



# Marine oil spill detection and segmentation in SAR data with two steps Deep Learning framework



Rubicel Trujillo-Acatitla<sup>a</sup>, José Tuxpan-Vargas<sup>a,d,\*</sup>, Cesaré Ovando-Vázquez<sup>b,c,d,\*\*</sup>, Erandi Monterrubio-Martínez<sup>a</sup>

<sup>a</sup> División de Geociencias Aplicadas, Instituto Potosino de Investigación Científica y Tecnológica A.C., Camino a la Presa de San José No. 2055, Colonia Lomas 4ta Sección, San Luis Potosí, San Luis Potosí C.P. 78216, Mexico

<sup>b</sup> División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica A.C., Camino a la Presa de San José No. 2055, Colonia Lomas 4ta Sección, San Luis Potosí, San Luis Potosí C.P. 78216, Mexico

<sup>c</sup> Centro Nacional de Supercómputo (CNS), Instituto Potosino de Investigación Científica y Tecnológica A.C., Camino a la Presa de San José No. 2055, Colonia Lomas 4ta Sección, San Luis Potosí, San Luis Potosí C.P. 78216, Mexico

<sup>d</sup> Cátedras-CONAHCyT, Consejo Nacional de Humanidades, Ciencias y Tecnologías, CDMX 03940, Mexico

## ARTICLE INFO

### Keywords:

Oil spill  
Sentinel-1 SAR  
Deep learning  
Classification  
Segmentation

## ABSTRACT

Marine oil spills pose significant ecological and economic threats worldwide, requiring effective decision-making tools. In this study, the optimal parameters, and configurations for Deep Learning models in oil spill classification and segmentation using Sentinel-1 SAR imagery were identified. First, a new Sentinel-1 image dataset was created. Ninety CNN configurations were explored for classification by varying the number of convolutional layers, filters, hidden layers, and neurons in each layer. For segmentation tasks, MLP and U-Net models were evaluated with variations in convolutional layers, filters, and incorporation of IoU and Focal Loss. The results indicated that a CNN model with six layers, 32 filters, and two hidden layers achieved 99 % classification accuracy. For segmentation, the U-Net model with more layers and filters using Focal Loss achieved 99 % accuracy and 96 % IoU. Therefore, a CNN and U-Net framework was proposed that achieves an overall accuracy of 95 % and an IoU of 90 %.

## 1. Introduction

Oil spills in the marine environment are a persistent global issue, with an estimated annual input of two million tons of oil due to accidents or failures from various sources, including oil platforms, pipelines, refineries, oil tankers, and small vessels (Brekke and Solberg, 2005; Briggs and Briggs, 2018; Huz et al., 2019; Kim et al., 2010). In addition, illegal discharges by different types of vessels seeking economic benefits from waste disposal contribute to the problem of marine pollution (Gauthier et al., 2007; Solberg, 2012). This behavior stems from the perception that, given the low probability of being detected, it is more advantageous to take the risk of paying fines than to incur the costs of compliance (Solberg, 2012).

These incidents can result in significant social, economic, and political consequences (Burgherr, 2007; Daly et al., 2016; Ivshina et al.,

