oceans.

3.2 Fishing Activity Characteristics

Different kinds of vessels are used in commercial, artisanal and recreational fishing[2]. The type of a vessel determines the fishing gear it has available and the characteristics of the observable pattern it produces when fishing. Fishing vessels are categorized into two broad types, trawlers and non-trawlers [20]. *Trawlers* are the most common type of commercial vessels and use trawler nets with suspended equipment for fishing under the water surface. While fishing, it usually maintains a constant, slow and steady speed to keep the strain on the dragged net. This creates a particular pattern of movements from the time the net is deployed for fishing until it is retrieved from the water. Non-Trawlers comprise four main categories. *Longlines* are equipped with multiple long fishing lines in series; each line (up to 54 NM or 100 km long) has hundreds of baited hooks attached. The speed of these vessels at the net setting time is a little bit lower than steaming speed. *Seiners* are used to catch pelagic species near the water surface using seine nets. They are fishing at higher speeds around 10 knots. *Fixed Gear* refers to vessels used for trapping or potting, and gillnetting. Figure 2 shows characteristic patterns for four main types of vessels.



(a) Purse Seiner (b) Longline

(c) Trawler (d) Fixed Gear

Figure 2: Fishing activity patters for different types of vessels

When considering the distinctive characteristics of fishing patterns depending on the type of fishing gear being used (as illustrated in Fig. 2), explicit knowledge of the vessel type is already encoded in the pattern itself; in fact, it is part of the "signature" a vessel leaves in its footprint. This insight has a clear advantage. Since the vessel type is linked to the MMSI, the information is unavailable when a vessel hides its MMSI, a common phenomenon known as "dark fishing", or in cases when the type information is inaccurate. In contrast to point-based detection methods, *FishNET* recognizes the type in a vessel's signature and takes this knowledge into account when identifying fishing activity based on invariant spatiotemporal features as explained in Sect. 4.1.

3.3 Problem Definition

We define the fishing activity detection problem and the scope of the proposed solution in abstract terms as a binary classification model for interpreting trajectories extracted from marine traffic data. While our analysis uses AIS data, the model would work equally well for other forms of data, including data from the Vessel Monitoring System (VMS), which is common in commercial fishing but rarely made available.

Let $v \in V$ denote a *vessel* with AIS equipment which can engage in fishing activities. Further, let τ_v denote a temporally ordered set of AIS reports, $\{r_{\tau_v}\}$, from a trajectory associated with vessel v. A report $_{\tau_v}$ is a vector of values for the AIS features listed in Section 3, Figure 1 . We then formulate our problem as follows:

> Given a set of fishing vessels $V = \{v_1, v_2, \ldots, v_m\}$ and their corresponding trajectories $T = \{\tau_{v_1}, \tau_{v_2}, \ldots, \tau_{v_m}\}$, fishing activity detection is the task of determining

[2]Any activity, other than scientific research conducted by a scientific research vessel, that involves the catching, taking, or harvesting of fish; or any attempt to do so; or any activity that can reasonably be expected to result in the catching, taking, or harvesting of fish and any operations at sea in support of it (FAO, 2020).

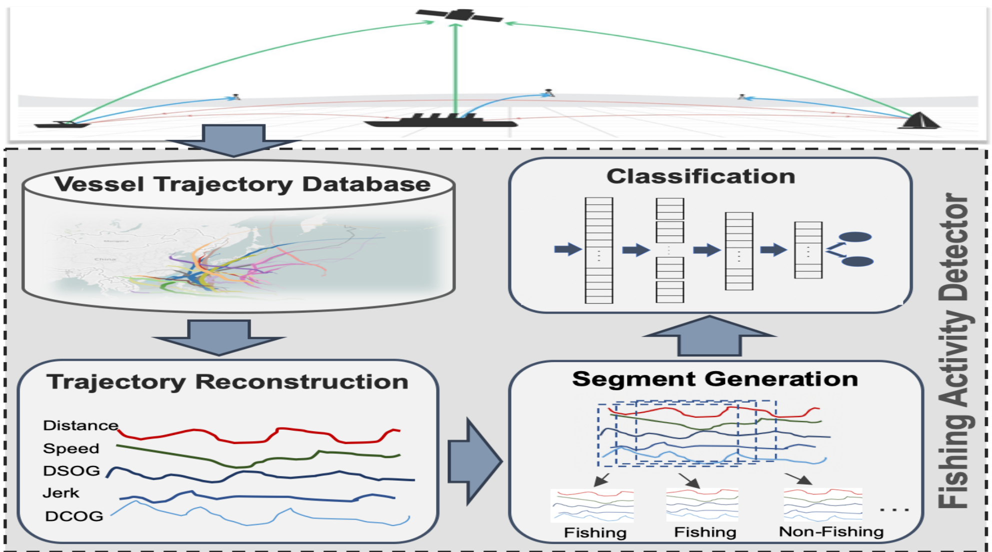


Figure 3: *FishNET*: Fishing activity detection framework

if a vessel $v_i \in V$ at time t was engaged in fishing or not by analysing δ consecutive reports from τ_{v_i}. This problem can be extended to identification of the whole trajectory representing an end-to-end fishing activity.

Throughout this paper, we use the notation $\tau_{v_i,t}$ to denote a specific time point t in a trajectory τ of some vessel v_i.

4 *FISHNET*: A FRAMEWORK FOR FISHING ACTIVITY DETECTION

This section presents *FishNET*, a framework to solve the fishing activity detection problem. *FishNET* overcomes the discussed challenges by utilizing a 3-stages algorithm to differentiate fishing activities from non-fishing ones. The detection framework is illustrated in Figure 3. First, the *Trajectory Reconstruction* component identifies a vessels' footsteps and generates a set of motion-related features to produce trajectories invariant to location and time.

