Characterization of Oil Spills Types Using Wavelet Analysis and Machine Learning Method in Satellite Images: Environmental Study

Sulaya Betsabe Bayancela-Delgado
sulaya.bayancela@espoch.edu.ec
https://orcid.org/0000-0003-4589-5488
Facultad de Recursos Naturales.
Escuela Superior Politécnica de Chimborazo – ESPOCH

Iván Mesias Hidalgo Cajo
ihidalgo@espoch.edu.ec
https://orcid.org/0000-0002-9059-0272
Escuela Superior Politécnica de Chimborazo – ESPOCH

Isabel Romane Peñafiel Moncayo
romane.peniafiel@espoch.edu.ec
https://orcid.org/0000-0002-2629-8576
Escuela Superior Politécnica de Chimborazo – ESPOCH
Sede Moróna Santiago

Ing. Oswaldo Geovanny Martínez Guashima
omartinez@espoch.edu.ec
https://orcid.org/0000-0003-4561-122X
Escuela Superior Politécnica de Chimborazo – ESPOCH

López González Wilmer Orlando
wilmerlopez@unae.edu.ec
https://orcid.org/0000-0002-6197-8665
Ciencias Experimentales, Universidad Nacional de Educación (UNAE)

Leonilo Durazno Delgado
ldurazno@uteq.edu.ec
https://orcid.org/0000-0002-9065-3026
Universidad Técnica Estatal de Quevedo
Facultad de Ciencias de la Ingeniería y Producción,
Carrera de Ingeniería en Alimentos

Background: In this research work a Genetic algorithm has been investigated to identify the occurrence of an oil spill disaster in the oceans. However, to know the type of oil spilled in the ocean is very important aspects which help in planning and fast cleanup process. Through the machine learning technique, it is very challenging to predict the type of oil in the ocean using SAR data. In this research, fifty satellite images with three classes namely crude oil, petroleum, and diesel were monitored and examined to identify the type of spill. The oil spills were identified using the KNN classifier algorithm. Features investigated as Color-based, Textural, Statistical, and Geographical features are extracted through various types of wavelets to determine the type of oil in the ocean.

Methods: For wavelet analysis (COIFLET analysis coif1, coif2, coif3, coif4, coif5 and machine learning (KNN classifier algorithm) applied to achieved feature sets using SAR data with 70% training and 30% testing.
Results: It also represents 97% accuracy for crude oil using coif4 analysis which indicates better characterization of oil spills. The study demonstrates that the Genetic Algorithm can be used as a good tool for oil spill detection using satellite images. As a result, it is clearly shown that the Genetic Algorithm can be used as a good tool for tracking oil spills using a satellite image with the highest covering spilled area and best classification method.

Keywords- Satellite Image, Oil Spill, Genetic Algorithm, Machine Learning Method, Wavelet Analysis, KNN classifier.

