Proceedings of the 2023 International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/27/20230088

Illegal, unreported and unregulated fishing detection with machine learning

Yuzhe Xu

American University, Washington, DC, United States, 20016

yx8756a@american.edu

Abstract. Illegal, unreported and unregulated fishing is a worldwide problem that causes local and global economic losses, depletes natural resources, alters our diverse ecosystems and takes an undue toll on fisheries. This study describes a machine learning-based strategy for response generation. Identifying data storage and processes has led to the initial development of a viable IUU fishing detection system that classifies vessels for IUU fishing by combining (1) the likelihood that a vessel is fishing using geospatially referenced signal data, and (2) whether or not it is classified The likelihood of fish activity is scored for ships to be within the area of interest, and (3) classification of whether the vessel is allowed to enter its habitable area of interest. In this paper, certain parts of the system were prototyped, including using logistic regression to develop highly predictive catch or no-fish classification models for longlines and trawlers, and identifying whether a vessel was within an area of interest process. In addition, many fishing vessel registries have been identified, which regulate the rights of specific vessels to fish in regulated areas of interest. The accuracy with which fishing models can predict the probability of fishing when vessels have longline or trawl gear is acceptable, and can predict vessels with seine gear, but additional research and analysis are needed. "In ROI" classification models should be extended to score their likelihood of being in ROI instead of outputting true/false judgments. Using machine learning and data analysis skills, the project aims to make further efforts to predict IUU fishing in order to enable law enforcement and ultimately significantly reduce or prevent IUU fishing.

Keywords: Machine learning, fishing, vessels, IUU, data analysis, track detection

1. Introduction

IUU (Illegal, Unreported, and Unregulated) fishing is a serious global problem that harms national interests, destroys marine ecosystems and is costly to manage. According to the United Nations, IUU fishing accounts for an estimated 11-26 million tones of fish caught annually, worth \$10 billion to \$23 billion. This illegal activity not only undermines sustainable fishing practices, but it also harms the livelihoods of legitimate fishers and the economies of coastal communities.

A key challenge in addressing Illegal, Unreported, and Unregulated fishing is the cost and time required to monitor, detect, and seize illegal fishing vessels. Traditional methods such as patrols and inspections are resource-intensive and can only cover a small portion of the ocean. Predictive knowledge on the likelihood of Illegal, Unreported, and Unregulated fishing activity in specific locations or by certain boats can improve this process. One way to gather this information is through the use of AIS

(Automatic Identification System) data, which is mandatory on international boats over 300 gross tonnages. AIS data includes information such as a vessel's position, direction, speed, and activity, and can be used to model vessel fishing behavior and estimate the probability of illegal fishing activity in a specific area. Predictive analytics based on AIS data may also help to identify boats that are authorized to fish in a regulated area [1].

This project aimed to develop an Illegal, Unreported, and Unregulated fishing detection architecture and identified data sources for a predictive analytics model to anticipate Illegal fishing activity. The goal was to use AIS data and geospatially-referenced sensor data to create and enhance an Illegal fishing detection technique, with a focus on MPAs (Marine Protected Areas). Marine Protected Areas are critical habitats for many fish species and are particularly vulnerable to Illegal fishing. By focusing on MPAs, this project aimed to protect these areas and the species that depend on them.

The project scope included researching and developing systems engineering diagrams, identifying public data sources, designing data analysis models, and creating models to identify boats with patterns of interest. The project used a data analytics lifecycle to deliver a fishing score model. This model would take into account various factors such as vessel behavior, location, and time to predict the likelihood of illegal fishing activity. By using this model, enforcement agencies could focus their resources on high-risk areas and vessels, increasing their chances of detecting and deterring IUU fishing [2].

Overall, the project used a combination of research, data analytics, and systems engineering to develop a method for detecting IUU fishing activity using AIS data and sensor data. The goal was to create a model that could predict illegal fishing activity in specific locations, using data from Global Fishing Watch's data repository. The project was constrained by the availability of data and time restrictions, and focused on a portion of the IUU architectural framework. However, the ultimate goal of this project is to provide a valuable tool for enforcement agencies and organizations to combat IUU fishing and protect marine ecosystems [3].

2. Machine-learning model

In order to do this task, a machine-learning model has been developed in Python. VS Code and Juypter Notebook will be used to write the code. The system may be constructed using free, real-time data provided by Global Fishing Watch.

