Indonesian Journal of Electrical Engineering and Computer Science
Vol. 37, No. 3, March 2025, pp. 1626~1636
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v37.i3.pp1626-1636

Machine learning based optimized sea vessel location detection to identify illegal fishing

Ajay Kumar, Kakoli Banerjee, Pradeep Kumar, Harsha K. G., Vinooth P., Pankaj Kumar, Saumya, Nishant, Satyam Verma, Vidushi Bhardwaj
Department of Computer Science and Engineering, JSS Academy of Technical Education, Noida, India

Article Info

Article history:

Received Jun 21, 2024
Revised Sep 23, 2024
Accepted Sep 30, 2024

Keywords:

Illegal fishing
Machine learning
Sea vessels location
Unregulated fishing
Unreported fishing

ABSTRACT

Illegal fishing is a pervasive and destructive global issue that poses a significant threat to maritime ecosystems and the resilience of fisheries. Illegal, unregulated, and unreported (IUU) fishing leads to the extinction of the fishing population. Many researchers have presented various approaches to detect illegal fishing, for example, using sensors, image recognition, and convolutional neural networks (CNNs) but each one has some limitations. Our research aims to compare different vessel gear types to select the best vessel container that can be easily monitored and less prone to illegal activities. To achieve this, our research proposed an optimization method that involves hyperparameter selection using a genetic algorithm instead of a grid search. Using the crossover method of the genetic algorithm our model is compatible with larger datasets and unknown search space which is not possible in the baseline algorithm i.e. grid search. Moreover, after applying the genetic hyperparameter optimization technique, the overall accuracy, recall, and F1 score is increased for all vessel types significantly. While comparing our optimized model with the existing model with different evaluation metrics, our model's performance is outstanding.

Corresponding Author:

Ajay Kumar
Department of Computer Science and Engineering, JSS Academy of Technical Education
C-20/1, Sector 62, Noida, Uttar Pradesh, 201301, India
Email: ajay.itcs@gmail.com

1. INTRODUCTION

Overfishing and illegal fishing are major dangers to marine ecosystems and the global security of food. Foreign fleets exploiting marine resources, particularly in regions such as West Africa, have a negative impact on the local economy and communities. The discovery that illegal, unreported, and unregulated (IUU) fishing is a significant problem for the region has increased worries about the sustainability of the fishery [1]. Because of methods for reconstructing previous captures and correcting misreporting, the actual quantity of fish that has been removed could be significantly larger than recorded [2]. Fisheries management requires rigorous monitoring and enforcement of rigid laws [3]. The fishing industry supported the Gulf's economy for generations, yet statistics on catch are still inaccurate today. Calculating these captures often yields dramatically different baselines for past catches, raising questions about how closely publicly reported data represent reality and the long-term sustainability of certain management decisions [4]. The present state of global fisheries provides an outlook on marine fisheries globally with a focus on mitigating the level of sustainability as well as industrial exploitation which results in unceasing economic growth [5]. Combating these issues requires creative strategies that make use of technology and data-driven solutions. This paper presents a framework of global fishing watch (GFW) to identify unlawful fishing in the ocean using automatic identification system (AIS) data. In the APFIC region, IUU fishing is a serious issue that is projected to cost up to US$23 billion a year [6]. The harmful impacts of unlawful foreign capturing on the viability and profitability of Somali fisheries are demonstrated in groundbreaking research entitled “Securing Somali fisheries” [7]. In order to solve the abovementioned challenge, Ball [8] provided a unique method for assessing IUU fish and bird by-catch. For providing empirically trustworthy data and confidence intervals, the approach uses an animated model that inculcates “noticed” and “unnoticed” traces of IUU fishing cases. Agnew [9] presented a detailed description of the 1999-adopted catch document scheme, which attempted to lower illicit fishing using trade-related measures. The methodology entails recording specifics about the catch, such as the vessel, the location, and the time of capture, and confirming that the fishing activity was lawful. But it requires collaboration from a lot of different people, such as importers, exporters, fishers, and processors. MRAG in 2016 [10] presented a Monte Carlo simulation to measure the extent of illicit fishing under the sea (IUU) linked to eleven primary risks in four risk categories: i) fishing without a license or authorization; ii) falsifying information; iii) violating other license requirements; and iv) risks arising after harvest.

