remote sensing

MDPI

Article

Oil-Contaminated Soil Modeling and Remediation Monitoring in Arid Areas Using Remote Sensing

Gordana Kaplan [1,*], Hakan Oktay Aydinli [2], Andrea Pietrelli [3], Fabien Mieyeville [3] and Vincenzo Ferrara [4]

1 Institute of Earth and Space Sciences, Eskisehir Technical University, Eskisehir 26555, Turkey
2 BePetro Consultancy, Çankaya, Ankara 06530, Turkey; hakanoktayaydinli@eskisehir.edu.tr
3 Ampere CNRS UMR 5005, Université de Lyon, UCBL1, INSA Lyon, ECL, F-69621 Villeurbanne, France; andrea.pietrelli@univ-lyon2.fr (A.P.); fabien.mieyeville@univ-lyon1.fr (F.M.)
4 Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome, 00184 Rome, Italy; vincenzo.ferrara@uniroma1.it
* Correspondence: gkaplan@eskisehir.edu.tr; Tel.: +90-536-697-5605

Abstract: Oil contamination is a major source of pollution in the environment. It may take decades for oil-contaminated soils to be remedied. This study models oil-contaminated soils using one of the world's greatest environmental disasters, the onshore oil spill in the desert of Kuwait in 1991. This work uses state-of-art remote sensing technologies and machine learning to investigate the oil spills during the first Gulf War. We were able to identify oil-contaminated and clear locations in Kuwait using unsupervised classification over pre- and post-oil spill data. The research area's pre-war and post-war circumstances, in terms of oil spills, were discovered by developing spectral signatures with different wavelengths and several spectral indices utilized for oil-contamination detection. Following that, we use this data for sampling and training to model various oil-contaminated soil levels. In addition, we analyze two separate datasets and used three modeling methodologies, Random Tree (RT), Support Vector Machine (SVM) and Random Forest (RF). The results show that the suggested approach is effective in detecting oil-contaminated soil. As a result, the location and degree of contamination may be established. The results of this analysis can be a valid support to the studies of an appropriate remediation.

Keywords: oil contamination; soil pollution; soil contamination; soil remediation; remote sensing; arid areas

Citation: Kaplan, G.; Aydinli, H.O.; Pietrelli, A.; Mieyeville, F.; Ferrara, V. Oil-Contaminated Soil Modeling and Remediation Monitoring in Arid Areas Using Remote Sensing. *Remote Sens.* **2022**, *14*, 2500. https://doi.org/10.3390/rs14102500

Academic Editors: Abdul M. Mouazen, Anne Gobin, Said Nawar and Yiyun Chen

Received: 26 April 2022
Accepted: 20 May 2022
Published: 23 May 2022

1. Introduction

An oil spill is described as uncontrollably spilled crude oil into the atmosphere, creating serious issues for humans, animals, plants, and the environment. In most cases, it is the consequence of human error or a system malfunction. Spills also might happen on purpose, as was observed during the Arabian Gulf War [1,2].

1.1. The Oil Spill in 1991

The major impact of the Arabian Gulf War on the Kuwait ecosystem was the uncontrolled release of oil into the marine and terrestrial environment [3,4]. Investigations were appropriately conducted after the Arabian Gulf War, and rehabilitation studies [5] for the affected areas were started almost immediately, in order to recover oil from the sea, but only later and recently to clean the beaches and the hinterland. The oil recovery was carried out by the Kuwait Oil Company and Kuwait Petroleum Corporation to minimize the impact of the spilled oil on the local environment. A significant amount of the lighter hydrocarbon proportion had evaporated, and the remaining portion in liquid form was partially reclaimed. The traces of oil leakage persisted as soot and tar, which are unrecovered denser hydrocarbon components [6].

The war activities in Kuwait started in August 1990. However, the oil spill started in the second part of January 1991, when Iraq released oil and initiated the burning of oil wells [1]. From 1990 to 1991, Kuwait's desert terrain was left with an iconic disaster [7]. When the forces withdrew in late 1991, they left behind smoldering wells and oil slicks from the stations that were built along the coast [4]. On 8 November 1991, emergency response teams struggled to extinguish the last well fire [8]. During this period, a large international team was already working on the clean-up, mainly by the physical removal of the oil. Kuwait had 810 actively operating wells, 730 were destroyed or burned down during the Arabian Gulf War. The emergence of oil lakes and hydrocarbon contaminated surfaces, which occurred in the Kuwait desert, can be described as the catastrophic environmental consequence of the Arabian Gulf War [6]. Kuwait is located in the northwestern part of the Arabian Gulf; the coverage area of Kuwait is 17,818 km^2 [9]. GIS analysis shows that 21.6 percent of Kuwait's area was affected by the Gulf War, of which 4.4 percent was due to oil pollution and 17.2 percent was due to remobilized sand sheet [7].

Crude oil spilled over the desert landscape gathered to form "oil lakes". After extensively examination, Kuwait's "oil lakes" were unique and regarded as one of the biggest environmental disasters in recent history [8,10–14].

1.2. Related Studies

Oil contamination in the producing fields was strictly monitored through the soil, water, atmosphere, and plants to determine if they were contaminated with oil [15,16]. Field studies for analysis were typically time-consuming and expensive, mainly when applied to broad or inaccessible regions. In addition to field analysis, the areas were monitored using satellite images. Satellite images cover vast distances simultaneously and regularly in a non-destructive way at reduced costs and less time, making it a viable alternative to the traditional approaches based on the ground [17–19].

Multispectral remote sensing, which incorporates gathering visible, near-infrared, and shortwave infrared images in several broad wavelength bands, was proven to be extremely useful in generating environmental monitoring capabilities [11,20]. In this regard, Abdunaser [21] studied Kuwait's oil-contaminated surfaces by applying time-series analysis on multi-temporal satellite data. As a result, high-resolution maps were created to delineate oil-contaminated zones. Kwarteng [7] noted that the Landsat TM imagery, obtained between 1987 and 1995, was employed to evaluate pre- and post-Arabian Gulf War conditions in the Burgan oil field in Kuwait. The images were processed according to the discrimination of the bright and dark cover types. It was emphasized that oil lakes, tar, soot, vegetation, and other unaffected sand sheet zones could be mapped to highlight the local characteristics in-depth. The total area of oil lakes was computed from the 1995 TM images as 38.93 km^2. Almost 90% of the oil lakes were observed in the Burgan oil field. The measured area decreased to about 20%, compared to the data calculated from the TM images in 1992. In their study over the Niger Delta, Ozigis et al. [22] emphasized that detecting the oil spill-impacted land cover and the oil-free areas can be done by combining different vegetation health indices with spectral wavelengths from visible, near-infrared, and shortwave infrared band, using machine learning. The areas were categorized with the help of a machine learning, Random Forest classifier. This study was used for determining the optimum characteristics of the oil-contaminated land from the clean areas. Individual land cover types performed better in identifying oil-free and oil-contaminated land covers when recognized separately than when the complete study area image, encompassing all of the land cover types, was assessed at once.

Applying specific indices on the satellite imageries with the machine learning approach effectively determines the oil spills in the terrestrial locations. Hence, this approach is used to map the oil-contaminated areas in Kuwait. The oil spill mapping model proposed in the current study is based on spectral band data, spectral indices, and machine learning algorithms. The proposed model served fruitfully as an effective oil spill detection method

and the exact location of the contaminated zones can be detected using its results. In the light of these findings, remediation studies can be conducted.

2. Materials and Methods

2.1. Study Area

The study area is represented by Kuwait's three most oil-contaminated areas due to the war in 1991 (Figure 1). Thus, three major onshore oil fields in Kuwait were investigated in the current study. The largest study area is Greater Burgan which comprises Burgan, Ahmadi, and Magwa; the second area contains Raudhatain, Ratqa, Sabriyah, and Behrah; and the third includes the Umm Gudair oil fields. The land cover classes in Kuwait can be simply divided into desert and urban areas [23]. With the rapid population growth in Kuwait, the urban areas have also increased in the last few decades. Thus, since the oil spill in 1991, the population has increased from approximately two million to over four million [24]. The desert soil in Kuwait may be classified as arid soils with weakly developed profiles, sandy compositions, and sandy textures. Soil profiles from numerous Burgan's oil-lake bed locations are produced at the edge of the Ahmadi calcareous limestone ridge and are divided into four distinct zones; (i) Lag gravel on the top of the soil profile mixed with Aeolian; (ii) Pebbly calcareous sandstone enriched gypsum and clay seams. Highly fractured; (iii) Coarse mixed sand layers of calcareous sandstone; (iv) Hard, impermeable and continuous caliche deposits [25]. The chemical properties of the oil samples from the Burgan area show that, in the range of 0–300 cm depths, the pH ranges from 7.6–8.1, EC (Electrical Conductivity) 3.9–26.5 (dS/m) [26], where the first 10 cm have the highest EC values, and the highest pH values were noticed in the range of 25–150 cm depth [25].