The system cannot be constructed without the free, real-time data provided by Global Fishing Watch. Try out the machine learning software called sky learn. You'll have to cope with fishing vessels that don't have GPS. By keeping an eye on shiny spacecraft in space. Algorithms can identify lights on ships ranging from "little wooden boats with simple lights" to "huge industrial ships with the equivalent of sports stadium lighting" using data from NOAA weather satellites. This would allow for the monitoring of an extra 100,000 - 200,000 vessels.

Keeping track of every ship at sea would be an arduous task. Time-consuming and inefficient, conventional radars and other monitoring equipment like cameras and drones provide limited situational awareness. Poaching of animals and fisheries might be uncovered using satellite and geolocation data. A staggering one-fifth of all fish caught in the globe is the result of illegal, undeclared, and unregulated fishing, which is worth an estimated $23.5 billion annually. The monitoring of illegal fishing might help alleviate a variety of environmental difficulties. Identification of fishing vessels might also help protect marine life.

There will be two primary phases to the project. The first step is to create a software that uses algorithms and machine learning to make predictions and judgments about fishing boat behavior and to generate a "heat map" of illicit fishing activity based on those predictions. The second phase involves using picture recognition software to spot other ships.

There is software that can generate heat maps of possible areas for illicit fishing. A database of fishing boats' current whereabouts that can be queried online. Meanwhile, it allows for the detection and tracking of illicit fishing boats based on their location and speed.

In addition to reducing the ecological harm caused by illicit fishing, the initiative can also successfully identify illegal fishing boats. The identification and safety of animals are only two of many

DOI: 10.54254/2755-2721/27/20230088

potential uses for vessel identification technology. The goal of this project is to develop a system for tracking fishing boats and identifying those that may be engaging in unlawful activity via the use of machine learning and image recognition. Framework for Basic Illegal Fishing Monitoring Software is a website we created to track vessel data. Then, we utilized the information given by Global Fishing Watch and our newly constructed machine learning model to keep tabs on illicit fishing boats. Exhibit the monitoring data using the web infrastructure and application. Algorithms can identify lights on ships ranging from "little wooden boats with simple lights" to "huge industrial ships with the equivalent of sports stadium lighting" using data from NOAA weather satellites. This allows for increased ship monitoring. The cost of detecting illegal fishing is still very high. We are trying to reduce the cost. However, there's still a 6% of the time that fishing ships are out of detection because of some units that shield signals.

3. Feasibility Discussion

3.1. Economic feasibility

From 2008 to 2020, it is estimated that the loss of fish catches due to illegal, unreported, and unregulated (IUU) fishing will be more than double the current annual fishing value. Additionally, the depletion of fish stocks from IUU fishing is so severe that, by 2020, catches in a scenario where IUU fishing is eliminated could be significantly higher than if it were to continue. It's important to note that the economic costs of IUU fishing are likely even higher than research indicates, as factors such as increased competition, poor data for management, and the negative impact of increasing jellyfish populations on tourism have not been fully accounted for [4].

3.2. Environmental

The depletion of fish stocks poses a major threat to the environment. Curbing illegal, unreported, and unregulated (IUU) fishing in the North Sea alone could boost fish populations and increase their value by over €4 billion by 2020. However, the actual environmental cost is likely much greater, as the calculation doesn't take into account other impacts such as the risk of extinction of targeted or accidental catch species like sharks and rays, or the effects on ecosystem services [4].

3.3. Manufacturability

The vessel score system was developed as a tool for engineering to help determine the likelihood of a fishing vessel engaging in a specific activity of interest. Key elements were identified as crucial for determining if further investigation was needed. These elements were chosen based on their ability to be incorporated into a multi-factor analytical model. In Spiral 2, the focus shifted to expanding the fishing model into an illegal, unreported, and unregulated (IUU) fishing detection model. This model identifies vessels in prohibited fishing areas and calculates the probability that they are engaging in illegal fishing activities. With this model, it may be possible to create a more precise method for identifying IUU fishing boats [1].