Aanes *et al.* [11] proposed a Norwegian Economic Zone’s satellite-based maritime tracking system that offers comprehensive data on individual vessel voyages. The main limitation of this system is its complex structure. Several machine learning techniques were used by Bellin *et al.* [12] to forecast the quantity and distribution of marine organisms. Models like random forests (RF), support vector machines (SVM), and neural networks were used in the study. The interactions between environmental factors and species distributions that are intricate and non-linear can be handled by these models. The main drawback of machine learning techniques is their heavy reliance on big, high-quality training datasets. Free *et al.* [13] proposed a study to assess the extent of illicit fishing and its negative impacts on a particular endangered fish species using a combination of quantitative and qualitative methodologies, offering important information for conservation efforts. A thorough evaluation of the literature was done by Li *et al.* [14] in order to gather and evaluate the current literature on maritime surveillance employing optical sensing methods. sensor fusion, object identification and tracking, picture processing, and surveillance. A ship detection technique that makes use of convolutional neural networks (CNNs) was presented by Huang *et al.* [15]. Deep learning architectures called CNNs are created especially for image processing applications. In order to teach a CNN model, the distinguishing characteristics of ships, a sizable dataset of tagged remote sensing photos was used as training data. The computational intensity of CNNs is a drawback, particularly when handling huge datasets and intricate image processing jobs. It frequently takes a lot of processing power and memory to train CNNs. Deep CNNs were used by Grace *et al.* [16] to recognize ships from high-resolution remote sensing photos. The process entailed using a sizable dataset of annotated remote sensing photos to train a CNN model in order to identify hierarchical characteristics typical of ships in various environmental settings. A drawback of deep neural network (DNN) models is their huge number of layers and parameters, which require a lot of processing power and memory when training on big datasets. The over feat system, which is built on CNNs for integrated recognition, localization, and detection tasks in computer vision, was suggested by Sermanet *et al.* [17]. The over feat system's unified architecture for object comprehension is one of its main benefits but its cost and complexity intricacy may be a drawback. A comprehensive assessment of the literature was carried out by Khatami *et al.* [18] to assess the state of the field regarding ship detection in optical remote sensing pictures. Studies, research papers, and technical reports that were published in scholarly journals and conference proceedings were included in the study. The integration of deep learning architectures and sophisticated machine learning algorithms, among other recent developments in ship identification approaches, was emphasized in the article. Mittendorf *et al.* [19] used machine learning approaches to determine sea conditions based on an analysis of in-service data obtained from a cargo vessel. Preprocessing the operating data from the vessel, including motion, acceleration, and other pertinent sensor readings, was part of the technique. The preprocessed data is used to train and assess a variety of machine learning models, including RF, SVM, and neural networks, with the purpose of classifying distinct sea states. RF algorithms were used by Wang *et al.* [20] to classify ships based on static information obtained from AIS data. Preprocessing AIS data to extract pertinent static properties such as ship class, length, breadth, and draft was part of the technique. The capacity of RF to produce high classification accuracy is a major benefit when it comes to ship classification. The low interpretability of RF algorithms in comparison to more straightforward models such as logistic regression or decision trees may be a drawback.

Zhong *et al.* [21] classified vessels based on space-based AIS data by utilizing RF algorithms. Preprocessing AIS data from satellite views was part of the technique to extract pertinent information such as vessel type, size, and speed. Computational complexity is a drawback that RF may have the computational complexity, particularly when working with big datasets and high-dimensional feature spaces. CNNs are used by Zhang *et al.* [22] to recognize ships from high-resolution remote sensing photos. A R-CNN model is trained using a sizable dataset of tagged photos of ships with background clutter as part of the process. High detection accuracy is an advantage, but significant quantity of labeled data is usually needed to train a R-CNN model.

Our study compares various types of vessel gear in order to identify the most readily monitored and less likely to be used for illicit purposes vessel container. The research we conducted suggested an optimization technique that selects hyperparameters via a genetic algorithm rather than a grid search in order to accomplish this. This approach is suitable with more extensive datasets and unknown search spaces, which is not achievable with the baseline technique, i.e. grid search, thanks to the genetic algorithm's crossover method. Furthermore, all vessel types experience a considerable increase in overall accuracy, recall, and F1 score following the use of the genetic hyperparameter optimization technique.

The work is organized as follows. Section 2 delineates the theoretical fundamentals of machine learning models employed in our research work. The proposed system methodology that has been chosen to satisfy the see vessel location detection data requirements is described in section 3. The implementation of a genetic algorithm to optimize the machine learning model’s accuracy and geographical map representation of actual and predicted sites to show that the illegal areas are described in section 4. Python Software is used to obtain results. The study's conclusions are presented in section 5.

2. THEORETICAL FUNDAMENTALS

This section discusses some of the theoretical fundamentals of conducting the experiment. Different algorithms are used to train the model and compared based on evaluation metrics like F1 score, recall and accuracy. Here is a brief description of the algorithms used.

2.1. Random forest

RF is a classifier that aggregates the results of several decision trees applied to various dataset subsets. It operates by generating many “decision trees” in the process of training. With the help of the insights from several trees, this cooperative approach to “decision-making” produces accurate and consistent outcomes [23]-[25]. In our research after the selection of hyperparameters, a model is trained using the RF algorithm which is further evaluated using different evaluation metrics. The architecture of the RF algorithm is given in Figure 1.

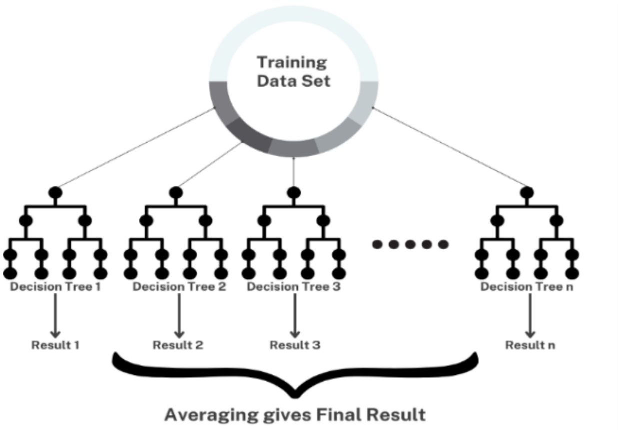

Figure 1. Random forest architecture

2.2. Gradient boosting classifier

Gradient boosting classifier (GBC) is a powerful enhancing method that uses gradient descent to train every fresh model to reduce the loss function of the preceding model, such as the mean squared error [26]-[30]. This approach trains a new weak model to reduce the slope of the loss function with respect to the predictions made by the current ensemble in each iteration. Then, the predictions of the new model are included in the group, and the procedure is repeated until a stopping condition is satisfied. In our research after the selection of hyperparameters, a model is trained using GBC algorithm which is further evaluated using different evaluation metrics [31]-[33]. The architecture of the RF algorithm is given in Figure 2.

