Feature Selection and Classification of Oil Spill from Vessels using Sentinel-1 Wide–Swath Synthetic Aperture Radar Data

L.W. Mdakane and W. Kleynhans,

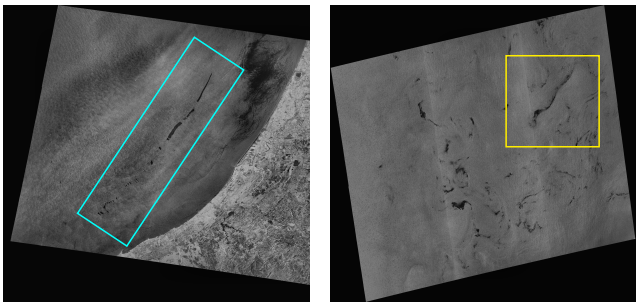
***Abstract*—Oil spills are often caused by vessels when dumping oily bilge wastewater at sea (also referred to as bilge dumping). In a SAR image, oil spills dampen the radar energy return and appear as linearly shaped dark regions. However, naturally occurring phenomena (e.g. natural seepage) known as oil spill look-alikes can also dampen energy return and occur more often compared to a real oil spill. The primary goal of the study is to develop a monitoring system dedicated to automatically detect oil spill events caused by ships (bilge dumping) in African Oceans. To achieve this goal, the knowledge of features that have a high probability of separating oil spills from look-alikes is of great importance. The study aimed to accomplish three things, a) to improve the lack of oil spill studies in Africa; b) to determine the critical features that yield the highest discrimination accuracy of oil spills caused by moving vessels from look-alikes; c) use these features to automatically detect and classify oil spill events. The study investigated the most common features used in literature for discriminating oil spills from look-alikes from SAR imagery. Multiple feature selection methods and the Gradient Boosting Decision Tree Classifier (GBT) were used to select, classify and determine the significant features for discriminating oil spills caused by moving vessels. The results showed that while some features vary, there are features that consistently have high and low significance across all methods.**

***Index Terms*—Oil spill, Bilge waste dumping, Synthetic Aperture Radar, Feature extraction, Object classification.**

I. Introduction

OIL spills from vessel discharges (bilge dumping) are regarded as a primary source of oil slicks and are estimated to be higher than accidental oil spills [1]. Depending on the amount and location, oil spill events can be very harmful to the sea ecosystem. Using spaceborne Synthetic aperture radar (SAR), illegal bilge dumping activities can be successfully monitored, due to its ability to capture large areas under most weather conditions, day or night [2].

In SAR images, bilge dump events appear as a linear dark area due to the vessel's movement, see Fig. (1a). However, this is not unique to oil spill events as naturally occurring phenomena (e.g. waves, natural seepage) can also be seen as linearly shaped dark regions. These phenomena are known as oil spill look-alikes and occur more often compared to the real oil spill, see Fig. (1b). Oil spill events can be detected either manually or using semi/fully-automated surveillance systems from a SAR image. The systems involve several steps, including: a) identifying possible bilge dumps; b) feature extraction; and c) the analyse of the extracted features and determine whether it is an oil spill or a look-alike phenomenon (classification). The study focuses on feature extraction for discriminating oil spills caused by moving vessels from look-alikes.



(a) Morocco (July 2016). (b) Morocco (March 2016)

Fig. 1. Sentinel-1A wide swath SAR images with oil spill event and oil spill look-alikes.

Feature extraction involves the selection/extraction of a vector of features that quantitatively describe relevant characteristics of the object. They extract possible oil spill and look-alike features from a SAR image and use that information to classify oil spill events. Although using all possible oil spill features can yield acceptable results, they are not always consistent. The inconsistency is because some features may be useful in one situation but can have zero contribution in another [3]. Topouzelis *et al.* [4], in a 2012 study concluded that there was not a single optimum feature combination out of the 25 most common combinations considered but several sets of combinations existed which contained at least some critical features. Recently, more work has been done on oil spill feature selection and feature ranking to determine features that yield the highest classification [5], [6]. However, the selection of features that yield the highest discrimination of oil spill from look-alikes is still a challenge. This was particularly challenging as it was difficult to compare or determine which features yield consistent performance based on the literature as they were not always be consistent.

Our primary goal is to develop a monitoring system dedicated to automatically detect oil spill events caused by ships (bilge dumping) in African Oceans. To achieve this goal, the knowledge of features that have a high probability of separating oil spill from look-alikes is of great importance. The study

L.W. Mdakane with the Spatial Information Systems, e-Government Next Gen Enterprises and Institutions, CSIR, Pretoria, South Africa.

