Engineered Science
DOI: https://dx.doi.org/10.30919/es1348

# ROSID: Remote Sensing Satellite Data for Oil Spill Detection on Land

Daniyar B. Nurseitov,[1, 2, 3] Galymzhan Abdimanap,[1, 3] Abdelrahman Abdallah,[4, 5,*] Gulshat Sagatdinova,[2] Larissa Balakay,[2] Tatyana Dedova,[2] Nurkuisa Rametov[2, 6] and Anel Alimova[1]

## Abstract

Oil spills on land pose significant environmental hazards, impacting ecosystems and human health. Effective detection and monitoring of these spills are critical for timely response efforts. This paper presents for the first time a ground truth dataset for onshore oil spill detection developed using Landsat imagery. The ground truth was implemented using aerial data. To demonstrate the utility of this dataset, we evaluated several state-of-the-art deep learning models, including DeepLabV3+, UNet, PSP-Net, DeepLabV3 and Mask2Former. Our experiments revealed significant insights into the models' capabilities and limitations. Mask2Former and DeepLabV3+, in particular, showed the highest performance metrics. On the validation data, Mask2Former achieved an intersection over Union (IoU) of 72.69% and a F-score coefficient of 84.18%, while DeepLabV3+ achieved an IoU of 67.6% and a F-score coefficient of 80.67%. These results demonstrate the effectiveness of our dataset as a crucial tool for enhancing oil spill detection methodologies and advancing the application of artificial intelligence in ecological preservation and disaster management.

*Keywords:* Remote sensing; Deep Learning; Terrestrial oil spill; Oil pollution dataset; Environmental monitoring; Image segmentation; Landsat imagery.
Received: 15 September 2024; Revised: 09 December 2024; Accepted: 18 December 2024.
Article type: Research article.

## 1. Introduction

Oil spills represent a serious environmental issue, causing significant damages to ecosystems.[1,2] These spills can occur at any stage of oil extraction, transportation, or processing, and their impacts may have toxic effects for years.[3–5] Prominent examples of global environmental pollution include the BP Deepwater Horizon oil spill in April 2010, which resulted in the loss of 4.9 million barrels of oil and billions of dollars spent on cleanup and compensation,[6,7] and the largest oil spill from the Lakeview Gusher well in the San Joaquin Valley, California, between March 1910 and September 1911, where 8.25 million barrels of crude oil were spilled.[8] While the consequences of the Lakeview spill were not as catastrophic as the BP Deepwater Horizon disaster, since the oil remained localized, thick layers of oil-saturated sand can still be found near the well in the desert.

Another significant incident occurred at the Tengiz oil field (Kazakhstan) in 1985 at well No. 37, where oil and gas erupted from a depth of over 4 kilometers. A fire broke out almost immediately and was extinguished only after 400 days. The flame had a diameter of 50 meters and reached heights of 200 meters. The temperatures around the well rose to 15,000°C, turning the soil into a glass-like mass. As a result, 3.4 million tons of oil, 1.7 billion cubic meters of gas, 850 tons of mercaptans, 900,000 tons of soot, and other combustion products were released into the atmosphere, contaminating the environment in a radius of 100–125 km around the oil field.[9] This incident caused significant damages to air, soil, and water

[1] *KazMunayGas Engineering LLP, Astana, 010000, Republic of Kazakhstan*
[2] *Institute of Ionosphere LLP, Almaty, 050020, Republic of Kazakhstan*
[3] *Software Engineering Department, Satbayev University, Almaty, 050013, Republic of Kazakhstan*
[4] *Department of Computer Science & DiSC, University of Innsbruck, 6020, Austria*
[5] *Information Technology Department, Assiut University, Assiut, 71515, Egypt*
[6] *Department of Geospatial Engineering, Satbayev University, Almaty, 050013, Republic of Kazakhstan*
**Email:* abdoelsayed2016@gmail.com (A. Abdallah)

resources.

When oil enters natural ecosystems, it causes both short-term and long-term negative effects. In water, oil forms films on the surface, disrupting gas exchange, reducing oxygen availability to aquatic organisms, and potentially leading to their death. Furthermore, oil is toxic, and direct contact with it can be lethal to aquatic and coastal organisms. On land, oil spills disrupt soil structure and chemical composition, suppress the functional activity of soil microorganisms, deteriorate water and air regimes, and alter ecosystem structure. In urban areas, oil spills can adversely affect human health and lead to the shutdown of water supply systems.[4,5] As light fractions of oil evaporate, harmful hydrocarbons and other compounds are released into the air, contaminating the atmosphere and posing risks to human and animal health. Long-term effects of oil spill include slow ecosystem recovery and, in some cases, the loss of biodiversity. Therefore, it is crucial to detect and mitigate oil spills quickly to minimize environmental damage.

Oil spill is a significant environmental concern for Kazakhstan. A substantial portion of soil pollution is concentrated in the Caspian region, the main oil production area. In severely polluted areas, the maximum concentration of petroleum products reaches 172,480 mg/kg, far exceeding the Kazakhstan's permissible limit of 100 mg/kg.[10] The majority of these polluted areas are considered historical and resulted from past activities during the Soviet era. In the past, these wastes were often improperly disposed of in open landfills or simply dumped on the ground, leading to long-term spill. In recent years, modern oil companies in Kazakhstan have actively engaged in the remediation and reclamation of historical pollution. According to the National Report on the State of the Environment and Use of Natural Resources of Kazakhstan for 2021,[11] work is underway to eliminate abandoned oil sludge waste (859.3 hectares) in oil fields in the Mangistau region, and plans include the disposal and processing of waste stored in 11 unauthorized oil sludge storage sites, with a total volume of 1.3 million cubic meters (annually 184.1 thousand cubic meters). Continuous monitoring and control of these activities are essential.

Remote sensing methods, including satellite observations and aerial photography, are effective tools for mapping oil spills on land and monitoring oil sludge waste reclamation.[12–15] These methods allow for the determination of the extent of oil spill and the progress of remediation efforts. For example, the analyzed data were obtained from the Sentinel-1, Sentinel-2, Landsat 5, Landsat 7, and Landsat 8 missions in Norilsk, along the Ambarnaya River, and Lake Pyasino.[16] The long-term analysis covered the data from 1984 to 2021. The short-term analysis focused on images acquired immediately before, during, and after the spill on May 29, 2020. This study examined the changes in vegetation, snow cover, and water levels before, during, and after the oil spill. The analysis confirms the effectiveness of remote sensing methods for mapping leaks. Modern data processing methods for remote sensing include machine learning and neural networks.[17–20] They enable the analysis of large datasets and the identification of patterns that can help improve the accuracy of oil spill detection in satellite images.[21]

Recent advancements in deep learning have significantly improved the detection of oil spills. Convolutional neural networks (CNNs), in particular, have outperformed traditional signal processing techniques, setting new standards in the field.[22–24] Unlike traditional methods, CNNs can be trained end-to-end, learning directly from examples to map input data to desired outputs. This approach eliminates the need for practitioners to design complex rules or hand-craft features, which are often biased and limited by human capabilities. Instead, CNNs automatically optimize and learn features from the data, making them more accurate and generalizable. Despite their effectiveness, these models require vast amounts of labeled data, which is often scarce and costly to obtain. Therefore, while deep learning represents a major leap forward in onshore and offshore oil spill detection, the challenge remains to gather sufficient labeled datasets to fully harness their potential.[25,26]

The creation of annotated datasets is a crucial task in object recognition on satellite images using artificial intelligence. This also applies to the identification of onshore oil contamination. However, after analyzing publicly available publications, the authors of this study did not find a dataset specifically designed for the recognition of onshore oil spills. Our study addresses this gap by publishing the ROSID dataset (https://github.com/GalymzhanAbdimanap/ROSID), which was created based on Landsat satellite imagery that has been processed and annotated. The accuracy of the labeled data regarding oil spill was verified using aerial survey results. The dataset includes annotated images containing 9 classes of different land surface areas and urban objects, enabling precise classification of the oil field area and identification of oil spill. The meticulous annotation process ensures that the dataset is highly reliable and representative of real-world conditions, making it an invaluable resource for training and evaluating machine learning models. To demonstrate the utility of our dataset, we evaluated several state-of-the-art deep learning models known for their effectiveness in semantic segmentation tasks. Specifically, we assessed the performance of DeepLabV3+,[27] UNet,[28] PSPNet,[29] DeepLabV3,[30] and

Mask2Former.[31] These models were selected based on their robust architectures and proven track records in handling complex segmentation challenges. Our experiments with these models provided significant insights into their capabilities and limitations, highlighting their strengths in accurately detecting and segmenting oil spills. This evaluation not only underscores the value of our dataset but also sets a clear benchmark for future research in environmental monitoring using advanced machine learning techniques.

## 2. Related work

Oil spill detection has been extensively studied in marine environments using a range of machine learning and neural network technologies.[32] In most of these studies, remote sensing data, particularly synthetic aperture radar (SAR) images, are used. On these images, oil slicks smooth the sea surface and appear as well-defined dark spots against the lighter, rougher water surface. For example, SAR images were used as input data to implement two artificial neural networks for oil spill detection, achieving 94% accuracy in detecting dark formations and 89% accuracy in detecting their look-alikes using a fully connected multilayer perceptron.[33] An optimized wavelet neural network was applied to polarimetric characteristics of SAR images, achieving overall accuracies of 96.55% and 97.67% for two datasets.[34] SegNet was utilized for oil spill segmentation in SAR images, reaching 93% accuracy under high noise conditions.[35] The polarimetric SAR filters and multistage autoencoders were employed for high-precision differentiation between crude oil, biogenic films, and clean sea water.[36] In addition to SAR images, optical monitoring data,[37] hyperspectral data,[38] synergistic data combining radar, optical, and thermal imagery,[39] and drone-based marine surface imaging[40] were also used for oil spill detection on the sea surface. The primary requirement for applying machine learning and neural network technologies is the creation of annotated datasets. The studies describe datasets containing annotated oil slick data from the Mediterranean Sea region, based on Sentinel-1 satellite imagery.[41,42] The first annotated dataset of RGB images were captured in a port environment.[40] However, these works focus on marine areas and cannot be applied to detect oil spills on land. There is a lack of work focused on modern research in artificial intelligence that classifies the detection of oil spills on land.

