Bima Journal of Science and Technology, Vol. 8(1B) Apr, 2024 ISSN: 2536-6041

DOI: 10.56892/bima.v8i1B.647

An In-Depth Study of Oil Spill Detection Using Various Machine Learning Techniques

Musa Suleiman Hamza[1]*, Suleiman Salihu Jauro[1], Yusuf Ali Sulaiman[2], Mohammed Ajuji[1]

[1]Department of Computer Science, Faculty of Science, Gombe State University, Nigeria
[2]Department of Computer Science, School of Science, Federal Polytechnic Kaltungo, Gombe State, Nigeria

Corresponding Author: musashamza@gmail.com

Abstract

Oil spills represent a significant threat to marine environments, causing ecological disasters and disrupting marine life cycles. This paper explores the environmental impact of oil spills, tracing their origins to the 1960s with the rise of supertankers. Despite stringent regulations, an annual transport of one million metric tons of oil into seas highlights the ongoing risk. To address this, the study proposes a structured approach for oil spill detection, encompassing data acquisition, preprocessing, feature extraction, and machine learning classification. The analysis emphasizes the diminishing occurrence of significant oil slicks, attributing this trend partly to regulatory efforts. The study highlights the growing role of deep learning in precise oil spill detection and calls for future research to address challenges, explore simplified detection techniques, and compare computational performance with traditional methods. This study summarizes the classification of modern ways for recognizing oil spills, as well as how machine learning techniques are applied to solve the problem through presentation and analysis. Conversely, discuss the benefits and drawbacks of these researches. In addition to indicating potential avenues for further investigation to advance the detection of oil spills.

Keywords: Oil slicks, environmental problem, petroleum output, spill detection.

INTRODUCTION

An external oil slick is the buildup of oil outside of a big river. The 1960s saw the advent of supertankers, which were capable of transporting over 500,000 metric tons of oil, as well as an increase in fuel production, prospecting on continental shelves, and other factors that contributed to the rise in marine oil slicks as a serious environmental problem. Dramatic oil slicks from sunken or sunk supertankers are becoming less common because to regulations governing transportation and the environment. Oil spills, whether little and huge, well leaks, and significant shipper responsibilities are all recorded each year. An estimated million metric tons of oil are yearly transported into the world's seas. The unintended or careless introduction of spent gas solvents and engine oils by companies and people exacerbates the mostly organic problem. Together with regular oil spills from the world's deep waters, these sources push global oil output up to 3.5-6 million metric tons per year.

One essential component of the potential to predict oil slicks is the detection and identification of oil spills. The precise location of oil slicks and the anticipation of their movements are beneficial to wildlife, fisheries, goal-setting, and the board's ability to monitor and manage the marine biological system. Over the past thirty years, there has been a notable increase in the extraction and transportation of oil from the ocean, mostly due to the increased oil consumption by various sectors (Seydi *et al.,* 2021). Both coastal and deep-ocean ecosystems are negatively impacted by the oil spill, which is one of the main sources of ocean pollution (Song *et al.,* 2020). The problem of oil leak detection has been extensively researched, however much of that research has been on very large offshore crude oil patches (De Kerf *et al.,* 2020). Satellites collect data quickly and cheaply over vast and remote ocean areas (Sun & Hu, 2019). Furthermore, the proliferation of remote sensing datasets especially open-access images has made it much easier to identify oil slicks in real time and on a regular basis.

Detecting and Knowledge about petroleum spills is a critical piece of possibility anticipating oil slicks. Exact discovery of oil slicks and expectation of their directions are useful to fisheries, natural life, obligation debate goals, and asset the board for checking and rationing the marine biological system.

GENERAL STRUCTURE

Figure 1 shows a general Structure of finding spills of oil. The preliminary portion is data acquisition take images as input in order to detect oil spill, the second phase is preprocessing where the input will be from data acquisition will be filter to remove noise that is unwanted properties, the third process, known as feature extraction, is crucial to the identification of oil spills. It provides a feature selection for classification and allows for the extraction and input of a collection of structures to differentiate oil spills from their lookalikes. Oil spills are classified using machine learning in the fourth step. Understanding and making connections between inputs and outputs from a complete set of representative training samples is a subset of artificial intelligence. Without assuming anything about the data's distribution, models of empirical and predictive categorization may be constructed from them. The last phase the oil spill will be detected from alike and real oil spillage area/portion.