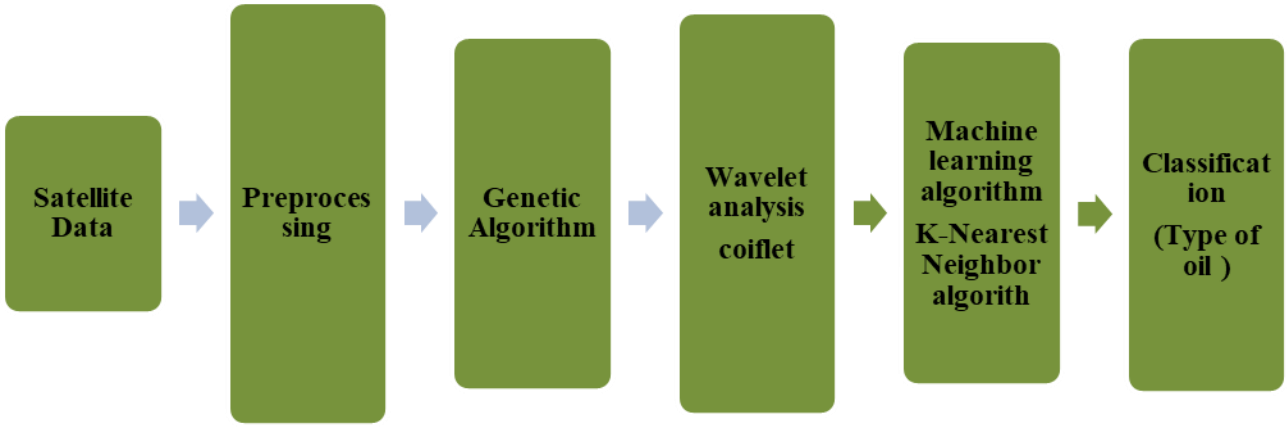
1. Introduction

Oil spills are one of the important research aspects to detect the oil spill area in the ocean. From the past years many disaster took place over the ocean, stills experts are researching the incident. For this research best source of detection is SAR data because it covers a larger area of disasters place. Satellite image works good on focusing oil spill compare to other sources. In the past years the research expert used MODIS-TERRA for data collection using SVM technique for extracting features for oil spills characterization and transformation for collection of data sources [1]. Similarly in this research, it used database on RADARSAT-1 using Technique DLA for extracting Features specificity and sensitivity, it tests the type of oil calculation using method DLA from different sources and regions [2]. Similarly this research also works on database of SAR using technique SVM to extract feature specificity and Sensitivity for oil spills detection using characterization method, over the oil spills disaster [3]. Another work is based on CPG data using techniques PARAFAC algorithm on oil fraction for detection of petroleum oil type [4]. By using synthetic aperture radar data for chemical analysis to detect oil spills types and stage of effects [5]. Another research took place using Acute ecotoxicity and a sample of sediments using technique advanced chemical analytical for determining fraction petroleum establishment [6]. The SAR image for classification type off oil spills using thematic mapping technique to identify the type, size, and thickness for crude oil [7]. Research worked on GC-MS imaginary model using molecular modeling technique for model detection [8]. By using GC-MS images for Biomarkers methods to extract the source of spilled oil and correlation observation, a researcher works on Petroleum with multivariate analysis [9]. The SAR image for oil spills characterization based on the thresholding method, resulting in an image represents crude oil with sediment concentration over periodic time and observation [10].

2. Methodology

In this research is the characterization of oil spills. It contains a wavelet transform and machine learning techniques. In this research, fifty satellite images were taken for oil spill characterizations using wavelet transform analysis & machine learning techniques. For wavelet analysis such as DWT coiflet families such as coif1, coif2, coif3, coif4, coif5 analysis, and machine learning such as k-nearest neighbor

algorithm applied to optimize the oil spill feature sets. Features included Color-based feature (RGB), statistical features (mean & standard deviation), Texture Feature (entropy, ellipticity, intensity, correlation coefficient), Geological feature (spreading and complexity). This experiment was conducted on RADARSAT-2 SAR images. The features were classified using the k-nearest neighbor algorithm. Seventy percent of features used for training and thirty percent for testing.

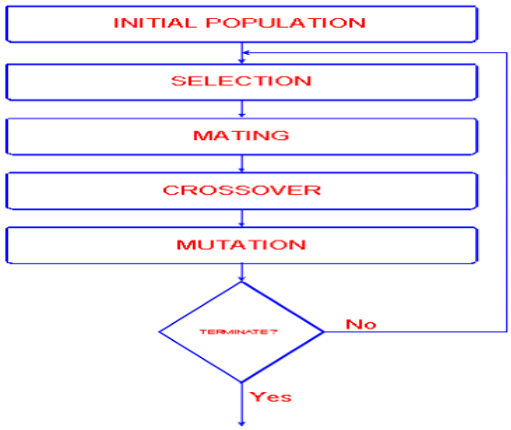


Flow chart 1. Oil spill characterization

3. Genetic Algorithm

It is an adaptive heuristic searching algorithm according to evolutionary ideas of genetics and natural selection. A genetic algorithm is intelligent exploitation of a random search that helps to find out solutions for optimization problems. Steps involved in genetic algorithms are initial population, selection process, mating, crossover, mutation.

Flow chart 2. for Genetic Algorithm



Population initialization

The genetic algorithm starts with the initial population. This initial population is randomly generated. The chromosomes help to represent populations. The chromosomes represented as strings character to encoded solutions to a problem.