A knowledge-based approach and its implementation system for predicting the consequences of emergency oil spills on land and in groundwater are presented.[12] The novelty of the proposed approach lies in its ability to comprehensively and systematically predict oil spills. The approach consists of components for modeling the geological environment, an oil spill prediction component, and a component for mitigating environmental pollution consequences. The temperature data obtained from the Landsat-8 satellite was used to determine the scale of oil spills in a region.[43] Spatial interpolation methods and gradient techniques were applied to determine the amounts of hydrocarbons mixed with soil, which helps to identify their accumulation in underground horizons. The study's results revealed 60 sites of thermal anomalies with temperature values ranging from 23.2 to 91.11°C in the studied area. After identifying the anomalies, it was possible to accurately determine the oil spills' locations and vegetation reflection was used to assess the level of total petroleum hydrocarbons (TPH) in soils.[14] Over 42 days, an experiment was conducted on Cenchrus alopecuroides (L.) in tropical conditions on soils contaminated with different TPH concentrations. It was found that the plants exposed to 5-19 g/kg TPH showed a stunted growth and reduced chlorophyll and carotenoid content in the leaves, while plants exposed to 1 g/kg TPH showed an increase in pigment content (hormesis effect). These changes affected the reflection of C. alopecuroides in the visible spectrum. Thirty-three vegetation indices were used to link biochemical and spectral responses to oil. This study opens prospects for monitoring the cessation of oil activities in the tropics, focusing on applying the method on a larger scale using satellite images. In the study, the authors employed cutting-edge remote sensing technologies and machine learning techniques to investigate oil spills during the First Gulf War, identifying oil-contaminated and clean areas in Kuwait through unsupervised classification of pre- and post-oil spill data.[15] The research focused on analyzing the spectral signatures of oil-contaminated soils and utilizing various indices for statistical analysis and prediction based on the following machine learning algorithms: support vector machine (SVM), random forest (RF), and random tree (RT). The results from the unsupervised classification were compared with orthophoto-based classifications. The primary objective of this study was to demonstrate the applicability of different remote sensing data, such as spectral bands and indices, for classifying the oil-contaminated soils at varying contamination levels. Unfortunately, according to the authors' statement in the article, the dataset is not public. The study focuses on exploring the potential for detecting oil-contaminated lands in the Niger River Delta region of Nigeria using the random forest machine learning classifier, based on Landsat 8 and vegetation indices.[44] Additionally, relevant indices and spectral bands of the imagery were combined and classified using the RF classifier to support the discrimination process. The classification operation was implemented at both

the macro level for the entire study area and the micro level for individual soil cover subsets. The results obtained from experiments, following sample site calibration and classification operations, demonstrated that the RF algorithm holds a potential for producing reliable maps of oil-free and oil-contaminated vegetation cover. The study most closely related to our work is the one, in which the authors tested the use of Sentinel satellite images for mapping land oil spills using machine learning.[13] At two sites in South Sudan, it was shown that the data from Sentinel-1 and -2 allows for mapping spills with an accuracy of over 90%. Accuracy increases (>95%) when considering multi-temporal information and spatial variables reflecting proximity to oil production infrastructure. This method may be an effective way to monitor areas affected by oil spills. Unfortunately, the authors did not publish the dataset they used in their research.

Therefore, the dataset we propose, which includes annotated data on oil-contaminated soils (land), represents a unique contribution to existing research on this topic. This dataset is innovative, as it provides highly accurate spatial and textural characteristics of oil spills that have not been systematically documented before.

## 3. Materials

This section provides a detailed description of the dataset used in our research for detecting oil spills on land. The dataset was constructed according to the following methodology: At the first stage, medium-resolution satellite images from Landsat-5 (2006, 2007, and 2008) were used and processed and annotated to create a ground truth dataset. The selection of Landsat-5 images and the specified time period was motivated by the results of aerial survey,[45] which took place during the same years and was used in our research to validate the annotated data. In the second stage, the images from Landsat-5, 8, and 9 (2009-2023) were utilized to expand the dataset. The validity of the annotated data was confirmed by comparing classified oil spill with data from previous years. The created dataset plays a crucial role in training and evaluating deep learning models, ensuring that they can accurately detect and segment oil spills across various terrains and environmental conditions.

### 3.1 Study area

The dataset was created using an oil field located in the Mangystau region of western Kazakhstan as the test area (see Fig. 1). Discovered in 1961, the field began the industrial development in 1965. This field is characterized by paraffinic oil. The ecological condition of the studied area has been developed over many years.[9] The main causes of oil spill in this field are emergency and technological emissions from wells and reservoirs within the field collectors, as well as leaks during the oil transportation through trunk pipelines. It is worth noting that until recently, oil well drilling was conducted using the “ambar” method, where wastewater and drilling mud were placed in pits adjacent to the wells. As a result, numerous oil lakes have been formed in the studied field.[46]

### 3.2 Data segmentation

In this study, the validation of the oil spill classified using Landsat-5 data was supported by the results of aerospace monitoring, covering the period from 2006 to 2008.[45] The resolution of the aerial survey was 0.38 meters per pixel, and for some oil sludge storage facilities, 0.14 meters per pixel. The processing of aerial images included the creation of orthophoto plans, mosaics, and further vectorization of contaminated areas using ArcGIS 9 software.[45,47] To eliminate shadows and other non-oil contaminants, 3D visualization ofareas with the potential oil spill was performed on stereoscopic images using the PHOTOMOD StereoDraw module based on the anaglyph effects.[48] Areas visually identified as oil-contaminated were reviewed on stereo images, and if confirmed as non-shadows, they were digitized. Thus, the results of this project, aimed at monitoring and inventorying oil-contaminated areas, formed the basis for constructing the described dataset as ground truth.

From the publicly available archive of medium and high-resolution satellite images for the given region during the 2006-2008 period, Landsat-5 images were selected. This choice was primarily driven by the fact that these were

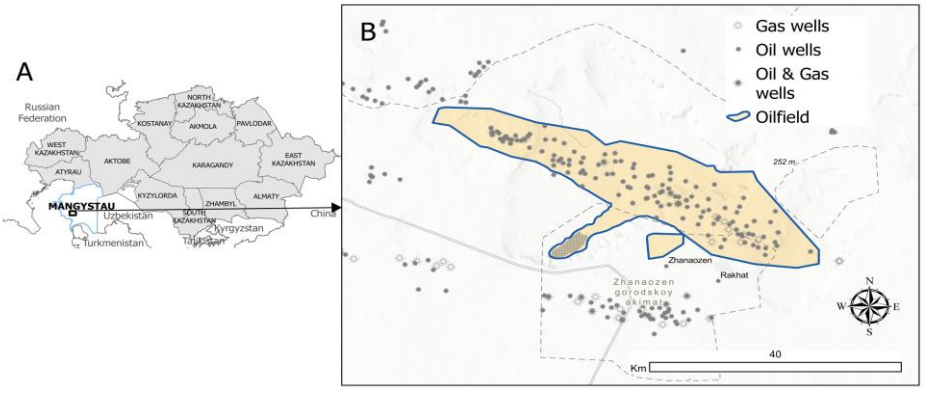


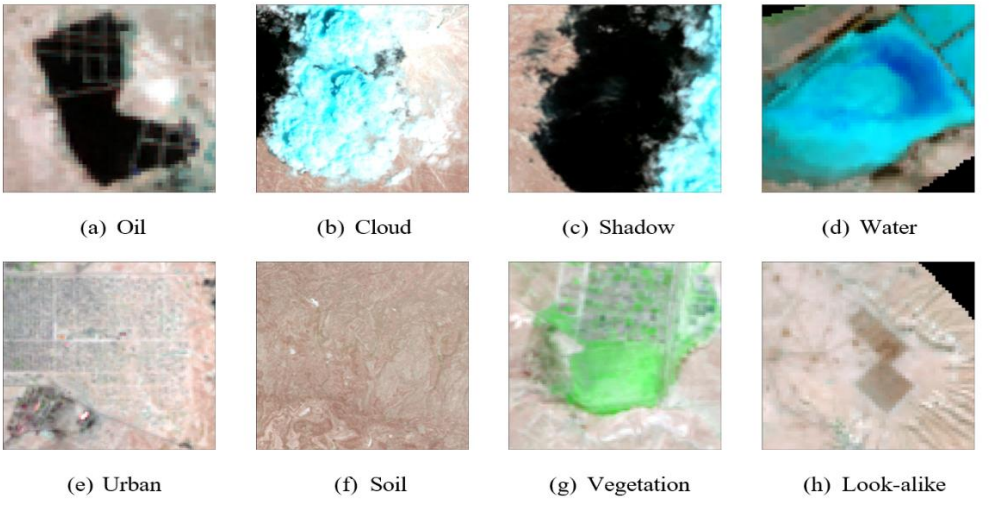

**Fig. 1** Map of the study area at an oil field in Western Kazakhstan.

virtually the only available data with 30-meter resolution that fully covered the study area during that period, with a frequency of at least twice a month. Secondly, there is an almost continuous time series of images from 2006 to 2023, allowing for their use without switching to other datasets. Thirdly, the Landsat satellite family, while evolving and adding new spectral channels, has maintained the parameters of older channels. Additionally, the next generation of Landsat satellites is expected to be released with 26 spectral channels and improved spatial resolution of 10 to 20 meters, making the developed technology applicable in the future. Thus, the dataset was created using Landsat-5, 8, and 9 satellite images from 2006 to 2023, with a spatial resolution of 30 meters and a swath width of 185 kilometers, covering data from 7 or 11 spectral channels.

The processing of satellite images to construct the dataset involved the following stages (see Fig. 2):

1. Selection of Landsat-5 images with cloud cover less than 10% for the period 2006-2008 using google earth engine.[49]
2. Creation of training samples by delineating oil-contaminated areas based on vector data from aerospace monitoring (2006–2008) uploaded to google earth engine. Other classes (water, vegetation, urban areas, clouds, shadows, false objects, and soil) were identified visually. In total, eight classes were identified (see Fig. 3). To refine class boundaries, the images were adjusted for cloud cover and seasonal variations (*e.g.*, autumn vegetation).
3. Supervised classification using the random forest method.
4. Export of classified layers (thematic layers) to Geo-TIFF format (see Fig. 4).
5. Conversion of Geo-TIFF raster data to vector format.
6. Accuracy assessment of the pre-classification data through joint analysis with Landsat images.
7. Manual correction of pre-classification data using ArcGIS.
8. Conversion of corrected vector data back to Geo-TIFF raster format.
9. After verification and corrections, 13 satellite images were selected and converted to three-channel images by extracting


**Fig. 2** Flowchart of data processing and dataset creation.



(a) Oil (b) Cloud (c) Shadow (d) Water

(e) Urban (f) Soil (g) Vegetation (h) Look-alike

**Fig. 3** Landsat satellite image fragments with examples of distinguished classes

spectral channels 7 (SWIR), 4 (NIR), and 3 (RED). These channels were chosen because they most effectively display oil spill on the land surface. Channel 7 (SWIR) is sensitive to moisture content and is used to detect liquids such as oil, particularly on land. Channel 4 (NIR) distinguishes vegetation from other objects such as oil or water due to its ability to capture changes in reflectance. Channel 3 (RED) helps to highlight contamination against soil and water, providing contrast between vegetation and oil-contaminated areas.

