Sustainable Computing: Informatics and Systems

Artificial Intelligence Approach for Prediction of Marine Pollution for Sustainable Ocean Health

--Manuscript Draft--

Manuscript Number:	
Article Type:	VSI:AI – Ocean Health _Tran [Open for new submissions until Oct 1st,2022]
Section/Category:	Special Issue Paper
Keywords:	Maritime traffic; Marine pollution; Plastic litter; buoy; artificial intelligence; Two-step clustering; machine learning; fully Convolutional Network
Corresponding Author:	Surjeet Dalal Teerthanker Mahaveer University Moradabad, Uttar Pradesh INDIA
First Author:	Edeh Michael Onyema
Order of Authors:	Edeh Michael Onyema
	Surjeet Dalal
	Natali Kryvinska
	Bijeta Seth
	Mohd Anas Wajid
	Kuldeep Kumar
Abstract:	Climate change has become a major source of concern to global community. The steady pollution of the environment including our waters is gradually increasing the effects of climate change. The disposal of plastics to the seas alters aquatic life. Marine plastic pollution poses a grave danger to the marine environment and long-term health of the ocean. Though technology is also seen as one of the contributors to climate change but many aspects of it are being applied to combat climate related disasters and to raise the awareness about the need to protect the planet. This study investigated the amount of pollution in marine and an undersea leveraging the power of Artificial intelligence to identify and categorise marine and undersea plastic wastes. Classification was done using two types of machine learning algorithms: Two-steps Clustering and fully Convolutional Network (FCN). The models were trained using Kaggle's plastic location data, which was acquired in-situ. An experimental test was conducted to validate the accuracy and performance of the trained models and the results were promising when compared to other conventional approach and models. The model was used to create and test an automated floating plastic detection system in the required time frame. In both cases, the trained model was able to correctly identify the floating plastic and achieved an accuracy of 90%. The technique presented in this study can be a crucial instrument for automatic detection of plastic garbage in the ocean thereby enhancing the war against marine pollution.
Suggested Reviewers:	Varun Malik varun.malik@chitkara.edu.in
	Meeta Singh meeta.fet@mriu.edu.in

12-07-2022

Dear Editor,

We wish to submit an original research article entitled "**Artificial Intelligence Approach for Prediction of Marine Pollution for Sustainable Ocean Health**" for consideration by Sustainable Computing: Informatics and Systems.

We confirm that this work is original and has not been published elsewhere, nor is it currently under consideration for publication elsewhere.

In this paper, this study investigated the amount of pollution in marine and an undersea leveraging the power of Artificial intelligence to identify and categorise marine and undersea plastic wastes. Classification was done using two types of machine learning algorithms: Two-steps Clustering and fully Convolutional Network (FCN). This is significant because the technique presented in this study can be a crucial instrument for automatic detection of plastic garbage in the ocean thereby enhancing the war against marine pollution..

We believe that this manuscript is appropriate for publication by Sustainable Energy Technologies and Assessments special issue 'Role of AI and Edge Computing in Sustainable Ocean Health'. We have no conflicts of interest to disclose.

Please address all correspondence concerning this manuscript to me at **profsurjeetdalal@gmail.com**.

Thank you for your consideration of this manuscript.

Sincerely,

Surjeet Dalal
College of Computing Science & IT,
Teerthanker Mahayeer University, Moradabad
Uttar Pradesh, India

AUTHORSHIP STATEMENT

Manuscript title: **Artificial Intelligence Approach for Prediction of Marine Pollution for Sustainable Ocean Health**

I declare that all persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in *Sustainable Computing: Informatics and Systems.*

Authorship contributions

Category 1

Conception and design of study: Edeh Michael Onyema, Surjeet Dalal;

Acquisition of data: Bijeta, Kuldeep Kumar;

Analysis and/or interpretation of data: Natali Kryvinska, Mohd Anas Wajid.

Category 2

Drafting the manuscript: Natali Kryvinska, Mohd Anas Wajid,

Revising the manuscript critically for important intellectual content: Surjeet Dalal, Bijeta, Kuldeep Kumar, Edeh Michael Onyema.

Category 3

Approval of the version of the manuscript to be published (the names of all authors must be listed):

Edeh Michael Onyema, Surjeet Dalal, Bijeta, Kuldeep Kumar, Natali Kryvinska, Mohd Anas Wajid

Acknowledgements

All persons who have made substantial contributions to the work reported in the manuscript (e.g., technical help, writing and editing assistance, general support), but who do not meet the criteria for authorship, are named in the Acknowledgements and have given us their written permission to be named. If we have not included an Acknowledgements, then that indicates that we have not received substantial contributions from non-authors.

Surjeet Dalal; Ph.D. (for all authors)

Dalal

Signature: Date: 12/07/2022

Artificial Intelligence Approach for Prediction of Marine Pollution for Sustainable Ocean Health

Edeh Michael Onyema[1], Surjeet Dalal[2], Natali Kryvinska[3], Bijeta Seth [4], Mohd Anas Wajid[5], Kuldeep Kumar [6]

[1]Department of Mathematics and Computer Science, Coal City University, Enugu, Nigeria
Adjunct Faculty, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India
mikedreamcometrue@gmail.com ORCID: https://orcid.org/0000-0002-4067-3256

[2]College of Computing Sciences & IT, Teerthanker Mahaveer University, Moradabad, U.P., India,
profsurjeetdalal@gmail.com

[3]Information Systems Department, Comenius University, Bratislava, Slovakia.
natalia.kryvinska@uniba.sk

[4]Dept. of Computer Science and Engineering, B. M. Institute of Engineering & Technology, Sonipat, Haryana, bijetaoberoi@gmail.com

[5]Department of Computer Science, Aligarh Muslim University, Aligarh, 202002, India.
anaswajid.bbk@gmail.com

[6]Lead Consultant BigData Engineer, Xebia, Gurugram, Haryana kuldeepaman@gmail.com

Corresponding Author: Surjeet Dalal [profsurjeetdalal@gmail.com]

Abstract

Climate change has become a major source of concern to global community. The steady pollution of the environment including our waters is gradually increasing the effects of climate change. The disposal of plastics to the seas alters aquatic life. Marine plastic pollution poses a grave danger to the marine environment and long-term health of the ocean. Though technology is also seen as one of the contributors to climate change but many aspects of it are being applied to combat climate related disasters and to raise the awareness about the need to protect the planet. This study investigated the amount of pollution in marine and an undersea leveraging the power of Artificial intelligence to identify and categorise marine and undersea plastic wastes. Classification was done using two types of machine learning algorithms: Two-steps Clustering and fully Convolutional Network (FCN). The models were trained using Kaggle's plastic location data, which was acquired in-situ. An experimental test was conducted to validate the accuracy and performance of the trained models and the results were promising when compared to other conventional approach and models. The model was used to create and test an automated floating plastic detection system in the required time frame. In both cases, the trained model was able to correctly identify the floating plastic and achieved an accuracy of 90%. The technique presented in this study can be a crucial instrument for automatic detection of plastic garbage in the ocean thereby enhancing the war against marine pollution.