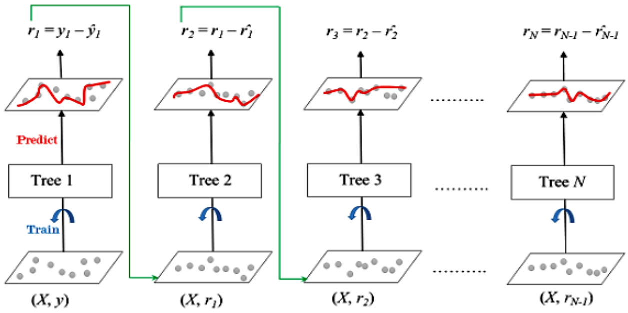

Figure 2. Gradient boosting classifier architecture

2.3. Genetic algorithm

Genetic algorithm is an iterative approach in which the population undergoes several modifications to find the best population [34]-[38]. This algorithm mainly focuses on finding the best individuals so that new offspring can be generated using them. It gradually evolves as the best solution after continuous improvement [39], [40]. In our research hyperparameters are selected and optimized using an iterative method of genetic algorithm to select the best hyperparameter before training our model. The flowchart of the genetic algorithm is given in Figure 3. The proposed methodology used in the research work is described in the next section.

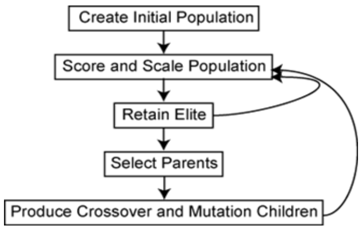

Figure 3. Flow chart of genetic algorithm

3. METHOD

This section describes the proposed research methodology in which step-wise procedure is given. As an illustration of the thorough approach used in the proposed technique, Figure 4 shows this process in graphic form, with each stage shown in a sequential flow. The methodical procedure for maximizing sea vessel position detection to detect illicit fishing is explained in this flow diagram, which can be used as a reference for implementation and comprehension.

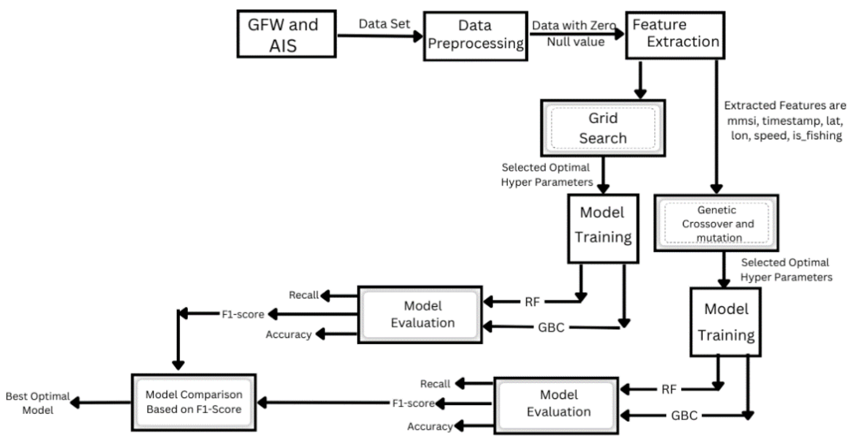

Figure 4. Proposed research methodology flow diagram

3.1. Data collection
Datasets are collected from GFW and AIS. GFW uses AIS data to track and examine fishing activity throughout the globe. The dataset consists of vessel identity using a unique maritime mobile service identity (MMSI) number, identity, course, speed, type, position (latitude, longitude), and navigational status. Further, the collected datasets are sent for data preprocessing.

3.2. Data preprocessing
Datasets were already labeled, cleaned and normalized. Some data with null value was present that was removed for further processes. It was done by the ‘pandas’ module. After data preprocessing all the data with null values is eliminated. Then, the preprocessed datasets are sent for data feature extraction.

3.3. Features extraction
In the data feature extraction phase, several key features were extracted from the datasets to aid in the analysis of sea vessel activities. These features include the MMSI number, timestamp, latitude (lat), longitude (lon), speed, an indicator of whether the vessel is fishing, distance from shore, and distance from port. These features provide critical information for assessing vessel behavior and identifying potential illegal fishing activities. With these features in hand, the next step is to optimize the performance of machine learning models by selecting the most effective hyperparameters. This optimization is achieved using two distinct methods: grid search and genetic algorithm. Grid Search involves a systematic, exhaustive search over a predefined set of hyperparameters to find the optimal combination, while the genetic algorithm employs evolutionary techniques to iteratively improve hyperparameters based on simulated natural selection. Both methods aim to enhance the model's ability to accurately detect illegal fishing activities by refining the hyperparameters to best fit the extracted features.

3.4. Grid search
Optimal hyperparameters are selected using the grid search method. It uses multiple combination techniques to find the best hyperparameter. It was done by the sklearn module. Hyperparameters with range were defined for both the RF and GBC models. For RF model: Param-grid = {'n-estimators': [100, 200, 300], 'max-depth': [10, 20, 30], 'min-samples-split': [2, 5, 10]}. For GBC model: Param-grid = {'learning-rate’: [0.05, 0.1, 0.2], 'max-depth': [2, 3, 5], 'min-samples-leaf': [3, 10, 20]}. Based on these ranges best hyperparameters are selected for respective models by grid search for further model training.