10. These 13 three-channel images were divided into patches of 320×320 pixels with a step of 106 pixels.
11. Segmentation of patches into subsets for training, validation, and testing.
12. Training of selected neural network models on these data.

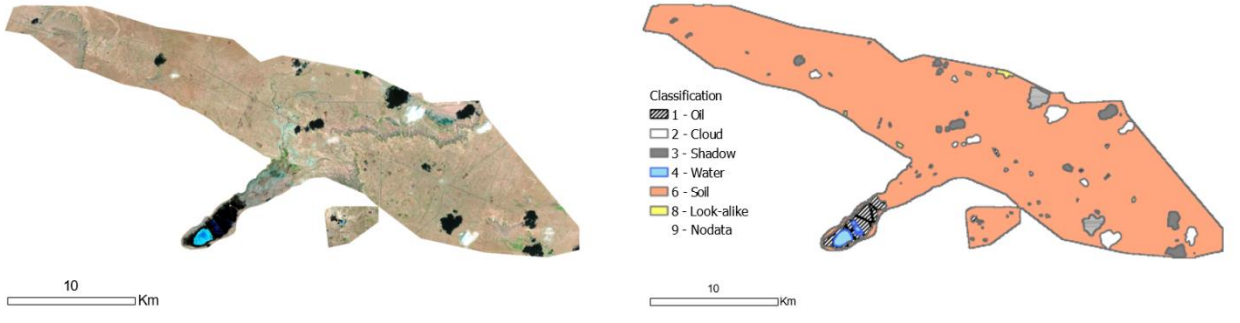
**Table 1.** Number of pixels for each class in the dataset.

| No | Class name | Number of pixels |
|---|---|---|
| 1 | oil | 782,947 |
| 2 | cloud | 216,133 |
| 3 | shadow | 84,053 |
| 4 | water | 249,060 |
| 5 | urban | 9,006 |
| 6 | soil / undefined | 47,081,945 |
| 7 | vegetation | 2,778 |
| 8 | look-alike | 121,921 |
| 9 | background | 73,111,465 |

Subsequently, the best-trained model, rather than manual classification as in the initial stage, was used to process and obtain pre-classification for 110 Landsat 8 and 9 images with cloud cover less than 10% for the period from 2009 to 2023. Initially, the spectral bands of Landsat 8 and 9 were adjusted using google earth engine to match the spectral bands of Landsat 5. Out of the 11 bands available in Landsat 8 and 9, only 7 were used and renamed to correspond to Landsat 5.

Steps 5-8 were then repeated, *i.e.*, the model results were converted into a GeoTIFF raster file with georeferencing, ensuring an accurate alignment with the original satellite images.

These results were used as preliminary annotations, which were provided to experts for manual correction using ArcGIS to obtain final labels. Manual correction of the preliminary annotation results was carried out sequentially based on the timestamps of the images, tracking changes in oil spill configuration year by year, initially documented during the period 2006-2008 (see Fig. 5). The results of the manual correction were also converted to raster format. After corrections, all images were incorporated into the main dataset, which included 123 Landsat images. Steps 9-12 were repeated, specifically, these data were converted into three-channel images by combining channels 7, 4, and 3. Fragmentation was performed using the OpenCV library with a sliding window of 106 pixels.[50] The fragments were divided into 7,475 patches of 320x320 pixels for training and validation, and 520 patches for testing. The model was retrained on the expanded dataset. Table 1 displays the number of pixels for each data class. The largest number of pixels, excluding the soil class, belongs to the oil class, constituting 0.64% (with the background class) or 1.61% (excluding the background class) of the total raster elements. However, our dataset exhibits a significant class imbalance, as reflected in the number of pixels associated with each class (see Table 1). For instance, the “oil” class contains only 782,947 pixels, which is considerably fewer than the 47,081,945 pixels for the “soil/undefined” class. This imbalance arises from the natural heterogeneity of the data, where classes associated with oil appear less frequently and cover smaller areas within the selected oil field, while the soil class occupies a much larger portion of the area, serving as a background component.




(a) Preprocessed satellite image (b) Labeled class samples

**Fig. 4** Classes in the dataset. (a) An example of a preprocessed satellite image of the study area; and (b) an example of marking of various classes: black – oil, white – cloud, gray – shadow of a cloud, blue – water, hit pink – soil, yellow – look-alike, blank – no data.

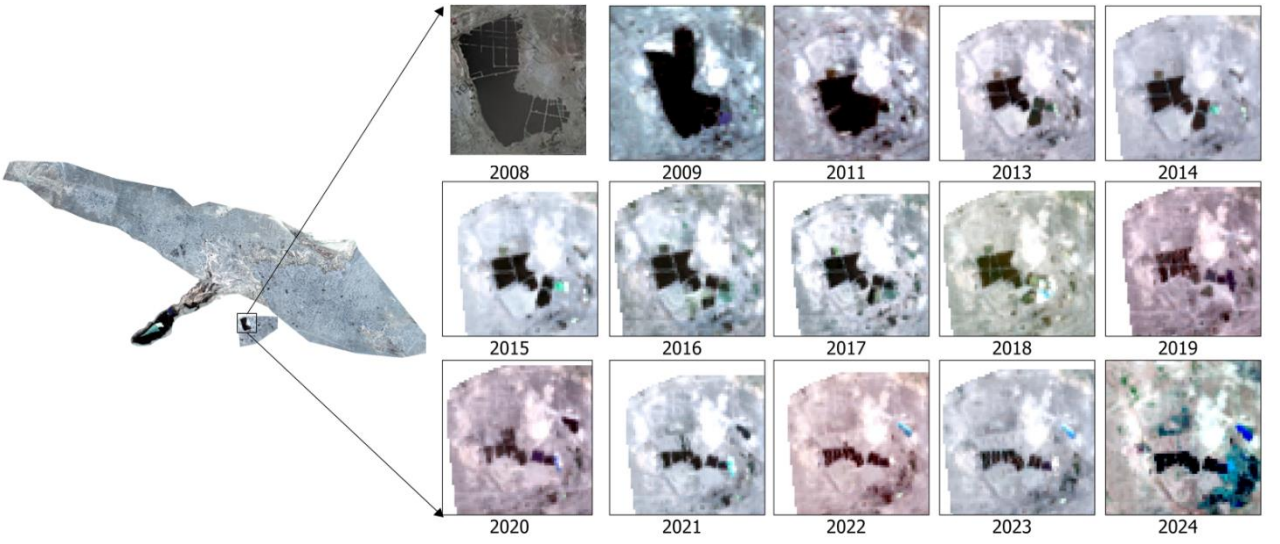



**Fig. 5** Evolution of oil spill at a technological site over time, shown with aerial photography and Landsat satellite images.

Attempts to artificially balance the dataset by increasing the representation of rare classes, such as "oil" and other rare classes through data augmentation or synthetic image generation could distort the original data characteristics and negatively impact model performance. For example, augmentation techniques involving rotations, shifts, or brightness adjustments could generate images that do not accurately reflect the real-world conditions, leading to a reduction in the model's ability to correctly segment oil in new data. Similarly, generating synthetic data could skew the class distribution, causing the model to train on unrealistic data and diminishing its generalization capability when applied to data from the same region. Therefore, artificially increasing data could create a false perception of class uniformity, whereas the model's actual challenge lies in correctly recognizing rare classes under the real-world imbalance. Ultimately, the observed class imbalance is an accurate reflection of the region's actual conditions and characteristics, justifying the use of the current approach without substantial modifications to class proportions.

Additionally, for the reliability of the test, an area with an oil reservoir was chosen, the pollution of which has been routinely eliminated for many years (See the boot-like shape area on Fig. 5).

## 4. Methodology

Oil spill detection on land requires advanced image processing techniques capable of accurately identifying and segmenting contaminated areas. In this study, we leverage state-of-the-art deep learning models known for their effectiveness in semantic segmentation tasks. Our approach involves using high-resolution aerial imagery and deep learning architectures to enhance the accuracy of oil spill detection. This section provides a comprehensive overview of five deep learning models utilized in our study: DeepLabv3, DeepLabv3+, PSPNet, U-Net, and Mask2Former. We begin by explaining the fundamentals of CNNs and their role in image processing. Subsequently, we delve into the specifics of each model, starting with DeepLabv3, which employs atrous convolutions for capturing multi-scale information. We then discuss DeepLabv3+, an enhanced version of DeepLabv3 with an encoder-decoder structure and depthwise separable convolutions. Following this, we explore PSPNet, which uses pyramid pooling to gather global context, and U-Net, a model designed for precise biomedical image segmentation that is also highly effective in our application. Finally, following the trend, we added a model with a transformer Musk2Former to the study

### 4.1 Deep learning

Deep learning, a subset of machine learning, involves neural networks with many layers (deep networks) that can learn representations of data with multiple levels of abstraction. Convolutional neural networks (CNNs) are a class of deep neural networks commonly used in computer vision tasks. They are particularly effective in analyzing visual imagery due to their ability to capture spatial hierarchies in data through convolution operations. Mathematically, a convolution operation on an image $I$ with a kernel $K$ is defined by Equation (1):

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n) \quad (1)$$

where $S(i, j)$ is the output feature map.

CNNs typically consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers apply the convolution operation, the pooling layers downsample the spatial dimensions, and the fully connected layers perform the final classification.

### 4.2 DeepLabv3

DeepLabv3 employs atrous (or dilated) convolutions to control the resolution, at which features are computed. Atrous convolution introduces holes (or zeros) between filter elements, allowing for a wider field of view without increasing the number of parameters. The atrous convolution operation is defined by Equation (2):

$$y[i] = \sum_k x[i + r \cdot k]w[k] \quad (2)$$

where $r$ is the atrous rate, $x$ is the input signal, $y$ is the output signal, and $w$ is the filter. This allows the network to effectively enlarge the field of view of filters without increasing the number of parameters or the amount of computation.

DeepLabv3 also utilizes atrous spatial pyramid pooling (ASPP) to capture multi-scale information by applying multiple atrous convolutions with different rates. ASPP uses parallel dilated convolutions with different rates to probe an incoming convolutional feature layer with filters at multiple sampling rates, capturing objects as well as image context at multiple scales:

$$ASSP(x) = concat[conv_{1\times1}(x), conv^{r}_{3\times3}(x), conv^{2r}_{3\times3}(x), conv^{3r}_{3\times3}(x)] \quad (3)$$

where $r$ denotes the atrous rate.