W. Kleynhans with the University of Pretoria, Department of Electrical, Electronic and Computer Engineering, South Africa.

aimed to accomplish three things, a) to improve the lack of oil spill studies in Africa; b) to determine the critical features that yield the highest accuracy for oil spills caused by moving vessels from look-alikes; c) to use these features to classify these oil spill events. The study investigated the most used (common) features used in literature for discriminating oil spill from look-alikes from SAR imagery. The most useful features for oil spill events from vessel discharges were determined using multiple feature selection methods. The features were evaluated and ranked according to importance using a Gradient Boosting Tree (GBT) Classifier.

II. Data Description

A large number of SAR images with oil spill examples is critical in developing and evaluating an automated oil spill monitoring system. Currently, there is a lack of studies for automated ocean monitoring applications in the African coastal areas. Therefore, the study areas covered oceans surrounding three African countries; A) Morocco, B) Mozambique and C) South Africa.

Large swath SAR imagery have been used successfully to detect bilge dumping activities at sea [7]. The data consisted of Sentinel-1 (SEN1) Ground Range Multi-looked Data (GRD) using two modes:

- Interferometric Wide Swath (IW) with 250 km swath, high (10 x 10 m) and medium (40 x 40 m) spatial resolutions.
- Extra-Wide Swath (EW) with 400 km swath, high (25 x 25 m) and medium (40 x 40 m) spatial resolutions.

The dataset had a total of 127 co-polarised (VV) images, where 16 were EW mode and 111 were IW mode. Only co-polarisation images were used for the study due to the stronger backscattering properties in oil spill studies [8]. All the images had either oil spills or look-alikes or both examples. The dataset was vertically co-polarised (VV) and was acquired from early 2015 to the end of 2016.

III. Methodology

Several features need to be extracted to successfully discriminate oil spill events from look-alikes from a SAR image [9]. The oil spill can be related to different events, and depending on those, have different geometries, which could influence which features were extracted and selected for the classification task. The study focused on the task of detecting of oil from moving vessels using only the most critical features derived from a SAR image.

The first step was to determine all available (or the most relevant) dark-spot and non-dark-spot features in literature. The features were tested on wide-swath Sentinel SAR data over three African countries. The dataset was separated into two classes (oil slicks and look-alikes) which were visually identified by an expert. The data had a total of 120 look-alike samples and 90 oil spill samples. All considered features were extracted for each class using 8-bit, calibrated and segmented images, see experiment flowchart in Fig. (2).

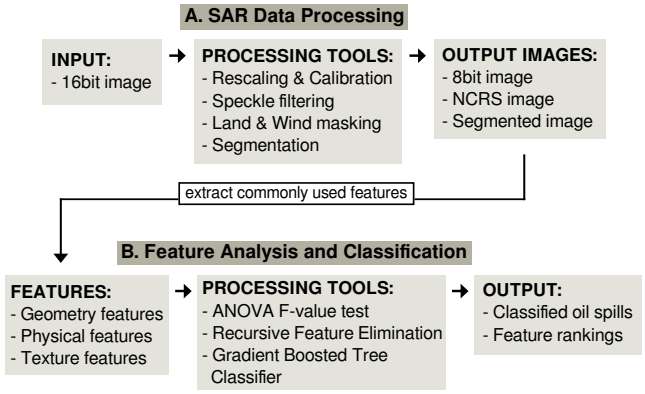


Fig. 2. Oil spill classification and feature ranking processing flowchart.

A. Data pre-processing

All images were geo-referenced by converting all pixel coordinates (x,y) to geographic (latitude, longitude) to global geodetic ellipsoid reference WGS84 from the tie point grids of the source product. A land masking operation was then applied to remove land areas to isolate ocean areas, any pixel on land was assigned a null value using a world land shapefile as mask. A Lee filter was used to reduce speckle while preserving edges and textural information [10]. The wind speed was estimated from each SAR image using the CMOD5 model [11]. All available SAR images were then visually inspected, and regions with wind speeds between 3–13 m/s containing oil spills or look-alikes were selected to create sub-images. Each sub-image was calibrated using σ_0 (also termed normalised radar cross section or NRCS) look-up table [12]) and segmented with an iterative threshold-based and region-based active contour model were used to segment linear dark-spots [7]. Fig. (3) shows an example of segmented SAR image.