3.5. Genetic crossover and mutation
Optimal hyperparameters are selected using a genetic algorithm. It uses operations such as selection, crossover, and mutation to repeatedly develop a population of potential solutions. Different combinations of parameters are selected and then crossover is performed. Crossover is nothing but the mixing of best guesses from the previous rounds. After crossover certain members of the population are subjected to random modifications to keep diverse search space this process is called mutation. In our case, the population size is set as 10 and the mutation rate as 0.1.

3.6. Model training
Different hyperparameter optimization strategies are used to train the RF and GBC models in order to maximize sea vessel location detection for identifying illicit fishing. Grid search is used to optimize the RF model by methodically assessing every conceivable combination of hyperparameters and identifying the model that performs the best.

3.7. Model evaluation
Evaluation metrics like accuracy, recall, and F1 score are used to evaluate the performance of each model. RF and GBC models trained using the grid search method and genetic method are evaluated separately by evaluation metrics. In the next step, a comparison of all the models is done using these model evaluation data.

3.8. Model comparison
Matplotlib is used to plot a graph based on the F1 score of each model. The best model and best hyperparameter selection technique are easily identified by reading the graph. After the detailed description of the proposed research methodology given above, the implementation results with discussion are provided in the next section.

4. RESULTS AND DISCUSSION

In this section, the implementation results are explained with the help of some performance parameters and geographical models. Using graphical representations of fishing tracks, such as tracks of fishing vessels and the estimated fishing locations and areas, vessel movements, and fishing patterns can be explained in a straighter forward and easy to follow manner. For a given MMSI in our research, two images of the tracks are generated, one indicating the actual fishing sites (in red) and another image with the predicted fishing sites (in green) as shown in Figures 5 and 6.

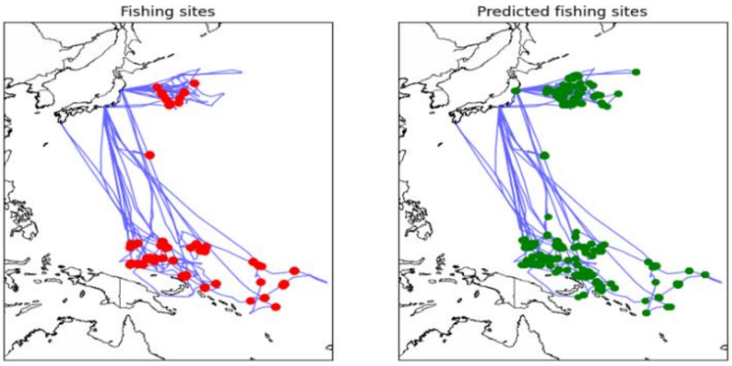

Figure 5. Geographical representation of actual and predicted sites

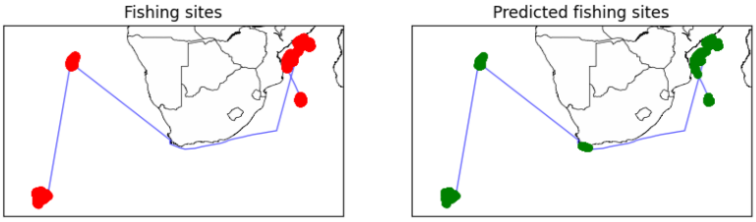

Figure 6. Zoomed geographical representation of actual and predicted sites

Our research proposed a model performance table based on the performance of different models after genetic hyperparameter optimization. Through a detailed analysis of these six model pairs' F1 scores and other pertinent metrics, as shown in Table 1, our research proposed which algorithms work best for classifying the various kinds of vessels in our tracking system GBC is the best algorithm for purse seines and for trawlers, RF is best. Concerning this table, the best model can be implemented on a large scale to improve accuracy and efficiency. The performance matrix of sea vessel models after genetic optimization is given in Table 1.

Table 1. The performance matrix

Model	Accuracy	Recall	F1-score
RFLongliners	0.892	0.906	0.912
GBCLongliners	0.882	0.890	0.902
RFTrawlers	0.945	0.957	0.934
GBCTrawlers	0.940	0.956	0.929
RFPurseSeines	0.895	0.814	0.986
GBCPurseSeine	0.897	0.840	0.991

4.1. Comparison with baseline algorithms

Previous research has utilized various machine learning methods, including gradient boosting and RF, to tackle the challenge of detecting illegal fishing activities. These models have proven effective but often face limitations in terms of accuracy and predictive capability. In our study, we enhance these traditional approaches by refining those using genetic algorithms, which introduce evolutionary optimization techniques to fine-tune hyperparameters and improve model performance. This methodological advancement aims to achieve more precise detection of fishing activities and better prediction of vessel movements. In this section, we conduct a comparative analysis between the performance of existing algorithms and our optimized models. This comparison highlights the improvements brought by the genetic algorithm-enhanced models, demonstrating their superior ability to accurately identify illicit fishing activities and predict vessel behavior more effectively than conventional methods.