3.4. Technical

Patrolling the waters for illegal activity is a significant task for many countries. In the past, this has been done primarily through air and sea patrols. However, with the use of technology such as satellites, GPS, and Vessel Monitoring Systems, it is becoming easier for authorities to track fishing boats and detect any illegal activity such as entering protected areas or other nations' waters. While human monitors are still used, modern cameras and AI systems are also being implemented to assist in the monitoring process [5]. The technical approach includes analyzing and creating system engineering diagrams to document the detection architecture, as well as using data analysis methods such as data preparation, descriptive analysis, predictive modeling, and model validation.

DOI: 10.54254/2755-2721/27/20230088

3.5. Validation

The data analysis team identified potential data for analysis and subsequently focused on creating models to predict fishing. Meanwhile, the systems engineering team employed a tiered approach for modeling Illegal, Unreported and Unregulated (IUU) fishing activity. The initial step was to model advanced fishing activity, followed by modeling IUU fishing activity. The systems engineering team's efforts centered on documenting the architecture using model-based systems engineering, specifically to understand the underlying behavior and activities of IUU fishing. Due to time constraints and the availability of data, the data analysis was limited to a subset of the IUU architectural framework. The detailed description of the systems engineering and data analysis methods can be found in the following chapters.

3.6. Ethical

All models generate predictions based on a large amount of data, and the accuracy of these predictions can impact the decisions we make. However, it's important to note that predictions are not always correct and using them to investigate fishing vessels, for instance, could lead to mistakes if non-violating boats are surveyed. Misjudgments can also cause pushback from fishermen. Additionally, while the Automatic Identification System (AIS) is widely used by larger ships, it is not as prevalent on smaller and medium-sized vessels. Therefore, the model may not be as effective in predicting and monitoring smaller fishing operations and individual behavior. This could lead to complaints from larger fishing companies about perceived unfairness. The model and software were created to prevent illegal activities and harm to the environment, and are completely non-profit, so there is no possibility of misuse.

3.7. Political

The model and software are provided for the government and are expected to be made available free of charge. Combating illegal fishing is an important issue that the government needs to face. It is not only related to the transaction of a large amount of illegal funds, but also related to the protection of natural resources in the country's MPA area.

4. Final Implementation

4.1. IUU Fishing Architecture

The IUU (Illegal, Unreported, and Unregulated) fishing architecture project was an important initiative to tackle unsustainable fishing practices. Its goal was to create a logical process for an analytical model to detect and track vessels involved in IUU fishing. To achieve this, general use cases were created to define the scope of the model, recognize key industry participants, and outline the procedures to be considered for each use case. These use cases served as the foundation for the model and were used to evaluate its effectiveness in identifying IUU fishing. Next, activity diagrams were developed to elaborate on the use cases and display the specific operations and flow between them for each scenario the analytical model would examine. These diagrams provided a visual representation of the fishing industry interactions and processes, and helped identify potential areas for IUU fishing. State machine diagrams were also created to show the specific states of participants such as vessels, crews and regulatory authorities, which helped to understand how these states may affect IUU fishing. After identifying data sources for the model, a high-level system architecture was developed to integrate and analyze data from multiple sources like vessel tracking systems, electronic reporting systems and other databases. Lastly, the project focused on creating a scoring system to determine the likelihood of a ship performing or not performing IUU fishing based on key identifiers like vessel registration numbers and weighted attributes like the vessel's compliance history (Figure 1).

In summary, the development of IUU fishing architecture was a complex and multi-faceted process that involved recognizing key participants and procedures, creating visual diagrams, integrating data from multiple sources, and generating a scoring system to effectively detect and track vessels engaged in IUU fishing.

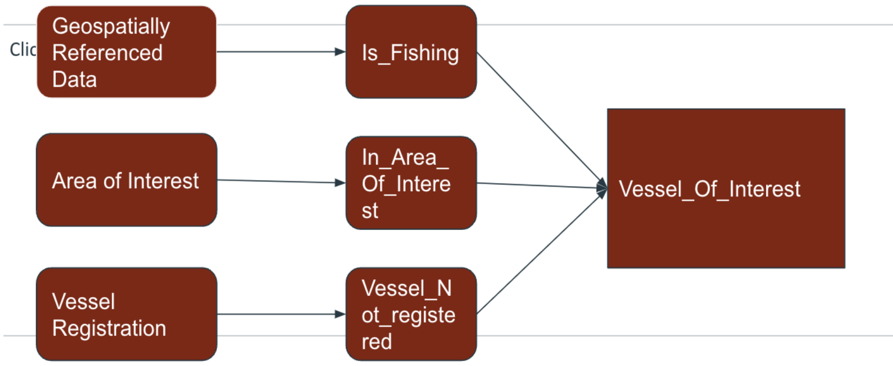

Figure 1. Project structure diagram

4.2. Data Preparation

Established to keep ships safe, AIS systems provide collision avoidance and allow maritime authorities to track and monitor the movements of ships. Each ship regularly reports information including the vessel's identity, type, location, course, speed, sailing status and other safety-related information. Ships equipped with AIS transceivers can be tracked by AIS base stations located on the shoreline, or outside the range of terrestrial networks, by a growing number of satellites equipped with special AIS receivers.