The next component, *Segment Generation*, prepares the *classification*'s input by utilizing a sliding window segmentation to partition trajectories into short segments, which are labeled based on their likelihood of showing fishing activity. In the end, the vessel status at each temporal point will be determined based on the majority labels assigned to the segments including the given point; this facilitates the end-to-end fishing trajectory identification process. A detailed description of each stage is presented in the following subsections.

4.1 Trajectory Reconstruction

Learning from AIS data is not an easy task; massive volumes of stream data need to be extracted, unified, and integrated. The trajectory reconstruction process handles a variety of challenges we are facing in preparing the data for the predictive analysis. For example, the data is exposed to different noise sources, like infeasible speed or location, which are cleaned or replaced. Also, the time sampling for AIS reports is sometimes irregular, which is incompatible with the presumptions for our predictive model. To reconstruct regular trajectories of the message sequences, we apply linear spatial interpolation relying on the *constant velocity* assumption for fishing attempts. However, the granularity level of reconstructed trajectories matters. For example, a fine granular sampling rate like one report per minute may decrease the performance, whereas a sampling rate of one report per day may limit the real-world practicability. Based on our analysis, resampling trajectories at a 10-minute time scale is logical, because it does not leave a big gap between two successive AIS reports, which helps to avoid potential interpolation errors.

A generalizable trajectory classifier is expected to find the movement patterns and trajectory shapelets independent of the exact time and space. In other words, we expect that shipping trajectories share the same basic sub-patterns inrrespective of geospatial characteristics. However, classifying trajectories based on the original features included in AIS reports, might lead the model to develop some undesired decision logic based on a vessel's location or fishing activity time. We thus transform the original features in the trajectories of each vessel, $\tau_{v,t}$, into a set of motion-related features $\zeta_{v,t}$ that are suitable for discovering dynamic behavioral patterns of fishing activities independent of spatiotemporal details.

$$\{\tau_{v,t}\}_{t=0}^{n} \Rightarrow \{\zeta_{v,t}\}_{t=0}^{n} \tag{1}$$

Generally, vessel motions as shown in Figure 4 are defined by six degrees of freedom: three components of translation (heave, sway, surge) and three components of rotation (pitch, roll, yaw) [21]. On this basis, a set of performance and movement related characteristics can be extracted [22], as described below:

Distance is the great-circle distance between two consecutive points, $\mathbf{D}_v(t_2, t_1)$, computed using Haversine formula [23].

Rectilinear speed is equivalent to SOG in the AIS reports. However, as AIS is impacted by noise, we calculate the rectilinear speed based on the vessel's position at the reported time as following:

$$\mathrm{S}_v\left(t_2, t_1\right)=\frac{\mathbf{x}_v\left(t_2\right)-\mathbf{x}_v\left(t_1\right)}{t_2-t_1} \tag{2}$$

where $\mathbf{x}_v(t_1)$ and $\mathbf{x}_v(t_2)$ show the location of vessel s at times t_1 and t_2, respectively. The reported SOG is replaced with this derived speed in case of observing any corrupted/infeasible speed value in the original report.

Rectilinear acceleration defines the rate of change of vessel velocity in a given time interval and is computed as follows:

$$\mathrm{A}_v\left(t_2, t_1\right)=\frac{\mathrm{S}_v\left(t_2\right)-\mathrm{S}_v\left(t_1\right)}{t_2-t_1} \tag{3}$$

where $\mathrm{S}_v(t_1)$ and $\mathrm{S}_v(t_2)$ are the rectilinear speed of vessel s at times t_1 and t_2, respectively.

Rectilinear Jerk is the second derivative of speed that indicates the ability of the vessel in changing its acceleration and is computed as follows:

$$\mathrm{J}_v\left(t_2, t_1\right)=\frac{\mathrm{A}_v\left(t_2\right)-\mathrm{A}_v\left(t_1\right)}{t_2-t_1} \tag{4}$$

where $\mathrm{A}_v(t_1)$ and $\mathrm{A}_v(t_2)$ show the rectilinear acceleration of vessel s at times t_1 and t_2, respectively.

Derivative of course implies the changes in a vessel's course at two consecutive time points:

$$\Psi_v(t_2, t_1)=\frac{\Phi_v\left(t_2\right)-\Phi_v\left(t_1\right)}{t_2-t_1} \tag{5}$$

where $\Phi_v(t_1)$ and $\Phi_v(t_2)$ show the COG of vessel s at times t_1 and t_2, respectively. The absolute value of *course* is usually unreliable and noisy due to frequent fluctuations in the vessel's direction. However, the derivative of course is an essential feature in identifying vessel activities as this value is directly related to the vessel's maneuverability and its capability of creating different movement patterns.

Thus, $\zeta_{v,t}$ is defined based on $\{\mathbf{D}_{v,t}, \mathbf{S}_{v,t}, \mathbf{A}_{v,t}, \mathbf{J}_{v,t}, \Psi_{v,t}\}$ as the moving behavioral characteristics of fishing vessels. Applying these intrinsic attributes as predictive features not only makes *FishNET* robust to differences in geographical locations and timestamps of trajectories but also eliminates the need for considering the fishing vessel types in the classification process.