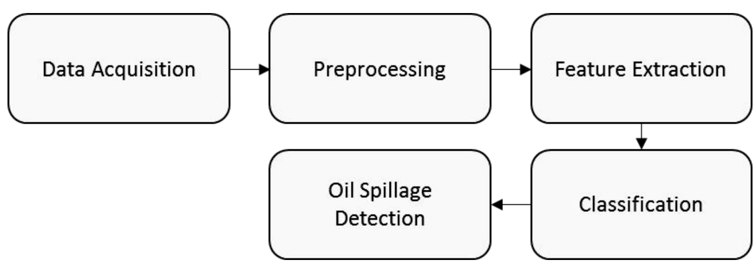

Figure 1: General flow of discovery of an oil spill.

LITERATURE REVIEW

This section reviews the work done in the area of finding oil spills. A survey on the various machines learning approach for oil spill is presented.

Chehresa *et al*., (2016) proposed a method for selecting the most appropriate set of features to extract from SAR images. The five phases of the proposed method are feature extraction, feature selection, dark spot detection, normalizing, and classification. The recommended to extract dark spots from SAR images that can indicate an oil leak, dark spot identification employs three methods: image enhancement, dark spot detection, and post processing. Image enhancement contain some filter (lee filter, local region filter, max-median filter and gamma transform) to eliminate noises, prepare it to the dark spot are detected by employing thresholding value determined the dark spot by pixel with lower value and post processing is used to eliminate error (regions that is mistakenly detect a dark spot or background). The proposed feature extraction used four classified group to differentiate oil spill from lookalike namely geometrical aspect, physical condition, context and texture feature of an oil spill. The recommended normalization strategy to avoid features with huge value ranges is the interquartile range (IQR). The suggested feature selection approach generates candidate feature subsets in the execution loop using an evolutionary algorithm. The proposed classification employs a Bayesian network with a Naive Bayes (NB) structure and rate achieved by classifying using an ideal set of characteristics, yielding an accuracy of around 93.19%.

Conceição *et al*., (2021) conducted research that devised a set of open-source procedures adept at managing oil spill scenarios using two random forest classifiers. The initial approach involves employing an ocean SAR image classifier to classify inputs, such as biofilm and multi-component oil. The purpose of this classifier was to address problems with similar-looking characteristics so that the gradients associated with the backscattering of these features could be statistically analyzed with confidence. The second technique, called Radar Image Oil Spill Seeker (RIOSS), uses Sentinel-1 SAR pictures to identify potential oil spill sites on marine surfaces. This technique helped to add features to a random forest, increasing the algorithm's accuracy by 90% and its capacity to provide better accuracy outcomes.

Tong *et al.,* (2019) classified polarimetric SAR data using Random Forest (RF) using a multi-feature-based ocean oil spill detection technique. This method combines seven more polarimetric characteristics with the self-similarity parameter. Radarsat-2 and UAVSAR oil spill datasets were used in two different types of situations to verify the viability of the suggested approach. Experiments show that the suggested method's accuracy is 92.99% and 82.25%, respectively.

Shaban *et al.,* (2021) recognized oil spills in the SAR dataset and proposed a two-stage DL framework, with the first phase categorized using a unique 23-layer Convolutional Neural Network (CNN). The second stage then used the U-Net structure to perform semantic segmentation. To decrease background speckle noise, the recommended framework first preprocesses the input, divides each image of 1250-650-3 into 64-64-3 patches, then applies a frost filter. 80% of the patches are used to train a new 23-layer CNN using the photos as input. At last, U-Net. When compared to comparable work during the implantation validation phase, the acquired results demonstrate accuracy of 92%, recall of 76%, precision of 84%, and dice80%.

Zeng and Wang (2020) presented the classification of SAR dark patches based on VGG-16, a moderately deep With 12 weight layers, the Oil Spill Convolutional Network (OSCNet) is a Deep Convolutional Neural Network (DCNN). Apart from that, the massive data set of 23,768 SAR dark patches collected from 336 SAR images is enhanced by applying the data augmentation technique to it. In terms of traits that the dataset allows for differentiation. OSCNet outperforms handcrafted features. As a result, this advanced machine learning classifier is different from AAMLP. Categorization performance is significantly improved with OSCNet. From 80.95%, 81.40%, and 92.50%, respectively, the precision, recall, and accuracy increased to 85.70%, 83.51%, and 94.01%.