B. Common Features

Relevant features should exhibit a clear relationship between the value of the selected feature and the probability of it being an oil spill [3]. The most common features of dark-spots and a limited area outside the dark-spot (oil spill free area) were extracted from the sub-images. These features can be categorised as geometric, physical and texture features. A brief discussion of the most common features for each category is presented below; detailed descriptions can be found in table (I, page 4) and from literature ([9] and [13]).

1) Geometric features: Describe the geometry and shape of the segmented dark-spots. A wide range of geometrical features have been proposed, where some features have been found more useful than others [13].

2) Physical features: Describe the backscatter values of the dark-spot and its surroundings. These are considered as the most important features for discriminating oil spill from look-alikes [3].

3) Textural features: Describe underlying texture of the dark-spot and region around it. These have been shown to be the least important features for oil slick discrimination [3],

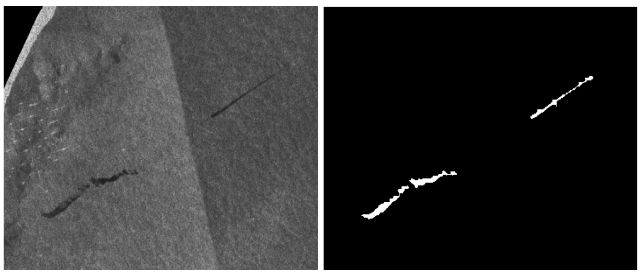


Fig. 3. Sentinel-1A Extra Wide mode SAR image and segmented sub-images with oil spills.

[13]. However, some studies have shown that Haralick textures [14], can be useful to a certain extent [15].

C. Important Features Selection

An important feature is a feature that enables the highest possible discrimination accuracy. There are numerous ways to determine important features, two methods were considered in the study.

1) Statistical analysis: The one-way analysis of variance (ANOVA) with F-value was used to determine whether there were any statistically significant features using differences between the means of two or more independent features. That is, for each feature the sample mean and sample variance was estimated and compared using the F-test taken from Snedecor F-table [16].

2) Recursive Feature Elimination (RFE): Features are recursively removed, and features that remain are evaluated. The most important feature set is one that shows highest classification accuracy.

D. Gradient Boosted Tree Classifier

Boosting refers to applying multiple weak estimators to produce a more powerful estimator. Gradient Boosted Tree (GBT) classifier builds an estimator by sequentially adding new trees learners to the expansion model. GBT sequence of tree expansions are of the form:

$$F_m = \sum_{m=1}^{M} T(x, \Theta_m). \tag{1}$$

where $T(x; \Theta)$ is the "weak tree estimator" of the input variables x, characterised by parameters Θ_m. For regression tree, the parameters Θ_m are the splitting variables, split locations and the terminal node of the individual trees [17].

Equation (1) is a minimisation problem that can be solved numerically using a steepest descent direction method [18]. The steepest descent direction is the negative gradient of the loss function evaluated at the current model (F_{m-1}):

$$F_m(x) = F_{m-1}(x) + \gamma_m T(x, \Theta_m), \tag{2}$$

where γ_m is the step length chosen using line search:

$$\gamma_m = \arg\min_{\gamma} \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \gamma T(x_i, \Theta_m)).$$

Equation (2) can be calculated for any differentiable loss function. For classification, the negative binomial log-likelihood (also known as deviance) is the commonly used loss function $L(y, F)$:

$$log(1 + e^{2yF}) \text{ for y } \in \{-1, 1\}.$$

E. GBT Parameters

1) Number of trees: This parameter controls the number of boosting stages to perform. GBT is relatively robust to over-fitting so a large number can be used without losing performance. However, oversized trees can substantially degrade performance and increase computation.

2) Shrinkage or Learning rate: Shrinkage controls how strongly each tree tries to correct the mistakes of the previous trees and is dependent on the number of trees. Shrinkage is implemented by scaling the contribution of each tree by a factor α when it is added to the current approximation (2):

$$F_m(x) = F_{m-1}(x) + \alpha\gamma_m T(x, \Theta_m).$$

3) Sub-sampling: This parameter controls the number of samples to be used for fitting the individual base learners. Not only does the sampling reduce the computing time, but can also produce a more accurate model.

4) Feature Ranking: GBT measures the importance of each feature by averaging feature ranking over several trees. It provides a more stable measure of feature importance by reducing the variance of feature selection methods.

IV. Experimental Results and Discussion

Verification of an oil spill can be costly and time-consuming. Therefore the highest possible detection accuracy is essential. The study objective was to classify oil spill from moving vessel and determine the most critical features that yield the highest accuracy.