4.1.1. RF vs RF optimized with genetic algorithm

Because the hyperparameters are tailored according to the unique data and job, optimized RF performs better (for example, higher accuracy, recall, and F1-score). Baseline RF is quicker to train since it does not require any hyperparameter searches. Simply train the model with the default values. Figures 7 and 8 demonstrates the metrics obtained before and after genetic algorithm optimization in model-1 purse seines. Baseline RF accuracy: 0.82600, optimized RF accuracy: 0.88740, baseline RF recall: 0.75532, optimized RF recall: 0.821435, baseline RF F1-score: 0.49477, optimized RF F1-score: 0.74957.

```
model_1
RF
purse_seines
course_norm_sin_cos + window_1800

================================================

Best parameters for model_1_RF_purse_seines are:
 {'max_features': 3, 'min_samples_split': 15, 'n_estimators': 100}

================================================
(data) | Accuracy | Recall | F1-Score |
------------------------------------------------
train | 0.98270 | 0.99080 | 0.98284 |
cross | 0.82600 | 0.75532 | 0.49477 |
================================================
```

Figure 7. Evaluation metrics of RF model

```
Genetic_model_1
RF
purse_seines
course_norm_sin_cos + window_1800

================================================

Best parameters for model_1_RF_purse_seines are:
 {'n_estimators': 100, 'max_features': 3, 'min_samples_split': 15}

================================================
(data) | Accuracy | Recall | F1-Score |
------------------------------------------------
train | 0.98450 | 0.99300 | 0.98463 |
cross | 0.88740 | 0.82143 | 0.74857 |
================================================
```

Figure 8. Evaluation metrics of RF model with GA optimization

4.1.2. GBC vs GBC optimized with genetic algorithm

The enhanced GBC with a genetic algorithm outperforms the baseline model, albeit at the cost of additional computing time and complexity. This trade-off is frequently justifiable in situations where the maximum potential performance is required. Figures 9 and 10 demonstrates the metrics obtained before and after genetic algorithm optimization in model-1 purse seines. Baseline GBC accuracy: 0.82860, optimized GBC accuracy: 0.88720, baseline GBC recall: 0.78191, optimized GBC recall: 0.98120, baseline GBC F1-score:0.50719, optimized GBC F1-score: 0.97612.

```
model_1
GBC
purse_seines
course_norm_sin_cos + window_1800

================================================

Best parameters for model_1_GBC_purse_seines are:
 {'learning_rate': 0.2, 'max_depth': 5, 'min_samples_leaf': 20}

================================================
(data) | Accuracy | Recall | F1-Score |
------------------------------------------------
train | 0.98540 | 0.99600 | 0.98555 |
cross | 0.82860 | 0.78191 | 0.50719 |
================================================
```

Figure 9. Evaluation metrics of GBC model

```
Genetic_model_1
GBC
purse_seines
course_norm_sin_cos + window_1800

================================================

Best parameters for model_1_GBC_purse_seines are:
 {'learning_rate': 0.2, 'max_depth': 5, 'min_samples_leaf': 20}

================================================
(data) | Accuracy | Recall | F1-Score |
------------------------------------------------
train | 0.98680 | 0.99660 | 0.98693 |
cross | 0.88720 | 0.98120 | 0.97612 |
================================================
```

Figure 10. Evaluation metrics of the GBC model with GA optimization

4.2. Overall model comparison

Our research proposed an overall model comparison between the existing methods i.e. grid search hyperparameter selection and the optimized method i.e. genetic hyperparameter selection. Concerning the graph, as shown in Figures 11 and 12, of F1 score it can be easily determined that after applying genetic algorithm the overall F1-score is increased for both GBC and RF methods significantly.

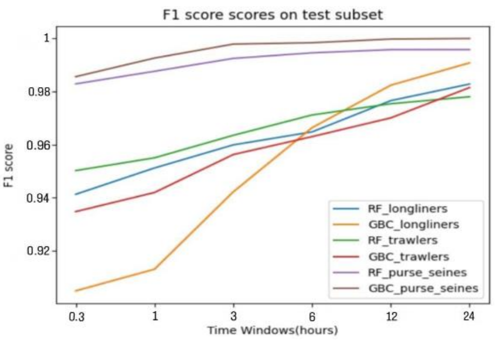

Figure 11. F1 score of different mode

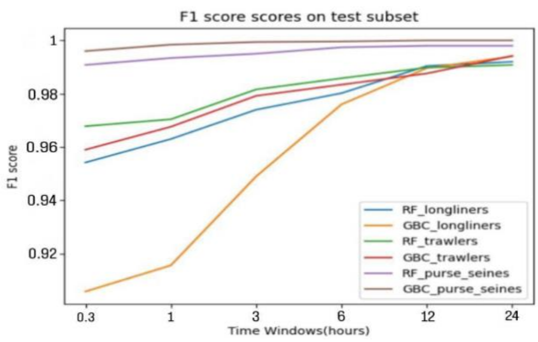

Figure 12. F1 score of different models after genetic optimization

5. CONCLUSION

The research concludes that the issues associated with unlawful and unsustainable fishing practices can be resolved by selecting best vessel container among different vessel gear types of sea vessels using machine learning algorithms. The modelling of geographical areas is used to predict correct fishing sites helps in combating IUU fishing. The proposed methodology uses two algorithms i.e. RF and gradient boosting to select best vessel container. Also, with genetic optimization of hyperparameter selection on the selected vessel model gives better results in terms of accuracy, F1 score and recall as compared to baseline hyperparameter selection techniques. The dataset used to do this research work is GFW AIS which is the larger datasets and compatible with our proposed algorithms. Through our research work, IUU fishing which is a significant problem of any government can be solved. We can easily predict the locations where illegal fishing is going to happen. This will be a great impact on the society as well due to the sustainability of fisheries. As a future work, in addition to AIS data, combining other sources like as satellite photographs, fisheries eyewitness reports, and vessel surveillance system (VMS) data can offer a more complete view of fishing operations and improve detection algorithm's accuracy.