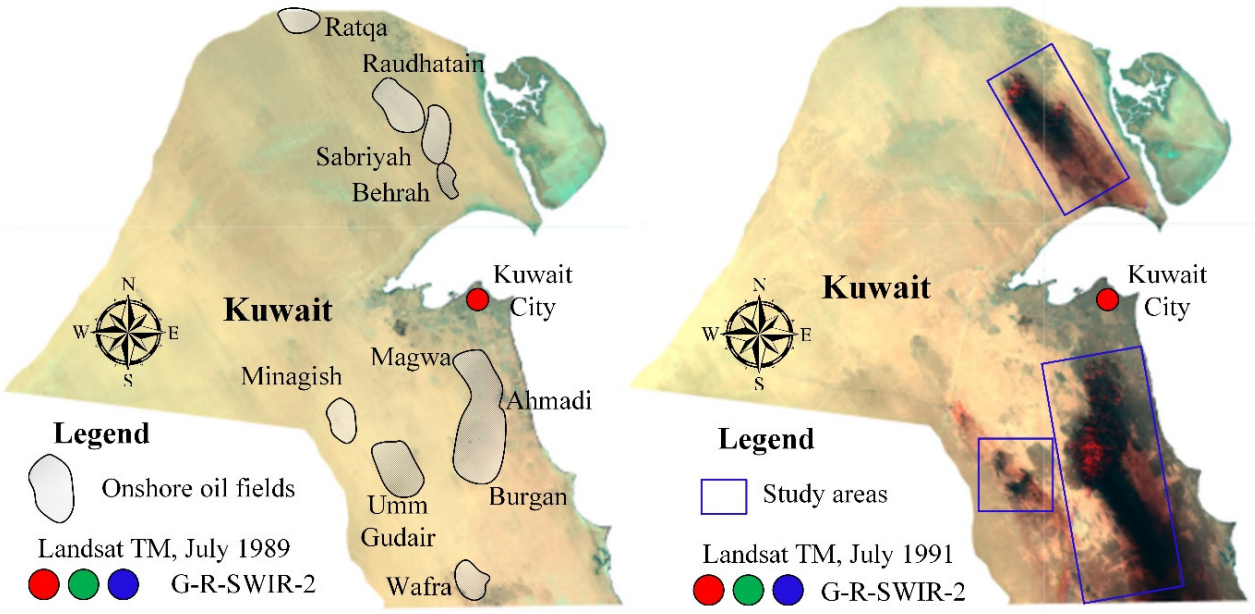

Figure 1. Study area; Left—Kuwait's onshore oil fields over Landsat TM (RGB, GRSWIR-2) before the oil spill; Right—Landsat TM (RGB, GRSWIR-2) over Kuwait after the oil spill.

2.2. Remote Sensing Data

According to the historical events in Kuwait, we have selected several investigation periods. As the oil spill event occurred in 1991, Landsat-5 satellite images from 1989 were selected as the pre-oil-spill data. Landsat-5 images from 1991 were considered the main images from the oil contamination, and the data from 1991 were considered as extremely contaminated. To draw a picture of what happened in 1991, we also investigated the spectral signatures of the oil-contaminated soil (OCS) in five different periods in 1991. Following the historical clean-up data, we used imagery from 1992 and considered it to be data from the highly OCS. Furthermore, we assessed the contamination level of the study

area in 1997 and 2002, using Landsat-5 data. Landsat-5 is a multispectral sensor with 30 m spatial resolution, and it is the longest-living and only satellite to provide open-source data for over forty years. As the wavelengths of the Landsat sensors are different, we evaluated Landsat-8 data separately and compared the spectral signatures between the clean and the OCS.

The image processing steps followed in this study were done within the Google Earth Engine (GEE) platform, a valuable cloud computing service that allows researchers to obtain free access to various satellite data [27]. After the study area and image period selection, images with less than 5% cloud cover were selected for further processing. Median values were used from image collection for the selected periods.

Although the oil-contaminated areas were visible on the 1991 satellite imagery (Figure 1), an unsupervised classification was conducted in GEE to distinguish the clean and oil-contaminated zones for further sample collection. Unsupervised classification is a method for grouping the pixels that have similar features. To discriminate the oil spill conditions of arid areas at pre-war (1990) and post-war (1991) periods in the study area, Landsat-5 satellite imageries were used. The 1991 classification results were further used in the sample collection process, where samples from two classes were considered; clean soil and OCS. Thus, over the 1991 satellite imagery, 300 samples per class were defined from the Burgan oil field for the oil-contaminated soil class. The samples for the clean soil class were selected in the west part of Kuwait, where there was no contaminated soil, and the area was not affected by the oil spill in 1991. The samples were collected randomly in the selected areas. Spectral signatures were created/generated in GEE for all of the samples.

Spectral Indices

Together with the spectral values from every Landsat band, we calculated and extracted values from several spectral indices that were significant in the literature for oil-spill mapping and monitoring. Thus, Kovalev and Tokareva [28] used Normalized Difference Vegetation Index (NDVI) to monitor the vegetation state in the oil production territory. Their findings showed abnormal changes due to oil contamination. The Green and Near Infrared ration (G/NIR) was successfully used to detect the oil pollution in the Niger Delta [29]. Using the Normalized Difference Moisture Index (NDMI), Balogun et al. [30] analyzed the oil spill impact on the coastal vegetation and wetlands. Zhao et al. [31] developed an oil slick detection index (OSDI) for Landsat-8 images, however, we also adjusted the index for the Landsat-5 images. A Fluorescence Index (FI) was used for the characterization of the oil spill at sea [32]. The Modified Normalized Difference Water Index (MNDWI) was one of the most important variables in mapping the terrestrial oil spills using the Random Forest algorithm [22]. The used spectral indices and their formulas are given in Table 1.

Table 1. Spectral indices used in this study.

Index	Formula	Reference
NDVI	NIR − RED/NIR + RED	[28]
G/NIR	G/NIR	[29]
NDMI	NIR − SWIR/NIR + SWIR	[30]
OSDI	GREEN − NIR/GREEN + NIR	[31]
FI	BLUE − RED/BLUE + RED	[32,33]
MNDWI	GREEN − SWIR/GREEN + SWIR	[22]

2.3. Oil-Contaminated Soil Modeling

The resolved data were further used for statistical analysis and prediction in Weka [34], a multi-purpose software developed to apply machine learning algorithms. Supported Vector Machine (SVM), Random Forest (RF), and Random Tree (RT) were evaluated for the oil-contaminated modeling. Thus, the spectral signature from all of the bands and spectral indices were extracted in GEE, and then imported into Weka. The samples were collected

from two classes, oil-contaminated and clean soil. The modeling was done using 70% of the samples, while the remaining 30% were separated for testing the models. Following the historical events, the modeling was done in different periods. For modeling the extremely oil-contaminated soil, satellite imagery from 1991 was used. Imagery from 1992 was used for the highly oil-contaminated soil modeling, while imagery from 2002 was used for the moderately oil-contaminated soil modeling.

SVM is a supervised non-parametric statistical learning approach, thus, no assumptions about the underlying data distribution are made. The approach is given a set of labeled data instances. The SVM training algorithm attempts to generate a hyperplane that divides the dataset into a definite specified number of classes that is consistent with the training examples [35].

Random Tree is a supervised classifier producing a random set of data for constructing a decision tree. Each node is split using the best split among all of the variables in the tree. Random trees are a collection of the tree predictors [36]. On the other hand, using predictions produced from an ensemble of decision trees, the RF classifier produces reliable classifications [37]. These three modeling techniques were selected due to their simplicity and the ability to employ the model with various data. The used models were evaluated using different accuracy assessment parameters.

The results were evaluated using Mean Absolute Error (MAE), error-based analysis between the predicted and the observed values used to assess the model's performance during the model processing [38]:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |O_i - P_i|}{n} \tag{1}$$

where, O_i is the observed and P_i is the predicted value.

Relative Absolute Error (RAE) was used to measure the performance of the predictive models [39]:

$$\mathrm{RAE} = \frac{\sum_{i=1}^{n} |O_i - P_i|}{\sum_{i=1}^{n} |\overline{O_i} - P_i|} \tag{2}$$

where $\overline{O_i}$ is a mean value of O_i.