Fitness Assignment

This section helps to evaluate the population in each chromosome. The results satisfied the criteria of fitness during evaluation and those criteria who do not match or satisfied will be discarded.

Selection process

In the genetic algorithm first step is the selection where the selection of the chromosome will satisfy the fitness function is done and this selection chromosome is utilized for a meeting after the genetic operation like mutation or crossover is utilized for selected parent for new offspring generation. After the selection crossover and mutation are applied.

Crossover

In the crossover, the process involved two chromosomes, and the swapping of the data of the two chromosomes during the crossover process new offspring generation exchange copy of the parents. The solution that is a chromosome that is more fit is selected and replaces the next fit solutions. This way the new generation of the population is created.

Mutation

It is used to find out the diversity among the population. It reduces slight changes among the population and representative of the evaluation approach. The process repeats until conditions satisfied. The new offspring are placed in the new population and the new optimal solution is

obtained on the bases of pre-specified criteria. Then it will terminate and written the best solution.

Algorithm

Step 1 *First step is to generate an initial population of individual*

Step 2 *Second steps are to evaluate fitness individual*

Step 3 *While condition*

3.1 Selection process individuals based on reproduction

3.2 Recombine individuals between them

3.3 Apply mutation process for individuals

3.4 Evaluate modified individuals based on fitness value

Step 4 *End while*

Step 5 *Terminate the process*

4. Feature Extraction

In this study, nine features including RGB, spreading, complexity, mean, standard deviation, entropy, ellipticity, intensity, correlation coefficient are extracted from the covariance matrix of SAR data. Here nine features have been investigated they are as follows.

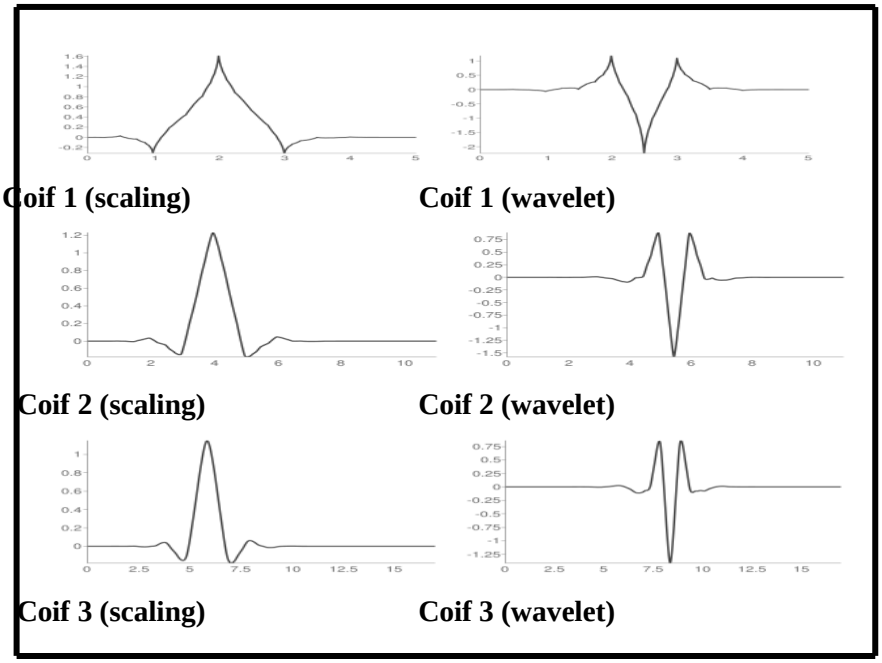
Table 1. Features Investigated

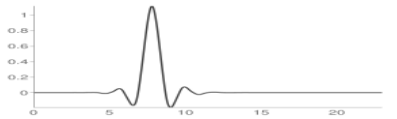
Features	**Definition**
Color-based feature RGB	True-color image (m-by-n-by-3 data array)
Geological Features	**Formula**
Spreading	$s = 100\lambda/(\lambda_1+\lambda_2)$
Complexity	$c = P/2\sqrt{\pi A}$
Statistical Features	**Formula**
Mean	Sum of all observation divided by several observati
Standard Deviation	Pixel intensity values belong to the object.
Thermal Features	$p_i = \frac{\lambda_i}{\sum_{j=1}^{3}\lambda_j}$
Entropy	
Ellipticity	$sin(2x) = -\frac{s_3}{ms_0}$
Intensity	S_{VV}^2
Correlation Coefficient	$Coh = \frac{\lvert < T_{12} > \rvert}{\sqrt{< T_{11} >< T_{22} >}}$