4.2 Segment Generation

As discussed earlier, fishing activity routinely extends over considerable time periods from several hours to days based on the fishing type [17]. Feeding a classifier model with such very long input sequences may not only misguide it and decrease the performance but will also increase the model's complexity and training time. Thus, to prepare the classifier's input, we extract trajectory fragments $\mathbf{w}_t$ out of the transformed trajectory ζ_t using a sliding window with size δ that moves across the signals and generates a set of fixed-length sequences.

$$\{\zeta_{v,t}\}_{t=1}^{n} \Rightarrow \{\{\mathbf{w}_t\}_{t=\delta}^{n} \mid \mathbf{w}_{v,t}=\{\zeta_{v,i}\}_{i=t-\delta-1}^{t}\} \tag{6}$$

Based on our experiments, on multiple values of δ we found that the setting of $\delta = 11$ helps the classifier to achieve the highest accuracy level, which intuitively makes sense; a sequence of about 2 hours is long enough to be informative and representative of a vessel's movement behavior.

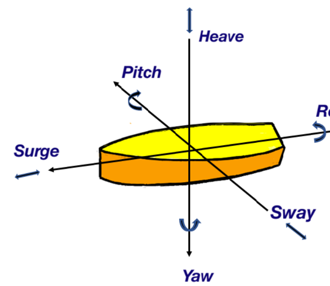


Figure 4: Vessel motion: six degrees of freedom

Next, a majority voting scheme is applied to assign a class label to the generated sequences. However, as the fishing vessels do not frequently switch their status, the majority of window sequences are pure fishing or non-fishing cases. To simplify labelling the entire sequence based on the majority rule,we considered odd values for δ.

4.3 CNN-based Classification

To classify the fixed-length multivariate segments generated in the previous phase, we design a one-dimensional convolutional neural network (1D-CNN). The output of this model is a binary label β assigned to any arbitrary input segment $\mathbf{w}$ to determine its *fishing/non-fishing* status:

$$\text{Classifier}(\mathbf{w}) \rightarrow \beta \tag{7}$$

The main reason for using a 1D-CNN is its significant performance in recent studies and its fairly low computational complexity. The model consists of two main modules: convolution and classification. The convolution module itself includes two main types of layers: *convolutional* and *pooling* layers. The former extracts in-depth features from the input sequences, while the latter filters the extracted features to highlight the most salient elements. In the end, the classification module, containing a set of fully connected layers, learns from the constructed internal representation how to classify the input segments.

Thus, as the kernel, conv $1D(.,.)$, in our $\text{Classifier}_{\text{CNN}}$ slides along with the 2D inputs w, it extracts the appropriate artificial features to distinguish fishing trajectories from the rest. Thus, as the Equation 8 shows, the input is convolved with the kernels and is executed through an activation function to produce the first layer's output feature map.

$$\mathbf{o}_j^l=b_j^l+\sum_{i=1}^{N_{l-1}} \operatorname{conv} 1D\left(\mathbf{k}_{ij}^{l-1}, \mathbf{w}_i^l\right) \mid l=1 \tag{8}$$

In general, the same process is followed through the next layers [24, 25]:

$$\mathbf{o}_j^l=f\left(\sum_{i \in M_j} \mathbf{o}_i^{l-1} * \mathbf{k}_{ij}^{l-1}+b_j^l\right) \mid l \in 2..L \tag{9}$$

where $\mathbf{w}_i^l$ is defined as the i^{th} element of input, L determines the number of convolutional layers, b_j^l is defined as the bias of the j^{th}

neuron at layer l, $\mathbf{o}_j^l$ is the output of the i^{th} neuron at layer $l-1$, $\mathbf{k}_{ij}^{l-1}$ is the kernel from the i^{th} neuron at layer $l-1$ to the j^{th} neuron at layer l.

In the next step, the final output of the convolution module is fed into the fully connected layers to map them to the final outputs of the network by determining the probability distribution over the set of target classes for the given input.

$$\mathrm{FCN}(\mathbf{o}^l) \rightarrow \alpha \mid l = L \tag{10}$$

Finally, the class with the maximum probability will be selected as the segment's *fishing/non-fishing* status ($\max(\alpha) \rightarrow \beta$).

5 EXPERIMENTAL EVALUATION

In this section we present the experimental evaluation, starting with describing the characteristics of datasets we used for experiments.

5.1 Data Characteristics

We use a labelled dataset to train our model and two unlabelled datasets to run longitudinal analysis, as described below.

Training Dataset. For training our model, we use a manually labelled data of fishing vessels, including 114 longlines, 28 purse seines, 35 fixed gears and 49 trawlers (226 MMSI). This dataset includes about 64K AIS data points representing over 46K hours of fishing and non-fishing activity.

U. S. AIS Dataset. The AIS data provided by the U. S. Coast Guard Services (available at www.marinecadastre.gov) covers all of the coastal waters of the United States and most of Canada from 2015 through 2018. AIS reports are filtered to one report per minute for each ship or boat, sending AIS signals and stored in geo-databases organized by Universal Transverse Mercator (UTM) Zones 1-11 and 14-20. In this period of four years, we are dealing with 579 vessels. 385 vessels were active in 2015 and the number of active vessels in 2016, 2017, and 2018 were 350, 350, and 436, respectively. Only 266 vessels were continuously active in every year of this period.

Denmark AIS Dataset. The AIS data provided by the Danish shore-based AIS system (available at dma.dk) covers vessel movement in Denmark's coastal waters for the period of 2015-2018. The total number of vessels in the Denmark dataset is 2432. This number is 1531 in 2015 and it increases to 1739 in 2018. 919 vessels were continuously active in every year of this period.

5.2 Experimental Design

The labelled dataset that we use in this study is a subset of the labelled dataset used to train *GFW*, and to the best of our knowledge this is the only among the state-of-the-art of fishing activity detection methods evaluated using large real-world data. *GFW* is a CNN-based approach with a basic architecture that adopts twelve features derived at each time point.