In addition, for the accuracy analyses, common evaluation statistics for binary classification were used. Namely, True Positives (TP) (a class correctly identified), False Positives (FP) (a class incorrectly identified; a commission error), and False Negatives (FN) (a class is missed; an omission error) parameters were taken into consideration. TP, FN, and FP indicate perfect identification, under-identification, and over-identification, respectively. Then the Precision (P), Recall (R), and F-score (F) were calculated. Precision (i.e., positive predictive value) describes the correctness of the detected oil-contaminated fields and how well the algorithm dealt with FP (Equation (3)), Recall (i.e., sensitivity) describes the building detection rate and how well the algorithm dealt with the FN (Equation (4)), and the F-score is the harmonic mean of Recall and Precision and reports the overall accuracy considering both commission and omission errors (Equation (5)) [40]. From the confusion matrix, the kappa statistics were also calculated:

$$\mathrm{P} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FP}) \tag{3}$$

$$\mathrm{R} = \mathrm{TP}/(\mathrm{TP} + \mathrm{FN}) \tag{4}$$

$$\text{F-score} = 2 \times ((\mathrm{P} \times \mathrm{R})/(\mathrm{P} + \mathrm{R})) \tag{5}$$

The data were separated, 70% for training the model and 30% for testing the model. As a result, the accuracy was evaluated using 180 independent points that were not part of the training set.

The methodology implemented in this study is presented in the flowchart in Figure 2.

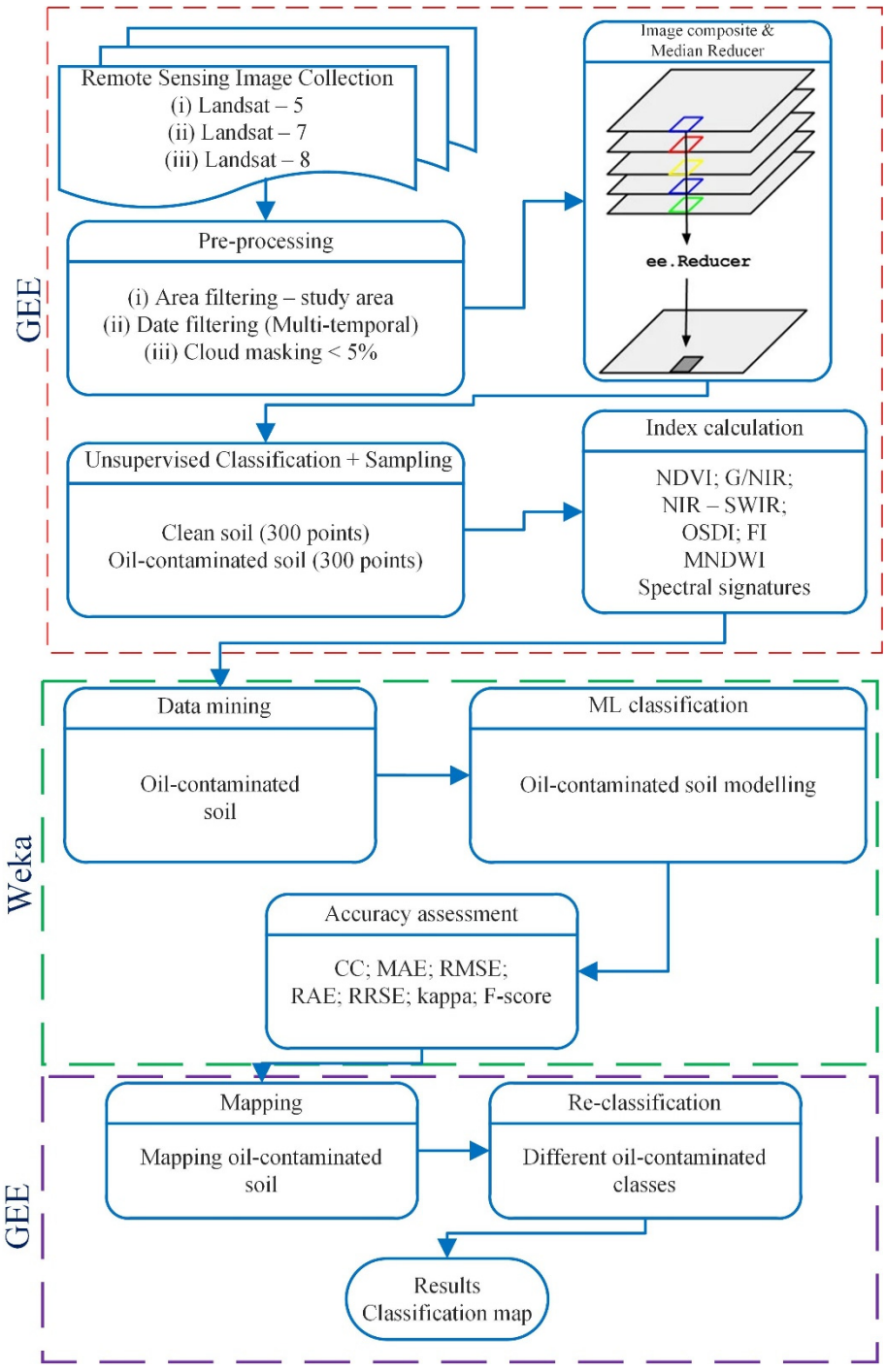

Figure 2. Flowchart of the implemented methodology.

3. Results

3.1. Unsupervised Classification and Spectral Signature Differences

The unsupervised classification results were compared to a 1:50,000 orthophoto classification [41] and the classes of 1991 were assigned accordingly. Thus, as shown in Figure 3, while the three classes in 1990 were assigned as soil, artificial, and roads, the three classes in 1991 after the oil spill were defined as clean soil, soot, and tar mat.

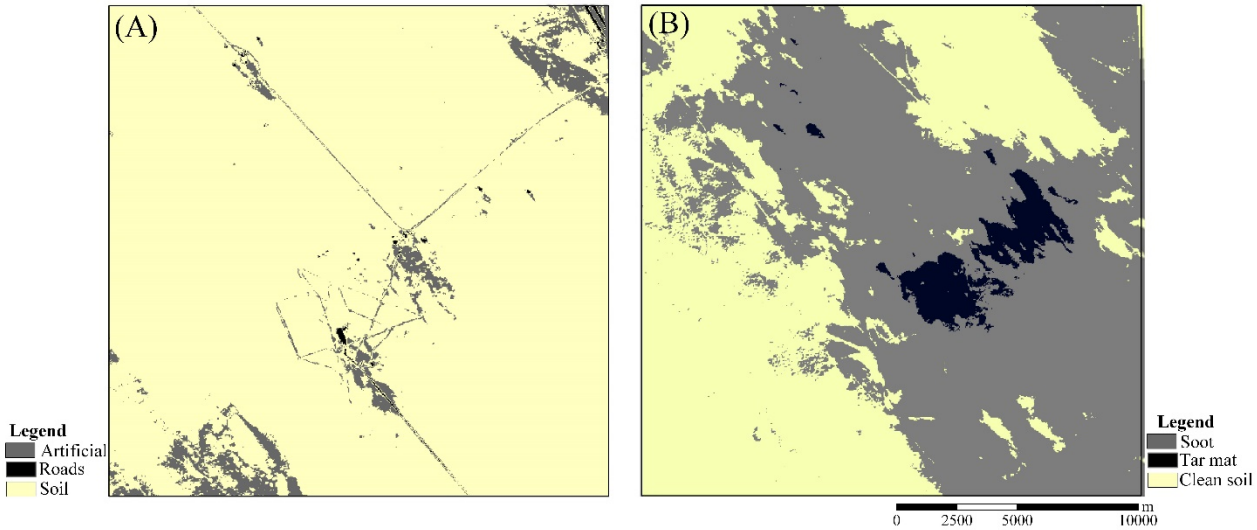

Figure 3. Unsupervised classification results over Umm Gudair; (**A**) 1990; (**B**) 1991.

The unsupervised classification results were taken as reference data. Soil samples were collected from the clean and contaminated soils. First, the spectral signatures of the oil-contaminated soil were investigated in 1991 in five different periods. The oil spill started in January 1991. The first comparison was made between clean soil and oil-contaminated soil in the period January–February (Figure 4). There is an obvious distinction between the examined groups, where the values in all of the bands have decreased dramatically. As the oil spill progressed, the spectral values of the oil-contaminated soil decreased, particularly in the SWIR-1 zone. The lowest readings were observed between June and November. The oil spill was finally stopped in November 1991. After the war, remediation studies launched with physical removal of the oil, and, as a result, the spectral values of the oil-contaminated soil increased toward the end of 1991 and the beginning of 1992. Figure 5 depicts the spectral fingerprints for the analyzed years (1997–2002).