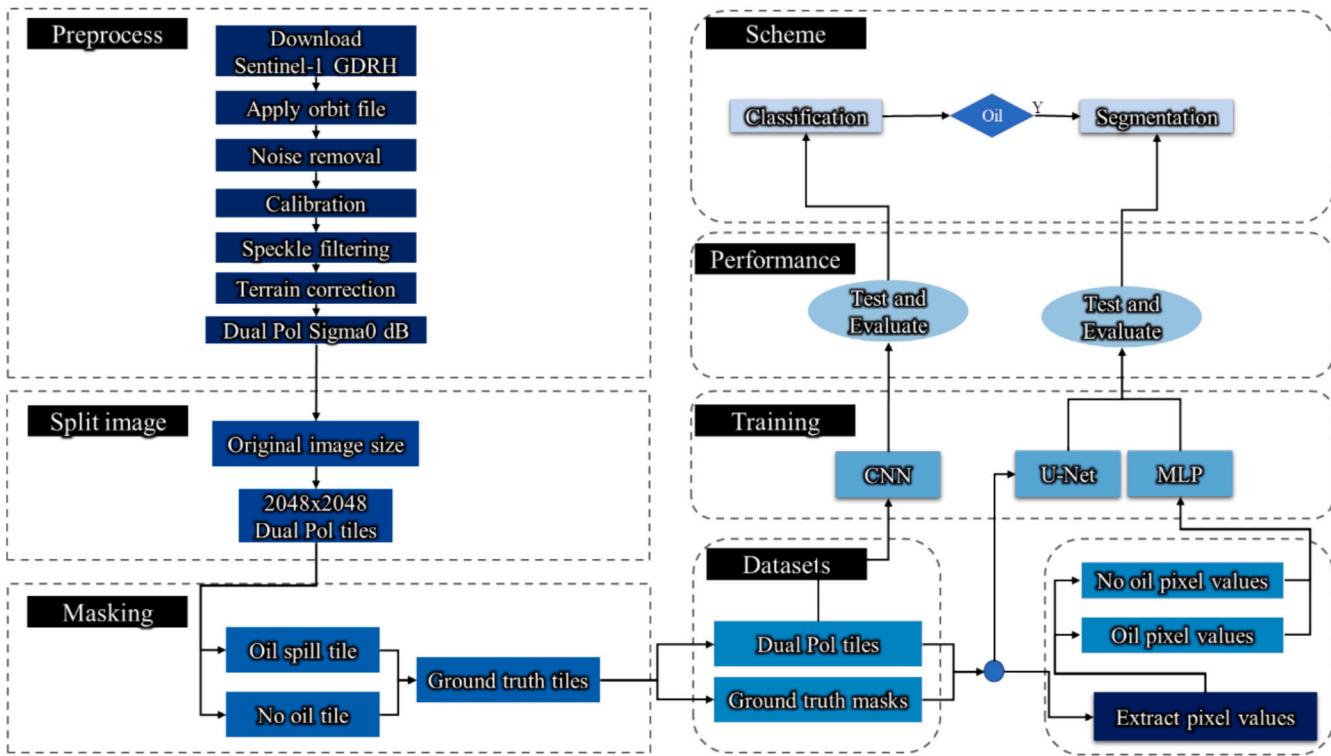
2015). Additionally, they can cause devastating impacts on coastal and marine ecosystems, often requiring decades of restoration (Solberg, 2012; White et al., 2012). Therefore, it is imperative to have an accurate detection method to enhance decision-making competence and effectiveness. This would facilitate a more expedient response to contain the contaminants spread, thereby safeguarding the marine ecosystem and preventing further deterioration (Solberg, 2012; Wang et al., 2023).

Remote sensing technology has been recognized for its efficacy in addressing disasters like oil spills (Fingas and Brown, 2015; Solberg, 2012). Passive sensor technology, which is used for spectral characterizations, has contributed to spill detection models but is constrained to daytime and clear skies (Al-Ruzouq et al., 2020; Leifer et al., 2012; Trujillo-Acatitla et al., 2022). In contrast, synthetic aperture radar (SAR), an active sensor, is excellent for observing the Earth's surface without being affected by atmospheric conditions or solar radiation

\* Correspondence to: J. Tuxpan-Vargas, Cátedras-CONAHCyT, Consejo Nacional de Humanidades, Ciencias y Tecnologías, CDMX 03940, Mexico.

\*\* Correspondence to: C. Ovando-Vázquez, División de Biología Molecular, Instituto Potosino de Investigación Científica y Tecnológica A.C., Camino a la Presa de San José No. 2055. Colonia Lomas 4ta Sección, San Luis Potosí, San Luis Potosí C.P. 78216, Mexico.

E-mail addresses: [jose.tuxpan@ipicyt.edu.mx](mailto:jose.tuxpan@ipicyt.edu.mx) (J. Tuxpan-Vargas), [cesare.ovando@ipicyt.edu.mx](mailto:cesare.ovando@ipicyt.edu.mx) (C. Ovando-Vázquez).