### 4.3 DeepLabv3+

DeepLabv3+ enhances DeepLabv3 by incorporating an encoder-decoder structure. The encoder extracts dense feature representations using atrous convolutions, while the decoder refines these features to improve segmentation accuracy at object boundaries. The encoder-decoder structure in DeepLabv3+ can be mathematically represented by Equation (4&5):

$$h_e = f_e(x; \theta_e) \quad (4)$$

$$y = f_d(h_e; \theta_d) \quad (5)$$

where $h_e$ represents the encoded features, $f_e$ and $f_d$ are the encoder and decoder functions respectively, and $\theta_e$ and $\theta_d$ are their parameters.

DeepLabv3+ also employs depthwise separable convolutions to reduce the number of parameters and computational complexity. A depthwise separable convolution is a two-step process: depthwise convolution, which applies a single filter per input channel, followed by pointwise convolution, which applies a 1 × 1 convolution to combine the outputs of the depthwise convolution.

$$y = PointwiseConv(DepthwiseConv(x)) \quad (6)$$

#### 4.3.1 Depthwise convolution

Depthwise convolution is a type of convolution where each input channel is convolved with a single depthwise filter. This operation can be mathematically represented as follows:

Let $X \in \mathbb{R}^{H\times W\times C}$ be the input feature map, where $H$ and $W$ are the height and width of the input, and $C$ is the number of channels. Let $K \in \mathbb{R}^{k\times k\times C}$ be the depthwise convolutional filter, where $k \times k$ is the spatial dimension of the filter.

The output of the depthwise convolution $Y \in \mathbb{R}^{H'\times W'\times C}$ can be computed as:

$$Y_{i,j,c} = \sum_{m=0}^{k-1}\sum_{n=0}^{k-1} X_{i+m,j+n,c} \cdot K_{m,n,c} \quad (7)$$

where $i$ and $j$ iterate over the spatial dimensions of the output feature map, and $c$ iterates over the channels. The depthwise convolution processes each channel independently.

#### 4.3.2 Pointwise convolution

Pointwise convolution, also known as $1 \times 1$ convolution, is applied after depthwise convolution to combine the outputs of the depthwise convolutions. It applies a $1 \times 1$ filter to combine the channels, effectively performing a linear combination of the input channels. This operation can be mathematically represented as follows:

Let $X \in \mathbb{R}^{H\times W\times C}$ be the input feature map, and let $W \in \mathbb{R}^{1\times1\times C\times C'}$ be the pointwise convolutional filter, where $C$ is the number of input channels, and $C'$ is the number of output channels.

The output of the pointwise convolution $Y \in \mathbb{R}^{H\times W\times C'}$ can be computed by Equation (8):

$$Y_{i,j,c'} = \sum_{c=0}^{C-1} X_{i,j,c} \cdot W_{1,1,c,c'} \quad (8)$$

where $i$ and $j$ iterate over the spatial dimensions of the input feature map, and $c'$ iterates over the output channels.

#### 4.3.3 Combined depthwise separable convolution

In depthwise separable convolution, these two operations are combined. First, a depthwise convolution is applied to each input channel independently. Then, a pointwise convolution is applied to combine the depthwise convolution outputs. This combination can be represented as (7) and (8). DeepLabv3+ utilizes these depthwise separable convolutions to efficiently capture spatial information while keeping the model lightweight and computationally efficient.

### 4.4 PSPNet

PSPNet addresses the need to understand global context in scene parsing by introducing a pyramid pooling module. This module aggregates context information from different regions of the image, helping the network to understand the overall scene structure. The pyramid pooling module applies pooling operations at different grid scales, producing feature maps of different sizes:

$$f_{pp} = [f_1, f_2, f_3, f_4] \quad (9)$$

where $f_1, f_2, f_3$, and $f_4$ are the pooled features at different scales (*e.g.*, $1x1, 2x2, 3x3$, and $6x6$). These pooled features are then upsampled and concatenated with the original feature map:

$$f_{out} = Concat([f_{pp}, f_{orig}]) \quad (10)$$

This multi-scale feature representation allows PSPNet to capture both local and global context, improving the

segmentation accuracy.

### 4.5 U-Net

U-Net is specifically designed for biomedical image segmentation and features a symmetric encoder-decoder architecture with skip connections. The encoder path captures context, while the decoder path enables precise localization. The U-Net architecture consists of a contracting path (encoder) and an expansive path (decoder). The contracting path follows the typical architecture of a convolutional network, with repeated applications of two $3x3$ convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) and a $2x2$ max pooling operation with stride 2 for downsampling:

$$h_{enc}^{l+1} = MaxPool\left(\sigma(W^l * h_{enc}^l + b^l)\right) \quad (11)$$

where $W^l$ and $b^l$ are the weights and biases of the $l$-th layer, and $\sigma$ is the ReLU activation function.

In the expansive path, the features are upsampled and combined with the high-resolution features from the contracting path via skip connections:

$$h_{dec}^l = Concat\left(h_{dec}^{l+1}, h_{enc}^l\right) \quad (12)$$

$$h_{dec}^{l-1} = \sigma\left(W_{up}^l * h_{dec}^l + b^l\right) \quad (13)$$

where $W_{up}^l$ represents the weights of the up-convolution (transposed convolution).

This architecture allows U-Net to utilize both local information from the contracting path and contextual information from the expansive path, making it highly effective for segmentation tasks that require precise boundary delineation.

### 4.6 Mask2Former

Mask2Former is an advanced meta-architecture for universal segmentation tasks, including image and video segmentation. It builds upon the concept of mask classification by dividing pixels into segments, with a transformer decoder as its key component. Mask2Former introduces masked attention (14), a variant of cross-attention that ensures computations remain efficient by restricting attention to specific segment locations.

$$X_l = Softmax(M_{l-1} + Q_l K_l^T)V_l + X_{l-1} \quad (14)$$

The mask $M_{l-1}$ in (15) controls the attention matrix, ensuring that only features within the foreground of predicted masks are attended to.

$$M_{l-1}(x,y) = \begin{cases} 0, & \text{if } M_{l-1}(x,y) = 1 \\ -\infty, & \text{otherwise} \end{cases} \quad (15)$$

where $M_{l-1}(x,y) \in \{0,1\}^{N \times H^l W^l}$ is the binarized output (thresholded at 0.5) of the resized mask prediction of the previous $(l-1)$-th transformer decoder layer. It is resized to the same resolution of $K_l$. $M_0$ is the binary mask prediction obtained from $X_0$.[31]

To optimize computational efficiency, the model uses multi-scale high-resolution features and avoids high-computation methods by employing efficient pyramid designs that reduce redundant computation. Moreover, the use of learned embeddings enhances spatial resolution while controlling the complexity of operations.

Mask2Former's advanced architecture enabled it to outperform other models in the segmentation of oil spills, achieving an Intersection over Union (IoU) of 72.69% and an F-score of 84.18% on the validation dataset. These metrics highlight its ability to accurately delineate oil spill boundaries, particularly in cases involving small or complex shapes, where other models like DeepLabV3+ and UNet showed limitations. The masked attention mechanism ensures the model focuses computational resources on relevant spatial areas, leading to a more precise segmentation.

To optimize computational efficiency, the model uses multi-scale high-resolution features and avoids high-computation methods by employing efficient pyramid designs that reduce redundant computation. Moreover, the use of learned embeddings enhances spatial resolution while controlling the complexity of operations.

## 5. Experiment setup

In this section, we outline the experimental setup used to train and evaluate the deep learning models for oil spill detection. We used PyTorch a popular open-source machine learning library known for its flexibility and efficiency in building and training neural networks.[51] The experiments were conducted on NVIDIA RTX 4070 Super GPUs, which are specifically designed for high-performance deep learning tasks, providing the necessary computational power to handle large datasets and complex model architectures.

### 5.1 Dataset preparation and preprocessing

The ROSID dataset was meticulously prepared to standardize input features and enhance model training. Images were resized to dimensions of 320×320 pixels to ensure uniform input sizes across all models. Normalization was applied to align pixel intensity distributions, with channel-wise means and standard deviations set to [123.675, 116.28, 103.53] and [58.395, 57.12, 57.375], respectively. These values correspond to typical natural image statistics, facilitating consistent model initialization.

Data augmentation techniques were employed to improve the generalization and robustness. Random resizing was performed with scaling factors ranging from 50% to 200% of the original size. Random cropping ensured that the focus remained on foreground objects, with a maximum category ratio of 0.75 to prevent overwhelming dominance of background classes. Additional augmentations, including random horizontal flips and photometric distortions, simulated

varying environmental conditions, further diversifying the training data.

### 5.2 Training configuration

The training configuration was carefully designed to optimize model performance while considering computational constraints. A batch size of 4 was chosen to balance memory limitations with convergence speed. To prevent overfitting and improve generalization, the InfiniteSampler (https://mmengine.readthedocs.io/en/v0.8.2/api/generated/mmengine.dataset.InfiniteSampler.html) was used to shuffle the training data continuously. The training pipeline also integrated preprocessing steps-such as resizing, augmentation, and annotation packing-directly into the data loader to streamline the data flow and minimize any bottlenecks in the pipeline.

The models were trained using the Adam optimizer with an initial learning rate of 0.001. The learning rate was adjusted dynamically using a scheduler, reducing the rate when the validation performance plateaued, ensuring that the models converged efficiently without overfitting. For the loss function, we used Cross-Entropy Loss, with class weights derived from the dataset distribution to account for the class imbalance, particularly the underrepresentation of oil spill regions.

**Table 2.** Training iterations and time for each model.

| Model | Iterations | Training Time |
|---|---|---|
| Mask2Former | 40,000 | 25 hours |
| DeepLabV3 | 40,000 | 26 hours 30 minutes |
| DeepLabV3+ | 40,000 | 27 hours |
| UNet | 40,000 | 1 hour 30 minutes |
| PSPNet | 40,000 | 2 hours 30 minutes |

Table 2 summarizes the number of iterations and corresponding training times for each model used in our experiments. The table provides a comparison of the training times for each model, all of which were trained for 40,000 iterations. These times were measured under similar hardware configurations, using an NVIDIA RTX 4070 Super GPU, ensuring the results are comparable in terms of computational resources.

### 5.3 Validation and testing configuration

Validation and testing pipelines were configured to provide unbiased assessments of model performance. A batch size of 1 was used to evaluate individual samples with precision, particularly important for edge cases and small objects. Input images were resized to 320 × 320 pixels while maintaining aspect ratios to ensure consistency with training. The DefaultSampler (https://mmengine.readthedocs.io/en/v0.8.2/api/generated/mmengine.dataset.DefaultSampler.html), without shuffling, preserved the order of samples, facilitating reproducibility and traceability in results. The same normalization and preprocessing steps as in training were applied to maintain uniformity across the datasets.