We present results for two scenarios: 1) vessel activity detection when we know the vessel types and 2) vessel activity detection when we do not know the vessel types. We compare *FishNET* results with this work. Please note that *GFW* does not present experimental results of fishing activity detection independent of fishing vessel types. Therefore, for *Scenario 2* we only compare *FishNET* with three baseline methods, in the lack of similar existing works:

- **Multi-layer perceptron (MLP)** is the basic class of feed-forward neural network. This model uses five hidden layers of 256 neurons with ReLu activation function and one sigmoid neuron in the output layer.
- **Random Forest** is an ensemble of decision trees. The result presented in Table 4 is achieved using the best parameters as follows: criterion = gini, max-depth = 9, min-sample-leaf = 7, min-sample-split = 7.
- **XGBoost** is an effective implementation of gradient tree boosting. The results presented in Table 4 are based on using the best parameters as follows: objective='binary:logistic', min-child-weight = 1, gamma = 2, subsample = 0.8, colsample-bytree = 0.8, max-depth = 3, n-estim. = 400,learning-rate = 0.02.

We use the labelled dataset for experimental evaluation, running 10-fold cross-validation over ten different randomly sampled sets. We use grid search cross validation technique to find the best parameters to optimize these three models. The performance of the evaluated methods is compared using different measures, including accuracy, recall, precision and F-measure.

1D-CNN component of *FishNET* has only five hidden layers, with zero padding. Its kernel size is three and has one fully connected layer. The number of neurons used in all hidden CNN layers is 256. Hidden fully connected layer has 128 neurons. CNN has better performance compare to other studied models, as each kernel filter is panned around the entire segment according to certain size and stride allowing kernels to find and match patterns of feature independent of where the pattern is located in each segment. The output layer contains a single neuron in order to make predictions. It uses the sigmoid activation function in order to produce a probability output in the range of 0 to 1 that can be converted to class values of fishing or non-fishing.

The number of training iterations of *FishNET* is 1000, and all convolutional layers used batch-normalization. This model tuned up with a grid search to find the best combination of layers and kernels. We achieved the highest accuracy by this configuration. We use Adam's version of stochastic gradient descent with default parameters and the binary-cross entropy loss function to train the network.

Also, to find the best value for the length for the sliding window, we applied odd values of $4 < \delta < 18$ As Table 2 shows the best accuracy is obtained using $\delta = 11$.

Table 1: Performance of *FishNET* for multiple window size

Window size	Accuracy	Precision	Recall	F1-Score
$\delta = 5$	91.34	91.21	90.61	90.19
$\delta = 7$	91.71	91.61	90.13	90.86
$\delta = 9$	92.42	91.86	90.35	91.10
$\delta = 11$	93.63	93.14	92.78	92.35
$\delta = 13$	91.83	90.73	89.92	90.32
$\delta = 15$	90.76	90.51	90.27	89.37
$\delta = 17$	89.32	88.80	87.11	87.94

5.3 Prediction Evaluation

Fishing activity detection applications are critical for IUU fishing control. While high precision is always desirable in machine learning solutions, high recall is more vital in solutions used for law enforcement. The reason is that the law enforcement agencies prefer to detect the illegal activities as much as possible even if they are forced to investigate a slightly larger search space. It is also a natural assumption when reducing the whopping cost of IUU fishing to the economy and environment is a priority.

Tables 2 and 3 show the overall performance of *FishNET* and *GFW*, respectively, for *Scenario 1* where we suppose the vessel type is given. For two vessels categories, longlines and trawlers, performance of *FishNET* and *GFW* in all evaluation measures are almost identical with slight differences. For longlines and in terms of accuracy and recall *FishNET* outperforms *GFW* by less than 1% and *GFW* outperforms *FishNET* by 2% and 1% with regard to precision and F1-Score. For trawlers and with regard to precision *FishNET* outperforms *GFW* by 2%. *GFW* outperforms *FishNET* in terms of recall by less than one percent. The results of both methods with regard to accuracy and f1-score is almost identical.

Interestingly, for purse seines and fixed gear categories *FishNET* outperforms *GFW* significantly. For purse seine the *FishNET* result is higher than *GFW* by 19%,14% and 16% with regard to accuracy, precision and F1-score, respectively. Both methods gain almost the same result in terms of recall. Eventually, for fixed gear *FishNET* outperforms *GFW* by 3%, 9% and 7% with regard to accuracy, precision and F1-score. The results of both methods in terms of recall are almost identical.

It is necessary to mention again that *GFW* classifies each AIS data point as fishing or non-fishing, while *FishNET* classifies trajectory segments into these two classes. In other words, *FishNET* gets the advantage of observing a longer sequence of AIS reports and learning the spatiotemporal dependencies to differentiate fishing and non-fishing patterns. Since engaging in fishing (with the gear in the water) is an activity which typically extends over several hours, employing a segment-based model like *FishNET* fits better with the realities of IUU fishing and makes it logistically suitable as an effective tool in enforcing fishing regulations.

Table 4 shows the experimental results for *Scenario 2* where the fishing activity detection model is applied on whole data blindly ignoring any side information about the vessel type. *FishNET* ranks first in all evaluation measures. *FishNET* achieves the accuracy of 94.31% outperforming all baseline methods. MLP works better than the other two methods with an accuracy of 91.25%. The accuracy of Random Forest is 80.17%, and XGBoost has the weakest performance gaining an accuracy of 78.93%.