**Fig. 1.** Method Overview. After acquiring Sentinel-1 images, the method involves pre-processing steps including calibration, denoising, and corrections to acquire Sigma0 images measured in decibels (dB). Next, the images are divided into  $2048 \times 2048$  sub-images for both polarizations (VV, VH). The dataset is generated by creating masks for each sub-image. Then, different experiments were performed with DL models such as CNN, MLP, and U-Net, varying the number of convolutional layers, feature maps, hidden layers, and neurons until the models with the best approximations were obtained. This resulted in a two-step process for detecting and segmenting oil spills.

constraints due to wavelength (Kim et al., 2010; Solberg, 2012). In this sense, X (2.4 to 3.75 cm) and C (3.75 to 7.5 cm) bands are optimal for oil spill detection because their scales are similar to those of Bragg waves (Solberg, 2012).

Oil spills disrupt water flow, leading to reduced backscatter and creating a contrast between oil and sea. This contrast is visible as smoother, dark patches in SAR images (Alpers et al., 2019; Fiscella et al., 2000; Li et al., 2018; Solberg, 2012; Wang et al., 2022). To monitor oil spills, dark spots are identified and classified as oil through a thresholding process based on backscatter values (Abou Samra and Ali, 2022, 2024; Babagolimatikolaei, 2022; Chaturvedi et al., 2020; Fiscella et al., 2000; Jones, 2023; Kim et al., 2010; Solberg, 2012; Solberg et al., 2007; Xu et al., 2017).

Nevertheless, certain natural phenomena, such as areas with low wind, algal blooms, or microconvective cells, can result in the formation of dark patches with backscatter values that are similar to those of oil. These are known as look-alikes and can cause false positives in monitoring systems (Alpers et al., 2019). To address this issue, oil spill detection algorithms have begun to incorporate feature extraction and transformations of areas with low backscatter values (Babagolimatikolaei, 2022; Brekke and Solberg, 2005; Fiscella et al., 2000; Xu et al., 2017). In this context, the focus has shifted to the use of artificial intelligence tools, especially Machine Learning (ML), to accurately classify dark patches as oil or not using feature engineering (Cantorna et al., 2019; Dasari et al., 2022; Guo and Zhang, 2014; Ma et al., 2023; Mera et al., 2017; Topouzelis and Psyllos, 2012; Wan and Cheng, 2013; Wang et al., 2023).

However, these methods are insufficient for feature extraction due to their limited nonlinear fitting capabilities, lack of contextual information, and reliance on human intervention, which makes them susceptible to inaccuracies (LeCun et al., 2015; Ma et al., 2023; Wang et al., 2023). To minimize human intervention, deep learning (DL) may be a suitable

option. Unlike ML, DL uses representation-learning methods to progressively transform raw input into higher-level, complex, and abstract representations without explicit design by human engineers (LeCun et al., 2015; Abadi et al., 2022; Vasudevan et al., 2021).

In oil spill analysis, DL models can learn distinctive features for differentiating oil from look-alikes. These key features encompass the regular shape of the spillage, particularly in the case of mobile sources; the quantity and type of oil; the temporal interval between the spillage and image capture, as factors such as wind, temperature, waves, and others can modify and degrade the oil (Abou Samra et al., 2021; Ivshina et al., 2015). The uniformity of both the dark spot and its surroundings is noteworthy. A dark spot situated within a uniform background is more likely to be indicative of oil than the same spot situated within a non-uniform environment (Brekke and Solberg, 2005; Solberg, 2012).

Studies have been conducted to detect oil in SAR images using DL, such as the multilayer perceptron (MLP) (Dasari et al., 2022; Del Frate et al., 2000; Guo and Zhang, 2014; Ma et al., 2014; Mera et al., 2017; Singha et al., 2012, 2013; Taravat and Del Frate, 2012), to more complex algorithms like CNN (Feinauer et al., 2022; Nieto-Hidalgo et al., 2018; Shaban et al., 2021; Zeng and Wang, 2020), Res-Net (Temitope Yeeken et al., 2020; Wang et al., 2023), Mask R-CNN (Temitope Yeeken et al., 2020), UNET (Basit et al., 2022; Dehghani-Dehcheshmeh et al., 2023; Krestenitis et al., 2019; Nieto-Hidalgo et al., 2018; Shaban et al., 2021; Zhai et al., 2022; Zhu et al., 2022), AlexNet (Wang et al., 2021), Faster R-CNN (Huang et al., 2022).