A. Classification Results

The study considered two feature selection methods, that is, ANOVA and RFE. They used all 29 commonly used features, see table I, found in literature as input and selected the most important features. The most common features of dark-spots and a limited area outside the dark-spot (oil spill free area) were extracted from a total of 120 look-alike samples, and 90 oil spill samples. A supervised Gradient Boosted Tree (GBT) classifier, with optimised parameters, was used to classify oil spill from look-alikes from the selected features. A k-fold cross-validation method ($k = 5$) to split training and testing set and to evaluate the overall classification accuracy for each feature set, see table II. The results in table II were achieved as follows:

1) GBT^*: All the 29 features were used as input and were evaluated with GBT classifier. Classification accuracy (GBT^*) using all features showed the lowest accuracy (77.4%).

TABLE I
FEATURES DESCRIPTION, CODES AND FEATURE IMPORTANCE MEASURE[a] (DESCRIBED BY THE SELECTION FREQUENCY AND AVERAGE RANKING).

	Description	Code	Selection Frequency		Average Ranking	
			Literature[b]	Study	Literature	Study
Geometry Features:						
1	Area	A	○	○	○	○
2	Perimeter	P	○○	●●●	○○	○○
3	Perimeter to Area ratio	P/A	○○	○○	○○	○○
4	Complexity	C	●●●	○○	○○	●●●
5	Shape Factor 1	SP1	○○	●●●●	●●●	●●●
6	Shape Factor 2	SP2	●●●●	●●●●	●●●●	●●●●
Physical Features:						
7	Object Mean	OMe	●●●	●●●●	○○	●●●
8	Background Mean	BMe	●●●	●●●●	●●●	●●●
9	Object Standard Deviation	OSd	●●●●	●●●	●●●●	●●●●
10	Background Standard Deviation	BSd	●●●●	●●●	●●●●	●●●●
11	Object Power to Mean Ratio	Opm	●●●●	●●●	●●●●	●●●●
12	Background Power to Mean Ratio	Bpm	●●●●	●●●	●●●●	●●●
13	Ratio of Standard Deviation	RaSd	●●●	○○	●●●	●●●●
14	Ratio of Means	RaMe	○○	●●●●	●●●	●●●
15	Ratio of Power to Mean Ratios	Opm/Bpm	●●●	○○	●●●	●●●
16	Max Contrast	ConMax	○	○	○	○
17	Mean Contrast	ConMean	○	○	○○	○○
18	Mean Contrast Ratio	ConRaMe	○○	○	○	○
19	Standard Deviation Contrast Ratio	ConRaSd	○○	○	○	○
20	Local Area Contrast Ratio	ConLa	●●●	●●●●	●●●	●●●●
21	Mean Border Gradient	GMe	●●●	○○	●●●●	●●●
22	Standard Deviation Border Gradient	GSd	●●●	●●●●	●●●●	●●●●
23	Max Border Gradient	Gmax	○○	○	○○	○○
24	Min Border Gradient	Gmin	○	○	○	○
25	Power to Mean Border Gradient	Gpm	○○	○	○○	○○
26	Mean Difference to Neighbours	Ndm	○○	○	○	○
Texture Features:						
27	Spectral texture	Tsp	○	○	○	○
28	Shape Texture	TSh	○	○	○	○
29	Mean Haralick Texture	THm	○○	○○	○○	●●●

[a]Measure symbols: Very low (○), Low (○○), Medium (●●●), High (●●●●).
[b]Literates studies: [3], [4], [9], [13], [19]–[21]

2) GBT^{**}: Important features that found most important in GBT^{*} results were determined and re-evaluation with GBT classifier. The least important features were sequentially removed, and features that remain were evaluated. For each feature removed, a classification (GBT^{**}) accuracy was determined and the smallest features set with the highest accuracy was 93.9% with nine features.

3) ANOVA: Input features were then ranked statistically according to their significant using F-test. The least important features were sequentially removed, and features that remain were evaluated with GBT classifier. For each feature removed a classification accuracy was determined and the smallest features set with the highest accuracy was 83.1% using six features.

4) RFE: Features were recursively removed from the input feature set and evaluated the features that remained. For each feature set (N combinations each with $n-k$ features, $\{k:1\ldots n-1\}$) a classification accuracy was determined and the smallest feature set with the highest accuracy was 87.8% with only four features. This process was more computation intensive as multiple combinations needed to be evaluated to find the best features.

TABLE II
CROSS-VALIDATION ACCURACY SCORES.