To develop a high-precision segmentation, identification, and detection model that can investigate texture and form in order to recognize and localize objects such as land

DOI: 10.56892/bima.v8i1B.647

areas, ships, lookalikes, and oil spills. By utilizing a Mask Region-based Convolutional Neural Network (Mask RCNN), Yekeen and Balogun (2020) created an instance segmentation DL model. Following different preprocessing steps, 2882 pictures were classified as 88% training and 12% testing. With values of 0.964, 0.969, and 0.968, respectively, precision, recall, and the F1-measure were used to assess the model's performance, and the results showed that it performed better than other existing models.

Ghorbani and Behzadan (2021) presented the VGG16 model to classify oil spills. Several CNN models were tested, validated, and trained for visual recognition with a 92% accuracy rate using an internal picture bank called Nafta, which was created by web mining and includes 1292 photos from three distinct angles (first-person, drone, and satellite). Then, PSPNet and mask-R-CNN models with a mean overlap across the Union (IoU) of 49% and 68%, respectively, were employed for oil spill segmentation (e.g., identifying the oil spill borders at the pixel level).

Table 1: Surveys about various techniques associated on oil spill detection.

Technique	Author	Description	Limitations
Neural Network (NNs)	Park *et al.* 2020	Employed an ANN architecture with 1000 epochs, a learning rate of 0.01 and a hidden layer with 8 neurons to classify oil spills from optical pictures.	Artificial neural networks (ANNs) have no set structure determined by any one rule.
	Lee *et al.* 2016	Developed a method for determining the temporal evolution of oil spills using high spatial resolution data from DubaiSat-2 and Landsat OLI. The proposed oil spill was classified as either relatively thick oil or film-like oil using a recursive neural network technique. In contrast to the non-recursive method, the technique shows significant differences in the area-based amount of oil spilled in the bay.	To training a recursive neural network (RNN) is a completely tough task.
Support vector machines (SVMs)	Hassani *et al.* 2020	Polarimetric Synthetic Aperture Radar (PolSAR) data has unique capabilities and informative qualities that can be very helpful for large-scale oil spill detection. Using PolSAR data, four categories were created for the detected oil spill in the ocean: heavy oil, thin oil, oil/water combo, and clean water. The excellent acquired classification accuracies of 90.21% and 85.41%, as well as the Kappa values of 0.8052 and 0.7905, illustrate the robustness of the proposed SVMs classifier.	SVM doesn't perform very well once the set of data contain more noise
Random forest	Tong et al. 2019	Using Radarsat-2 and UAVSAR polarimetric SAR datasets, a random forest was presented to increase the accuracy of oil spill detection; moreover, accuracy of the proposed technique reaches 92.99% and 82.25%, respectively, with two datasets. Therefor improve the discrimination ability between look-alikes and oil slicks.	Once there are too many trees in the random forest, the method becomes too sluggish and unreliable for real-time prediction.

Yang *et al.,* (2017) suggested a technique for detecting oil spills from SAR pictures that is based on the Hilbert spectrum of the bidimensional empirical mode decomposition (BEMD). The BEMD is used to handle nonstationary and nonlinear signals by using a bidimensional IMF (BIMF) and residue signal to determine the area of interest (ROI) from which the feature vector may be constructed. Consequently, the ROI image consisted of three IMFs and residual. To transform the 2-D data into 1-D data, find the Hilbert spectrum and the Hilbert marginal spectrum. The amplitude and frequency data will be easy to extract, and texture analysis will be performed by creating a feature vector. Suggested Mahala-Nobis distances are used for classification with above-average accuracy rates.

Singha *et al.*, (2016) developed a chain of processing that uses coherent dual-polarimetric terreSAR-X images to identify oil spills. The dark spot segmentation, extracted features, and final output were the three processes in the recommended methodology. Consequently, the approach makes use of traditional calibration and processing. An adaptive thresholding approach was used for segmentation in order to exclude dark items inside the region of interest (ROI). From every segmented dark spot, a set of polarimetric features and a set of conventional characteristics were extracted. Between speckle reduction and resolution preservation, polarimetric feature extraction was found. Furthermore, feature parameters obtained from support vector machines (SVMs) are used for training and calibration; the methodology's classification accuracy shows that the recommended approach correctly detects 90% of oil spills and 80% of lookalikes from dataset with an overall accuracy of 89%.