### 5.4 Evaluation metrics

To evaluate our models, the commonly used metrics we employed for semantic segmentation, ensuring a comprehensive assessment of model performance. These metrics include IoU, accuracy, precision, recall, and F-score.

#### 5.4.1 Intersection over Union (IoU)

IoU measures the overlap between the predicted segmentation and the ground truth. It is defined as the ratio of the intersection area to the union area of the predicted and ground truth masks.

$$IoU = \frac{|Prediction \cap Ground\ Truth|}{|Prediction \cup Ground\ Truth|} \tag{16}$$

#### 5.4.2 Accuracy

The accuracy represents the proportion of correctly classified pixels out of the total number of pixels. This metric provides a general sense of the model's performance across all classes.

$$Accuracy = \frac{\sum_{i=1}^{N} TP_i}{\sum_{i=1}^{N}(TP_i + FP_i + FN_i + TN_i)} \tag{17}$$

where *TP* – true positive: Correctly predicted positive cases. These are instances where the model correctly identifies a positive outcome.

*TN* – true negative: Correctly predicted negative cases. These represent instances where the model correctly identifies a negative outcome.

*FP* – false positive: Incorrectly predicted positive cases. These occur when the model predicts a positive outcome, but the actual outcome is negative.

*FN* – false negative: Incorrectly predicted negative cases. These occur when the model predicts a negative outcome, but the actual outcome is positive.

#### 5.4.3 Precision and recall

Precision is the ratio of true positive predictions to the total predicted positives, while recall is the ratio of true positives to the actual positives. These metrics help to understand the model's ability to correctly identify the oil spills without generating excessive false positives or negatives.

$$Precision = \frac{TP}{TP + FP} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

#### 5.4.4 F-score

The F-score is the harmonic mean of precision and recall, providing a single metric that balances both aspects of model performance.

$$F\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{20}$$

The computational cost of each model is critical for assessing its practical applicability, particularly for real-time oil spill detection. Table 3 summarizes the FLOPs and

inference times for DeepLabv3, DeepLabv3+, PSPNet, U-Net, and Mask2Former.

**Table 3.** Computational cost analysis.

| Model | FLOPs (GFLOPs) | Inference Time (ms) |
|---|---|---|
| DeepLabv3 | 45 | 12 |
| DeepLabv3+ | 60 | 15 |
| PSPNet | 65 | 18 |
| U-Net | 25 | 8 |
| Mask2Former | 120 | 25 |

### 5.5 Computational analysis

DeepLabv3 and DeepLabv3+, which use atrous convolutions and encoder-decoder structures, exhibit computational costs of approximately 45 and 60 GFLOPs, respectively. PSP-Net, due to its pyramid pooling module, incurs a cost of around 65 GFLOPs. U-Net, being lighter with its symmetric encoder-decoder design, requires about 25 GFLOPs for similar inputs. Mask2Former, leveraging a transformer-based architecture with masked attention, is the most computationally intensive, requiring approximately 120 GFLOPs. Inference times, measured on an NVIDIA RTX 4070 GPU, range from 8 milliseconds for U-Net to 25 milliseconds for Mask2Former, highlighting the trade-offs between computational efficiency and segmentation accuracy. For scenarios where real-time performance is paramount, U-Net and DeepLabv3 offer a balanced trade-off between cost and accuracy. However, Mask2Former, while more demanding, provides a superior accuracy, making it suitable for high-stakes applications.

## 6. Results

The performance of different semantic segmentation models was evaluated using high-resolution aerial imagery for the detection of oil spills on land. The models evaluated include DeepLabV3, DeepLabV3+, PSPNet, and UNet. The evaluation metrics used were IoU), Accuracy (Acc), F-score, Precision, and Recall. The results for each model on both test and validation datasets are presented in the following subsections.

### 6.1 DeepLabV3 results

The performance of the DeepLabV3 model on test and validation datasets is summarized in Tables 4 and 5. For the test data, the model had an IoU of 62.55 for oil, indicating a moderate level of overlap between the predicted and actual oil spill regions. The model achieved an accuracy of 76.83, with a F-score coefficient of 76.96, suggesting a good segmentation performance. Precision and Recall for oil were 77.09 and 76.83, respectively. On the validation data, the model was improved and an IoU of 64.07 was achieved for oil. The accuracy was 74.74, with a F-score coefficient of 78.1. Precision and Recall for oil were 81.78 and 74.74, respectively, indicating better segmentation and detection performance.

### 6.2 DeepLabV3+ results

The performance of the DeepLabV3+ model on test and validation datasets is summarized in Tables 6 and 7. On the test data, DeepLabV3+ achieved an IoU of 66.26 for oil, with an accuracy of 81.76. The F-score coefficient was 79.7, with Precision and Recall values of 77.75 and 81.76, respectively. For the validation data, the model achieved an IoU of 67.6 for oil, with an accuracy of 78.72. The F-score coefficient was 80.67, with Precision and Recall values of 82.72 and 78.72, respectively, showing a significant improvement over the test data.

**Table 4.** Performance metrics for DeepLabV3 on test data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 62.55 | 76.83 | 76.96 | 77.09 | 76.83 |
| cloud | 37.84 | 41.47 | 54.91 | 81.2 | 41.47 |
| shadow | 44.22 | 63.75 | 61.32 | 59.07 | 63.75 |
| water | 76.31 | 86.73 | 86.57 | 86.4 | 86.73 |
| urban | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| soil | 98.34 | 99.32 | 99.16 | 99.01 | 99.32 |
| vegetation | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| look a like | 29.05 | 34.54 | 45.02 | 64.62 | 34.54 |
| background | 99.57 | 99.77 | 99.79 | 99.81 | 99.77 |

**Table 5.** Performance metrics for DeepLabV3 on validation data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 64.07 | 74.74 | 78.1 | 81.78 | 74.74 |
| cloud | 67.95 | 89.1 | 80.92 | 74.11 | 89.1 |
| shadow | 37.16 | 66.65 | 54.18 | 45.65 | 66.65 |
| water | 74.15 | 85.79 | 85.16 | 84.54 | 85.79 |
| urban | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| soil | 98.44 | 99.29 | 99.22 | 99.14 | 99.29 |
| vegetation | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| look a like | 33.63 | 39.17 | 50.33 | 70.36 | 39.17 |
| background | 99.56 | 99.75 | 99.78 | 99.81 | 99.75 |

**Table 6.** Performance metric for DeepLabV3+ on test data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 66.26 | 81.76 | 79.7 | 77.75 | 81.76 |
| cloud | 32.32 | 35.71 | 48.86 | 77.31 | 35.71 |
| shadow | 34.47 | 49.18 | 51.27 | 53.55 | 49.18 |
| water | 77.77 | 86.28 | 87.49 | 88.74 | 86.28 |
| urban | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| soil | 98.52 | 99.4 | 99.25 | 99.11 | 99.4 |
| vegetation | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| look a like | 33.9 | 40.47 | 50.64 | 67.64 | 40.47 |
| background | 99.64 | 99.79 | 99.82 | 99.84 | 99.79 |

**Table 7.** Performance metrics for DeepLabV3+ on validation data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 67.6 | 78.72 | 80.67 | 82.72 | 78.72 |
| cloud | 63.05 | 84.87 | 77.34 | 71.04 | 84.87 |
| shadow | 30.27 | 66.66 | 46.47 | 35.39 | 66.66 |
| water | 74.35 | 84.28 | 85.29 | 86.32 | 84.28 |
| urban | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| soil | 98.49 | 99.25 | 99.22 | 99.13 | 99.25 |
| vegetation | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| look a like | 35.62 | 41.91 | 52.53 | 70.36 | 41.91 |
| background | 99.62 | 99.78 | 99.81 | 99.84 | 99.78 |

**Table 8.** Performance metrics for PSPNet on test data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 63.12 | 76.48 | 77.39 | 78.32 | 76.48 |
| cloud | 42.93 | 46.57 | 60.07 | 84.6 | 46.57 |
| shadow | 41.29 | 61.93 | 58.44 | 55.33 | 61.93 |
| water | 76.76 | 87.43 | 86.85 | 86.27 | 87.43 |
| urban | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| soil | 98.36 | 99.33 | 99.17 | 99.01 | 99.33 |
| vegetation | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| look a like | 29.33 | 36.69 | 45.35 | 59.36 | 36.69 |
| background | 99.59 | 99.76 | 99.79 | 99.82 | 99.76 |

**Table 9.** Performance metrics for PSPNet on validation data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 64.54 | 74.28 | 78.45 | 83.1 | 74.28 |
| cloud | 66.37 | 89.01 | 79.78 | 72.29 | 89.01 |
| shadow | 33.49 | 67.74 | 50.18 | 39.85 | 67.74 |
| water | 74.82 | 85.76 | 85.60 | 85.44 | 85.76 |
| urban | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| soil | 98.45 | 99.29 | 99.22 | 99.14 | 99.29 |
| vegetation | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| look a like | 35.7 | 42.5 | 52.62 | 69.05 | 42.5 |
| background | 99.56 | 99.73 | 99.78 | 99.82 | 99.73 |

**Table 10.** Performance metrics for UNet on test data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 57.87 | 73.75 | 73.31 | 72.88 | 73.75 |
| cloud | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| shadow | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| water | 54.79 | 59.03 | 70.79 | 88.41 | 59.03 |
| urban | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| soil | 98.35 | 99.67 | 99.17 | 98.67 | 99.67 |
| vegetation | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| look a like | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| background | 99.66 | 99.82 | 99.83 | 99.83 | 99.82 |

**Table 11.** Performance metrics for UNet on validation data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 57.9 | 69.92 | 73.34 | 77.11 | 69.92 |
| cloud | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| shadow | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| water | 42.14 | 46.4 | 59.29 | 82.11 | 46.4 |
| urban | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| soil | 97.8 | 99.64 | 98.89 | 98.15 | 99.64 |
| vegetation | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| look a like | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| background | 99.55 | 99.71 | 99.78 | 99.85 | 99.71 |

**Table 12.** Performance metrics for Mask2Former on test data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 66.66 | 85.0 | 79.99 | 75.54 | 85.0 |
| cloud | 0.17 | 0.17 | 0.34 | 82.93 | 0.17 |
| shadow | 3.67 | 3.86 | 7.08 | 43.03 | 3.86 |
| water | 81.72 | 89.09 | 89.94 | 90.8 | 89.09 |
| urban | 33.96 | 36.96 | 50.71 | 80.74 | 36.96 |
| soil | 98.67 | 99.49 | 99.33 | 99.17 | 99.49 |
| vegetation | 27.2 | 59.08 | 42.77 | 33.52 | 59.08 |
| look a like | 33.54 | 65.47 | 50.23 | 40.74 | 65.47 |
| background | 99.69 | 99.78 | 99.84 | 99.9 | 99.78 |