Note that in comparison with *Scenario 1* the learning task in *Scenario 2* is more difficult where we need to detect fishing activities without knowing the vessel type. *FishNET* performs weaker in *Scenario 2* but gaining the result of 93.63%, 94.14%, 92.78% and 92% with regard to accuracy, precision, recall and F1-score, respectively, is a promising results with reasonable potential in combating IUU fishing. From a quantitative perspective, all the evaluations listed under Scenarios 1 and 2 show that *FishNET* outperforms the peer methods. It is a proven fact that 1-D CNN is very effective in extracting features from fixed-length sequential data. Especially, we believe the potential state dependency in the trajectories is extracted by the kernels in the CNN (considering the granularity level of the trajectories).

Table 2: Performance of *FishNET* for fishing activity detection of various vessel types (*Scenario 1*)

Vessel Type	Accuracy	Precision	Recall	F1-Score
Longlines	92.63	92.23	91.78	92.00
Purse Seines	97.58	96.21	95.34	95.77
Fixed Gears	97.62	97.35	97.36	97.35
Trawlers	98.27	96.13	95.18	95.65

Table 3: Performance of *GFW* for fishing activity detection of various vessels types (*Scenario 1)*

Vessel Type	Accuracy	Precision	Recall	F1-Score
Longlines	92	94	91	93
Purse Seines	78	81	95	79
Fixed Gears	95	88	97	90
Trawlers	98	94	96	96

Table 4: Performance of *FishNET* and baseline methods for fishing activity detection of various vessels types (*Scenario 2)*

Vessel Type	Accuracy	Precision	Recall	F1-Score
FishNET	93.63	93.14	92.78	92.35
MLP	91.25	89.13	90.27	89.69
Random Forest	80.17	69.53	87.83	77.62
XGBoost	78.93	68.24	86.47	76.28

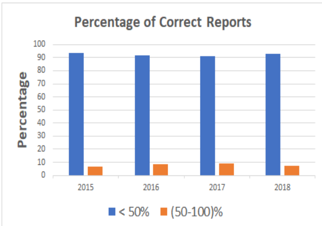
6 LONGITUDINAL ANALYSIS OF FISHING VESSELS BEHAVIOR

IUU fishing detection is a necessary but challenging task. *FishNET* can contribute to operational and daily IUU fishing detection with the aim of detecting anomalous activities and enforcing the fishing regulations. Moreover, *FishNET* can be used to learn long-term patterns and trends to help policy-makers in making more informed and proactive decisions by providing accurate and unbiased information. This section presents a longitudinal analysis of fishing vessel behavior using sizable AIS data from the U.S. and Denmark.

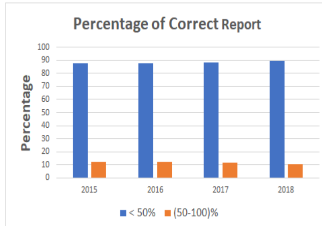
6.1 Fishing Effort Detection

AIS reports include a status field for reporting vessel activity such as fishing and harboring (see Sect. 3). Intuitively, a vessel involved in IUU fishing will try to conceal its fishing activity, which results in misreporting or under-reporting of their status. Now, using *FishNET* we can detect if a vessel was fishing at a given time and thus compare the vessel's real activity with what it reports.

Figures 5a and 5b, respectively, show the fraction of vessels that report their fishing activity status correctly in more or less than 50% of cases in the Denmark and U.S. datasets in different years of the studied period. Surprisingly, in both datasets and all years only about 10% of all vessels report their fishing activity status

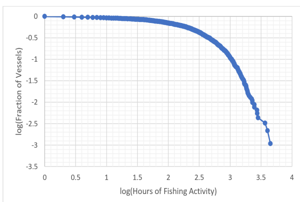


(a) Denmark Dataset

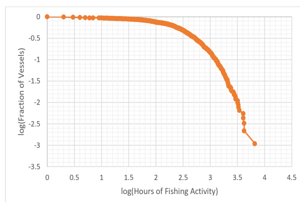


(b) U.S. Dataset

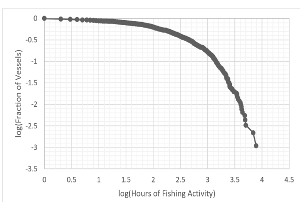
Figure 5: Ratio of AIS reports that correctly represent vessels fishing activity in the U.S. AIS dataset



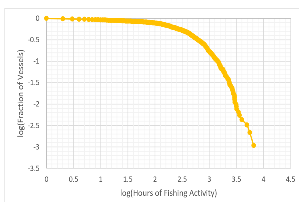
(a) Year 2015



(b) Year 2016



(c) Year 2017



(d) Year 2018

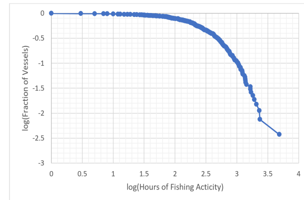
Figure 6: Fishing activity in the seas around Denmark (log scale): Fraction of vessels (y), Hours of fishing activity (x)

correctly in more than 50% of the cases. In other words, a very small group of vessels is committed to transmitting their fishing status correctly. This can be partly due to hiding fishing activity, but the more important reason is that the majority of vessels use AIS only for navigation safety.

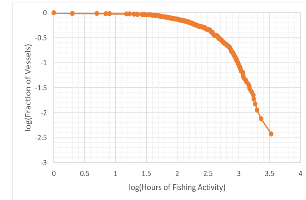
While estimating fishing effort is an essential requisite for the sustainable management of marine fisheries, the exploration confirms the unreliability of the self-reported status in the AIS data for identifying a vessel's fishing effort. This motivates us to use *FishNET* to learn more about the fishing effort of individual vessels.