Although efforts have been made, the issue of look-alikes persists. Therefore, a mechanism is needed to eliminate or minimize this aspect (Ma et al., 2023; Temitope Yeeken et al., 2020; Zhang et al., 2020). Several studies have analyzed the potential for a two-stage mechanism using DL to detect and segment oil spills to eliminate false positives caused by look-alikes (Nieto-Hidalgo et al., 2018; Shaban et al., 2021). However, these studies are limited in their generalizability because of

their focus on specific regions and the use of sensors.

The objective of this study is to develop a mechanism that can reduce the number of false positives caused by look-alikes using a DL detection and segmentation framework. Sentinel-1 SAR images were used because they are widely available for monitoring different parts of the Earth's surface. While there are numerous applications of DL in the field of spill monitoring, most of them rely on the transfer learning method, which requires the availability of data with the same dimensions as those used to train the original model. Nevertheless, this is not always feasible option due to the nature of the data or the necessity to create additional channels. The process entails an exhaustive search for the channel that provides the necessary potential for the model to achieve high accuracy, therefore, a pre-trained model was not used. To develop an accurate and reliable framework for detecting and segmenting oil spills, computational experiments were conducted to test various configurations and combinations of parameters for the MLP, CNN, and U-Net model architectures. Only polarizations (channels) present in Sentinel-1 (VV, VH) were used, which eliminates the need for an extra channel and human intervention in the pre-processing part. The results offer guidance on configurable parameters to improve spill detection and contribute to existing knowledge.

## 2. Methods

**Fig. 1** shows the general overview of the methodology. The process starts with the acquisition of Sentinel-1 SAR images, which are then pre-processed to obtain Sigma0 images in decibels (dB). Each image was then divided into  $2048 \times 2048$  pixel size sub-images for both polarizations (VV, VH). A ground truth (gt) mask was created for each sub-image, resulting in a complete dataset of Sentinel-1 Dual-Pol Sigma0 images in dB along with their respective gt. Several experiments were performed using DL models. In the classification part, we used a Convolutional Neural Network (CNN) and experimented with the number of convolutional layers and feature maps (kernels). The segmentation experiment was carried out with different numbers of neurons and hidden layers in a MLP. In addition, the U-Net model was experimented with by varying the number of convolutional layers, features, and loss functions. The models were trained, validated, and tested to develop an effective system for oil spill detection and segmentation using the best performing models.

### 2.1. Sentinel-1 image acquisition, pre-processing, and dataset generation

We obtained 685 full Sentinel-1 SAR level-1 GRD high-resolution images in dual-polarization (VV, VH) IW mode. The data were retrieved from the ASF DAAC website (<https://search.asf.alaska.edu/>), processed by ESA, and distributed worldwide (see Fig. S1). The imaging criterion was based on the presence of oil spills, as reported by the National Oceanic and Atmospheric Administration (NOAA) Raw Incident Data database (<https://incidentnews.noaa.gov/raw/index>) and detections made by the European Maritime Safety Agency (EMSA) CleanSeaNet (<https://www.emsa.europa.eu/>). These databases provide XY coordinates for the spills, which enabled us to confirm the presence of dark areas as oil spills.

Each image was pre-processed using the Sentinel Application Platform SNAP (ESA, 2018), as described by Filippini (2019). Specifically, the Refined Lee Filter ( $5 \times 5$ ) was used for speckle filtering, and images were stored in Sigma0 in dB values with a resolution of 10 m, as shown in Fig. S2. The images were divided into  $2048 \times 2048$  pixel size sub-images for each polarization (VV, VH) using GDAL version 4.1.1 (GDAL/OGR contributors, 2021) and Python Version 3.8 (Van Rossum and Drake, 2009).