**Table 13.** Performance metrics for Mask2Former on validation data.

| Class | IoU | Accuracy | F-score | Precision | Recall |
|---|---|---|---|---|---|
| oil | 72.69 | 83.43 | 84.18 | 84.95 | 83.43 |
| cloud | 62.83 | 66.19 | 77.17 | 92.51 | 66.19 |
| shadow | 44.23 | 56.78 | 61.33 | 66.67 | 56.78 |
| water | 80.79 | 88.5 | 89.37 | 90.27 | 88.50 |
| urban | 41.97 | 47.04 | 59.13 | 79.56 | 47.04 |
| soil | 98.74 | 99.52 | 99.37 | 99.21 | 99.52 |
| vegetation | 40.35 | 67.97 | 57.50 | 49.82 | 67.97 |
| look a like | 46.72 | 68.08 | 63.68 | 59.82 | 68.08 |
| background | 99.69 | 99.79 | 99.84 | 99.90 | 99.79 |

### 6.3 PSPNet results

The performance of the PSPNet model on test and validation datasets is summarized in Tables 8 and 9. On the test data, PSPNet achieved an IoU of 63.12 for oil, with an accuracy of 76.48. The F-score coefficient was 77.39, with Precision and Recall values of 78.32 and 76.48, respectively. For the validation data, PSPNet achieved an IoU of 64.54 for oil, with an accuracy of 74.28. The F-score coefficient was 78.45, with Precision and Recall values of 83.1 and 74.28, respectively.

### 6.4 UNet Results

The performance of the UNet model on test and validation datasets is summarized in Tables 10 and 11. On the test data, UNet achieved an IoU of 57.87 for oil, with an accuracy of 73.75. The F-score coefficient was 73.31, with Precision and Recall values of 78.88 and 73.75, respectively. For the validation data, UNet achieved an IoU of 57.9 for oil, with an accuracy of 69.92. The F-score coefficient was 73.34, with Precision and Recall values of 77.11 and 69.92, respectively.

### 6.5 Mask2Former results

The performance of the Mask2Former model on test and validation datasets is summarized in Tables 12 and 13. For the test data, the model achieved an IoU of 66.66 for oil, indicating a good overlap between the predicted and actual oil spill regions. The model's accuracy was 85.00, with a F-score coefficient of 79.99, suggesting a solid segmentation performance. Precision and Recall for oil were 75.54 and 85.00, respectively.

On the validation data, the model was improved significantly, achieving an IoU of 72.69 for oil. The accuracy reached 83.43, with a F-score coefficient of 84.18. Precision and Recall for oil were 84.95 and 83.43, respectively, indicating better segmentation and detection results on the validation set.

### 6.6 Comparison of models

The evaluation of DeepLabV3, DeepLabV3+, PSPNet, UNet, and Mask2Former reveals insights into their respective strengths and weaknesses in oil spill detection on land. Among these, Mask2Former consistently outperformed others, achieving the highest IoU and F-score coefficients, particularly on validation datasets. Its superior performance aligns with prior findings, such as those reported by Cheng *et al.*,[31] where transformer-based architectures demonstrated state-of-the-art results in universal segmentation tasks due to their ability to capture global contextual information effectively.

DeepLabV3+ also performed strongly, especially in generalizing to unseen data, as evidenced by its high accuracy and precision scores. This result is consistent with the findings in Chen *et al.*,[27] where the encoder-decoder structure and depthwise separable convolutions were shown to improve segmentation performance while maintaining the computational efficiency. PSPNet's performance highlights the effectiveness of pyramid pooling in capturing the global context, as first demonstrated by Zhao *et al.*[29] While it achieved competitive results, its precision and recall scores were slightly lower than those of Mask2Former and

DeepLabV3+. This may be attributed to its relative sensitivity to small-scale features, which is less suited for the fine-grained segmentation required in oil spill detection.

UNet, originally designed for biomedical applications,[28] proved to be a robust choice for segmentation tasks with a balanced trade-off between computational cost and accuracy. However, its reliance on local information through skip connections may limit its performance in scenarios requiring strong global context awareness. DeepLabV3, while effective in capturing multi-scale information via atrous spatial pyramid pooling (ASPP), exhibited a moderate performance in this application compared to its successors, DeepLabV3+ and Mask2Former. This observation aligns with the incremental improvements reported in the literature[30] when transitioning from DeepLabV3 to DeepLabV3+.

Our findings demonstrate that while traditional CNN-based models like UNet and DeepLabV3 remain competitive, transformer-based architectures such as Mask2Former are better equipped for complex segmentation tasks involving intricate boundaries and diverse scales. These results fill a gap in the literature by providing a focused comparison of these models in the context of oil spill detection, an application area that has seen limited exploration in prior studies.

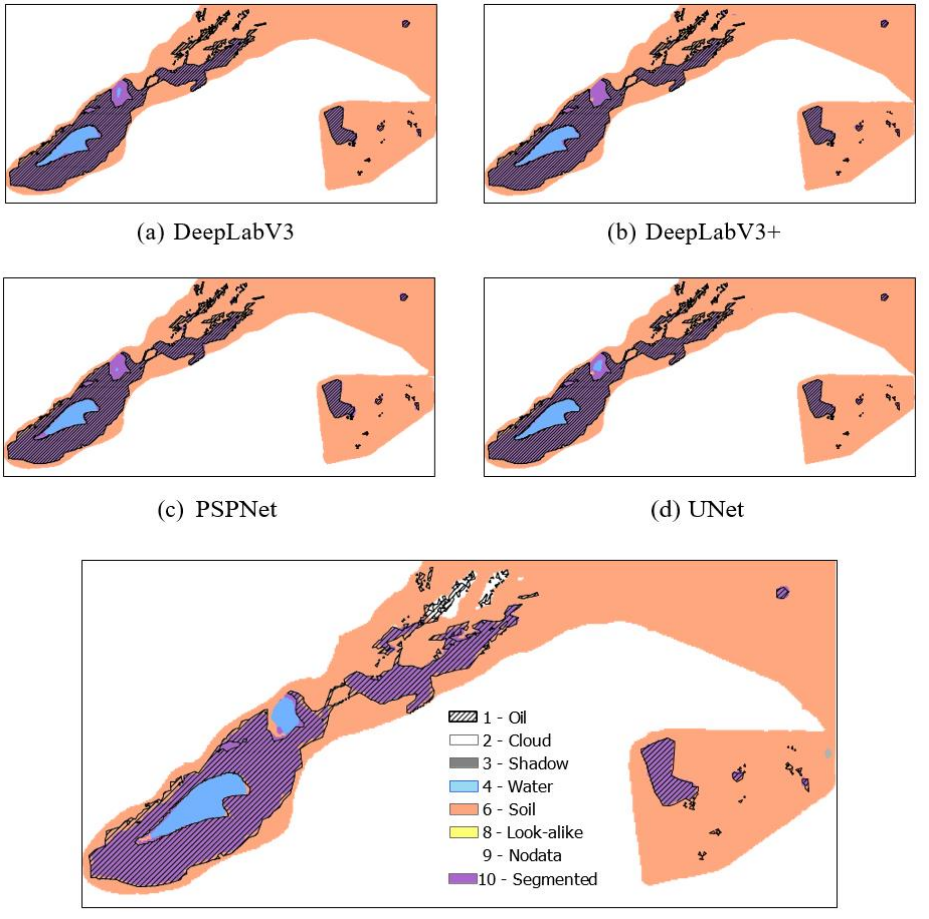
### 6.7 Statistical analysis of model performance

**Table 14.** Standard deviations for key performance metrics across five trials.

| Model | Standard Deviation (mFscore) | Standard Deviation (mIoU) | Standard Deviation (Time per Image) |
|---|---|---|---|
| Mask2Former | 0.0 | 0.0 | 0.2077 |
| DeepLabV3+ | 0.0 | $7.1 \times 10^{-15}$ | 0.0005 |
| DeepLabV3 | 0.0 | 0.0 | 0.0005 |
| UNet | 0.004 | 0.0 | 0.0002 |
| PSPNet | 0.0 | 0.0 | 0.0003 |

To assess the reliability of the findings, we conducted five trials for each model and computed the standard deviation for key performance metrics. The models were evaluated based on the mean F-score (mFscore), mean Intersection over Union (mIoU), and average time per image. Table 14 presents the standard deviations for these metrics.

The table shows that most models exhibited minimal




(a) DeepLabV3 (b) DeepLabV3+

(c) PSPNet (d) UNet

(e) Mask2Former

**Fig. 6** Comparison of segmentation results for different models. Purple color indicates oil segmentation for the corresponding model.

variability in their performance, with standard deviations for the mFscore and mIoU close to zero. The only model with a non-zero standard deviation for mFscore was UNet, which indicates a slight variability in its performance across different trials. The standard deviation for the time per image is consistent across the models, with Mask2Former showing the highest variation, which could be attributed to the increased computational complexity of the model.

Fig. 6 provides a visual comparison of the segmentation results produced by the DeepLabV3, DeepLabV3+, PSP Net, UNet, and Mask2Former models. Each figure illustrates the model's capability to detect and segment oil spill regions, highlighted in red. DeepLabV3 (Fig. 6(a)) utilizes atrous convolution to maintain resolution while capturing multi-scale context. The resulting segmentation shows a moderate accuracy, with oil spill regions detected, but with some boundaries being less precise. DeepLabV3+ (Fig. 6(b) incorporates an encoder-decoder structure and depthwise separable convolutions, leading to an improved segmentation accuracy. This model effectively captures smaller oil spill regions with more refined boundaries, indicating a better handling of spatial details. PSPNet (Fig. 6(c)) uses a pyramid pooling module to gather contextual information at multiple scales. The segmentation results focus on broader areas, effectively identifying large oil spill regions. However, the precision at boundaries may be slightly lower compared to Mask2Former and DeepLabV3+. With its symmetric encoderdecoder architecture and skip connections, UNet (Fig. 6(d)) excels in the precise boundary delineation. However, compared to Mask2Former and DeepLabV3+, UNet falls behind in terms of accuracy and the ability to capture fine details, especially in complex or smaller oil spill regions. Mask2Former (Fig. 6(e)) excels due to its transformer-based architecture that allows for a better global context understanding and refined boundary segmentation. The transformer framework enables more effective feature aggregation across scales, making it highly adept at capturing both large-scale structures and fine details. Mask2Former's precision is particularly evident in the clear delineation of boundaries and the correct identification of smaller oil spill regions, outperforming other models, making it particularly effective for detailed segmentation of oil spills (see comparison result on Fig. 6).