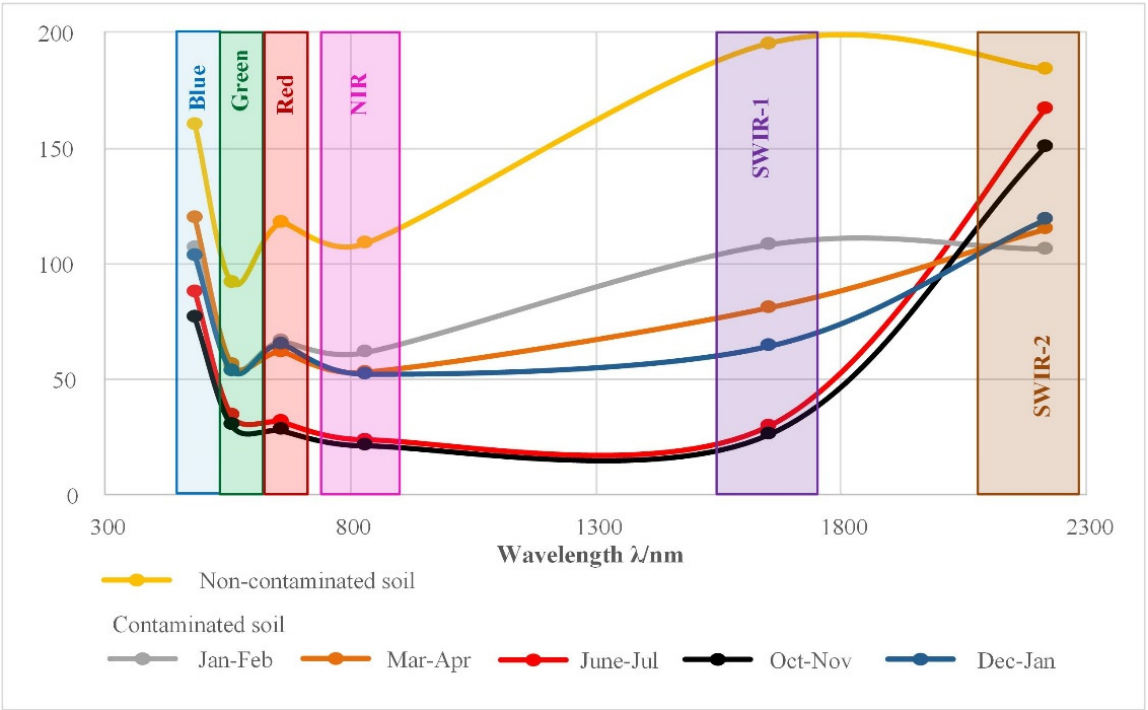

Figure 4. Spectral signatures of non-contaminated and contaminated soil in different periods (1991–1992).

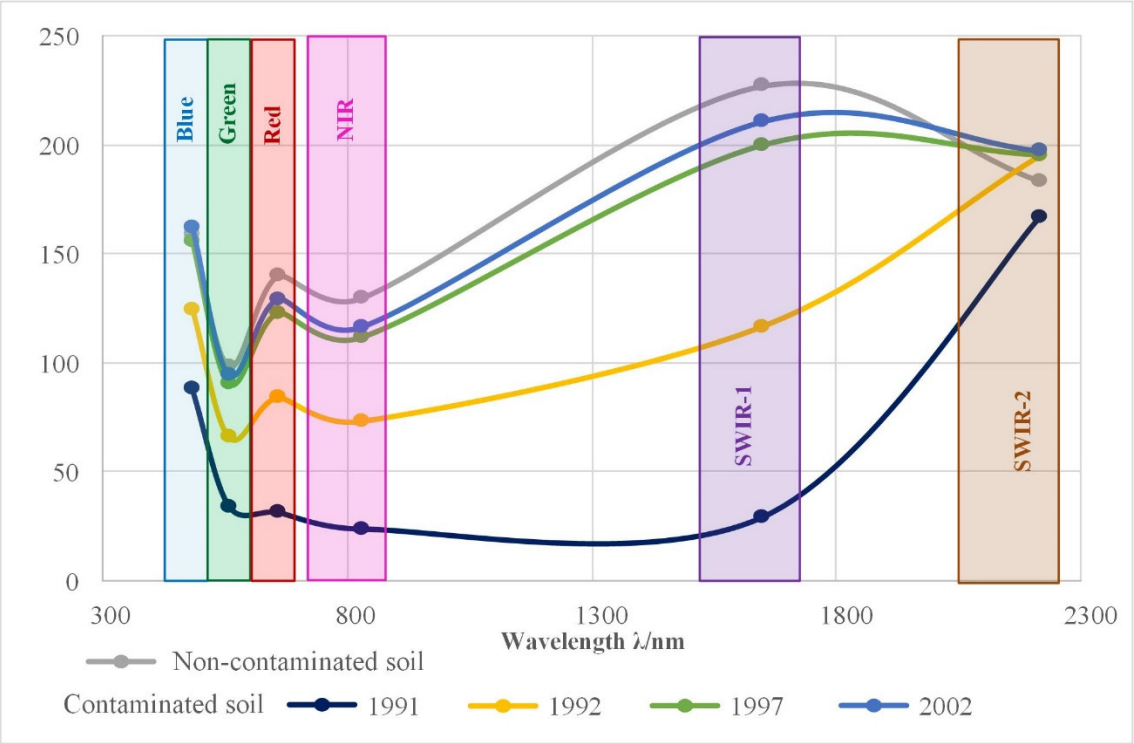

Figure 5. Spectral signatures of non-contaminated and contaminated soil 1991–2002.

3.2. Modeling Results

SVM, RT, and RF were applied to the samples by employing Weka. During the Gulf War, the vast majority of the land was transformed into pools of oil lakes, tar, oil mist, and soot from the burning oil wells. The oil lakes were nearly indistinguishable from the surrounding landscape [13]. Consequently, based on the spatial resolution of the satellite imagery, the oil-contaminated soil areas from 1991 were considered extremely contaminated. Two separate datasets were used in the modeling. All of the spectral bands and indices were used in Dataset I, whereas in Dataset II only the spectral indices were considered.

The extremely contaminated soils were simply modeled with a two-node tree using the Green band. Thus, the Green < 56 values were considered as oil-contaminated, and the Green > 56 as clean soils. Although the results for both models and datasets were similar, the extremely oil-contaminated soil was the most efficiently separated from the clean soil using RT. The RT model with Dataset II was constructed of two nodes, where FI was used as a threshold value (Figure 6). The results revealed a clear distinction between the investigated classes, where values less than 0.3 represented clean soil, while the values greater than 0.3, represented OCS.

B1-B7 represents the bands from Landsat-5, respectively. The accuracy assessment for the extremely OCS modeling is given in Table 2.

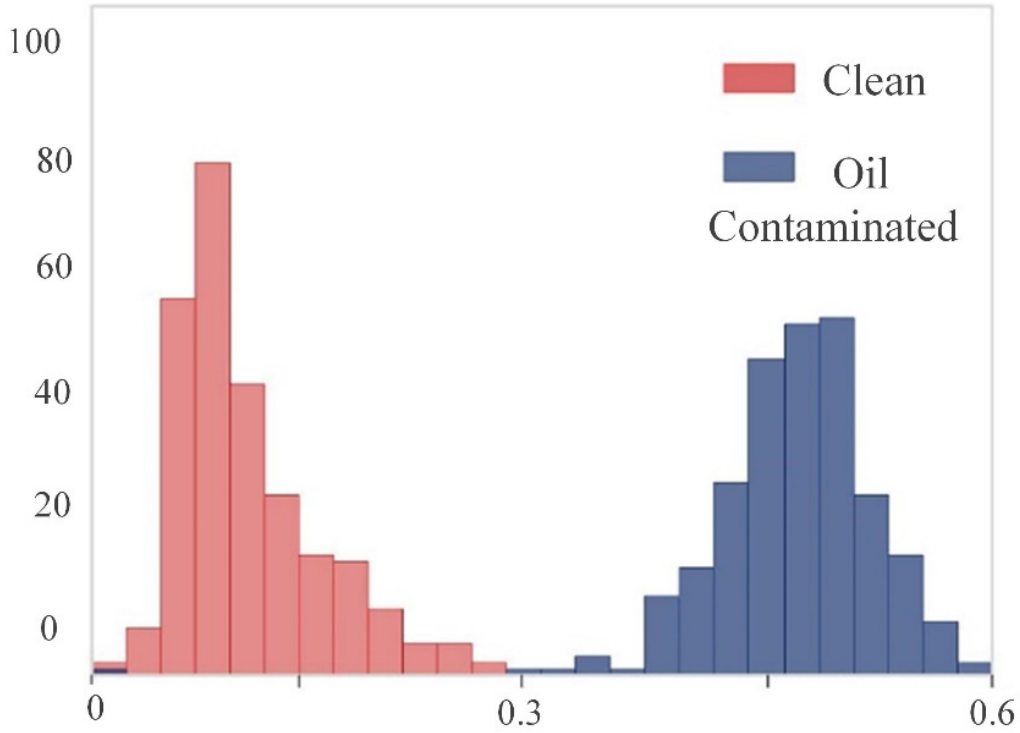

Figure 6. FI values histogram of clean and OCS in June-July 1991.

Table 2. Validation results for the extremely OCS modeling.