This approach generates 20 million data points per day. In this project, we will have data that has already been labeled by different experts and using crowdsourcing methods. Data features include the type of vessel (tug, drifting line, fixed gear) and a time series of each vessel's position, speed, and course. The time interval of data collection is not fixed. There may be a gap when the ship is not within the "line of sight" of the satellite. It takes 90 minutes for a low-orbiting satellite to circle the Earth, so we can expect gaps of up to an hour. When there is good coverage, the time interval still varies from 2 seconds to 2 minutes due to differences in satellite communications and signal interference around the world.

Table Schema [6]

- mmsi: Anonymized vessel identifier
- timestamp: Unix timestamp
- distance_from shore: Distance from shore (meters)
- distance from port: Distance from port (meters)
- speed: Vessel speed (knots)
- course: Vessel course
- lat: Latitude in decimal degrees
- Ion: Longitude in decimal degrees
- is fishing: Label indicating fishing activity.
  - 0 = Not fishing
  - >0 = Fishing. Data values between 0 and 1 indicate the average score for the position if scored by multiple people.
  - -1 = No data

Vessel type of Trawler’s speed between is fishing and Not fishing show there are huge difference, as shown in Figure 2.

DOI: 10.54254/2755-2721/27/20230088

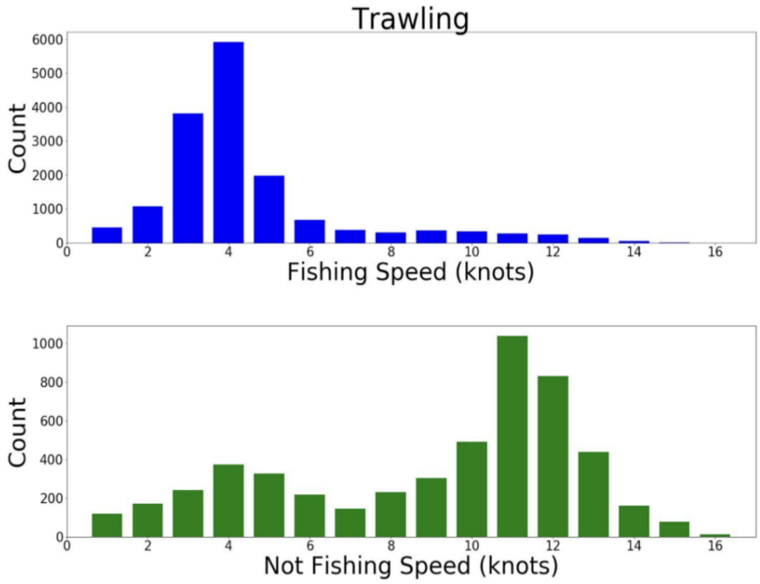

Figure 2. Fishing speed between is fishing and not fishing

4.3. Predict model

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both classification and regression problems in ML. It is based on the concept of ensemble learning, which is the process of combining multiple classifiers to solve complex problems and improve model performance (Figure 3).

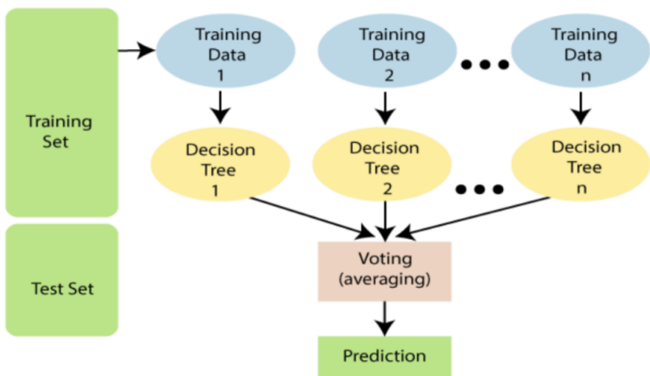

Figure 3. Random Forest model [7]

This paper categorized the data by vessel type and implemented the best fit machine learning model for training. It can be found that Longliners and Trawlers are more suitable for training with random forest models, but Purse-seines are more suitable for GB models (Gradient Boosting).