Figures 6a, 6b, 6c, and 6d show the cumulative fishing effort distribution of vessels in the sea around Denmark for each of the years 2015-2018. Likewise, Figures 7a, 7b, 7c, and 7d show the cumulative fishing effort distribution of vessels in U.S. coastal waters for the years 2015-2018, respectively. As one can see, fishing effort for both regions and all years follows a *power law distribution.* Most vessels show less fishing effort compared to very few vessels that engage distinctively more in fishing.

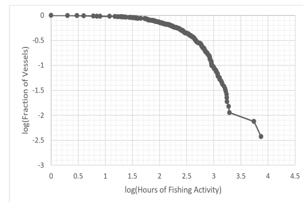
Having a power-law distribution for fishing effort over a longer time period is consistent with the conclusion of an experimental



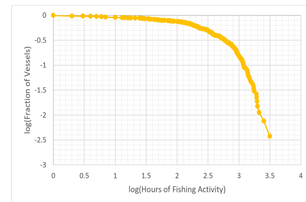
(a) Year 2015



(b) Year 2016



(c) Year 2017



(d) Year 2018

Figure 7: Fishing activity in North America's coastal areas (log scale): Fraction of vessels (y), Hours of fishing (x)

study in this field [26]. In this study, the authors show that fishing effort distribution of vessels in different fishing grounds over a fixed period follows a power-law distribution similar to our finding in this work. While many human-made phenomena follow power-law distributions and our result is also consistent with the limited available knowledge, we consider this problem open and subject to further scientific and experimental investigation.

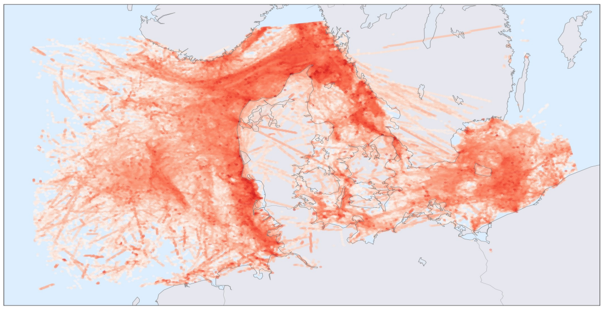
A more fine-grained analysis of fishing effort shows that over a period of one year each vessel in Denmark and U.S. waters, has on average its gear in the water for 345.6 and 391.2 hours, respectively. In a similar study [18], *GFW* detected 40 million hours of fishing activity for more than 70,000 vessels over a five-year period. In other words, this model detects 114 hours of fishing activity on average per vessel per year.

6.2 Spatiotemporal Distribution of Fishing Effort

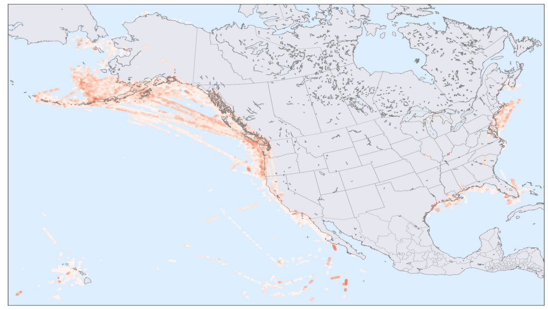
Spatiotemporal dynamics of fishing effort influences species distributions on a broad scale. Modern industrial fishing, using sophisticated electronic tracking devices, made it possible for humans to locate fish and catch them in large volumes. As a result, fisheries are exploiting adult prey at higher rates than any other predators [27]. On the other hand, the knowledge about Spatial distribution is limited, which makes resource management and policy decision-making difficult for fisheries management.

Spatial Distribution of Fishing Effort. Catch per unit effort (CPUE) is an index for assessing fish population dynamics widely used in fishery management. CUPE-based abundance analysis is not possible without understanding fishing effort spatial distribution. To address this issue, we create a grid over the studied area and then analyze vessels' fishing behavior in detail.

The geographical area of the sea around Denmark is from longitude 2 to longitude 18 (around 993 km) and latitude 53 to latitude 59 (around 667 km). We divide this area into grids of 2X2 km creating



(a) The seas around Denmark



(b) North America's coastal areas

Figure 8: Spatial distribution of fishing effort

about 100 thousand of them were in our area of interest. In only 60.6% of grids, *FishNET* detects fishing activity. Note that in both studied regions grids which have no fishing activity can potentially be located on land. On average, there are 12.6 vessels in each grid engaged in fishing activities in the year 2015 and changes to 12.4, 7.9, and 11.3 for years 2016, 2017, and 2018, respectively. On average, fishing effort by all vessels in each grid is 6.3, 7.3, 8.2, and 8.3 hours in the years 2015, 2016, 2017, and 2018, respectively. On average, the fishing effort in each grid are 6.3, 7.3, 8.2, and 8.3 hours in the years 2015, 2016, 2017, and 2018, respectively. Figure 8a is the heatmap of fishing effort in the sea around Denmark.

As mentioned in Section 5.1, the U.S. dataset covers the geographical area from longitude -63 to longitude -180 and latitude 0 to latitude 84. We segment this area into grids of 2x2 Km, and as a result, about 700 thousand of such grids are in our area of interest. In only 3.4% of such grids, *FishNET* detects fishing activity. On average, there are 3.4 vessels in each grid of interest that were engaged in fishing activities in 2015 which is reduced to 3.1, 3.3, and 2.4 for years 2016, 2017, and 2018, respectively. On average, fishing effort by all vessels in each grid is 24.7, 23.6, 25, and 26.3 minutes in years 2015, 2016, 2017, and 2018, respectively. On average, the fishing effort in each grid are 24.7, 23.6, 25, and 26.3 minutes in years 2015, 2016, 2017, and 2018, respectively. Figure 8b is the heatmap of fishing effort around North America.