A selection of sub-images was made, including those showing oil spills reported by NOAA and EMSA, oil-free images, and images with dark areas that do not correspond to reported spills, indicating look-alikes. Each full image could generate two or more sub-images,

depending on the coverage of the object of interest. A gt mask was generated for each sub-image using the Labkit labeling tool (Arzt et al., 2022). The background was assigned a value of 0, and the area corresponding to the oil spill (foreground) was labelled as 1. For oil-free and look-alike sub-images, the mask contained only a value of 0 for the background. Both were labelled as "No oil" (Fig. S3).

All labelled images were saved together with their gt, resulting in a new dataset of 1200 Oil images and 1200 No Oil images (divided into oil-free and look-alikes), a total of 2400 images with their respective gt, which were used for training and validation. In addition, 450 images with the same dimensions were obtained for model testing. These images consisted of 150 oil images, 150 oil-free images, and 150 look-alike images. Overall, the complete dataset consists of 2850 images with their respective gt and a spatial resolution of 10 m.

For the DL implementations, the sub-images from the training set were used as input to the model. To optimize the computational cost, the sub-images were resized to  $512 \times 512$  for both channels (VV, VH). The image set was split into a training (66 %) and a validation set (34 %) using the Scikit-learn Train Test Split module (Pedregosa et al., 2011). Subsequently, data augmentation was applied to each set to train and validate the models by incorporating rotation, translation, and flip operations. This significantly increased the number of examples.

## 2.2. Deep learning models

### 2.2.1. Fine-tuning CNN for classification task

Convolutional Neural Networks (CNNs) are specialized DL models designed to process matrix data, especially images. These networks map features for classification using convolutional, pooling, and fully connected layers (Fig. S4) (Indolia et al., 2018; O'Shea and Nash, 2015; Yamashita et al., 2018). The convolutional layer extracts features by generating filters and feature maps using movable kernels of various sizes (e.g.,  $3 \times 3$ ,  $5 \times 5$  or  $7 \times 7$ ) (Indolia et al., 2018; Taye, 2023). The outputs of this layer were then passed through a nonlinear activation function, such as ReLU (Albawi et al., 2017; Alzubaidi et al., 2021).

The pooling layer employs operations such as Max-Pooling with different kernel sizes (e.g.,  $2 \times 2$ ) to reduce complexity and parameters by subsampling feature maps while retaining the most significant or relevant information (Alzubaidi et al., 2021; Taye, 2023). After extracting the features, they are flattened and connected to a fully connected network consisting of neurons with nonlinear functions. The output layer contains the same number of neurons as the classes. For a binary problem, the final layer employs a sigmoid activation function, which yields probability values between 0 and 1. In most binary classification problems, the 0.5 cutoff is utilized, whereby values exceeding 0.5 correspond to one class and values below 0.5 to another. Since this is a backpropagation model, an optimization method such as ADAM was used (Fig. S4) (Albawi et al., 2017; Kingma and Ba, 2014).

To determine the optimal architecture of the CNN model, several tests were performed by varying the number of convolutional and pooling layers from 1 to 6 and the number of feature maps (filters) at 16, 32, and 64. A kernel size of  $3 \times 3$  and a pooling size of  $2 \times 2$  were used. For the dense layer, 1 and 2 hidden layers with 10 to 20 neurons each were explored. The ReLU activation function was used for both the convolutional and dense layers, while the sigmoid activation function was used for the output layer.

The accuracy of the model was used to measure the training progress, indicating the proportion of correctly predicted outcomes (Elkan, 2012; Goutte and Gaussier, 2005). The loss function used was Binary Cross-Entropy, which penalizes misclassified data, reducing the loss value and increasing the number of well-classified examples (Ruby and Yendapalli, 2020). A total of 90 CNN models were trained and validated using the TensorFlow library version 2.10 (Abadi et al., 2022) in Python 3.8 (Van Rossum and Drake, 2009).

### 2.2.2. Fine-tuning MLP for semantic segmentation

Semantic segmentation involves labelling at the pixel level, resulting in detailed class maps. Multilayer perceptron's (MLP) are a useful tool for this task due to their flexibility in both simple and complex configurations. They can be used as both ML and DL models (Mo et al., 2022). MLP includes layers for input, hidden, and output. The input layer contains raw data, while the hidden layers process and transform it. The output layer provides a solution to the classification issue, which can be a binary classification problem (Mehlig, 2019). To analyze an image, it is necessary to flatten the matrix data, which results in a vector or tensor. This type of model is trained using the backpropagation mechanism, which utilizes different optimization methods, such as ADAM, loss functions, and metrics (Fig. S5) (Kingma and Ba, 2014).