5. Coiflet Analysis

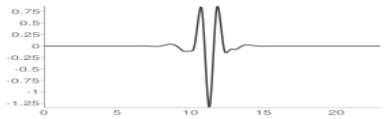
It is a symmetrical wavelet, it contains COIFLET families such as coif1, coif2, coif3, coif4, coif5 analysis. It is part of discrete wavelet transform techniques. In COIFLET families, coiflet1,2,3,4 and 5 have near-symmetrical, biorthogonal, and orthogonal properties.

Figure 1. Coiflet Analysis

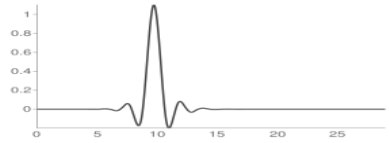




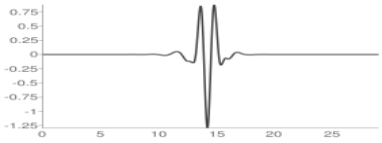
Coif 4 (scaling)



Coif 4 (wavelet)



Coif 5 (scaling)



Coif 5 (wavelet)

6. K-Nearest Neighbor Algorithms

This is done using K-NN classifier, class membership is the output. On the training data set, machine learning algorithms have excellent performance than neural networks. Based on pre-training, machine learning algorithms such as k-nearest neighbors have a stronger capability to achieve the optimized solution of the problems. The results show that oil spill classification achieved by wavelet transforms and machine learning algorithms outperformed very well with similar parameter settings, especially with 70% training data and 30% testing data using a confusion matrix. The performance was analyzed by K-Nearest Neighbor (KNN) classifiers. KNN assigns a class based on the predominant class among the k nearest neighbors. The value of k was chosen as the number of classes used for classification. In this research work, the features derived from the collected data set with 70% and 30% for all oil spilled images. Then the 70% features were used for training the classifier and 30% features were used for testing. The testing and training features belonged to random subjects and varied in each run of the program.

7. Result And Discussion For Coiflet Analysis

For coiflet analysis, the test sample classified as petrol, diesel, or crude oil, where k value is 4 because the data point is more attracted to each other than other k values comparatively. The object assigned to the class of a single nearest neighbor. The given table represents coiflet analysis with three classes' petrol, diesel and crude oil using the confusion matrix have been examined.

Table 2: Overall Accuracy result of Color-based, statistical, geological and other features results using COIFLET analysis with K-Nearest Neighbor classifier

	Accuracy Petrol (%)	Accuracy Diesel (%)	Accuracy Crude (%)	Total Accuracy (%)
Coiflet 1	83	92	81	85
Coiflet 2	95	70	84	83
Coiflet 3	88	96	79	88
Coiflet 4	69	89	97	85
Coiflet 5	75	92	92	86

The table represents oil spills characterization; using coiflet analysis, it also illustrates that coiflet 4 with 97% provides higher accuracy of crude oil.