## 7. Discussion

The findings of this study provide significant insights into the detection of oil spills on land using the state-of-the-art deep learning models and annotated datasets. The high-performing models, such as Mask2Former and DeepLabV3+, demonstrated exceptional segmentation accuracy and robustness, making them highly suitable for operational deployment in various environmental settings. This section explores the broader implications of these findings, particularly their potential applications, limitations, and contributions to policymaking and environmental monitoring (Fig.7).

### 7.1 Real-time monitoring applications

The integration of the developed ground truth dataset and high-performing models into real-time monitoring systems presents a promising avenue for practical applications. The accuracy and efficiency of these models make them ideal candidates for deployment on edge devices or cloud-based platforms. Such integration can enable real-time processing of satellite or drone imagery, facilitating a quicker detection of oil spills. This capability would significantly improve response time, enabling immediate containment efforts and minimizing the environmental damage caused by oil spills. Future research should focus on optimizing the computational efficiency of these models to ensure that they meet the performance requirements of real-time systems.

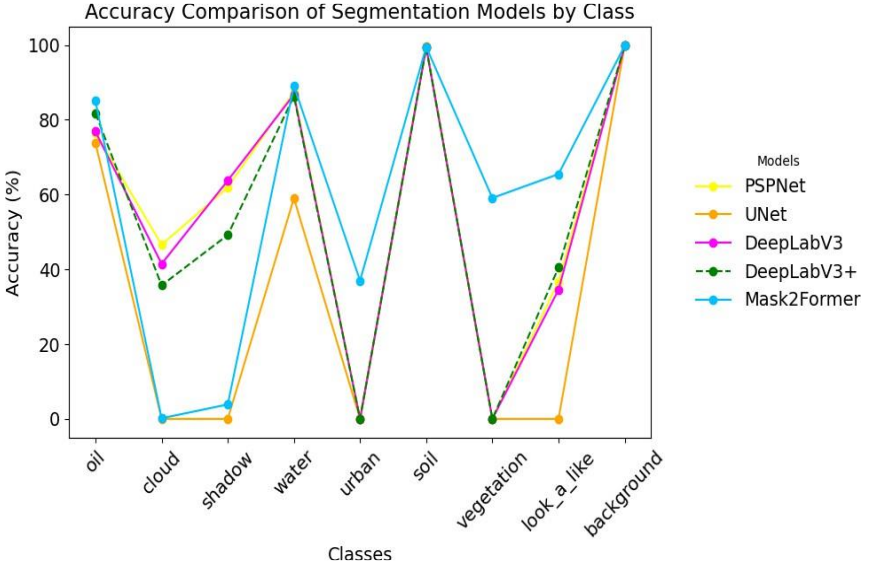



**Fig. 7** Comparison of the accuracy of all models for various classes from the dataset: yellow – PSPNet, orange – UNet, red – DeepLabV3, pink – DeepLabV3+, blue – Mask2Former.

### 7.2 Policy implications

The insights gained from this research have critical implications for policymakers. The annotated ROSID dataset and reliable detection capabilities of the models provide a robust framework for assessing the effectiveness of current oil spill response strategies. Policymakers can leverage these findings to develop data-driven regulations and practices, enhancing environmental protection measures. For instance, the models can be used to identify high-risk areas, inform the placement of response resources, or evaluate the performance of cleanup efforts. Moreover, the public availability of the dataset promotes transparency and benchmarking, enabling regulatory bodies to establish standardized detection protocols and encourage the adoption of advanced monitoring technologies.

### 7.3 Integration with drone and unmanned aerial vehicle (UAV) technologies

The methodologies developed in this study, combined with the advancements in drone and unmanned aerial vehicle (UAV) technologies, present significant opportunities for enhancing oil spill monitoring and response efforts. Drones and UAVs offer the advantages of rapid deployment, real-time data acquisition, and accessibility to remote or hazardous areas, making them an ideal complement to the models and dataset proposed in this work. By equipping drones with high-resolution cameras and sensors, such as multispectral or hyperspectral imagers, real-time data can be collected and processed using models like Mask2Former and DeepLabV3+. The high segmentation accuracy and robustness of these models allow for the precise identification of oil-contaminated regions, even in challenging environments. Additionally, the lightweight nature of drone platforms enables their deployment across vast and inaccessible terrains, providing continuous surveillance and situational awareness during oil spill incidents.

### 7.4 Limitations and future work

While the proposed approach offers substantial advancements, certain limitations warrant further investigation. The computational intensity of transformer-based models like Mask2Former poses challenges for real-time deployment in resource-constrained environments. Future work could explore lightweight adaptations or model compression techniques to address this limitation. Additionally, the study focused on a single dataset. Extending the evaluation to multi-temporal and multi-spectral datasets could improve the models' adaptability to diverse environmental conditions.

Another potential limitation is the inherent variability of land cover and spill patterns across different geographic regions. While the ROSID dataset captures a range of spatial and textural characteristics, further efforts are needed to validate the generalizability of the models to other regions with varying ecological and industrial conditions. Collaborations with local agencies and international organizations could provide access to additional datasets, enabling a broader evaluation of the proposed approach.

Future work will focus on expanding the dataset to include more diverse terrains and varying conditions, further refining the deep learning models to enhance their robustness and generalizability. Additionally, exploring other deep learning architectures and integrating multi-temporal and multi-spectral data could provide even greater accuracy and reliability in oil spill detection. This ongoing research aims to contribute significantly to environmental preservation and disaster management efforts by providing effective tools for timely and accurate oil spill detection.

## 8. Conclusion

In this study, we presented a dataset specifically designed for on-land oil spill detection using Landsat satellite imagery, with Ground-Truth data validated using the high-resolution aerial imagery. The development of this dataset aimed to enhance the accuracy and reliability of oil spill detection methods, providing a robust foundation for training and evaluating the deep learning models. Given the scarcity of publicly accessible datasets on onshore oil spills, we recognized the need to address this gap by publishing our research data (dataset) on GitHub. Alongside the dataset, we have provided benchmarking codes to facilitate accurate and consistent comparisons.[52] We evaluated several state-of-the-art deep learning models, including DeepLabV3, DeepLabV3+, PSPNet, UNet, and Mask2Former, to benchmark the effectiveness of the proposed dataset. Our experiments demonstrated that these models could significantly improve the oil spill detection and segmentation accuracy, with Mask2Former and DeepLabV3+ showing the most promising results due to their advanced architectures and capability to capture detailed spatial information. The results indicate that our dataset substantially enhances the training and validation processes of deep learning models, leading to more precise and efficient oil spill detection. This improvement underscores the potential of combining high-quality groundtruth datasets with advanced deep-learning techniques to address environmental challenges associated with land-based oil spills. The high-performing deep learning models, such as Mask2Former and DeepLabV3+, coupled with the annotated ROSID dataset, present significant opportunities for real-time oil spill monitoring systems. The segmentation accuracy and robustness demonstrated by these models make them suitable for deployment in operational environments where quick decision-making is critical. By integrating these models into edge computing devices or cloud-based systems, real-time processing of satellite or drone imagery can be achieved. This capability would enable a quicker detection of oil spills, facilitating faster containment and remediation efforts, and thereby minimizing environmental damages. Future work will

focus on optimizing the computational efficiency of these models for real-time deployment and testing their performance under diverse environmental conditions to ensure reliability and scalability.

**Acknowledgements**

We would like to thank the following people for helping with this research project: Daniker Chepashev and Alena Yelisseyeva. This work was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No AP14872458 "Development of a methodology for automated space monitoring of oil spills based on neural network technologies").

**Conflict of Interest**

There is no conflict of interest.

**Supporting Information**

Not applicable.

**References**

[1] Z. Asif, Z. Chen, C. An, J. Dong, Environmental impacts and challenges associated with oil spills on shorelines, Journal of *Marine Science and Engineering*, 2022, **10**, 762, doi: 10.3390/jmse10060762.

[2] B. L. Chilvers, K. J. Morgan, B. J. White, Sources and reporting of oil spills and impacts on wildlife 1970–2018, *Environmental Science and Pollution Research*, 2021, **28**, 754-762, doi: 10.1007/s11356-020-10538-0.

[3] K. Bostanbekov, D. Nurseitov, D. Kim, Risk assessment model of technogenic pollution of the environment from oil spill in the northern caspian sea, *International Journal on Advanced Science, Engineering and Information Technology*, 2018, **8**, 37-43, doi: 10.18517/ijaseit.8.1.3190.

[4] R. Lovindeer, S. Mynott, J. Porobic, E. A. Fulton, S. E. Hook, H. Pethybridge, S. E. Allen, D. Latornell, H. N. Morzaria-Luna, J. Melbourne-Thomas, Ecosystem-level impacts of oil spills: a review of available data with confidence metrics for application to ecosystem models, *Environmental Modeling & Assessment*, 2023, **28**, 939-960, doi: 10.1007/s10666-023-09905-1.

[5] I. A. Silva, F. C. G. Almeida, T. C. Souza, K. G. O. Bezerra, I. J. B. Durval, A. Converti, L. A. Sarubbo, Oil spills: impacts and perspectives of treatment technologies with focus on the use of green surfactants, *Environmental Monitoring and Assessment*, 2022, **194**, 143, doi: 10.1007/s10661-022-09813-z.

[6] R. C. Bishop, K. J. Boyle, R. T. Carson, D. Chapman, W. M. Hanemann, B. Kanninen, R. J. Kopp, J. A. Krosnick, J. List, N. Meade, R. Paterson, S. Presser, V. K. Smith, R. Tourangeau, M. Welsh, J. M. Wooldridge, M. DeBell, C. Donovan, M. Konopka, N. Scherer, Putting a value on injuries to natural assets: the BP oil spill, *Science*, 2017, **356**, 253-254, doi: 10.1126/science.aam8124.

[7] L. C. Smith, M. Smith, P. Ashcroft, Analysis of environmental and economic damages from British petroleum's deepwater horizon oil spill, *SSRN Electronic Journal*, 2011, **74**, 563–585, doi:10.2139/ssrn.1653078.

[8] W. P. Sims, W. G. Frailing, Lakeview pool, midway-sunset field, *Journal of Petroleum Technology*, 1950, **2**, 7-18, doi: 10.2118/950007-g.

[9] M.D. Aldakova, A.U. Sabirov, D.U. Sugirov, E.M. Khamanova. Modern methods of processing and disposal of oil-containing waste (in Russian). *Scientific Journal Mechanics and Technologies*, 2021, **1**, 137–143, 2021. doi: 10.55956/YRKQ9448.