Keywords: Maritime traffic, Marine pollution, Plastic litter, buoy, artificial intelligence, Two-step clustering, machine learning, fully Convolutional Network

1. Introduction

There is a growing concern about climate change and its devastating effects. In recent times, the world has witnessed many disasters which can be traced to climate change. Waters are not spared as they have become a disposal point for all kinds of waste especially plastic wastes (Egunjobi and Onyema, 2020). This has resulted to serious water pollution. Ocean pollution has caused a range of environmental problems including organic matter enrichment in confined waters, oil contamination, and sedimentation caused by land-based operations like as dredging (Nunes andLeston, 2020). Dissolved Oxygen (DO) and microbial concentration levels are significant markers of the health of coastal water; however pollution from organic and inorganic sources has increased over time. Untreated home and industrial waste contributes to coastal pollution by causing bacteria to consume large amounts of oxygen from the saltwater, lowering the concentration of dissolved oxygen in the water and negatively impacting aquatic life. Aquatic lives are now endangered more than before (Minchew et al, 2012). Technology offers some possibilities that can assist environmentalists and relevant agencies to deal with the surfacing problem of plastic waste in water. For instance, Artificial intelligence is being deployed for ocean observation and proper monitoring of underwater things and spillage (Song et al, 2020). This study is directed towards that regards with a view to automate the process and approach to automatically search and to locate undersea plastic wastes. Leveraging the power of artificial intelligence as suggested in the current would go a long way to support the long term goal of achieving sustainable ocean health both for animals and humans.

This paper is designed with aiming of prediction of Marine pollution in two perspectives i.e. Coast Pollution and underwater trash. The section 2 defines the various existing works carried out in this problem domain. The section explains the about datasets used for experimental purpose. In this paper, 2 different types of datasets are being taken. Section 4.1 demonstrated coast pollution assessment through modified Two-step clustering algorithms and section 4.2 highlighted underwater trash detection with YOLO algorithm. Next section 5 demonstrate the result followed by conclusion of the paper in section 6.

2. Related works

Minchew et al. (2012) examined that the dampening of the sea wave terrible portions by the oil and a compelling fall in dielectric steady caused by the mix of 65-90 percent water with oil in the surface layer are to blame for the difference in backscatter over the fundamental smooth. As instrument noise rises over the instrument commotion floor, the anisotropy, A, border exhibits substantial variation throughout the oil spill and a large-reach subordinate sign.

Espeseth et al. (2020) introduced two strategies that are correlative as far as recognizing transient replaces inside an oil spill. In contrasting ways, the two approaches show how various people view oil spills. As an intermediate for increasing oil thickness, the primary technique detects regions within the smooth that demonstrate a determinedly high damping percentage (the distinction between clean ocean and oil power). With each new photo that is included in the computation, this approach updates the scene's age in addition to the original age. With the next approach, you may create a transient float and track the changes in damping and copolarization ratios between two scenes, as well as thickening and emulsification intermediates in between. To see this development and its persistence in both broad settings, one must examine the outcomes of these two approaches.

Seydi et al. (2021) grew new OSD system in view of a profound learning calculation for optical RS symbolism. The proposed strategy depended on a multiscale multi-layered leftover part convolution brain organization. The proposed technique explored the profound elements by the two-layered multiscale leftover blocks and, then, at that point, used them at one-layered multiscale remaining blocks. Landsat-5 satellite imagery from the Gulf of Mexico was used to evaluate the suggested approach in this study. Overall, the suggested technique has an accuracy rate of over 95% and a miss location and deception rate of less than 5%, suggesting its significant potential for OSD. In addition, it was found that the proposed method would be wise to implement compared to other OSD estimates that were examined in this research.

Stable water maps will soon be available thanks to a recurrence-based technique presented by Meng and colleagues (2020) to distinguish between hydroponics water and conventional water. Each year, Landsat Level-2 images from 1984 to 2018 were used to construct yearly 30 m target water items, which were then used to investigate the spatial-worldly changes in the Taihu Lake area. Furthermore, each big graphical modification was linked to a specific event in the actual world at the time. The outcomes propose that human exercises impact surface water than environment changes in the Taihu Lake locale, and affirm the adequacy of biological security strategy in keeping up with the strength of the aggregate sum of normal water in the beyond couple of many years. The spatial-transient unsettling influence of hydroponics likewise gave one more point of view and a solid proof of past investigations because of human exercises on the eutrophication interaction of Taihu Lake.

Walden et al. (2020) featured that Waterway flotsam and jetsam can affect numerous parts of marine conditions remembering marine route and sporting looking for waterways. Stream plastic flotsam and jetsam contamination is happening at such a huge scope universally that it is grave to follow and measure the sheer amount of lost plastics. Advanced picture handling is one compelling method for observing stream flotsam and jetsam, yet the related complex cycles have intrinsic difficulties. This study examines the effects of evaluating the weight and volume of plastic jug flotsam and jetsam from various waste scenarios.

Egunjobi and Onyema (2020) assessed the variation of Ocean Wave velocity with Ocean Depth Based on a Cubic Polynomial Fit Expression. The study established the correlation between the ocean wave velocity and the ocean depth

Chen et al. (2021) expressed that diminishing ozone harming substance emanations turns into a first concern on the planet with the development of an unnatural weather change and natural issues. Different sustainable power sources show up during the last many years. Sea catches and stores colossal measures of energy, which could fulfill multiple times of world energy interest. Because of innovation constraints and monetary contemplations, marine ebb and flow energy seems the most alluring decision contrasted and the other sea energy structure. Albeit the current mastery and innovation in seaward wind energy transformation framework can be to some degree moved to marine flow energy change framework because of the comparative design, there are as yet numerous mechanical difficulties to survive. In the meantime, the framework works under the water will definitely decidedly affect the general climate. In this paper, it shows the interest and the rule of the marine current energy, and furthermore talks about the benefits and weaknesses. The ecological effects around the gadgets, the mechanical difficulties, and the fundamental help structures are introduced also. The cutting edge level pivot turbines and their general innovations and the most recent ventures are portrayed at last. This survey