ML Algorithm	Correctly Classified Instances	MAE	RAE (%)	Kappa	TP	FP	Precision	Recall	F-Score
RT-I	100%	0	0	1	1	0	1	1	1
RT-II	100%	0	0	1	1	0	1	1	1
SVM-I	96%	0.04	8.8	0.91	0.96	0.05	0.96	0.96	0.96
SVM-II	100%	0	0	1	1	0	1	1	1
RF-I	100%	0	0.1	1	1	0	1	1	1
RF-II	99%	0	0.6	0.99	0.99	0	0.99	0.99	0.99

All of the algorithms performed successfully in separating oil-contaminated and clean soil. RT, unlike SVM and RF, provides a model. The RT model for the highly oil-contaminated soil using Dataset I was constructed of nine nodes using the B3, B5, B1, and G/NIR index from Landsat-5. On the other hand, the RT model using Dataset II, also consists of nine nodes, using OSDI, FI, and G/NIR indices. The RT models are demonstrated in Figure 7. RT Dataset I presented highly dedicated results that are similar to 1991. The accuracy assessment results are given in Table 3. Visual results of the classification for extremely and highly OCS are given in Figure 8. As can be seen from the results, in the highly OCS, the impervious areas that have similar spectral values with the contaminated areas are being mixed with the OCS. However, this is not the case for the results from 1991, where the contaminated areas were clearly classified.

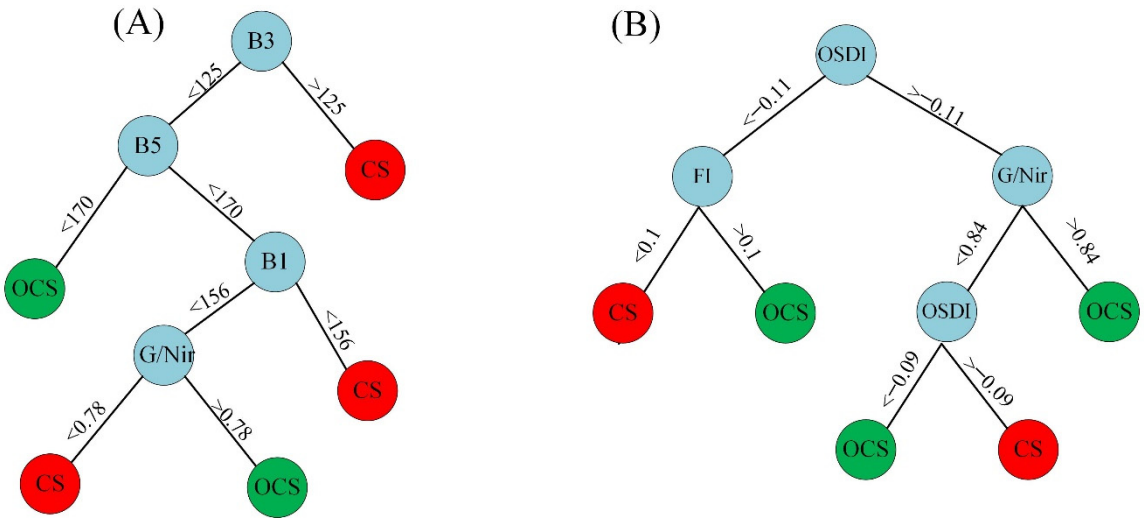

Figure 7. RT models for 1992, highly oil-contaminated soil using; (**A**) Dataset I; (**B**) Dataset II. CS stands for Clean Soil, while OCS stands for Oil-Contaminated Soil.

Table 3. Validation results for the highly OCS modeling.

ML Algorithm	Correctly Classified Instances	MAE	RAE (%)	Kappa	TP	FP	Precision	Recall	F-Score
RT-I	100%	0	0	1	1	0	1	1	1
RT-II	99%	0	1.1	0.99	0.99	0	0.99	0.99	0.99
SVM-I	97%	0.03	5.5	0.94	0.97	0.03	0.97	0.97	0.97
SVM-II	97%	0.03	6.6	0.93	0.97	0.04	0.97	0.97	0.97
RF-I	100%	0	0.1	1	1	0	1	1	1
RF-II	99%	0	1.2	0.99	0.99	0	0.99	0.99	0.99

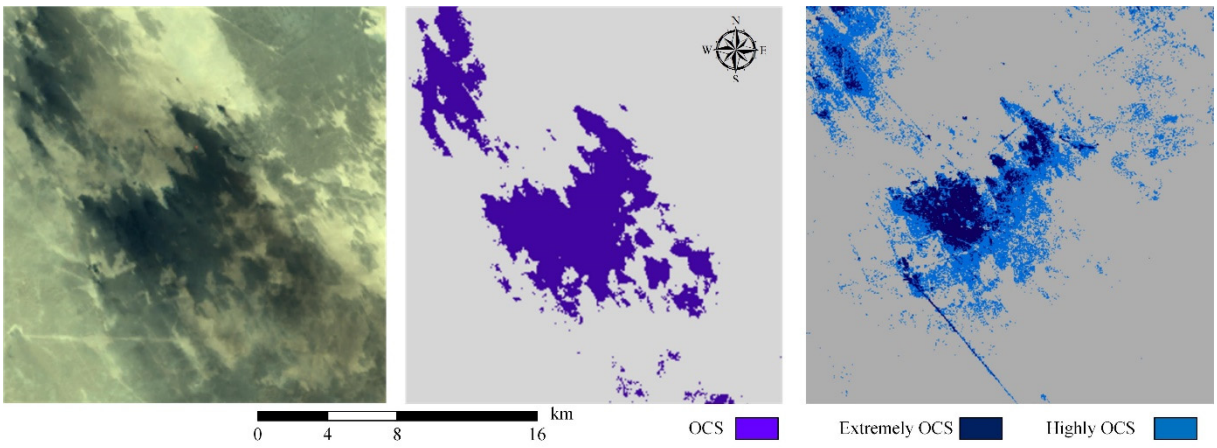

Figure 8. Classification results in Umm Gudair; **Left**—Landsat-5 image; **Middle**—Extremely OCS (1991); **Right**—OCS in 1992.

Compared with the extremely and highly contaminated soil, the accuracy for modeling the moderately oil-contaminated soil was significantly lower, especially in the SVM models. The RT models produced higher correlations. In addition, the error values from the SVM were significantly higher, compared to RT and RF (Table 4). The number of nodes in the tree was 27 for Dataset I and 23 for Dataset II. Similar to the previous contamination levels, RT and RF with Dataset I performed best. In Figures 9 and 10 are shown the RT models of Dataset I and Dataset II, respectifuly. Figure 11 depicts the visual findings of the moderately oil-contaminated soil classification.

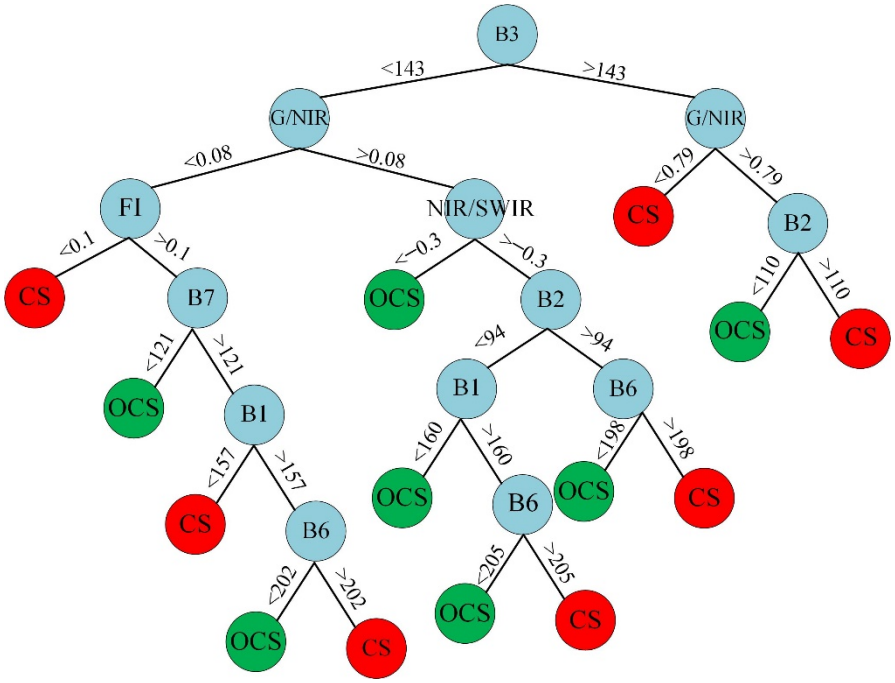

Figure 9. RT model for moderately oil-contaminated soil (2002) using Dataset I.