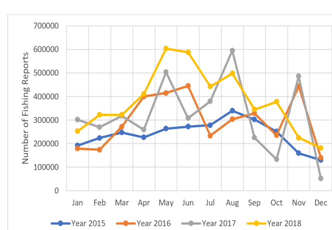
Temporal Distribution of Fishing Effort. Complexity of fisheries management due to the heterogeneity of this phenomenon. In addition to spatial, temporal variability is also playing an important role in catch size and fishing effort distribution. Analyzing vessels activity at a high temporal resolution makes our image of global fishing more complete. Figures 9a and 9 show the temporal distribution of fishing effort for the sea around Denmark and North America's Coastal Areas, respectively, for the studied period. Comparing the fishing effort pattern in both regions, we can see a similar trend: lower level of fishing in colder months of the year and a higher level of fishing effort in warmer months of the year, as expected.

In comparing of yearly temporal fishing effort distributions of the sea around Denmark, we can see many similar but not identical patterns. Interestingly, the yearly fishing effort distributions of North America's coastal area have a strong similarities. A fishing effort spatiotemporal distribution analysis contributing to fisheries management needs to be fine-grained and focused in close collaboration with domain experts. The analysis presented here is not sufficient for such a purpose as it is not the main objective of this study, which we leave as future work.

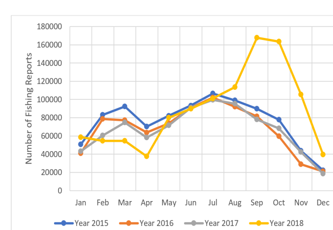
6.3 Anomalous Fishing Activity Detection

Fisheries authorities and regional fisheries management organizations face many difficulties in detecting and combating IUU fishing. One of the fundamental obstacles is the lack of systematic methods for measuring vessels' fishing effort. In the presence of uncertainty, predictive modeling, trained using expert-labelled data, reduces the size of the search space and increases the chances of disrupting IUU fishing. As a practical application of *FishNET* here, we discuss how successful fishing activity detection can be used to overcome this difficulty. The output of *FishNET* provides input for intelligence-based profiling and ranking of vessels linked to observable behavior that diverts from routine operations.

A small improvement in the success rate of quantification and identification of IUU fishing means a lot in absolute numbers when dealing with an enormous fleet of ships and boats that operate across vast coastal areas and open oceans, considering the limited surveillance resources constrain maritime domain awareness and security [28]. The very first group of vessels susceptible to IUU fishing are the ones in the top of a list ordered by fishing hours. Here, we focus on the top 10% of most active fishing vessels. In the sea around the Denmark area, the top 10% of most active vessels are fishing 1791 hours on average six times more than an average



(a) The seas around Denmark



(b) North America's coastal areas

Figure 9: Temporal distribution of fishing effort

vessel in this area. 153, 160, 177, and 142 vessels are in this list for the years 2015, 2016, 2017, and 2018 respectively. There are only 10 vessels which are continuously active for all the four years. Among these 10 vessels, 7 of them are industrial fishing vessels with the length of more than 20 meters, and more specifically 4 of them are trawlers. While continuous fishing activity is assumptive, such phenomenon for a small vessel is strange and needs close investigation by monitoring agencies. Although the U.S. dataset contains a fewer number of vessels, the average hours of fishing for the top 10% most active fishing vessels is 1490 hours. 38, 35, 35, and 44 vessels are in this list for the years 2015, 2016, 2017, and 2018, respectively. Only three vessels are on this list for all four years.

Fishing in marine protected areas can cause significant threats to already endangered species. To extract patterns of fishing in rare spots, we compute for each grid the ratio of the total fishing hours to the number of vessels for that grid. Grids with a higher ratio are the ones with more fishing done by a fewer number of vessels and need closer investigation. In the sea around Denmark, we found 45 areas with a high ratio, seven of them are relatively far from the shoreline, which makes them even more suspicious. In North American coastal waters, we found 61 areas, four of these are further from the shoreline, and 3 of them are around Alaska. As mentioned earlier such analysis can be more meaningful and applicable in collaboration with experts monitoring fishing activity.

7 CONCLUSIONS

Illegal, unreported and unregulated fishing leads to overfishing at a scale acutely threatening marine ecosystems and economies worldwide. Stricter enforcement of compliance with regulations by closely watching fisheries and their activities needs effective approaches to utilize data and information gathered extensively through maritime intelligence. *FishNET* is a contribution towards automatic fishing activity detection frameworks to assist marine authorities and law enforcement in their routine operations.

Solving seemingly intractable real-world problems is naturally challenging. Lack of large expert-labelled data, common deficiencies of AIS reports, incomplete signal coverage and inaccessibility of other available sources like VMS data together with notorious efforts to hide fishing activity while operating across vast geographic areas are just some of the factors adding more dimensions to the problem. Still, the proposed approach shows data mining and machine-learning have already a lot to offer for extracting actionable knowledge from data and, with current trends in AI, even more though in the future.

Despite the remarkable progress of machine learning approaches for many commercial applications, unfortunately, some of the most important problems tied to essential human needs and necessities, including the one addressed in this work, did not get sufficient attention from the scholars in this field. Although the accuracy of close to 94% on real-world data is encouraging, the framework presented here clearly leaves room for further improvements regarding its accuracy, robustness and scalability. Considering what's at stake and that time is running out, we will continue our work in this direction but also invite others to contribute efforts to fight illegal fishing with advanced AI-based methods.