paper gives the valuable data about the fascination and challenge of the marine current energy, and the freshest advancement of the innovations and tasks.
Xue et al. (2021) examined whether convolutional brain organizations can recognize the distinctions of flotsam and jetsam and normal remote ocean climate, to really accomplish remote ocean garbage distinguishing proof. Initial, a genuine remote ocean flotsam and jetsam pictures dataset is developed for additional grouping research in view of a web-based remote ocean trash data set. Moreover, five normal Convolutional neural networks (CNNs) structures are likewise utilized to execute the order cycle. At long last, the distinguishing proof examinations are done to approve the presentation of the proposed strategy. The outcomes exhibit that the proposed strategy is better than the cutting edge CNN technique and has the potential for remote ocean trash recognizable proof.
Shi et al. (2021) developed Wheatstone span circuits to improve the reaction to garbage and lessen the impedance of oil temperature and thickness. A postprocessing circuit with sifting and intensifying capacities is embraced to further develop precision. The sensor yield attributes and trash position impact are concentrated on through the investigation. The complete examination of the identification brings about two modes can further develop sensor versatility and conquer the lack of low dependability in light of a solitary discovery technique. This sensor can give more precise garbage data to the shortcoming analysis of water driven hardware and is of incredible importance for smart upkeep.
Luccio et al. (2020) first present the idea of the Internet of Floating Things (IoFT), which stretches out the IoT to the sea situation. Then, we present our most recent execution of the DYNAMO (Distributed relaxation Yachts sensor Network for Atmosphere and Marine Observations) framework, a structure for seaside information assortment from sensors and gadgets sent in marine gear. To show the significance of IoFT information assortment in reality natural science setting, we think about a logical work process for seaside water quality. The chose application centerson foreseeing the spatial and transient example of ocean poisons and their conceivable presence and season of perseverance nearby mussel ranch regions in the Bay of Pozzuoli in Italy. The contaminations are straightforward Lagrangian particles, so the sea elements assume a significant part in the reenactment. Our outcomes show that coordinating publicly supported bathymetry information in the work process mathematical model arrangement works on the exactness of the end-product, considering a more itemized spatial circulation example of the ocean ebb and flow driving the Lagrangian tracers.
Gao et al. (2018) fostered a sea object float forecast model. "High goal surface flows, Stokes float, and winds" were handled, and a progression of model examinations were built. Predicted paths for the items on display were similar to observed directions for the floating objects. Many of the items that were tracked ended up at Reunion Island, Mauritius, and Tanzania with probabilities of 5 percent, 5 percent, and 19 percent, respectively, after drifting north and then west. Eventually, most of the reenactment's components were located in the western Indian Ocean, about latitude 10°S. Possibly owing to the influence of southeast trade winds, there were substantial differences in the results of several room factor explorations. Since several pieces of rubbish have been found on Africa's east bank and Australia's west shoreline has been sparse, it is likely that the accident site will be located in the underwater pursuit area's north section because of this.
Melody et al. (2020) proposed Convolutional brain organization (CNN) is fit for mining spatial component from huge informational index consequently. CNN's deep multi-facet component extraction has inspired us to develop an innovative oil slick-detection proof technique in this research. The PolSAR information is immediately converted into a 9-channel information block for CNN's use. A 5-layer CNN is then used to extract two important levels of information from the initial data set. RBF-SVM, a support vector machine approach with an outspread premise work component, is utilised for order. RADARSAT-2 totally polarimetric SAR data was used in this study to endorse the suggested technique. Findings reveal that the suggested technique has significant effects on grouping accuracy and kappa coefficient generally. Other advantages include a reduction in deception rates and the ability to distinguish an oil slick from a biogenic smooth.
Han et al. (2020) present a meta-heuristic whale optimization algorithm (WOA), which assists ships with tracking down a low-energy-utilization and safe course in an enormous scope complex marine climate. A few strategies have been proposed to tackle this issue, however there are a few weaknesses, for example, and no thought of the impact of wind bearing, wind speed and wave. The consequences of our recreation tests show that WOA is more cutthroat than other best in class calculations for course arranging.
Plag et al. (2020) represents the information on the wide open concerning the danger marine flotsam and jetsam to the marine biosphere and humankind stays at low levels. Because of the lack of suitable local area-focused virtual entertainment, it is difficult to work with the framework of an open cross-sectoral local area that may foster greater collaboration and correspondence among all parties involved in marine rubbish. There is a prototype Marine Debris Virtual Community Centre (MD-VCC) where partners may learn about the numerous sources of marine flotsam and jetsam and the paths waste enters the marine ecosystem.This space gives valuable open doors to individuals to foster the abilities to take part in the discourse about marine flotsam and jetsam and ways of handling this issue. The MD-VCC is made to further develop correspondence and support connected with marine garbage.

Xue et al. (2021) lays out an effective remote ocean flotsam and jetsam identification technique with fast utilizing profound learning strategies. Initial, a genuine remote ocean garbage location dataset (three dimensional dataset) is laid out for additional exploration. Material, fishing net and rope, glass, metal, normal rubbish, elastic, and plastic are all included in the dataset. Another option is the ResNet50-YOLOV3 remote ocean flotsam and jetsam location system. The recognition cycle of distant ocean debris also includes eight high-level identification models. Finally, testing are performed to verify the exhibit of ResNet50-YOLOV3. These trials also show the relevance and feasibility of ResNet50-YOLOV3 in detecting ocean flotsam and jetsam from a distance.

Li et al. (2019) proposed a various levelled structure for coal and gangue recognition in view of profound learning models. First, the Gaussian pyramid guideline is utilised to create staggered prepared information, resulting in a variety of scales of coal and gangue image highlights. Next, neural networks (CNNs) are created to recognise coal and gangue objects in discrete up-and-comer regions. Three distinct datasets were used to test our method. Coal and gangue object localization precision improved by 0.8 percent in comparison to previous methods, reaching up to 983.33 percent with the suggested strategy. We also present a methodology that makes it possible to see a large number of coal and gangue items at once and addresses the problem of lining requirements in existing methods.

3 Datasets

Following is the list of datasets being considered.

- **Coast Pollution**

This data represents some of the data gathered in a year of research, on 3 coasts which represent a pollution gradient. These rows and columns make it easy for others to get started by describing how scholar acquired the data and what time period it represents, too. The label for this research should be the pollution level (3 levels, when 0 is clean and 2 is the most polluted) and the analysis should divide the data into test and train data manually.

- **Underwater Trash**

From the J-EDI marine debris dataset, we obtained this information. That dataset has a wide range of video quality, depth, objects in the scene, and camera settings. They show a wide range of maritime detritus in various levels of decay, opacity, and overgrowth as they were photographed in real-world locations. The water's purity and light's quality might vary greatly from one video to the next. This dataset consists of 5,700 photos annotated with bounding boxes on instances of garbage, biological things such as plants and animals, and ROVs, which were extracted from the processed movies. The ultimate objective is to build onboard garbage detecting technologies that are both efficient and accurate. This information is being made public in the hopes that it will help the marine robotics community work toward a solution to the pressing issue of autonomous garbage identification and removal.

4 Methods

There are two major goals of our research: developing a system that can identify Underwater Trash on the coast and underwater. Research tools were used to gather and evaluate data from diverse sources. FCN and Twi-step clustering algorithms are employed as our study framework. Clustering is a highly effective method for detecting similarities across various groups or clusters. According to the features of plastic pollution (space, shape, etc.), these algorithms examine the hidden information in our research.

4.1 Coast Pollution Assessment through Two-step clustering

The pre-clustering, parsing of typical data kinds, and clustering steps are all performed using the two-step clustering method. It is chosen whether or not to begin a new cluster during pre-clustering after each set of data has been analysed and evaluated. Table 1 highlights the model specifications of Two-steps clustering as given below:

Table 1: Model Specifications

Minimum Number of Regular Clusters		2
Maximum Number of Regular Clusters		15
Adaptive Feature Selection		On
Feature Importance Method		Information Criterion
Information Criterion		Bayesian Information Criterion (BIC)
Distance Measure		Log Likelihood
Final Model	Number of Regular Clusters	2
	Number of Outlier Clusters	0
	Continuous Inputs	Month
		Season

	Shore
	Pollution Level
	Organic matter%
	OC
	Water Salinity
	P
	Total dissolved solids
	PP
	Conduction
	ORP
	Specific resistance

This choice is made based on how far apart the data are. The Euclidean distance and the log-probability distance are two distance metrics. Feature importance for this algorithm are being displayed in fig.1.