REFERENCES

[1] S. C. Clarke *et al.*, "Global estimates of shark catches using trade records from commercial markets," *Ecology Letters*, vol. 9, no. 10, pp. 1115–1126, Oct. 2006, doi: 10.1111/j.1461-0248.2006.00968.x.
[2] J. Hentati-Sundberg, J. Hjelm, and H. Österblom, "Does fisheries management incentivize non-compliance? Estimated misreporting in the Swedish Baltic Sea pelagic fishery based on commercial fishing effort," *ICES Journal of Marine Science*, vol. 71, no. 7, pp. 1846–1853, Oct. 2014, doi: 10.1093/icesjms/fsu036.
[3] D. Zeller, M. Ansell, V. Andreoli, and K. Heidrich, "Trends in Indian Ocean marine fisheries since 1950: synthesis of reconstructed catch and effort data," *Marine and Freshwater Research*, vol. 74, no. 4, pp. 301–319, Jan. 2023, doi: 10.1071/MF22148.
[4] K. L. Astles, M. G. Holloway, A. Steffe, M. Green, C. Ganassin, and P. J. Gibbs, "An ecological method for qualitative risk assessment and its use in the management of fisheries in New South Wales, Australia," *Fisheries Research*, vol. 82, no. 1–3, pp. 290–303, Dec. 2006, doi: 10.1016/j.fishres.2006.05.013.
[5] D. Pauly, *5 Easy Pieces : How Fishing Impacts Marine Ecosystems*. Waashington D.C.: Island Press, 2010.
[6] F. Smith and Lee, "Combating and eliminating illegal, unreported and unregulated (iuu) fishing in the asia-pacific region," Bangkok, Thailand, 2015. [Online]. Available: https://www.fao.org/3/ca4509en/ca4509en.pdf.
[7] S. M. Glaser, P. Roberts, R. Mazurek, K. Hurlburt, and L. Kane-Hartnett, "Securing Somali Fisheries," *Secure Fisheries report* 2015.
[8] I. Ball, "An alternative method for estimating the level of illegal fishing using simulated scaling methods on detected effort," *CCAMLR Science*, vol. 12, pp. 143–161, 2005.
[9] D. J. Agnew, "The illegal and unregulated fishery for toothfish in the southern ocean, and the CCAMLR catch documentation scheme," *Marine Policy*, vol. 24, no. 5, pp. 361–374, 2000, doi: 10.1016/S0308-597X(00)00012-9.
[10] MRAG Asia Pacific, *Towards the Quantification of Illegal, Unreported and Unregulated (IUU) Fishing in the Pacific Islands Region*, no. February. MRAG Asia Pacific, 2016.
[11] S. Aanes, K. Nedreaas, and S. Ulvatn, "Estimation of total retained catch based on frequency of fishing trips, inspections at sea, transhipment, and VMS data," *ICES Journal of Marine Science*, vol. 68, no. 8, pp. 1598–1605, 2011, doi: 10.1093/icesjms/fsr117.
[12] N. Bellin, G. Tesi, N. Marchesani, and V. Rossi, "Species distribution modeling and machine learning in assessing the potential distribution of freshwater zooplankton in Northern Italy," *Ecological Informatics*, vol. 69, p. 101682, 2022, doi: 10.1016/j.ecoinf.2022.101682.
[13] C. M. Free, O. P. Jensen, and B. Mendsaikhan, "A mixed-method approach for quantifying illegal fishing and its impact on an endangered fish species," *PLoS ONE*, vol. 10, no. 12, p. e0143960, 2015, doi: 10.1371/journal.pone.0143960.
[14] B. Li, X. Xie, X. Wei, and W. Tang, "Ship detection and classification from optical remote sensing images: a survey," *Chinese Journal of Aeronautics*, vol. 34, no. 3, pp. 145–163, 2021, doi: 10.1016/j.cja.2020.09.022.
[15] J. Huang, Z. Jiang, H. Zhang, and Y. Yao, "Ship object detection in remote sensing images using convolutional neural networks," *Beijing Hangkong Hangtian Daxue Xuebao/Journal of Beijing University of Aeronautics and Astronautics*, vol. 43, no. 9, pp. 1841–1848, 2017, doi: 10.13700/j.bh.1001-5965.2016.0755.