Method	No. of Features	Mean (+/- Std)	Min	Max	TP[†]	FP[††]
GBT^{*}	All	77.4 (13.8)	61.9	98.4	78	63
GBT^{**}	6	91.6 (6.36)	82.5	99.2	84	80
	9	93.9 (5.02)	87.7	99.6	81	80
	15	85.7 (11.3)	70.2	100	79	76
ANOVA	6	83.1 (7.37)	72.6	95.2	77	70
	9	81.0 (9.95)	69.4	96.4	74	60
	15	75.8 (11.53)	64.5	95.6	77	63
RFE	4	87.8 (9.89)	75.0	99.6	84	77
	9	87.1 (7.58)	72.2	92.9	81	66
	15	83.6 (12.3)	69.4	99.2	78	67

[†]True Positive, [††]False Positive.

B. Important Features Selection

The most commonly used oil spill features in literature were investigated and found that the studies often did not use the same number of features. The disparity in the number of features was because that the features are not of equal importance, as a particular feature may be critical in one study but removing the same feature improves the results of another study.

We measured how often features were selected as critical (selection frequency) to determine a more general important feature set. The selection frequency was based on literature (i.e., [3], [4], [9], [13], [19]–[21]) and selection methods proposed in the study (i.e., GBT, ANOVA and RFE). All the measurements were normalised and divided into very low to high selection frequency (SF), see table I.

The most frequently selected feature sets across the literature studies had 13 medium to high SF and 5 high SF, of the 29 features, see table I. The study showed the most frequently selected feature sets between GBT, ANOVA and RFE had 12 medium to high SF and 7 high SF, of the 29 features, see table I. The number of important features for each feature selection method was presented in table II.

C. Important Features Rankings

The most critical feature sets were determined based on frequency selection measure. However, features are not of equal importance, each feature in the important feature sets had a different ranking. Feature rankings was measured based on how important each feature was compared to other features in a given feature set. The feature ranking were measure from literature and in the proposed study. All the measurements were normalised and divided into very low to high feature ranking, see table I.

1) Literature rankings: Numerous studies have presented essential features using SAR; these features vary in importance and rankings. An average ranking of each feature was determined based on previous studies in literature [3], [4], [9], [13], [19]–[21], see table I. The literature showed that 7 of the 29 standard features were of high rankings, these include *SP2, OSd, BSd, Opm, Bpm, GMe* and *GSd* features.

2) GBT rankings: GBT can reduce the variance of feature selection methods by averaging them over several trees and is thus considered a reliable measure. The GBT feature importance rankings including outputs from the ANOVA and RFE methods. ANOVA rankings were not used as they were not based on oil spill discrimination but only on feature separation. The proposed study showed that 5 of the 29 standard features were of high rankings, these include *SP2, OSd, BSd, Opm, Conla* and *GSd* features.

D. Discussion

From a total number of 29 commonly used features investigated, the result showed that using all available features is not the most efficient way to classify oil spill from look-alikes. Only half of the commonly used features was needed for acceptable results. Their respective rankings determined the critical features. Seven features were determined as the most important with five being consistent across of methods and two being less consistent. The five features included one geometry feature based on object shape (*SP2*), the rest were physical features (*OSd, BSd, Opm, GSd*), and no texture features. The shape factor (*SP2*) was measured the general shape of the object. The *OSd* and *BSd* measured the standard deviation of the intensity values of the pixels belonging to the object and background. The object power to mean ratio (*Opm*) measured the ratio between the standard deviation (*OSd*) and the mean (*OMe*) values of the object. The last significant feature, Standard deviation border gradient (GSd) measured the backscatter values spreading of the border gradient of the object. The results showed the most consistent important features with minimum variations can be achieved across literature and the proposed study.

V. Conclusion

Oil spill feature selection and classification of oil spill events from look-alikes using only SAR data can be a difficult task. This is due to many factors which could influence which features were extracted and selected for the classification task. In this study, oil spill events were classified from moving vessels, and the most important features were determined. The classification task used an optimised Gradient Boosting Tree Classifier (GBT). The most used oil spill features in literature were determined and extracted. The most important features were determined using ANOVA and RFE feature selection methods. The study ranked critical features with a high selection frequency and less critical features in high accuracy discrimination tasks. From the study, we can conclude with high confidence that the 15 important features can be the new standard feature set for oil spill discrimination tasks, particularly for the oil spill caused by moving vessels. While some features vary in selection frequency and importance there are a few that consistently have high and low importance across all methods. Consistently high valued feature include geometry and physical features, and less critical features include contrast, area and texture features. Consistently high valued feature include shape description, the object and background standard deviation measure, the ratio object standard deviation and mean values and the object border standard deviation. The less essential features include contrast, area and texture features. Future work includes further performance investigations for spatial resolution dependent (e.g., Gme and Gsd) and instrument-dependent (band, polarisation) parameters (e.g., OSd, BSd, Opm, Bpm) is also required.