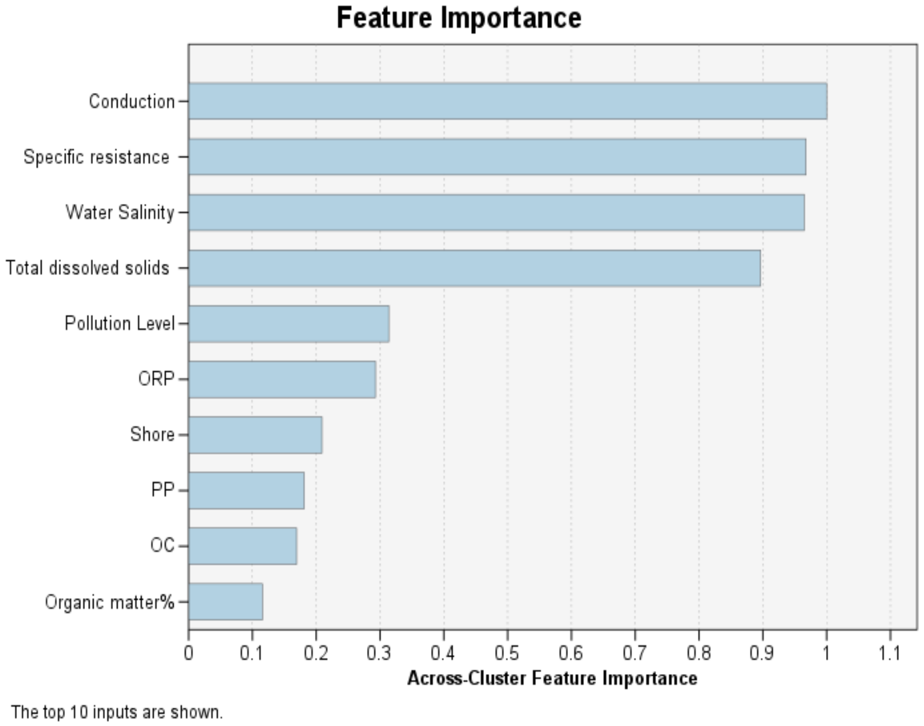

Fig. 1 Across-Cluster Feature Importance

It is common for data that cannot be clustered to be assessed during the data analysis step. The data is segregated as external if the inclusion cannot be achieved despite all attempts to incorporate it. Table 2 highlighted the model quality in the help of Goodness and Importance factor.

Table 2 Model Quality

Cluster	Number of Records	Goodness	Importance
Cluster-1	186	0.34	1.00
Cluster-2	33	0.32	1.00

In the cluster stage, a tree structure is created. All data starts to be distributed from root to leaves. Table 3 show the Cluster-1 Profilein term of Within-Cluster Feature Importance.

Table 3 Cluster-1 Profile

Input	Center	Within-Cluster Feature Importance[a]
Conduction	0.37	1.00
Specific resistance	-0.37	0.97
Water Salinity	0.38	0.96

Total dissolved solids	0.38	0.90
Pollution Level	-0.21	0.31
ORP	-0.28	0.29
Shore	0.00	0.21
PP	0.19	0.18
OC	0.18	0.17
Organic matter%	0.11	0.12
P	0.05	0.11
Month	0.07	0.07
Season	-0.08	0.01

Cluster centers show modes for categorical inputs, and means for continuous inputs.

a. This is the importance of an input to a particular cluster.

Each piece of information is connected to a nearby branch. When the ideal number of groups has been obtained, the cluster is linked to another cluster in another branch that meets the distance criterion's requirements. Table 4 show the Cluster-2 Profile in term of Within-Cluster Feature Importance.

Table 4 Cluster-2 Profile

Input	Center	Within-Cluster Feature Importance[a]
Conduction	-2.09	1.00
Specific resistance	1.92	0.97
Water Salinity	-2.11	0.96
Total dissolved solids	-1.98	0.90
Pollution Level	1.21	0.31
ORP	1.60	0.29
Shore	-0.01	0.21
PP	-0.98	0.18
OC	-0.95	0.17
Organic matter%	-0.68	0.12
P	-0.27	0.11
Month	-0.40	0.07
Season	0.46	0.01

Cluster centers show modes for categorical inputs, and means for continuous inputs.

a. This is the importance of an input to a particular cluster.

To repeatedly decide the appropriate number of clusters, BIC (Schwarz's Bayesian Information Criterion) or AIC (Akaike's Information Criterion) procedures are used.

4.2 Underwater Trash Detection with YOLO algorithm

A real-time object identification technique known as YOLO stands for "You Only Look Once." All of YOLO's layers are convolutional, creating a neural network (FCN). Using skip connections and upsampling layers, there are 75 convolutionary levels in this algorithm. A convolution layer with a stride of 2 is utilised to down sample the feature maps without the usage of pooling. It learns from entire photos and improves detection performance by doing so directly. In comparison to other object-detection algorithms, this model offers a few advantages.

- YOLO is very quick.
- YOLO takes the whole image while training and test time so it completely encodes related data about classes and their appearance.
- YOLO learns generalized representations of objects and thus outperforms other detection methods.

YOLO is a real-time object identification method based on neural networks. Because of its speed and precision, this algorithm has become a popular choice among users. It's been put to good use spotting traffic lights, people, parking metres, and animals, among other things.Computer vision's phenomena of object detection includes seeing certain things in digital photos or movies. Humans, automobiles, chairs, stones, structures, and even animals have all been spotted.Two fundamental concerns are being addressed by this phenomenon:

- What is the item? This question asks you to name the thing you see in a particular picture.
- Where did you find it? With this question, you're trying to pin down exactly where in the image the object is located.

Detection methods include R-CNN, Retina-Net, and Single-Shot MultiBox Detector. Object detection (SSD). A single algorithm run is unable to find things, even though these techniques have overcome the limitations of limited data and model-based object recognition. The YOLO algorithm has become popular because it outperforms the other object detection methods. Figures 2-5 depict the inference results.

Figure 2 Inference results 1

Following satisfactory training results, our model is ready for inference. Using test-time augmentations (TTA), we may increase the accuracy of our predictions even further: each picture is enhanced (horizontal flip and 3 different resolutions) and the final prediction is an ensemble of these augmentations.