4.4. Detecte In_Area_of_Interest(processing)

Our team utilized the Marianas Trench Marine National Monument (MPA) - which encompasses Guam - as the designated region of interest for the "in_area_of_interest" code (Figure 4). This location was selected due to its exceptional characteristics, such as its depth and biodiversity, making it an ideal location for scientific research and conservation. To establish if a point fell within the area of interest, we employed the ray casting method. This method entails drawing a horizontal ray from the point of interest and counting the number of times it intersects a line segment that outlines the polygon. If the number of intersections is odd, the point is considered to be within the area of interest, while an even number of intersections implies that the point is outside the polygon [8].

The Python code for this method was rigorously tested and compared to other Python-based methods for determining if a point is inside a polygon, and it was found to be highly precise and consistent. This implementation of the ray casting method is a straightforward and efficient way to determine if a point is within a polygon, and it can easily be applied to other regions of interest with similar features.

It's worth noting that while the ray casting method only provides information on whether a point is inside or outside the polygon, it demonstrates the suitability of the geospatial data contained in an Environmental Systems Research Institute (ESRI) shapefile for this algorithm. Additionally, it has the potential to be applied to more advanced algorithms that can also determine the distance from the polygon and the distance inside the polygon, providing a more comprehensive analysis of the area of interest [9].

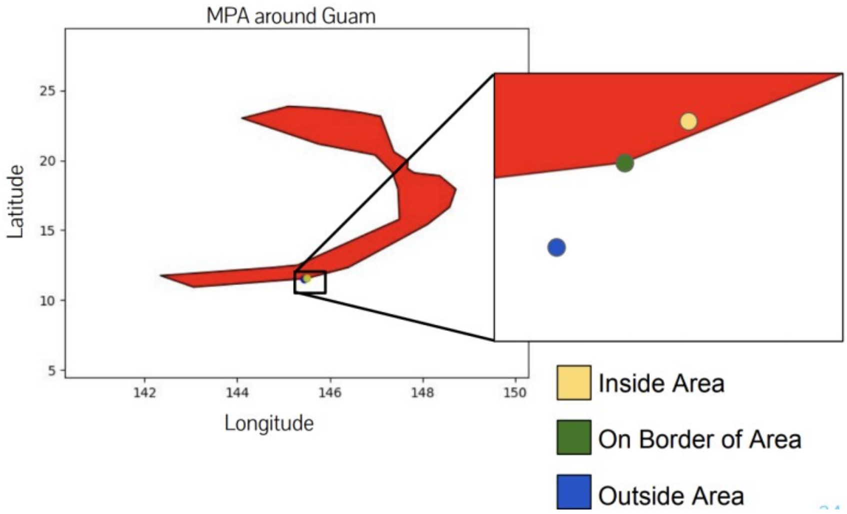

Figure 4. Interest area model output

5. Results

5.1. Model performance

Figure 5 shows the summary of the model performances.

	Accuracy	F1-score
Longliners (RF + 24h windows)	0.99	0.95
Trawlers (RF + 24h windows)	0.98	0.87
Purse Seines (GB + 6h windows)	0.88	0.45*

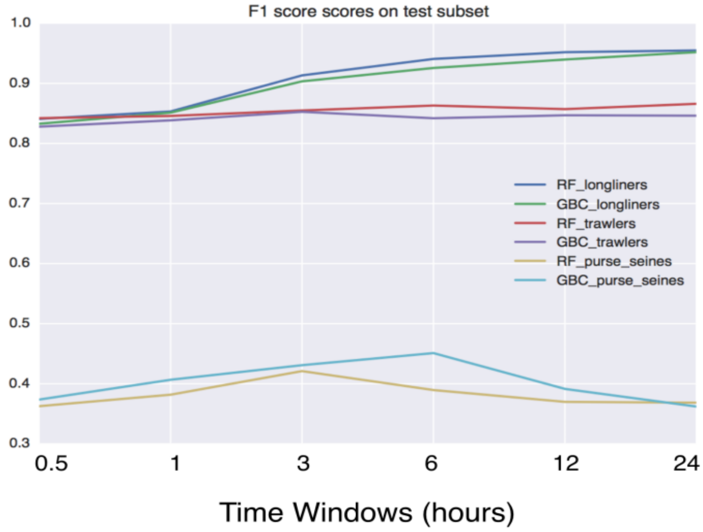

Figure 5. Summary of model performances.