[16] L. K. J. Grace, K. R. Reddy, K. P. Kalyan, P. Asha, and G. Kalaiarasi, "Ship detection using high resolution remote sensing image," *AIP Conference Proceedings*, vol. 2446, 2022, doi: 10.1063/5.0108608.
[17] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "OverFeat: integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, Dec. 2013.
[18] A. Khatami, M. M. Fouladgar, and A. Roodan, "Ship detection in optical remote sensing images: a survey," *IEEE Geoscience and Remote Sensing Magazine*, vol. 6, no. 4, pp. 70–91, 2018.
[19] M. Mittendorf, U. D. Nielsen, H. B. Bingham, and G. Storhaug, "Sea state identification using machine learning—a comparative study based on in-service data from a container vessel," *Marine Structures*, vol. 85, p. 103274, 2022, doi: 10.1016/j.marstruc.2022.103274.
[20] Y. Wang, L. Yang, X. Song, and X. Li, "Ship classification based on random forest using static information from AIS data," *Journal of Physics: Conference Series*, vol. 2113, no. 1, p. 012072, 2021, doi: 10.1088/1742-6596/2113/1/012072.
[21] H. Zhong, X. Song, and L. Yang, "Vessel classification from space-based AIS data using random forest," in *2019 5th International Conference on Big Data and Information Analytics (BigDIA)*, Jul. 2019, pp. 9–12, doi: 10.1109/BigDIA.2019.8802792.
[22] S. Zhang, R. Wu, K. Xu, J. Wang, and W. Sun, "R-CNN-based ship detection from high resolution remote sensing imagery," *Remote Sensing*, vol. 11, no. 6, p. 631, 2019, doi: 10.3390/RS11060631.
[23] X. Li, Z. He, and C. Wu, "Ship recognition in optical satellite images based on deep CNN," *Remote Sensing*, vol. 11, no. 15, p. 1813, 2019.
[24] F. Chollet, "Xception: deep learning with depthwise separable convolutions," *arXiv preprint arXiv:1610.02357*, Oct. 2016, [Online]. Available: http://arxiv.org/abs/1610.02357.
[25] W. J. Lee *et al.*, "Detection and tracking for the awareness of surroundings of a ship based on deep learning," *Journal of Computational Design and Engineering*, vol. 8, no. 5, pp. 1407–1430, 2021, doi: 10.1093/jcde/qwab053.
[26] G. Matasci *et al.*, "Deep learning for vessel detection and identification from spaceborne optical imagery," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 5, no. 3, pp. 303–310, 2021, doi: 10.5194/isprs-annals-V-3-2021-303-2021.
[27] I. L. Huang, M. C. Lee, C. Y. Nieh, and J. C. Huang, "Ship classification based on AIS data and machine learning methods," *Electronics (Switzerland)*, vol. 13, no. 1, p. 98, 2024, doi: 10.3390/electronics13010098.
[28] Z. Guo, H. Qiang, S. Xie, and X. Peng, "Unsupervised knowledge discovery framework: from AIS data processing to maritime traffic networks generating," *Applied Ocean Research*, vol. 146, p. 103924, 2024, doi: 10.1016/j.apor.2024.103924.
[29] J. Liu, G. Shi, and K. Zhu, "Vessel trajectory prediction model based on ais sensor data and adaptive chaos differential evolution support vector regression (ACDE-SVR)," *Applied Sciences (Switzerland)*, vol. 9, no. 15, p. 2983, 2019, doi: 10.3390/app9152983.
[30] Z. Chu, R. Yan, and S. Wang, "Vessel turnaround time prediction: a machine learning approach," *Ocean and Coastal Management*, vol. 249, p. 107021, 2024, doi: 10.1016/j.ocecoaman.2024.107021.
[31] C. Zhang, X. Zou, and C. Lin, "Fusing XGBoost and SHAP models for maritime accident prediction and causality interpretability analysis," *Journal of Marine Science and Engineering*, vol. 10, no. 8, p. 1154, 2022, doi: 10.3390/jmse10081154.
[32] H. Arbabkhah, A. Sedaghat, M. Jafari Kang, and M. Hamidi, "Automatic identification system-based prediction of tanker and cargo estimated time of arrival in narrow waterways," *Journal of Marine Science and Engineering*, vol. 12, no. 2, p. 215, 2024, doi: 10.3390/jmse12020215.
[33] R. Yan, S. Wang, J. Cao, and D. Sun, "Shipping domain knowledge informed prediction and optimization in port state control," *Transportation Research Part B: Methodological*, vol. 149, pp. 52–78, 2021, doi: 10.1016/j.trb.2021.05.003.
[34] Y. Chen, M. Huang, K. Song, and T. Wang, "Prediction of ship traffic flow and congestion based on extreme learning machine with whale optimization algorithm and fuzzy c -means clustering," *Journal of Advanced Transportation*, no. 1, p. 7175863, 2023, doi: 10.1155/2023/7175863.
[35] H. Zhang, Y. Xiao, X. Bai, X. Yang, and L. Chen, "GA-support vector regression based ship traffic flow prediction," *International Journal of Control and Automation*, vol. 9, no. 2, pp. 219–228, 2016, doi: 10.14257/ijca.2016.9.2.21.
[36] W. Xing, J. Wang, K. Zhou, H. Li, Y. Li, and Z. Yang, "A hierarchical methodology for vessel traffic flow prediction using Bayesian tensor decomposition and similarity grouping," *Ocean Engineering*, vol. 286, 2023, doi: 10.1016/j.oceaneng.2023.115687.
[37] C. H. Chen, L. P. Khoo, Y. T. Chong, and X. F. Yin, "Knowledge discovery using genetic algorithm for maritime situational awareness," *Expert Systems with Applications*, vol. 41, no. 6, pp. 2742–2753, 2014, doi: 10.1016/j.eswa.2013.09.042.
[38] X. Zhou, Z. Liu, F. Wang, Y. Xie, and X. Zhang, "Using deep learning to forecast maritime vessel flows," *Sensors (Switzerland)*, vol. 20, no. 6, p. 1761, 2020, doi: 10.3390/s20061761.
[39] X. Wang and Y. Xiao, "A deep learning model for ship trajectory prediction using automatic identification system (AIS) data," *Information (Switzerland)*, vol. 14, no. 4, p. 212, 2023, doi: 10.3390/info14040212.
[40] D. A. Kroodsma *et al.*, "Tracking the global footprint of fisheries," *Science*, vol. 359, no. 6378, pp. 904–908, 2018, doi: 10.1126/science.aao5646.