Mera *et al.*, (2016) developed a generic and systematic methodology based on feature selection (FS) approaches for choosing a condensed and applicable collection of attributes to improve oil spill detection. The proposed techniques contrasted five feature selection (FS) strategies: Correlation-based feature selection (CFS), Consistency-based Filter, Information Gain, Relief, and Recursive Feature Elimination for Support Vector Machine (SVM-RFE). Furthermore, the proposed methodologies were utilized to choose a relevant set of features with excellent discriminating control between oil spills and lookalikes. Consequently, feature selection (FS) was used to acquire and minimize the feature vector. The methods were evaluated using SVM classifier, and an overall accuracy of 87.1% was obtained.

DOI: 10.56892/bima.v8i1B.647

Table 2: An overview of the numerous studies on the detection of oil spills

Authors	Dataset	Description	Models	Accuracy	Advantage	Disadvantage
Conceição et al. (2021)	SAR	Using two random forest classifiers enhances accuracy in categorizing oil spill inputs and detecting spill targets.	Random Forest/ML	90%	Adopt a novel and open-source method to oil leak detection.	There are still some hazards when there is no data in a radar image's outer regions.
Tong et al., (2019)	SAR	Random Forest classification for oil spill detection in polarimetric SAR data.	ML/random forst	92.99% and 82.25%	Tests reveal that the approach outperforms the three popular approaches in these datasets in terms of accuracy.	The suggested method's performance could be impacted by severe sea conditions or varying degrees of weathering slicks.
Shaban et al. (2021)	SAR	A two-step DL framework using CNN and U-Net for oil spill detection.	DL/CNN	92%	In comparison to related work, provide a comparable increased precision and Dice score.	When solving a multiclass problem, it is useless.
Zeng and Wang (2020)	AAMLP	OSCNet, a DCNN with 12 layers, improves SAR dark patch classification.	DL/ DCNN	94.01%	The final two steps of the processing framework are implemented using a three-step procedure that comprises picture segmentation, feature extraction, and target categorization	Most people focus on pixel-level deep learning classification techniques.
Yekeen and Balogun (2020)	SAR	Instance segmentation DL model with Mask RCNN.	DL/ Mask RCNN	-	The model can use shape and texture information to detect and locate oil spills, ships, and land areas.	Creating a comprehensive oil spill SAR image database is essential to enhance model accuracy.
Ghorbani and Behzadan (2021)	In-house image called Nafta	The VGG16 model was used for oil spill classification and PSPNet and mask-R-CNN models for segmentation.	DL/CNN	92%	Used a dataset generated by web mining.	When there is insufficient contrast, it is difficult to distinguish between an oil spill and its surroundings.

Direction of Research

Future prospects for oil spill detection research have been proposed in a number of ways. This section presents numerous suggestions for future machine learning-based oil spill detection research directions:

- Dataset: Several requirements are included in this term's dataset requirement, including:
  - Volume of the dataset: In order to maximise the model's accuracy, a sizable database of photos of oil spills must be provided.
  - Dataset availability: No publicly available or industry-standard oil spill databases exist. In the future, researchers could create a dataset that is available to all developers.
  - Standard criteria for datasets: An industry standard dataset is necessary to ensure that the results drawn by each study team have a uniform measuring standard.
  - Various datasets were utilized to train the model, including SAR images, shipborne radar, and drone-based Nafta dataset. In a different way, researchers may focus their efforts on creating a dataset moment of the SAR image.
- Scalability: DL models often continue to function well as dataset sizes increase.
- Method of classification: To improve accuracy, consider label 0 instead of labels 0, 1, 2, and so on. This avoids classifying land, sea, and look-alikes as a single large category.
- Detection context: The model is often run in a standalone context for oil spill detection; however, a cloud computing environment is a wise choice to enhance results and offer real-time detection.
- Overarching structure: Examine well-known deep learning approaches for Framework Content and test them on alternative datasets.

CONCLUSION

The world's population is growing faster than ever, and industrialization and contemporary transportation methods indicate that there is a greater need for oil, which has led to an increase in oil spills. Environmental danger will so increase as well. This study set out to examine many modern approaches to oil spill detection and tracking. In fact, a wide range of approaches have been developed to pinpoint the issue of oil spills. In order to serve as a reference for those new to this topic, the article only concentrates on machine learning approaches and classifies them. Evaluated each study's benefits and drawbacks in further detail. However, it also highlighted a popular topic in the domains where scholars may work. Although there are alternative methodologies, our study was limited to machine learning investigations. However, our future work aims to expand the paper's content to include additional statistical studies.