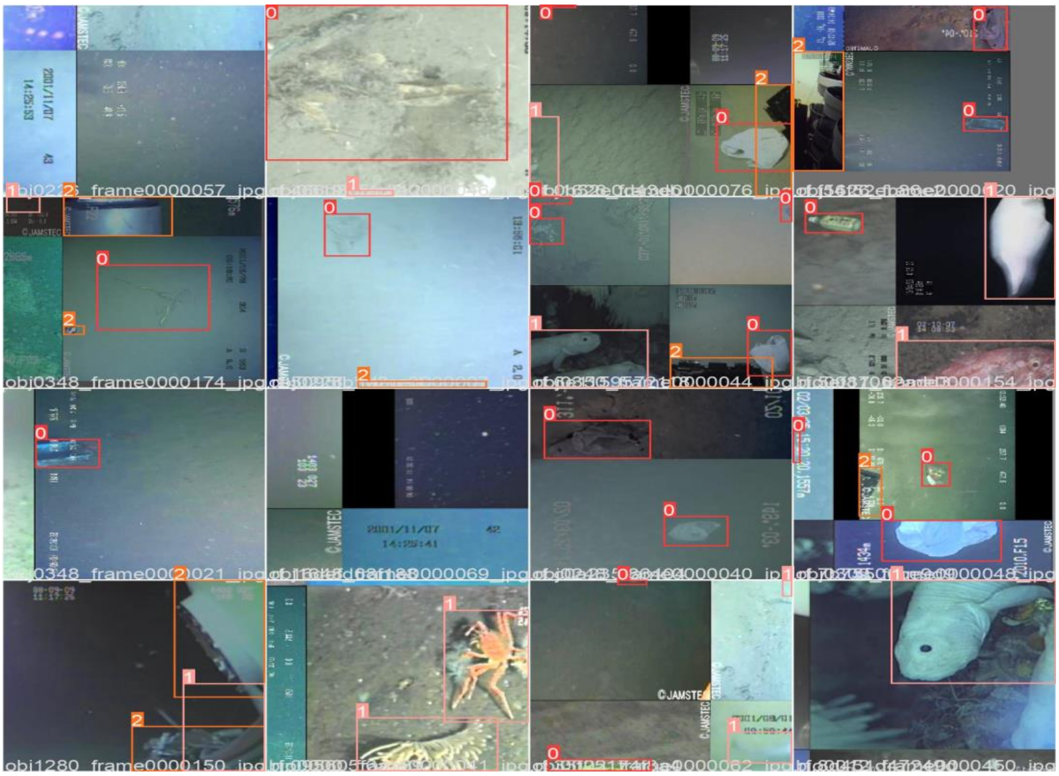

Figure 3 Inference results 2

If we’re tight on the Frames-Per-Second (FPS) rate, we’ll have to ditch the TTA since the inference with it is 2–3 times longer.

Figure 4 Inference results 3

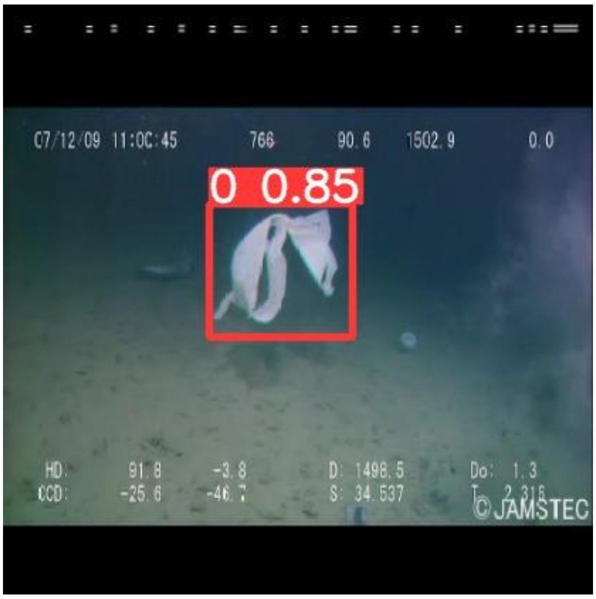

Figure 5 Inference results 4

YOLO is a sophisticated CNN for real-time object identification called a convolution neural network (CNN). YOLO's popularity is due to its ability to operate in real-time and its high level of accuracy. The following figure 6 shows the feature extraction.

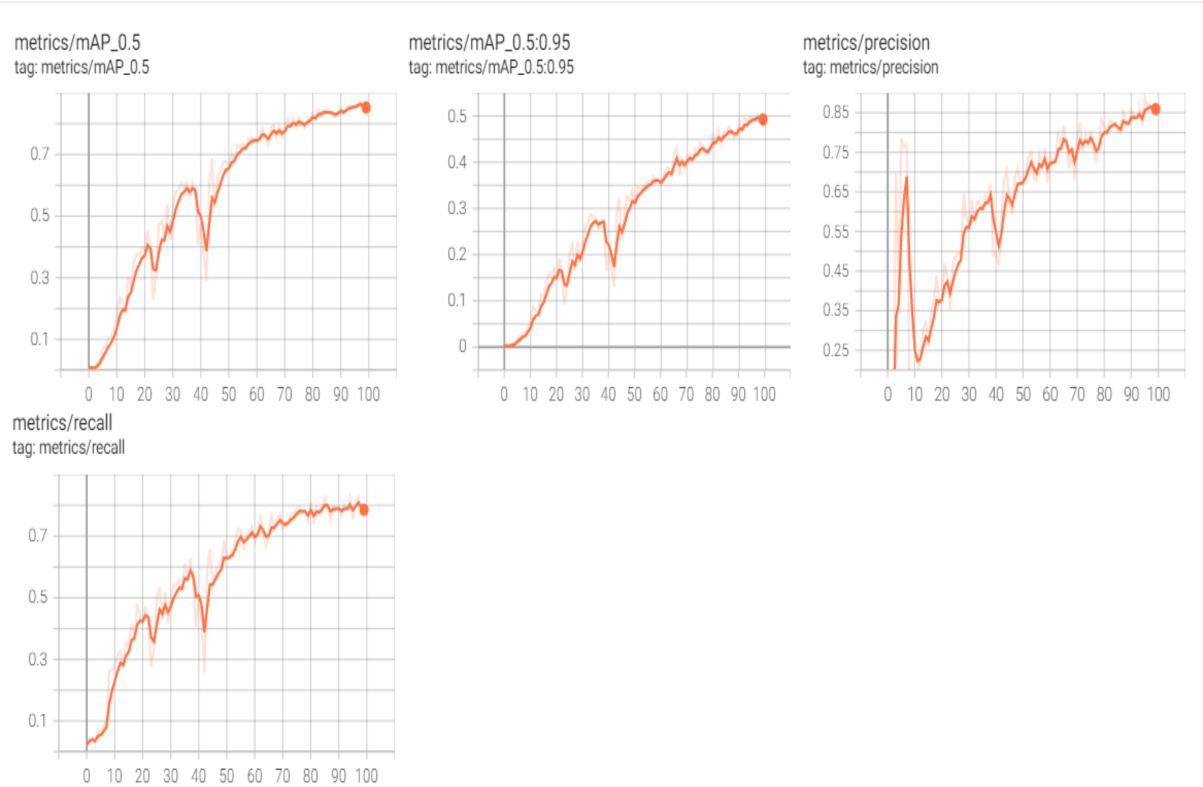

Figure 6 Results of 'feature extraction'

5 Results

In order to fine-tune the model, we may unfreeze it completely and retrain it on our data with an extremely low learning rate. By progressively modifying the pre-trained features to the fresh data, this may be able to produce considerable gains. The following figure 7 shows the results of training.