For the oil spill segmentation task, we selected 50 sub-images from the training and validation image datasets (Section 2.1). We flattened these sub-images, along with their respective masks, to obtain a vector for each channel and *gt*. To address the imbalance in the number of pixels between classes, a random subsampling process was performed using the oil class with fewer samples as the base. The dataset for each channel contained 61,598,228 samples, with equal representation (30,799,114) of the Oil and No Oil classes. The dataset was split into a training set (66 %) and a validation set (34 %). This split was performed using the Train Test Split of Scikit-learn (Pedregosa et al., 2011).

To determine the optimal architecture for the MLP model, a series of experiments were carried out. Initially, a basic model was trained with one input layer and one output layer using the sigmoid activation function (this model can be interpreted as a logistic regression). Then other models were explored with different numbers of hidden layers (one, two, and three layers) and neurons (ranging from 5 to 500). The hyperbolic tangent (*tanh*) activation function, which is a commonly used nonlinear function in this context, was applied to the hidden layers. In the output layer, the sigmoid activation function was used, and the binary cross-entropy loss function was employed. For testing, several images from a given test set (Section 2.1), which were flattened to match the dimensions of the training data. The resulting vectors were then converted back to matrices to match the original image dimensions (Fig. S5).

### 2.2.3. Fine-tuning U-Net for semantic segmentation

In some cases, segmentation tasks require the inclusion of contextual and spatial information in addition to the values present in the images. Models that rely solely on pixel values are unable to achieve this goal, as important contextual information is lost during the flattening process. The U-Net model proposed by Ronneberger et al. (2015) is a contextual information-based model used for image segmentation tasks. Its architecture consists of two parts: contraction and expansion.

The contraction part uses a CNN architecture that doubles the number of features in each convolution operation. Each feature has an activation function and undergoes a pooling operation. In the expansion part, an upsampling process is performed first, followed by concatenation with the corresponding feature map of the contraction part. Then, two convolutional layers were applied each with an activation function. In the final layer, a  $1 \times 1$  convolutional layer is used to map each feature vector to the corresponding class.

To address the challenge of oil spill segmentation, a U-Net model was systematically fine-tuned through a series of experiments. First, the original U-net model (Ronneberger et al., 2015) was used as base, maintaining the number of convolutions, kernel size ( $3 \times 3$ ), number and order of feature maps (64, 128, 256, 512, 1024), and the pooling ( $2 \times 2$ ). For the expansion part, the Conv2DTranspose operation of Keras (Chollet et al., 2015) was implemented with the ReLU activation function and the sigmoid function for the output. The same loss function (Binary Cross-Entropy) and metric (Accuracy) were maintained. The model was trained using two-channel images, specifically considering polarizations (VV, VH), from the designated dataset for training and validation (Section 2.1) and the Keras Normalize module was applied to

**Table 1**

Accuracy scores and loss of the five models with the highest scores. The model with 6 convolutions with 32 filters and two hidden layers with 20 neurons obtained the best results (in bold).

Model	Accuracy (%)		Loss	
	Training	Validation	Training	Validation
Only output 5Conv 32 filters	99.68	98.88	0.015	0.15
One-layer (10) 5Conv 16 filters	99.84	96.47	0.0088	0.1679
One-layer (20) 5Conv 64 filters	99.84	98.53	0.0085	0.0979
Two layers (10) 5Conv 32 filters	99.69	97.35	0.0171	0.0964
Two layers (20) 6Conv 32 filters	<b>99.99</b>	<b>98.24</b>	<b>0.018</b>	<b>0.0573</b>

normalize all images.