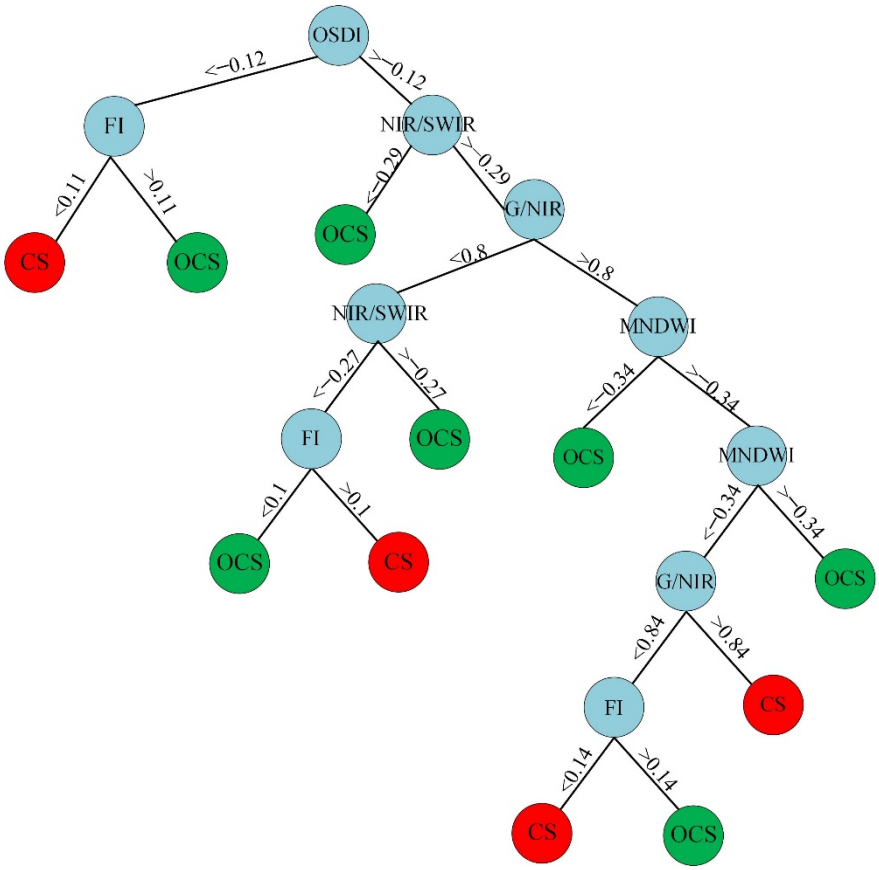

Figure 10. RT model for moderately oil-contaminated soil (2002) using Dataset II.

Table 4. Validation results for the moderately OCS modeling.

ML Algorithm	Correctly Classified Instances	MAE	RAE (%)	Kappa	TP	FP	Precision	Recall	F-Score
RT-I	99%	0	0	0.99	0.99	0	0.99	0.99	0.99
RT-II	97%	0.03	6.6	0.93	0.96	0.03	0.96	0.96	0.96
SVM-I	97%	0.03	5.5	0.94	0.97	0.03	0.97	0.97	0.97
SVM-II	78%	0.2	44.4	0.56	0.78	0.21	0.78	0.78	0.77
RF-I	98%	0.03	6.0	0.97	0.98	0.01	0.98	0.98	0.98
RF-II	97%	0.03	6.9	0.94	0.97	0.03	0.97	0.97	0.97

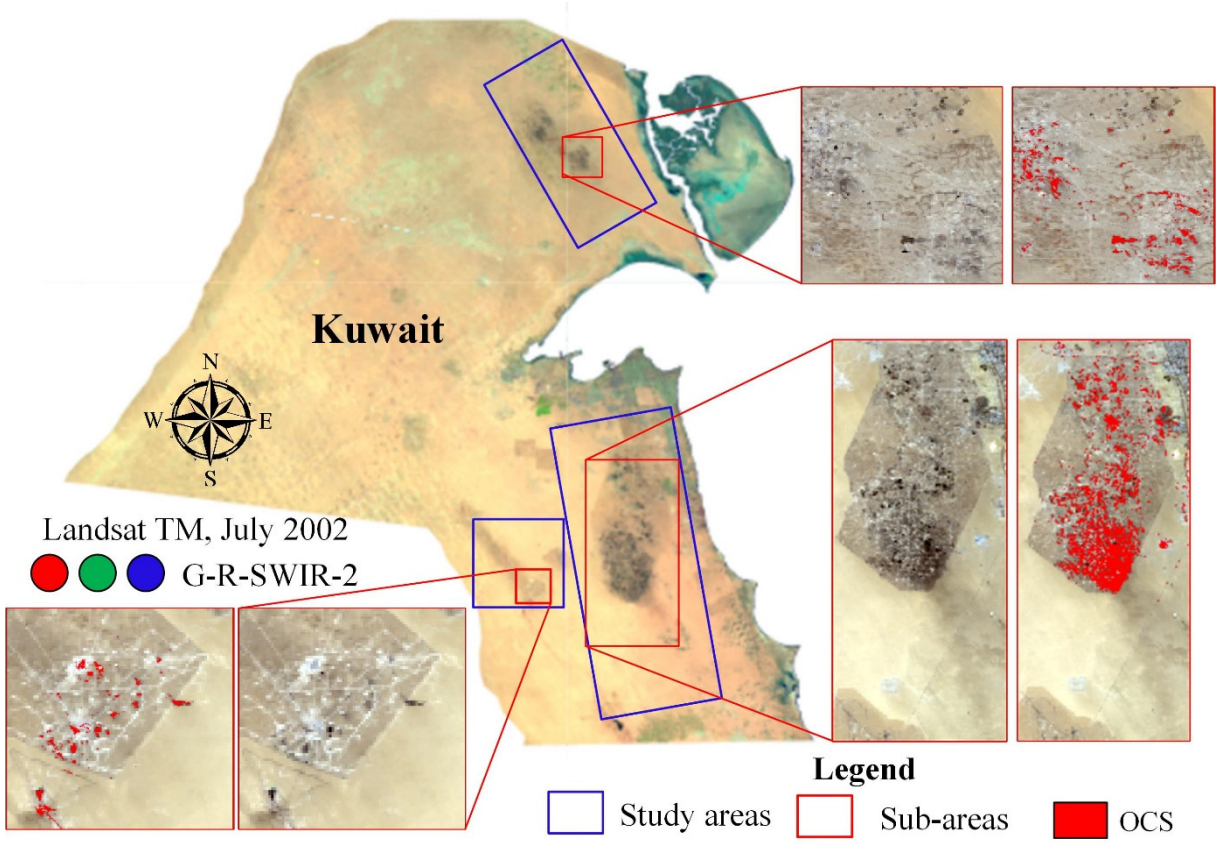

Figure 11. OCS modeling results from 2002.

4. Discussion

4.1. Oil-Contaminated Soil Classification

The unsupervised classification results in 1991 showed that the clean soil and the oil-contaminated soil distinguished each other successfully. Using this classification, we have further collected samples from both investigated classes and investigated the spectral indices in different periods. The data were classified as extremely (1991), highly (1992), and moderately oil-contaminated (2002) soil, following their spectral signatures. In addition, the spectral signatures of the clean soil and oil-contaminated soil in several periods in 1991 were evaluated by employing Landsat-5 images. The results of the spectral values from 1991 correspond to the historical events. Thus, as the oil spill started in January 1991, the spectral values of the contaminated areas were significantly lowered. The most significant difference can be noticed in the SWIR region. The findings are consistent with the results obtained in Kwarteng [22], where the oil lakes and oil-polluted surfaces were investigated in the same research area. Using hyperspectral satellite data, Pelta et al. [1] conducted research to identify the clean and oil-contaminated soil on the Evrona Nature Reserves. The results were matched for different wavelengths.

4.2. Comparison of Oil-Contaminated Classification Performance

The differences between the clean and oil-contaminated soil are clearly visible in the following years. The remediation from the study area causes the major difference between 1991 and 1992. As shown in Figure 5, the spectral differences between 1997 and 2002 are not significantly high. Landsat-5 has not obtained data over the study area since 2002, thus, further research could not be carried out. Spectral signature differences from Landsat-8 show that there are still spectral traces of oil contamination in 2017 and 2021 (Figure 12). As can be seen, there are still considerable differences amongst the investigated classes. It should be noted that this difference could be due to new oil contaminations and that cleaning of the contaminated areas continues [7,13]. Although some studies [13] discuss the resilience and remediation of the study area, the spectral differences show that some of the oil-contaminated areas in 1991 still exist.