REFERENCES

[1] D. Pauly and D. Zeller, "Comments on faos state of world fisheries and aquaculture (sofia 2016)," *Marine Policy*, vol. 77, pp. 176–181, 2017.
[2] "Illegal, Unreported and Unregulated (IUU) Fishing - International Fisheries." [Online]. Available: https://dfo-mpo.gc.ca/international/isu-iuu-eng.htm
[3] A. Shaver and S. Yozell, "Casting a wider net," *The Stimson Center, January*, 2018.
[4] T. N. P. Bondaroff, W. Werf, and T. Reitano, "The global initiative against transnational organized crime and the black fish," *The Global Initiative Against Transnational Organized Crime*, 2015, accessed: 2020-06-20.
[5] "Presidential task force on combating illegal, unreported, and unregulated fishing and seafood fraud. U.S. national oceanic and atmospheric administration (NOAA)," https://www.iuufishing.noaa.gov, 2018, Accessed: 2020-06-20.
[6] "Fisheries and Oceans Canada. Illegal, unreported and unregulated (IUU) fishing," http://www.dfo-mpo.gc.ca/international/isu-iuu-eng.htm, 2019, 2020-06-20.
[7] Food and Agriculture Organization of the United Nations, "The State of World Fisheries and Aquaculture 2018," http://www.fao.org/3/i9540en/I9540EN.pdf, 2018, Accessed: 2020-06-20.
[8] K. Cutlip, "IUU-illegal, unreported, unregulated fishing. Global fishing watch," http://globalfishingwatch.org/fisheries/iuu-illegal-unreported-unregulated-fishing, 2016, Accessed: 2020-06-20.
[9] Y. Ye and N. L. Gutierrez, "Ending fishery overexploitation by expanding from local successes to globalized solutions," *Nature Ecology & Evolution*, vol. 1, no. 7, p. 0179, 2017.
[10] "Sustainability through transparency. Global fishing watch." [Online]. Available: https://globalfishingwatch.org/
[11] S.-K. Chang and T.-L. Yuan, "Deriving high-resolution spatiotemporal fishing effort of large-scale longline fishery from vessel monitoring system (vms) data and validated by observer data," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 71, no. 9, pp. 1363–1370, 2014.
[12] H. Ljunggren, "Using deep learning for classifying ship trajectories," in *2018 21st International Conference on Information Fusion (FUSION)*. IEEE, 2018, pp. 2158–2164.
[13] J. Venskus, P. Treigys, J. Bernatavičienė, G. Tamulevičius, and V. Medvedev, "Real-time maritime traffic anomaly detection based on sensors and history data embedding," *Sensors*, vol. 19, no. 17, p. 3782, 2019.
[14] D. Nguyen, R. Vadaine, G. Hajduch, R. Garello, and R. Fablet, "A multi-task deep learning architecture for maritime surveillance using ais data streams," in *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2018, pp. 331–340.
[15] X. Jiang, D. L. Silver, B. Hu, E. N. de Souza, and S. Matwin, "Fishing activity detection from ais data using autoencoders," in *Canadian Conference on Artificial Intelligence*. Springer, 2016, pp. 33–39.
[16] X. Jiang, E. N. de Souza, X. Liu, B. H. Soleimani, X. Wang, D. L. Silver, and S. Matwin, "Partition-wise recurrent neural networks for point-based ais trajectory classification." in *ESANN*, 2017.
[17] E. N. de Souza, K. Boerder, S. Matwin, and B. Worm, "Improving fishing pattern detection from satellite ais using data mining and machine learning," *PloS one*, vol. 11, no. 7, p. e0158248, 2016.
[18] D. A. Kroodsma, J. Mayorga, T. Hochberg, N. A. Miller, K. Boerder, F. Ferretti, A. Wilson, B. Bergman, T. D. White, B. A. Block *et al.*, "Tracking the global footprint of fisheries," *Science*, vol. 359, no. 6378, pp. 904–908, 2018.
[19] "Automatic Identification Systems (AIS)," International Maritime Organization. [Online]. Available: https://www.imo.org/en/OurWork/Safety/Navigation/Pages/AIS.aspx
[20] "MarineTraffic - the most popular online service for vessel tracking | AIS Marine Traffic." [Online]. Available: https://www.marinetraffic.com/en/p/ais-station-operator-contest-2018
[21] J. N. Newman, "The theory of ship motions," in *Advances in applied mechanics*. Elsevier, 1979, vol. 18, pp. 221–283.
[22] M. McDonald, "SHF SATCOM terminal ship-motion study," Naval Command Control and Ocean Surveillance Center, San Diego, CA, Tech. Rep., 1993.
[23] G. Van Brummelen, *Heavenly mathematics: The forgotten art of spherical trigonometry*. Princeton University Press, 2012.
[24] S. Kiranyaz, T. Ince, O. Abdeljaber, O. Avci, and M. Gabbouj, "1-d convolutional neural networks for signal processing applications," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8360–8364.
[25] J. Bouvrie, "Notes on convolutional neural networks," *In Practice*, pp. 47–60, 2006. [Online]. Available: http://cogprints.org/5869/
[26] N. Ellis and Y.-G. Wang, "Effects of fish density distribution and effort distribution on catchability," *ICES Journal of Marine Science*, vol. 64, no. 1, pp. 178–191, 2007.
[27] C. T. Darimont, C. H. Fox, H. M. Bryan, and T. E. Reimchen, "The unique ecology of human predators," *Science*, vol. 349, no. 6250, pp. 858–860, 2015.
[28] D. A. Goward, "Maritime domain awareness: The key to maritime security," *Legal challenges in maritime security. Leiden: Martinus Nijhoff*, 2008.