5.2. Model summary

model_6

RF

longliners

course_norm_sin_cos + window_1800 + window_3600 + window_10800 + window_21600 + window_43200 + window_86400

===

Best parameters for model_6_RF_longliners are:

{'max_features': 7, 'min_samples_split': 15, 'n_estimators': 50}

===

(data) | Accuracy | Recall | F1-Score |

----

train | 0.98090 | 0.99260 | 0.98112 |

cross | 0.89860 | 0.92619 | 0.92354 |

===

top features for best classifier:

measure_coursestddev_43200

measure_coursestddev_43200_log

measure_speedstddev_43200

measure_coursestddev_86400_log

measure_speedstddev_86400_log

measure_speedstddev_21600

measure_speedavg_1800

measure_coursestddev_86400

measure_speedavg_86400

Measure_speedstddev_43200_log

5.3. Vessel route forecast and fishing forecast

Figure 6 shows the Simple output of vessel track classification and prediction.

DOI: 10.54254/2755-2721/27/20230088

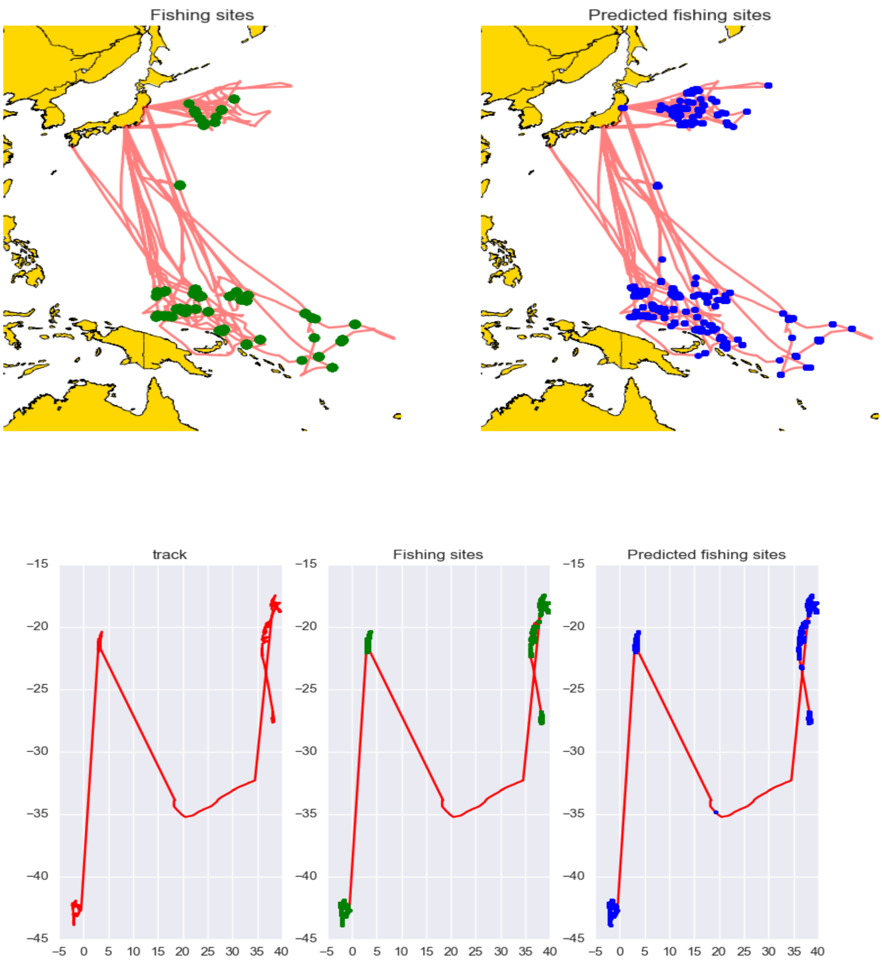

Figure 6. Simple output of vessel track classification and prediction.