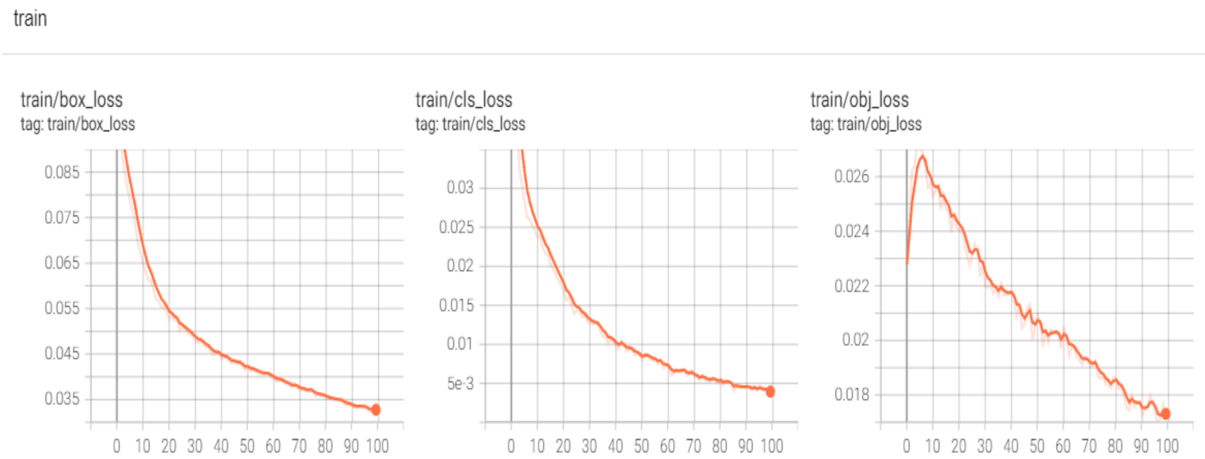

Figure7 Results of ‘Training’

The hyperparameters-configurations file may be used to change the learning rate parameter. Hyperparameters specified in the built-in 'hyp.finetune.yaml' file will be used for the instructional example because they have a considerably lower learning rate than the default. The stored weights from the previous step will be used to re-initialize the weights.

As a convenience for this lesson, we'll utilise the YOLOv5s6 model, which has a relatively modest number of parameters. In this section, we'll discuss some of the most often utilised training methods, as well as a few others. Here, in this section, we'll discuss some of the most often utilised training methods in figure 8.

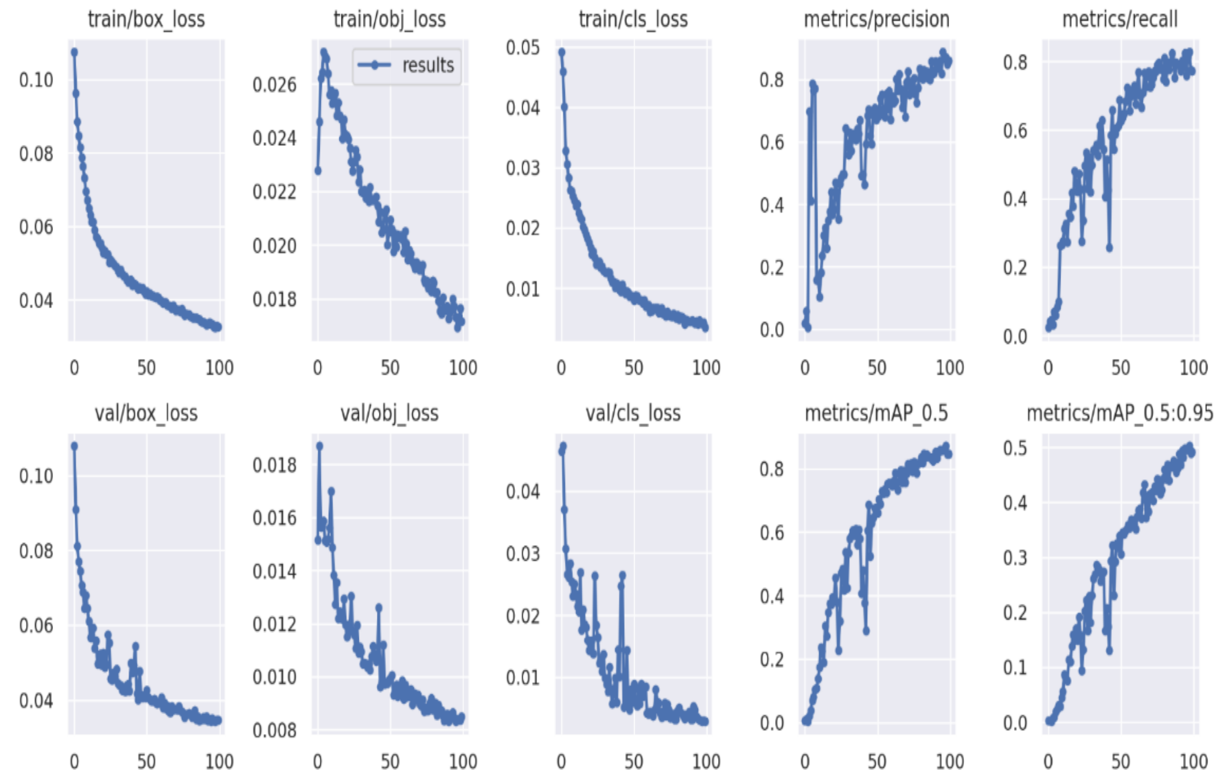

Figure 8 Results of ‘feature extraction’ training

The backbone layers of a model function as a feature extractor, whereas the head layers are responsible for calculating the output predictions. We'll utilise the same backbone as the pre-trained COCO model and simply train the model's head in order to compensate for a short dataset size. Backbone YOLOv5s6 has 12 layers, and the 'freeze' parameter fixes them all. The confusion matrix created is shown in figure 9.