REFERENCES

Chehresa, S., Amirkhani, A., Rezairad, G.-A., & Mosavi, M. R. (2016). Optimum Features Selection for oil Spill Detection in SAR Image. *Journal of the Indian Society of Remote Sensing, 44(5), 775–787.*

Conceição, M. R. A., de Mendonça, L. F. F., Lentini, C. A. D., da Cunha Lima, A. T., Lopes, J. M., de Vasconcelos, R. N., ... & Porsani, M. J. (2021). Sar oil spill detection system through random forest classifiers. *Remote Sensing*, *13*(11), 2044.

De Kerf, T., Gladines, J., Sels, S., & Vanlanduit, S. (2020). Oil spill detection

Ghorbani, Z., & Behzadan, A. H. (2021). Monitoring offshore oil pollution using multi-class convolutional neural networks. Environmental Pollution, 289, 117884.

Hassani, B., Sahebi, M. R., & Asiyabi, R. M. (2020). *Oil Spill Four-Class Classification Using UAVSAR Polarimetric Data. Ocean Science Journal.* 55, 1–11.

Lee, M.-S., Park, K.-A., Lee, H.-R., Park, J.-J., Kang, C.-K., & Lee, M. (2016). Detection and Dispersion of Thick and Film-Like Oil Spills in a Coastal Bay Using Satellite Optical Images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(11), 5139–5150.*

Mera, D., Fernández-Delgado, M., Cotos, J. M., Viqueira, J. R. R., & Barro, S. (2016). Comparison of a massive and diverse collection of ensembles and other classifiers for oil spill detection in SAR satellite images. *Neural Computing and Applications, 28(S1), 1101–1117.*

Park, S.-H., Jung, H.-S., & Lee, M.-J. (2020). Oil Spill Mapping from Kompsat-2 High-Resolution Image Using Directional Median Filtering and Artificial Neural Network. *Remote Sensing, 12(2), 253.*

Seydi, S. T., Hasanlou, M., Amani, M., & Huang, W. (2021). Oil Spill DetectionBased on Multiscale Multidimensional Residual CNN for Optical RemoteSensing Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*,10941–10952. Shaban, M., Salim, R.,Khalifeh, H. A., Khelifi, A., Shalaby, A.,El-Mashad, S., Mahmoud, A., Ghazal.

Shaban, M., Salim, R., Abu Khalifeh, H., Khelifi, A., Shalaby, A., El-Mashad, S., ... & El-Baz, A. (2021). A deep-learning framework for the detection of oil spills from SAR data. *Sensors*, *21*(7), 2351.

Singha, S., Ressel, R., Velotto, D., & Lehner, S. (2016). A Combination of Traditional and Polarimetric Features for Oil Spill Detection Using TerraSAR-X. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(11), 4979–4990.*

Song, D., Zhen, Z., Wang, B., Li, X., Gao, L.,Wang, N., Xie, T., & Zhang, T. (2020). A Novel Marine Oil Spillage Identification Scheme Based on Convolution Neural Network Feature Extraction from Fully Polarimetric SAR Imagery. *IEEE Access*, *8*, 59801–59820.

Sun, S., & Hu, C. (2019). The Challenges of Interpreting Oil-Water Spatial and Spectral Contrasts for the Estimation of Oil Thickness: Examples From Satellite and Airborne Measurements of the Deepwater Horizon Oil Spill. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(5), 2643–2658.

Tong, S., Liu, X., Chen, Q., Zhang, Z., & Xie, G. (2019). Multi-feature based ocean oil spill detection for polarimetric SAR data using random forest and the self-similarity parameter. Remote Sensing, 11(4), 451..

using machine learning and infrared images. *Remote Sensing*, *12*(24), 1–13.

Yekeen, S. T., & Balogun, A. L. (2020). Automated marine oil spill detection using deep learning instance segmentation model. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 1271-1276.

Zeng, K., & Wang, Y. (2020). A deep convolutional neural network for oil spill detection from spaceborne SAR images. Remote Sensing, 12(6), 1015.