BIOGRAPHIES OF AUTHORS

Ajay Kumar Assistant Professor in the Department of Computer Science and Engineering at JSS Academy of Technical Education, Noida. He has received his M.Tech. (Computer Engineering) from YMCA University of Science and Technology, Faridabad, India, and B.Tech. (Computer Engineering) from University Institute of Engineering and Technology, M.D.U. Rohtak. He has more than 12 years of academic experience. He has contributed 11 research papers in international journals and 12 research papers in international/national conferences/proceedings and edited books. He has added 02 patents with his name out of which 01 has granted. His areas of research are machine learning, IoT, and network security. He has organized various workshops and FDPs in these areas. He is the lifetime member of International Association of Engineers (IAENG). He can be contacted at email: ajaykrverma@jssaten.ac.in.

Dr. Kakoli Banerjee HOD of Computer Science and Engineering Department, JSS Academy of Technical Education, Noida. She graduated from IET, Kanpur – B.Tech. (CSE), did her Post Graduation from MNNIT, Allahabad – M.Tech. (CSE) and her Doctorate (Ph.D.) from Shobhit University, Saharanpur. She has 22 years' experience in academics and three years in industry. She has published around 50 research papers in international journals of repute indexed in SCI, Scopus, ESCI, Google Scholar, and other eminent databases. Currently, she is working for a funded project, sanctioned by the Collaborative Research and Innovation Program (CRIP) under TEQIP- 3 by Dr. A. P. J. Abdul, Kalam Technical University Uttar Pradesh, Lucknow. She is a member of IEEE and other professional societies such as ISTE, and CSI. She has been a reviewer and guest editor of several reputed journals and books. She has also served in many conferences as session chair. She has four granted patent to her name and eleven published. She has developed video content for Dr. A.P.J. Abdul Kalam Technical University, Lucknow and Swayamprabha Chanel of Government of India for subjects like Operating System and Database Management System. She has published many books including - "Industry 4.0: research trends, challenges and future of AI in data science", CRC Press, Taylor and Francis Group. She can be contacted at email: kakolibanerjee@jssaten.ac.in.

Dr. Pradeep Kumar is Ph.D. computer engineering and engineering at Department of Computer Engineering Shobhit Institute of Engineering and Technology (Deemed-to-be University), Meerut, 250110. He has obtained his M.Tech. in Computer Science and Engineering Department of Computer Engineering Shobhit Institute of Engineering and Technology (Deemed-to-be University), with first class. He obtained his B.Tech. in computer engineering and engineering degree from College of Engineering Roorkee, India in 2006 with first class. He can be contacted at email: pradeep8984@jssaten.ac.in.

Harsha K. G. received the M.Tech. degree in Computer Science and Engineering from VTU, Karnataka, India, in 2017. He is currently working as an Assistant Professor with the JSS academy of technical education, Noida, U.P, India. AKTU. His research interests include blockchain security, cloud security, IoT and cyber security. He is a member of IAENG. He can be contacted at email: harshakg.official@gmail.com.

Vinooth P. received the B.E. degree in Information Science Engineering from VVIET, and M.Tech. in CSE from VTU-RC, Karnataka. He is Working as an Assistant Professor in the Department of CSE, JSS University, Noida from past 2.5 years. Currently pursuing Ph.D. in Cyber security at JSSSTU, Mysuru. He can be contacted at email: vinooth@jssaten.ac.in.

Pankaj Kumar is undergraduate final year B.Tech. (Computer Science and Engineering) student in JSS Academy of Technical Education, Noida. Over his time as an undergraduate, he studied a wide range of computer science topics, including programming languages, data structure and algorithms. In his third year it was his first experience in the research field, when he collaborated with a professor on a project of machine learning techniques. He is a passionate software developer who pursues his research efforts in parallel with them. With the use of several frameworks and full-stack technologies, he has contributed to a variety of projects. As a System Engineer at Tata Consultancy Services (TCS), he is about to start a new chapter in his career in upcoming months. He can be contacted at email: pankajverma4950@gmail.com.

Saumya is undergraduate final year B.Tech. (Computer Science and Engineering) student in JSS Academy of Technical Education, Noida. She has shown a great talent for programming logic building across her academic efforts, laying a solid base for her future goals. She involved herself in the complexities of practical coding problems throughout her software development internship, sharpening her talents. As a System Engineer at Tata Consultancy Services (TCS), she is about to start a new chapter in her career in upcoming months. With her determination, she is well positioned to establish a prosperous and satisfying career in the dynamic field of software development. She can be contacted at email: saumyasinghdeshwal@gmail.com.

Nishant is undergraduate final year B.Tech. (Computer Science and Engineering) student in JSS Academy of Technical Education, Noida. His research interests include machine learning (ML), with a focus on developing innovative algorithms and applications. He is actively developing his knowledge of API integration and design. He is a valuable addition to any production environment because of his technical knowledge and collaborative nature, which promote teamwork and creative problem-solving to achieve goals. He can be contacted at email: nishantnehra03@gmail.com.

Satyam Verma is undergraduate final year B.Tech. (Computer Science and Engineering) student in JSS Academy of Technical Education, Noida with a solid background in C++ and proficiency in Flutter for mobile app development. Competent in creating cross-platform apps and effective algorithms, and very motivated to solve problems and develop software. enthusiastic in using technology to develop meaningful answers. He can be contacted at email: dellvenue7satyam11@gmail.com.

Vidushi Bhardwaj a recent computer science engineering graduate, has a keen interest in machine learning and software development. Throughout her academic journey, she has developed strong technical skills and a passion for problem-solving in these fields. Eager to expand her knowledge, Vidushi is continuously seeking opportunities to apply her expertise in real-world applications and further advance her career in cutting-edge technologies. She can be contacted at email: vidushibh211269@gmail.com.