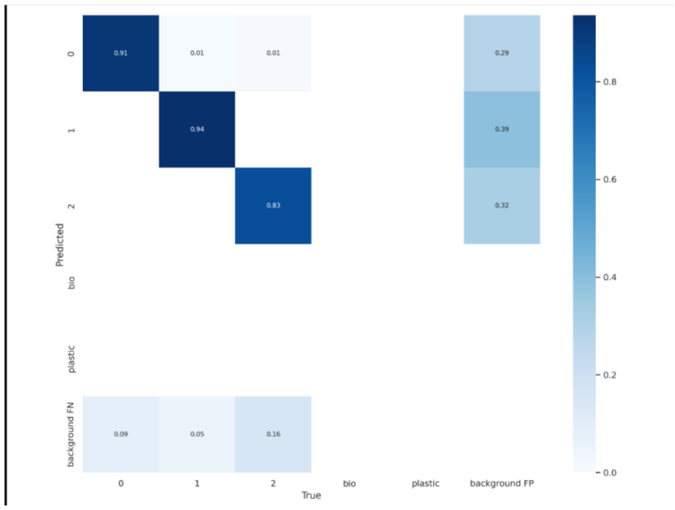

Figure 9 Confusion Matrix

The validation script will be used to test our model. The 'task' option can be used to adjust the divides between the training, validation, and test datasets for evaluating performance. The following figure 10 is an evaluation of the test dataset split. The following is an evaluation of the test dataset split:

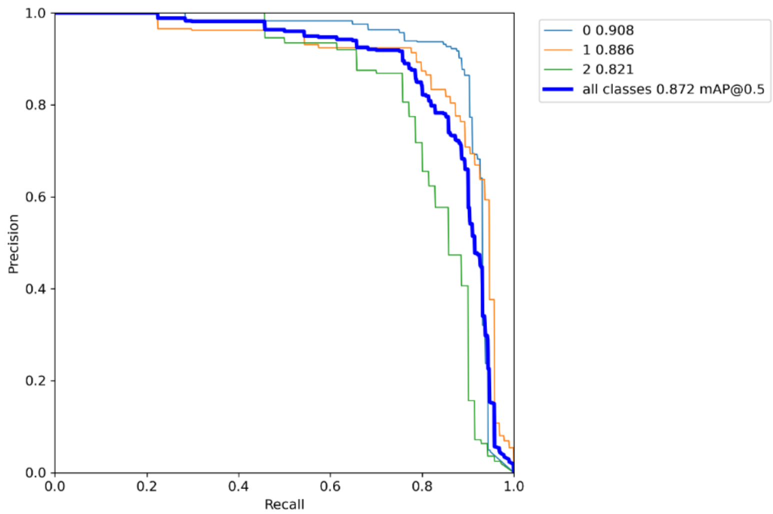

Figure10 Precision- Recall Curve

In the sense that it only takes one forward propagation run through the neural network to create predictions, the algorithm "just looks once" at the picture. Multi-class probabilities and bounding boxes may be predicted concurrently using YOLO. Precision-confidence curve is shown in figure 11.

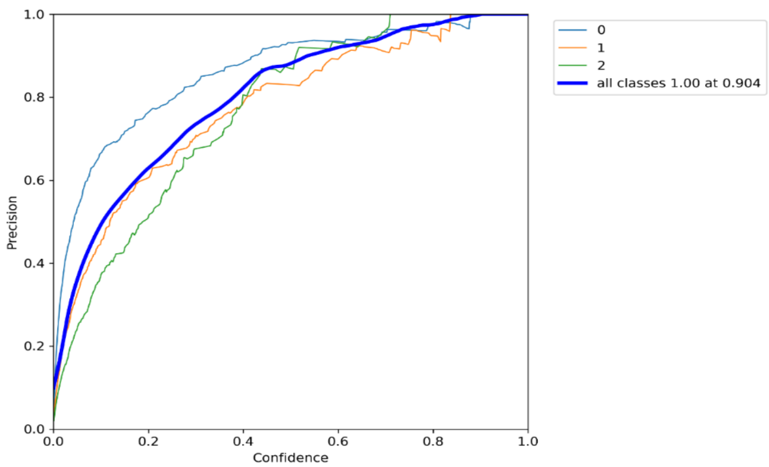

Figure11 Precision-Confidence Curve

Recall measures how much of the true bbox was successfully predicted (Real positives / (True positives + True Negatives).) Precision measures how much of the true bbox was accurately predicted. MAP 0.5 is the average precision (mAP) for an intersection over union threshold of 0.5-0.6-0.5. An IoU threshold range of 0.5 to 0.95 is used to calculate the average mAP. Figure 12 shows confidence curve.

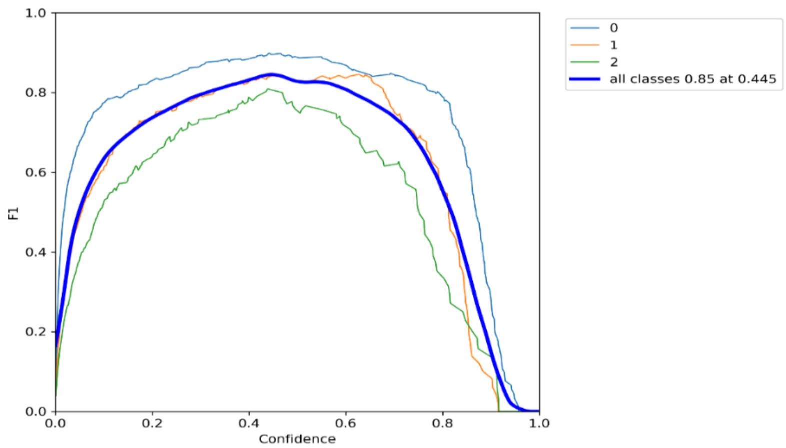

Figure12 F1-Confidence Curve

It was trained on synthetic data to examine the performance of object detectors and estimate the need for more real data for two-step clustering and FCN, which are the methods discussed in this work. Aside from that, the suggested sensor system's computing needs were studied using FCN.

The results of FCN were utilised to improve the design of the proposed system as a proof of concept. FCN and a projected multi-sensor system will provide a more complete situational awareness system. The suggested sensor system for training FCN is now being used to acquire additional real-time, actual data. An ocean basin containing exclusively floating plastic samples will be used for data collection in the near future, as will data collected from a pending research vessel excursion in southern North Sea waters. The study would add to the growing works (Onyema et al, 2021, Chen et al, 2021) on powers of artificial intelligence in prediction of Ocean polution and related events.

6 Conclusion

The study presented a technique that could help reduce water pollution and indeed detection and removal of pollutants. The study proves the power of Artificial intelligence in marine garbage detection. The YOLOv5 is a successful method even with a small number of datasets, particularly underwater. The current system was able to achieve a higher accuracy rate in testing can be implemented in wider scale. In the future, we intend to collaborate with relevant authorities to adopt this solution to complement efforts geared towards water protection and indeed preservation of our waters. Mores so, we will also develop an AI-driven app for promotion of Ocean health awareness.

Acknowledgement

The authors are grateful for the support provided by Information Systems Department, Faculty of Management, Comenius University Bratislava, Odbojárov 10, 82005 Bratislava 25, Slovakia

References

[1]. Nunes M., Leston S. (2020) Coastal Pollution: An Overview. In: Leal Filho W., Azul A.M., Brandli L., Lange Salvia A., Wall T. (eds) Life Below Water. Encyclopedia of the UN Sustainable Development Goals. Springer, Cham. https://doi.org/10.1007/978-3-319-71064-8_9-1

[2]. M. Vikas, G.S. Dwarakish, Coastal Pollution: A Review, Aquatic Procedia, Volume 4, 2015, Pages 381-388, ISSN 2214-241X, https://doi.org/10.1016/j.aqpro.2015.02.051.