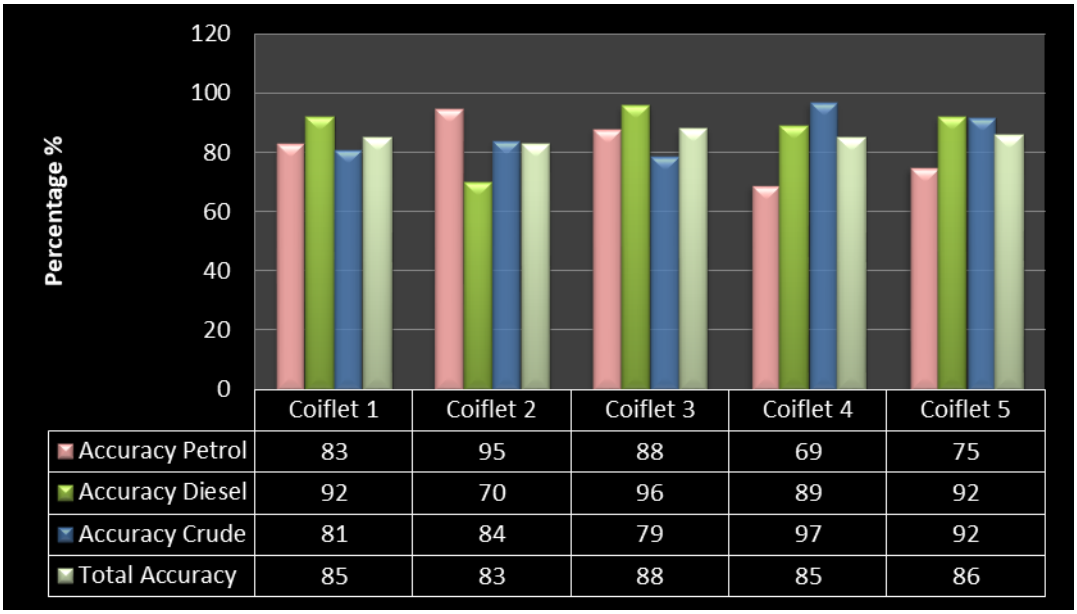


Figure 2: Result using Coiflet Analysis

The figure represents oil spills characterization; using coiflet analysis, it also illustrates that coiflet 4 with 97% provides higher accuracy of crude oil than compared to other measures.

Conclusion

In this research work, a Genetic algorithm has been investigated to identify the occurrence of an oil spill disaster in the oceans. However, to know the type of oil spilled in the ocean is very important aspects which help in planning and fast cleanup process. Through the machine learning technique, it is very challenging to predict the type of oil in the ocean using SAR data. In this research, fifty satellite images with three classes namely crude oil, petroleum, and diesel were monitored and examined to identify the type of spill. The oil spills were identified using the KNN classifier algorithm. Features investigated as Color-based, Textural, Statistical, and Geographical features are extracted through various types of wavelets to determine the type of oil in the ocean. For wavelet analysis (COIFLET analysis coif1, coif2, coif3, coif4, coif5 and machine learning (KNN classifier algorithm) applied to achieved feature sets using SAR data with 70% training and 30% testing. It also represents 97% accuracy for crude oil using coif4 analysis which indicates better characterization of oil spills. The study demonstrates that the Genetic Algorithm can be used as a good tool for oil spill detection using satellite images. As a result, it is clearly shown that the Genetic Algorithm can be used as a good tool for tracking oil spills using satellite images with the highest covering spilled area and best classification method.

References

1. Harlow, C.,and Argialas, D., Computational image interpretation models. Photogram. Engg. and Remote Sens., 1990.
2. Betti, M., Barni, M., and Mecocci, A., A fuzzy logic approach for detection of oil spills on SAR images. In Interl. Geosci. and Remote Sens. (Symposium), 1995.
3. Wahl, T., Bern, T.I., Olsen R., and Anderssen, Satellite based oil spill detection in SAR data. (proceedings) Symposium on ERS-1France, Cannes, 1992.
4. Brekke, C. and Solberg, A.H.S. Detection of oil spills by using satellite remote sensing process. Remot.Sens. (Environment), 2005.
5. Canny, J., Edge detection using computational approach . IEEE Transa. of Pattern Analysis and (Intelligence) Machine, 1986.
6. Satellite SAR images of Cearac database , http://cearac.poi.dvo.ru/en/db/ (July 2006 accessed) 2003.
7. . Available at #ESA online: Vese, L. and Chan, T. Without edges an active contour model. Proceedings of second Conference (International) on scale space, Greece, 1999.
8. Change, L.Y., Multiresolutionapproach with A multiplayer to detect slicks using ERS SAR images. Proceedings seventeen Asian Conference of Remote Sens., Sri Lanka, 1996.

9. Chen, C.F, Satellite imagery uses and monitoring coastal environment in Taiwan. Geoscience and Remote Sens. (International) Symposium, 1997.

10. Del Frate A., Calabresi, G. and Lichtenegger, J., Neural based Networks for detection of oil spills using ERS-SAR data. IEEE Transa. On Geoscience and Remote Sens., 2000.