6. Conclusion

The project aims to address the problem of illegal, unreported, and unregulated (IUU) fishing, which poses a significant threat to the sustainability of fish stocks and the overall health of marine ecosystems. It calls for individuals and governments to take responsibility for protecting seafood populations and oceans through the use of advanced technology and cross-sector collaboration.

To combat IUU fishing, the project suggests utilizing tools such as satellite imagery, vessel tracking systems, and other monitoring devices. These technologies enable real-time detection and documentation of illegal fishing activities, and organizations like Global Fishing Watch also play a vital role.

By identifying the areas where IUU fishing is taking place, law enforcement agencies can focus their efforts and resources on these specific regions, instead of conducting wide-ranging, time-consuming investigations. Furthermore, the detailed records of a vessel's activities provided by these technologies, can aid in the prosecution of illegal fishers.

The project also emphasizes the importance of cooperation between governments, non-governmental organizations, and the private sector to effectively combat IUU fishing. By working together, all parties can share information, coordinate enforcement efforts and develop policies to protect oceans and seafood resources.

DOI: 10.54254/2755-2721/27/20230088

In the future studies, it will examine how SAR, a satellite tool, can help with monitoring illicit fishing practices. The combination of optical and SAR sensors can gather information on ships and their activities even when there aren't AIS signals, increasing the accuracy of ship detection [10].

Overall, the project aims to raise awareness about the gravity of IUU fishing and the need for action to protect our oceans and seafood resources. It proposes the use of advanced technology for detection and deterrence, and the importance of collaboration among all sectors to effectively address the problem.

References

[1] Jarred Byrnes, Jonathan Gessert, Edward Kerrigan, Jonathan Matteson, IUU Fishing Detection 2.3, 4.5, 2017, http://seor.vse.gmu.edu/~klaskey/Capstone/IUU_Fishing/assets/deliverables/final-report 2017spr.pdf.

[2] Jarred Byrnes, Jonathan Gessert, Edward Kerrigan, Jonathan Matteson IUU Fishing Detection 2.4, 2017, http://seor.vse.gmu.edu/~klaskey/Capstone/IUU_Fishing/assets/deliverables/final-report-2017spr.pdf.

[3] Jarred Byrnes, Jonathan Gessert, Edward Kerrigan, Jonathan Matteson IUU Fishing Detection 3.0, 2017, http://seor.vse.gmu.edu/~klaskey/Capstone/IUU_Fishing/assets/deliverables/final-report-2017spr.pdf.

[4] PEW Environment Group The costs of IUU fishing to the EU, 2008, https://www.pewtrusts.org/~/media/legacy/uploadedfiles/peg/publications/report/iuu20briefing20englishpdf.pdf.

[5] Seafood Watch, Stop illegal fishing, 2022, https://www.seafoodwatch.org/seafood-basics/sustainable-solutions/stop-illegal-fishing.

[6] Global Fishing Watch Anonymized AIS training data, https://globalfishingwatch.org/data-download/datasets/public-training-data-v1.

[7] Abouheaf M., Qu S., Gueaieb W., Abielmona R. and Harb M., "Responding to Illegal Activities Along "the Canadian Coastlines Using Reinforcement Learning," in IEEE Instrumentation & Measurement Magazine, vol. 24, no. 2, pp. 118-126, April 2021, doi: 10.1109/MIM.2021.9400967.

[8] Marzuki M.I., Garello R., Fablet R., Kerbaol V. and Gaspar P., "Fishing gear recognition from VMS data to identify illegal fishing activities in Indonesia", OCEANS 2015 - Genova, 2015, pp. 1-5, doi: 10.1109/OCEANS-Genova.2015.7271551.

[9] Bellemare, M.G., Candido, S., Castro, P.S. et al. Autonomous navigation of stratospheric balloons using reinforcement learning, Nature 588, 77–82 (2020). https://doi.org/10.1038/s41586-020-2939-8.

[10] Kurekin A. A., Loveday B.R., Clements O., Quartly G.D., Miller P. I., Wiafe G., Adu Agyekum K.,

[11] Operational Monitoring of Illegal Fishing in Ghana through Exploitation of Satellite Earth Observation and AIS Data. Remote Sensing. 2019; 11(3):293. https://doi.org/10.3390/rs11030293.