The model underwent a modification of the loss function used during training by incorporating the Focal Loss function. The aim was to address the class imbalance between the foreground and background during training of the segmentation task. The Focal Loss function introduces a modulation factor  $(1 - p_i)^\gamma$  to the cross-entropy, which focuses the training process and prevents the overwhelming impact of numerous easy-to-classify negatives, which typically constitute the majority of the loss. The focusing parameter  $\gamma$  can take values between 0 and 5 (Lin et al., 2017).

Furthermore, instead of using Accuracy to monitor the model training, Intersection Over Union (IoU) was used as a metric, since the accuracy tends to be biased by the class imbalance present in the segmentation problems, and they may end up predicting that all pixels are background pixels (Rahman and Wang, 2016). In the other hand, IoU measures the overlap between the prediction and *gt*, yielding a value between 0 and 1, where 1 indicates a perfect match (Niwattanakul et al., 2013; Rahman and Wang, 2016). Later, in the expansion phase, Upsampling2D from Keras (Chollet et al., 2015) was used to modify the usage of the Conv2DTranspose module. The filters were adjusted to 16, 32, 64, 128 and 256. Then, a model was then trained using a feature map scheme with values of 16, 32, 64, 128, 256, 512, and 1024. The input images had dimensions of  $512 \times 512$ . Focal Loss was used as the loss function for these models, and the IoU metric was employed to monitor their training and performance.

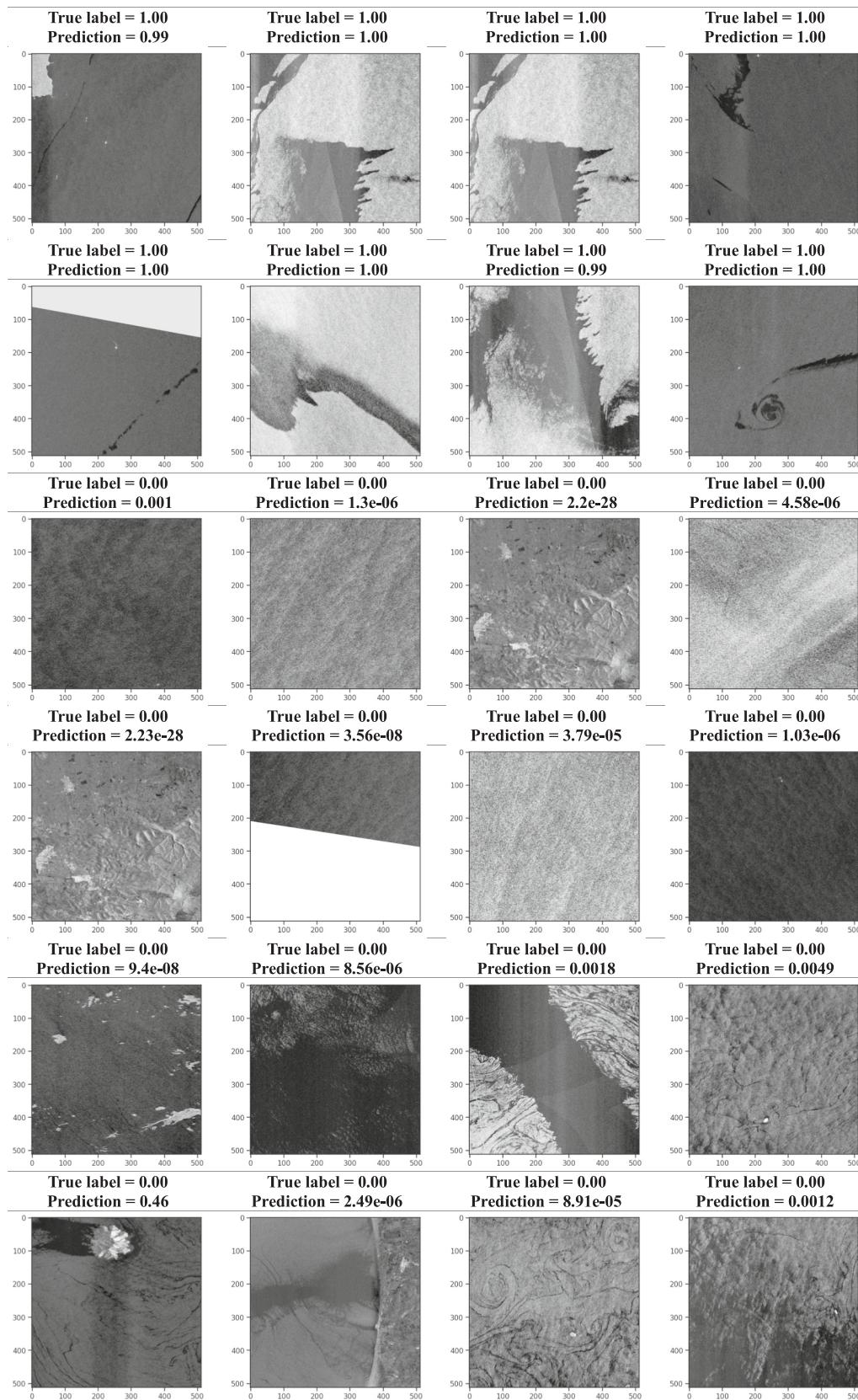
After achieving the best results with the U-Net architecture, we proceeded to train and evaluate other models based on it. These models were trained considering both the architecture and the loss function, as well as the metrics that showed reliable performance with the U-Net model. The models analyzed were Attention U-Net (Att Unet) (Oktay et al., 2018), which has the same architecture as the original U-net, with the addition of attention modules in the expansion phase. The other model was Unet++ (Zhou et al., 2018), which is a nested variant of the U-net model. The TensorFlow library version 2.10 (Abadi et al., 2022) in Python 3.8 (Van Rossum and Drake, 2009) was used to implement the training and validation of the U-Net models.