[3]. Jerry R. Schubel, Chapter 9 Coastal Pollution and Waste Management National Research Council. 1994. Environmental Science in the Coastal Zone: Issues for Further Research. Washington, DC: The National Academies Press. doi: 10.17226/2249.

[4]. Thiel Martin, Luna-Jorquera Guillermo, Álvarez-Varas Rocío, Gallardo Camila, Hinojosa Iván A., Luna Nicolás, Miranda-Urbina Diego, Morales Naiti, Ory Nicolas, Pacheco Aldo S., Portflitt-Toro Matías, Zavalaga Carlos, Impacts of Marine Plastic Pollution From Continental Coasts to Subtropical Gyres—Fish, Seabirds, and Other Vertebrates in the SE Pacific, Frontiers in Marine Science, Volume 5, 2018, DOI=10.3389/fmars.2018.00238

[5]. Silhadi, M.A., Refes, W. & Mazouzi, S. Assessment of coastal ecosystems vulnerability to pollution: Algiers coast, Algeria. Environ Sci Pollut Res 27, 42670–42684 (2020). https://doi.org/10.1007/s11356-020-10123-5

[6]. Minchew B, Jones CE, Holt B. Polarimetric analysis of backscatter from the deepwater horizon oil spill using l-band synthetic aperture radar. *IEEE Trans Geosci Remote Sens*. 2012;50(10 PART1):3812-3830. doi:10.1109/TGRS.2012.2185804

[7]. Espeseth MM, Jones CE, Holt B, Brekke C, Skrunes S. Oil-Spill-Response-Oriented Information Products Derived from a Rapid-Repeat Time Series of SAR Images. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2020;13:3448-3461. doi:10.1109/JSTARS.2020.3003686

[8]. Seydi ST, Hasanlou M, Amani M, Huang W. Oil Spill Detection Based on Multiscale Multidimensional Residual CNN for Optical Remote Sensing Imagery. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2021;14:10941-10952. doi:10.1109/JSTARS.2021.3123163

[9]. Meng Y, Du P, Wang X, Bai X, Guo S. Monitoring Human-Induced Surface Water Disturbance around Taihu Lake since 1984 by Time Series Landsat Images. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2020;13:3780-3789. doi:10.1109/JSTARS.2020.3005135

[10]. Walden K, Mehrubeoglu M. Quantifying Plastic Bottle Debris in Waterways Using Image Processing. *Proc - 2020 Int Conf Comput Sci Comput Intell CSCI 2020*. Published online 2020:1658-1663. doi:10.1109/CSCI51800.2020.00305

[11]. Chen J, Chen S, Fu R, et al. Remote Sensing Estimation of Chlorophyll-A in Case-II Waters of Coastal Areas: Three-Band Model Versus Genetic Algorithm-Artificial Neural Networks Model. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2021;14:3640-3658. doi:10.1109/JSTARS.2021.3066697

[12]. Xue B, Huang B, Wei W, et al. An Efficient Deep-Sea Debris Detection Method Using Deep Neural Networks. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2021;14:12348-12360. doi:10.1109/JSTARS.2021.3130238

[13]. Shi H, Xie Y, Zhang H, et al. An Ultrasensitive Debris Microsensor for Oil Health Monitoring Based on Resistance-Inductance Parameter. *IEEE Trans Instrum Meas*. 2021;70. doi:10.1109/TIM.2021.3105255

[14]. Luccio D Di, Riccio A, Galletti A, et al. Coastal Marine Data Crowdsourcing Using the Internet of Floating Things: Improving the Results of a Water Quality Model. *IEEE Access*. 2020;8:101209-101223. doi:10.1109/ACCESS.2020.2996778

[15]. Gao J, Mu L, Bao X, Song J, Ding Y. Drift analysis of MH370 debris in the southern Indian Ocean. *Front Earth Sci*. 2018;12(3):468-480. doi:10.1007/s11707-018-0693-0

[16]. Song D, Zhen Z, Wang B, et al. A Novel Marine Oil Spillage Identification Scheme Based on Convolution Neural Network Feature Extraction from Fully Polarimetric SAR Imagery. *IEEE Access*. 2020;8:59801-59820. doi:10.1109/ACCESS.2020.2979219

[17]. Han Q, Yang X, Song H, Sui S, Zhang H, Yang Z. Whale Optimization Algorithm for Ship Path Optimization in Large-Scale Complex Marine Environment. *IEEE Access*. 2020;8:57168-57179. doi:10.1109/ACCESS.2020.2982617

[18]. Plag HP, Jones K, Garello R. A Virtual Center for the Community Addressing the Challenge of Marine Debris. *Ocean Conf Rec*. 2021;2021-Septe. doi:10.23919/OCEANS44145.2021.9705762

[19]. Xue B, Huang B, Chen G, Li H, Wei W. Deep-sea debris identification using deep convolutional neural networks. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2021;14:8909-8921. doi:10.1109/JSTARS.2021.3107853

[20]. Li D, Zhang Z, Xu Z, et al. An image-based hierarchical deep learning framework for coal and gangue detection. *IEEE Access*. 2019;7:184686-184699. doi:10.1109/ACCESS.2019.2961075

[21]. Egunjobi, K.A. and Onyema, E.M. (2020). Variation of Ocean Wave Velocity with Ocean Depth Based on a Cubic Polynomial Fit Expression. COAL City University Journal of Science, 1 (1), 65-71.

[22]. E. M. Onyema, P. K. Shukla, S. Dalal, M. N. Mathur, M. Zakariah, and B. Tiwari, “Enhancement of patient facial recognition through deep learning algorithm: ConvNet,” Journal of Healthcare Engineering, vol. 2021, Article ID 5196000, 8 pages, 2021.

Conflicts of Interest Statement

Manuscript title: **Artificial Intelligence Approach for Prediction of Marine Pollution for Sustainable Ocean Health**

I declare that the authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Author names:

Edeh Michael Onyema[1], Surjeet Dalal[2], Natali Kryvinska[3], Bijeta Seth[4], Mohd Anas Wajid[5], Kuldeep Kumar[6]

[1]Department of Mathematics and Computer Science, Coal City University, Enugu, Nigeria

Adjunct Faculty, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India mikedreamcometrue@gmail.com ORCID: https://orcid.org/0000-0002-4067-3256

[2]College of Computing Sciences & IT, Teerthanker Mahaveer University, Moradabad, U.P., India, profsurjeetdalal@gmail.com

[3]Information Systems Department, Comenius University, Bratislava, Slovakia. natalia.kryvinska@uniba.sk

[4]Dept. of Computer Science and Engineering, B. M. Institute of Engineering & Technology, Sonipat, Haryana, bijetaoberoi@gmail.com

[5]Department of Computer Science, Aligarh Muslim University, Aligarh, 202002, India. anaswajid.bbk@gmail.com

[6]Lead Consultant BigData Engineer, Xebia, Gurugram, Haryana kuldeepaman@gmail.com

Corresponding Author: Surjeet Dalal [profsurjeetdalal@gmail.com]

Surjeet Dalal; Ph.D. (for all authors)

Dalal

Signature: Date: 12/07/2022