References

[1] C. McLaughlin, D. Falatko, R. Danesi, and R. Albert, "Characterizing shipboard bilgewater effluent before and after treatment," *Environmental Science and Pollution Research*, vol. 21, no. 8, pp. 5637–5652, 2014.

[2] A. H. S. Solberg, "Remote sensing of ocean oil-spill pollution," *Proceedings of the IEEE*, vol. 100, no. 10, pp. 2931–2945, 2012.

[3] S. Singha, T. J. Bellerby, and O. Trieschmann, "Satellite Oil Spill Detection Using Artificial Neural Networks," *IEEE Journal of selected topics in applied earth observations and remote sensing*, vol. 6, no. 6, pp. 2355–2363, 2013.

[4] K. Topouzelis and A. Psyllos, "Oil spill feature selection and classification using decision tree forest on SAR image data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 68, pp. 135–143, mar 2012.

[5] D. Mera, V. Bolon-Canedo, J. M. Cotos, and A. Alonso-Betanzos, "On the use of feature selection to improve the detection of sea oil spills in sar images," *Computers & Geosciences*, vol. 100, pp. 166–178, 2017.

[6] P. Genovez, N. Ebecken, C. Freitas, C. Bentz, and R. Freitas, "Intelligent hybrid system for dark spot detection using sar data," *Expert Systems with Applications*, vol. 81, pp. 384–397, 2017.

[7] L. W. Mdakane and W. Kleynhans, "An image-segmentation-based framework to detect oil slicks from moving vessels in the southern african oceans using sar imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 6, pp. 2810–2818, June 2017.

[8] M. Fingas and C. Brown, "Review of oil spill remote sensing," *Marine pollution bulletin*, vol. 83, no. 1, pp. 9–23, 2014.

[9] D. Stathakis, K. Topouzelis, and V. Karathanassi, "Large-scale feature selection using evolved neural networks," *Proceedings of SPIE, image and signal processing for remote sensing*, vol. 12, p. 636513, 2006.

[10] J.-S. Lee, "Refined filtering of image noise using local statistics," *Computer graphics and image processing*, vol. 15, no. 4, pp. 380–389, 1981.

[11] H. Hersbach, *CMOD5: An improved geophysical model function for ERS C-band scatterometry*. European Centre for Medium-Range Weather Forecasts, 2003.

[12] N. Miranda, P. Meadows, D. TYPE, and T. NOTE, "Radiometric calibration of s-1 level-1 products generated by the s-1 ipf," *Viewed at https://sentinel. esa. int/documents/247904/685163/S1-Radiometric-Calibration-V1. 0. pdf*, 2015.

[13] K. Topouzelis, D. Stathakis, and V. Karathanassi, "Investigation of genetic algorithms contribution to feature selection for oil spill detection," *International Journal of Remote Sensing*, vol. 30, no. 3, pp. 611–625, 2009.

[14] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[15] P. Liu and C. Zhao, "Oil Spill Identification in Marine SAR Images Based on Texture Feature and Fuzzy Logic System," in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, 2009, pp. 433–437.

[16] M. B. Brown and A. B. Forsythe, "Robust tests for the equality of variances," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 364–367, 1974.

[17] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*. CRC press, 1984.

[18] J. Friedman, T. Hastie, and R. Tibshirani, *The elements of statistical learning*. Springer series in statistics New York, 2001, vol. 1.

[19] V. Karathanassi, K. Topouzelis, P. Pavlakis, and D. Rokos, "An object-oriented methodology to detect oil spills," *International Journal of Remote Sensing*, vol. 27, no. 23, pp. 5235–5251, dec 2006.

[20] G. Suresh, G. Heygster, C. Melsheimer, and G. Bohrmann, "Natural oil seep location estimation in sar images using direct and contextual information," in *Geosc. and Remote Sens. Symp., (IGARSS), IEEE Int.* IEEE, 2014, pp. 1678–1681.

[21] C. Brekke and A. H. S. Solberg, "Oil spill detection by satellite remote sensing," *Remote sens. of env.*, vol. 95, no. 1, pp. 1–13, 2005.