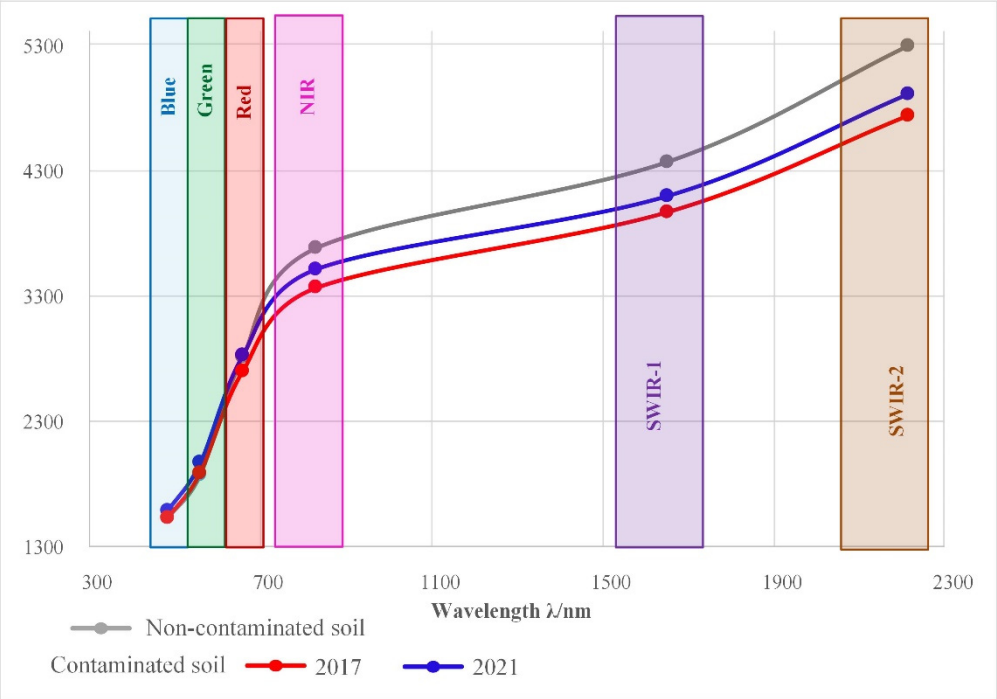

Figure 12. Spectral signatures by using Landsat-8 of non-contaminated and contaminated soil for 2017 and 2021.

As the investigation of the spectral values showed significant differences, we used the sample data for the oil-contaminated soil modeling. In the previous studies, thermal data were used for oil-contaminated soil investigation in the literature. However, considering the low resolution of the thermal band and the use of the spectral difference in the other parts of the spectrum, in the present study single bands and spectral indices for oil-contaminated soil modeling were used. Similar studies [22] have used Random Forest as a classifier. We employed RT, SVM, and RF models, and we have visualized the RT model to better understand the structure of the model and the effect of the dataset. Dataset I, which includes all of the Landsat-5 bands and calculated spectral indices, presented better findings for all of the cases. Since the differences in the spectral values between the clean and extremely oil-contaminated soil are higher than highly and moderately contaminated soil, the classification model is simple and more accurate. Thus, with a simple threshold of a single index, in this case, FI, or band (Green), we were able to separate clean from extremely oil-contaminated soil. We were able to classify the highly oil-contaminated soils with a relatively small Random Tree model (nine nodes for both Datasets). However, the moderately oil-contaminated soils were difficult to classify, and the Random Tree model was quite complicated, more than 25 nodes for both Datasets.

All of the results obtained from the different modeling methods are close to each other. This is because of the spectral differences between the clean and contaminated soil in the spectral bands and in the indices derived from the bands. In addition, the used modeling methods were compared in other research supporting the results obtained in this paper. Thus, Baek and Jung [42] compared RF and SVM in oil spill, and while both methods performed accurately, the RF method slightly outperformed the SVM method.

4.3. General Overview of the Proposed Methodology

A scrutinized classification should be needed to determine the degree of the oil-contamination such as dry oil lake, oil lake, oil pit, soot, tar mat, etc. [43]. Detailed classification can only be done using high-resolution imagery and a field survey. Similarly, aerial photography and satellite imagery were used with the addition of collateral maps and documents and field studies [43]. Classification of the dry and wet oil lakes were evaluated using 0.5 m resolution satellite imagery [44].

Many of the previous studies focused on the investigation of oil-spills on water. In the present study, several spectral indices were used successfully used to detect oil slicks, and promising results were obtained for modeling oil-spill in arid areas. Besides, several studies in the literature investigate the remediation of the field conditions, by using a limited number of satellite imageries after the oil spill in Kuwait [6,13]. With the aid of GEE, we were able to investigate a large quantity of data. Thus, we drew the historical event in 1991 with spectral values (Figure 4) and historical events between 1991 2002. In addition, we investigated the spectral values in more recent years (2017 and 2021), and the findings revealed that there are still spectral differences between the clean and oil-contaminated soils.

In the light of these findings, our model allows researchers and organizations to monitor the active oil spill case much earlier. The severity of the spill can be delineated by monitoring the existing conditions. Furthermore, oil spill remediation studies should be carried out by selecting the appropriate spill kit. The degree of the spill and its adverse consequences should be assessed to build the right treatment strategy and select the proper material (chemical, solvents, etc.) as the spill kit. Remotely sensed satellite data enable the creation of a set of data that may be used to assess the degree of spill in inaccessible places.

Overall, the study's findings show that; the proposed methodology can be efficient in detecting different levels of terrestrial oil-contaminated areas. The RT and RF approach were superior in comparison to the SVM models. In addition, the simplicity of the RT is beneficial. The usage of GEE has considerably decreased the time required to process and analyze large volumes of data.

The results from this study can be significant in future clean-up studies, as field measurements are costly and time-consuming. The proposed methodology can be used for detecting the oil-contaminated soil, and, afterward, using higher remote sensing imagery, a detailed investigation can be performed.

5. Conclusions

The oil spill in 1991 in Kuwait was labeled as one of the greatest ecological disasters. The mapping and monitoring of oil spills and oil-contaminated areas are vital for clean-up and remediation studies. Remote sensing data and technologies can help decision-makers and can also be of great importance by saving both time and costs. In the current study, remote sensing data were employed for the spectral signature investigation of the OCS for the ecological disaster in Kuwait in 1991. The major findings demonstrated the differences between the spectral values of clean and oil-contaminated soils. These differences were used to build a model. The main aim of this study is to demonstrate the applicability of different remote sensing data, such as bands and spectral indices, in classifying the oil-contaminated soil at different contamination levels, using machine learning techniques. The significant findings of the current study were highly correlated with the pre-war and post-war period, which included remediation studies. It is concluded from the study that,

by using remote sensing data, it is possible to track the dimensions and directions of the previous oil-spills. In addition to that, more accurate results can be achieved by employing both spectral bands and indices. Although our study presents significant findings for the designated area, more detailed classification can be conducted with high-resolution satellite data.

Author Contributions: Conceptualization, G.K. and H.O.A.; methodology, G.K.; software, G.K.; validation, G.K. and H.O.A.; investigation, G.K., H.O.A., A.P., F.M. and V.F.; writing—original draft preparation, G.K., H.O.A., A.P. and F.M.; writing—review and editing, A.P., F.M. and V.F.; visualization, G.K.; supervision, F.M. and V.F.; project administration, A.P.; funding acquisition, A.P. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by the European Cooperation in Science and Technology, COST Action CA19123.

Data Availability Statement: Not applicable.