[10] National report on the state of the environment and the use of natural resources of the republic of Kazakhstan for 2022 (in Russian), 2023.

[11] National report on the state of the environment and the use of natural resources of the republic of Kazakhstan for 2021 (in Russian), 2022.

[12] D. Kalibatiene, A. Burmakova, V. Smelov, On knowledge-based forecasting approach for predicting the effects of oil spills on the ground, *Digital Transformation*, 2021, 44-56, doi: 10.38086/2522-9613-2020-4-44-56.

[13] F. Löw, K. Stieglitz, O. Diemar, Terrestrial oil spill mapping using satellite earth observation and machine learning: a case study in South Sudan, *Journal of Environmental Management*, 2021, **298**, 113424, doi: 10.1016/j.jenvman.2021.113424.

[14] G. Lassalle, A. Credoz, R. Hédacq, G. Bertoni, D. Dubucq, S. Fabre, A. Elger, Estimating persistent oil contamination in tropical region using vegetation indices and random forest regression, *Ecotoxicology and Environmental Safety*, 2019, **184**, 109654, doi: 10.1016/j.ecoenv.2019.109654.

[15] G. Kaplan, H. Aydinli, A. Pietrelli, F. Mieyeville, V. Ferrara, Oil-contaminated soil modeling and remediation monitoring in arid areas using remote sensing, *Remote Sensing*, 2022, **14**, 2500, doi: 10.3390/rs14102500.

[16] P. Tysiąc, T. Strelets, W. Tuszyńska, The application of satellite image analysis in oil spill detection, *Applied Sciences*, 2022, **12**, 4016, doi: 10.3390/app12084016.

[17] R. Rousso, N. Katz, G. Sharon, Y. Glizerin, E. Kosman, A. Shuster, Automatic recognition of oil spills using neural networks and classic image processing, *Water*, 2022, **14**, 1127, doi: 10.3390/w14071127.

[18] M. Medelbekov, M. Nurtas, Machine learning methods for phishing attacks: survey, 2023 IEEE International Conference on Smart Information Systems and Technologies (SIST). May 4-6, 2023, Astana, Kazakhstan. IEEE, 2023.

[19] D. Sultan, S. Mussiraliyeva, A. Toktarova, M. Nurtas, Z. Iztayev, L. Zhaidakbaeva, L. Shaimerdenova, O. Akhmetova, B. Omarov, Cyberbullying and hate speech detection on Kazakh-language social networks, 2021 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). May 15-17, 2021, NY, USA. IEEE, 2021.

[20] N. Van Hung, L. K. Loc, N. T. T. Huong, N. Van Dung, N. M. Quy, Applied machine learning and deep learning to predict oil and gas production. D. V. K. Huynh, A. M. Tang, D. H. Doan,

P. Watson, eds. Lecture Notes in Civil Engineering. Singapore: Springer Singapore, 2022.
[21] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, F. Fraundorfer, Deep learning in remote sensing: a comprehensive review and list of resources, *IEEE Geoscience and Remote Sensing Magazine*, 2017, **5**, 8-36, doi: 10.1109/MGRS.2017.2762307.
[22] M. Krestenitis, G. Orfanidis, K. Ioannidis, K. Avgerinakis, S. Vrochidis, I. Kompatsiaris, Early identification of oil spills in satellite images using deep cnns. In MultiMedia Modeling: 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8-11, 2019.
[23] M. Nurtas, Z. Baishemirov, V. Tsay, M. Tastanov, Zh. Zhanabekov, Convolutional neural networks as a method to solve estimation problem of acoustic wave propagation in poroelastic media, News of the National Academy of Sciences of the Republic of Kazakhstan. Series: Physics and Mathematics, 2020.
[24] B. Omarov, A. Tursynova, O. Postolache, K. Gamry, A. Batyrbekov, S. Aldeshov, Z. Azhibekova, M. Nurtas, A. Aliyeva, K. Shiyapov, Modified UNet model for brain stroke lesion segmentation on computed tomography images, *Computers, Materials & Continua*, 2022, **71**, 4701-4717, doi: 10.32604/cmc.2022.020998.
[25] A. Zheng, A. Casari. Feature engineering for machine learning: principles and techniques for data scientists. O'Reilly Media, Inc. 2018.
[26] Y. Bengio, A. Courville, P. Vincent, Representation learning: a review and new perspectives, 2012.
[27] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018.
[28] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015.
[29] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, 2017.
[30] L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation, 2017.
[31] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, R. Girdhar, Masked-attention mask transformer for universal image segmentation, 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022.
[32] R. N. Vasconcelos, A. T. C. Lima, C. A. D. Lentini, J. G. V. Miranda, L. F. F. de Mendonça, J. M. Lopes, M. M. M. Santana, E. C. B. Cambuí, D. T. M. Souza, D. P. Costa, S. G. Duverger, W. S. Franca-Rocha, Deep learning-based approaches for oil spill detection: a bibliometric review of research trends and challenges, *Journal of Marine Science and Engineering*, 2023, **11**, 1406, doi: 10.3390/jmse11071406
[33] K. Topouzelis, V. Karathanassi, P. Pavlakis, D. Rokos, Detection and discrimination between oil spills and look-alike phenomena through neural networks, *ISPRS Journal of Photogrammetry and Remote Sensing*, 2007, **62**, 264-270, doi: 10.1016/j.isprsjprs.2007.05.003.
[34] D. Song, Y. Ding, X. Li, B. Zhang, M. Xu, Ocean oil spill classification with RADARSAT-2 SAR based on an optimized wavelet neural network, *Remote Sensing*, 2017, **9**, 799, doi: 10.3390/rs9080799.
[35] H. Guo, G. Wei, J. An, Dark spot detection in SAR images of oil spill using segnet, *Applied Sciences*, 2018, **8**, 2670, doi: 10.3390/app8122670.
[36] Y. Li, Y. Zhang, Z. Yuan, H. Guo, H. Pan, J. Guo, Marine oil spill detection based on the comprehensive use of polarimetric SAR data, *Sustainability*, 2018, **10**, 4408, doi: 10.3390/su10124408.
[37] S.-H. Park, H.-S. Jung, M.-J. Lee, W.-J. Lee, M.-J. Choi, Oil spill detection from PlanetScope satellite image: application to oil spill accident near ras Al zour area, Kuwait in August 2017, *Journal of Coastal Research*, 2019, **90**, 251-260, doi: 10.2112/si90-031.1.
[38] J.-F. Yang, J.-H. Wan, Y. Ma, J. Zhang, Y.-B. Hu, Z.-C. Jiang, Oil spill hyperspectral remote sensing detection based on DCNN with multi-scale features, *Journal of Coastal Research*, 2019, **90**, 332-339, doi: 10.2112/si90-042.1.
[39] D. Blondeau-Patissier, T. Schroeder, G. Suresh, Z. Li, F.I. Diakogiannis, P. Irving, C. Witte and Steven, A.D., Detection of marine oil-like features in Sentinel-1 SAR images by supplementary use of deep learning and empirical methods: Performance assessment for the Great Barrier Reef marine park. *Marine Pollution Bulletin*, 2023, **188**, 114598, doi: 10.1016/j.marpolbul.2023.114598.
[40] T. De Kerf, S. Sels, S. Samsonova, S. Vanlanduit. Oil spill drone: A dataset of drone-captured, segmented RGB images for oil spill detection in port environments, *arxiv preprint arxiv*, 2024, **2402**, 18202, doi: 10.48550/arXiv.2402.18202.
[41] M. Krestenitis, G. Orfanidis, K. Ioannidis, K. Avgerinakis, S. Vrochidis, I. Kompatsiaris, Oil spill identification from satellite images using deep neural networks, *Remote Sensing*, 2019, **11**, 1762, doi: 10.3390/rs11151762.
[42] S. Ahmed, T. ElGharbawi, M. Salah, M. El-Mewafi. Deep neural network for oil spill detection using sentinel-1 data: application to Egyptian coastal regions, *Geomatics, Natural Hazards and Risk*, 2023, **14**, 76-94, doi: 10.1080/19475705.2022.2155998
[43] D. U. K. Abbas, L. E. George, The detection of oil spill onshore using the thermal band of landsat-8, TELKOMNIKA, *Telecommunication Computing Electronics and Control*, 2022, **20**, 383-391, doi: 10.12928/telkomnika.v20i2.22462.
[44] M. S. Ozigis, J. D. Kaduk, C. H. Jarvis, 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, *Environmental Science and Pollution Research*, 2019, **26**, 3621-3635, doi: 10.1007/s11356-018-3824-y.
[45] B.M. Mirkarimova, E.A. Zakarin, L.A. Balakay, L.A. Dedova, and N.B. Tuseeva. Aerospace Monitoring of Oil and Gas Facilities (in Russian), chapter Aerospace environmental monitoring of the Kazakhstan sector of the Caspian Sea to address

issues of the oil and gas industry, pages 301–317. Scientific World Publishing House, Moscow, 2012.
[46] S. N. Dosbergenov. Ecological problems of oil contaminated soils in oil production areas of Western Kazakhstan and ways to solve them (in Russian). Hydrometeorology and Ecology, 2010.
[47] Environmental Systems Research Institute. ArcGIS Desktop: Release 9.3. Redlands, CA, 2009.
[48] I. Balenović, H. Marjanović, D. Vuletić, E. Paladinić, M. Z. Sever and K. Indir, Quality assessment of high density digital surface model over different land cover classes, *Periodicum biologorum*, 2015, **4**, 117, doi: 10.18054/pb.v117i4.3452.
[49] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, R. Moore, Google Earth Engine: Planetary-scale geospatial analysis for everyone, *Remote Sensing of Environment*, 2017, **202**, 18-27, doi: 10.1016/j.rse.2017.06.031.
[50] A. Zelinsky, Learning OpenCV: -computer vision with the OpenCV library (bradski, G.R. et Al.; 2008)[on the shelf, *IEEE Robotics & Automation Magazine*, 2009, **16**, 100, doi: 10.1109/MRA.2009.933612.
[51] R. Rousso, N. Katz, G. Sharon, Y. Glizerin, E. Kosman and A. Shuster, Automatic recognition of oil spills using neural networks and classic image processing, *Water*, 2022, **14**, 1127, doi: 10.3390/w14071127.
[52] D. Nurseitov, G. Abdimanap, A. Abdallah, G. Sagatdinova, L. Balakay, T. Dedova, N. Rametov, A. Alimova. ROSID: Remote Sensing Satellite Data for Oil Spill Detection on Land, 2024.