Acknowledgments: This article is based on COST Action CA19123—Protection, Resilience, Rehabilitation of damaged environment, supported by COST (European Cooperation in Science and Technology) https://www.cost-phoenix.eu/ (accessed on 19 May 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. El-Baz, F.; Makharita, R.M. *The Gulf War and the Environment*; Routledge: Bingdon-on-Thames, UK, 2016.
2. Pelta, R.; Carmon, N.; Ben-Dor, E. A machine learning approach to detect crude oil contamination in a real scenario using hyperspectral remote sensing. *Int. J. Appl. Earth Observ. Geoinfor.* **2019**, *82*, 101901. [CrossRef]
3. Al-Sarawi, M.; Massoud, M.; Al-Abdali, F. Preliminary assessment of oil contamination levels in soils contaminated with oil lakes in the greater burgan oil fields, kuwait. *Water Air Soil Pollut.* **1998**, *106*, 493–504. [CrossRef]
4. Khordagui, H.; Al-Ajmi, D. Environmental impact of the gulf war: An integrated preliminary assessment. *Environ. Manag.* **1993**, *17*, 557–562. [CrossRef]
5. Almutairi, M.S. An assessment of remediation strategies for kuwaiti oil lakes. *Environ. Geotech.* **2017**, *5*, 345–355. [CrossRef]
6. Al Dousari, A.; Literathy, P. Evidence of hydrocarbon contamination from the burgan oil field, kuwait—Interpretations from thermal remote sensing data. *J. Environ. Manag.* **2008**, *86*, 605–615.
7. Kwarteng, A.Y. Multitemporal remote sensing data analysis of kuwait's oil lakes. *Environ. Int.* **1998**, *24*, 121–137. [CrossRef]
8. El-Baz, F.; Abuelgasim, A.; Lambin, E.; Al-Doasari, A.; Marr, P.; Ryherd, S.; Morency, R. Detection by satellite images of environmental change due to. In *The Gulf War and the Environment*; Center for Remote Sensing Boston University: Boston, MA, USA, 1994; p. 1.
9. Ali, A. Study of characterization of oil contaminated soil in kuwait hera ag ambiental company soil remediation project kuwait. *IJISET-Int. J. Innov. Sci. Eng. Technol.* **2016**, *3*, 7–18.
10. Al-Besharah, J.; Salman, M.; Al-Matrook, F. *Characterization and Quantification of Reclaimable Oil from Oil Lakes Formed in Kuwait Oil Fields*; Report No. KISR4081; Kuwait Institute for Scientific Research: Kuwait City, Kuwait, 1992.
11. Al-Ajmi, D.; Misak, F.; Khalaf, F.; Al-Sudairawi, M.; Al-Dousari, A. *Damage Assessment of the Desert and Coastal Environment of Kuwait by Remote Sensing (vt001c)*; Report No. KISR; Kuwait Institute for Scientific Research: Kuwait City, Kuwait, 1994; Volume 4405.
12. Salam, A. Process of the Symposium on Restoration and Rehabilitation of the Desert Environment. In *The Oil Lakes Environment Disaster*; Elsevier: Amsterdam, The Nertherland, 1996.
13. Kwarteng, A.Y. Remote sensing assessment of oil lakes and oil-polluted surfaces at the greater burgan oil field, kuwait. *Int. J. Appl. Earth Observ. Geoinfor.* **1999**, *1*, 36–47. [CrossRef]
14. Al-Doasari, A.E. *Analysis the Changes in the Tarcrete Layer on the Desert Surface Kuwait Using Satellite Imagery and Cell-Based Modeling*; Boston University: Boston, MA, USA, 2001.
15. Ren, X.; Zelenay, P.; Thomas, S.; Davey, J.; Gottesfeld, S. Recent advances in direct methanol fuel cells at los alamos national laboratory. *J. Power Sources* **2000**, *86*, 111–116. [CrossRef]
16. Zhu, L.; Zhao, X.; Lai, L.; Wang, J.; Jiang, L.; Ding, J.; Liu, N.; Yu, Y.; Li, J.; Xiao, N. Soil tph concentration estimation using vegetation indices in an oil polluted area of eastern china. *PLoS ONE* **2013**, *8*, e54028. [CrossRef]
17. Kooistra, L.; Salas, E.; Clevers, J.; Wehrens, R.; Leuven, R.; Nienhuis, P.; Buydens, L. Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains. *Environ. Pollut.* **2004**, *127*, 281–290. [CrossRef]
18. Austin, J.M.; Mackey, B.G.; Van Niel, K.P. Estimating forest biomass using satellite radar: An exploratory study in a temperate australian eucalyptus forest. *For. Ecol. Manag.* **2003**, *176*, 575–583. [CrossRef]

19. Mason, D.C.; Davenport, I.J.; Neal, J.C.; Schumann, G.J.-P.; Bates, P.D. Near real-time flood detection in urban and rural areas using high-resolution synthetic aperture radar images. *IEEE Trans. GeoSci. Remote Sens.* **2012**, *50*, 3041–3052. [CrossRef]
20. Koch, M.; El-Baz, F. Identifying the effects of the gulf war on the geomorphic features of kuwait by remote sensing and gis. *Photogramm. Eng. Remote Sens.* **1998**, *64*, 739–746.
21. Abdunaser, K. Spatio-temporal analysis of oil lake and oil-polluted surfaces from remote sensing data in one of the libyan oil fields. *Sci. Rep.* **2020**, *10*, 20174. [CrossRef]
22. Ozigis, M.S.; Kaduk, J.D.; Jarvis, C.H. Mapping terrestrial oil spill impact using machine learning random forest and landsat 8 oli imagery: A case site within the niger delta region of nigeria. *Environ. Sci. Pollut. Res.* **2019**, *26*, 3621–3635. [CrossRef] [PubMed]
23. Alghais, N.; Pullar, D. Modelling the impacts current patterns urban form expansion in kuwait with the use abm and gis. In Proceedings of the 21st International Congress on Modelling and Simulation, Broadbeach, QLD, Australia, 29 November–4 December 2015.
24. Algharib, S.M. *Spatial Patterns of Urban Expansion in Kuwait City between 1989 and 2001*; Kent State University: Kent, OH, USA, 2008.
25. Mohammed, A.-S.; Wahba, S. Oil lakes and contamination of soil in kuwait. In *Assessments And Remediation of Oil Contaminated Soils*; New Age International (P) Ltd.: New Delhi, India, 1999; pp. 91–100.
26. Dana, B.; Al-Duwaisan, D.; Al-Naseem, A. Characterization oil contaminated soil. In Proceedings of the 2nd International Conference on Environmental Science and Technology IPCBEE, Singapore, 26–28 February 2011; pp. 439–442.
27. Brovelli, M.A.; Sun, Y.; Yordanov, V. Monitoring forest change in the amazon using multi-temporal remote sensing data and machine learning classification on google earth engine. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 580. [CrossRef]
28. Kovalev, A.; Tokareva, O.S. Using modis ndvi products for vegetation state monitoring on the oil production territory in western siberia. In Proceedings of the MATEC Web of Conferences. Space Engineering—Les Ulis, 2016, Tomsk, Russia, 12–14 April 2016; p. 5003.
29. Adamu, B.; Tansey, K.; Ogutu, B. Using vegetation spectral indices to detect oil pollution in the niger delta. *Remote Sens. Lett.* **2015**, *6*, 145–154. [CrossRef]
30. Balogun, A.-L.; Yekeen, S.T.; Pradhan, B.; Althuwaynee, O.F. Spatio-temporal analysis of oil spill impact and recovery pattern of coastal vegetation and wetland using multispectral satellite landsat 8-oli imagery and machine learning models. *Remote Sens.* **2020**, *12*, 1225. [CrossRef]
31. Zhao, D.; Cheng, X.; Zhang, H.; Zhang, H. An oil slick detection index based on landsat 8 remote sensing images. In Proceedings of the 2018 International Workshop on Big Geospatial Data and Data Science (BGDDS), Wuhan, China, 22–23 September 2018; pp. 1–4.
32. Loos, E.; Brown, L.; Borstad, G.; Mudge, T.; Álvarez, M. Characterization oil slicks at sea using remote sensing techniques. In Proceedings of the 2012 Oceans, Hampton Roads, VA, USA, 14–19 October 2012; pp. 1–4.
33. Zhao, D.; Cheng, X.; Zhang, H.; Niu, Y.; Qi, Y.; Zhang, H. Evaluation of the ability of spectral indices of hydrocarbons and seawater for identifying oil slicks utilizing hyperspectral images. *Remote Sens.* **2018**, *10*, 421. [CrossRef]
34. Holmes, G.; Donkin, A.; Witten, I.H. Weka: A machine learning workbench. In Proceedings of the ANZIIS'94-Australian New Zealnd Intelligent Information Systems Conference, Brisbane, QLD, Australia, 9 November–2 December 1994; pp. 357–361.
35. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
36. Asha Kiranmai, S.; Jaya Laxmi, A. Data mining for classification of power quality problems using weka and the effect of attributes on classification accuracy. *Prot. Control Mod. Power Syst.* **2018**, *3*, 1–12. [CrossRef]
37. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
38. Pham, B.T.; Jaafari, A.; Nguyen-Thoi, T.; Van Phong, T.; Nguyen, H.D.; Satyam, N.; Masroor, M.; Rehman, S.; Sajjad, H.; Sahana, M. Ensemble machine learning models based on reduced error pruning tree for prediction of rainfall-induced landslides. *Int. J. Digit. Earth* **2021**, *14*, 575–596. [CrossRef]
39. Moayedi, H.; Jamali, A.; Gibril, M.B.A.; Kok Foong, L.; Bahiraei, M. Evaluation of tree-base data mining algorithms in land used/land cover mapping in a semi-arid environment through landsat 8 oli image; shiraz, iran. *Geom. Nat. Hazards Risk* **2020**, *11*, 724–741. [CrossRef]
40. Kaplan, O.; Kaplan, G. Response spectra-based post-earthquake rapid structural damage estimation approach aided with remote sensing data: 2020 samos earthquake. *Buildings* **2021**, *12*, 14. [CrossRef]
41. Roy, W.; Asem, S. Application of gis for mapping oil-contaminated soil in kuwait. In *AMCIS 2007 Proceedings*; Kuwait Institute for Scientific Research: Kuwait City, Kuwait, 2007; p. 484.
42. Baek, W.-K.; Jung, H.-S. Performance comparison of oil spill and ship classification from x-band dual-and single-polarized sar image using support vector machine, random forest, and deep neural network. *Remote Sens.* **2021**, *13*, 3203. [CrossRef]
43. Omar, S.; Grealish, G.; Roy, W. Types and extent of soil contamination in greater al-burqan oil field, kuwait. *Kuwait J. Sci. Eng.* **2006**, *33*, 89.
44. Mostagab, H.; Senosy, M.; Al Rashed, A.; Salem, M.E. The impact of hydrocarbon pollution on soil degradation using gis techniques and soil characterization in burgan oil field, south kuwait. *J. Environ. Prot.* **2018**, *9*, 699. [CrossRef]