To evaluate the performance of the models, tests were conducted in different scenarios. The first scenario included only images of oil spills, whereas the second scenario included images of both oil spills and oil-free spills. This second scenario was planned because some Sentinel-1 full images contain areas of land and open sea without any surfactant. The third scenario was more realistic because it included look-alikes. The purpose of these tests was to evaluate the performance of the models and to determine the feasibility of using only the segmentation model or adjusting the need for the classification model to eliminate oil-free and look-alike images before segmentation.

## 3. Results

### 3.1. CNN for oil spill classification

The accuracy and loss for the different CNN configurations



**Fig. 2.** The results of the tests performed with CNN. The top section of each image displays its True label with a value of 0.00 for No Oil and 1.00 for Oil along with its predicted probability value. These values range from 0 to 1, with values  $<0.5$  corresponding to one class and values  $>0.5$  corresponding to another class. Our observations indicate that the images corresponding to Oil have a probability value  $>0.9$ , whereas the images of No oil have probability values between  $2.2e-28$  and 0.46. It should be noted that the No oil class includes look-alike images.

**Table 2**

Accuracy and loss of the configurations with the highest scores. The results indicate that the logistic model achieved the lowest accuracy scores. The model with the highest accuracy scores was the one comprising three hidden layers with 20 neurons, which also had the lowest loss scores. The best scores are shown in bold.

Model	Accuracy (%)		Loss	
	Training	Validation	Training	Validation
Logistic	93.31	93.30	0.1959	0.1959
One layer (500)	94.54	94.40	0.1451	0.1449
Two layers (20)	95.13	95.12	0.1331	0.1332
Three layers (20)	<b>95.14</b>	<b>95.13</b>	<b>0.1329</b>	<b>0.1329</b>

implemented for oil spill classification are illustrated in Fig. S6. Table 1 shows the configurations that yielded the highest scores. These include a configuration with no hidden layers and five convolutional layers with 32 filters; one hidden layer with 10 neurons and five convolutional layers with 16 filters; one hidden layer with 20 neurons and five convolutional layers with 64 filters; two hidden layers with 10 neurons and five convolutional layers with 32 filters; and two hidden layers with 20 neurons and six convolutional layers with 32 filters.

The CNN configuration with six convolutional layers (6Conv) with 32 filters and two hidden layers of 20 neurons achieved the highest scores with a training accuracy of 99.99 %, validation accuracy of 98.24 %, training loss of 0.0018, and validation loss of 0.0573 (Fig. S7).

The tests were carried out using the model with the best scoring architecture (Fig. S7), and the images for the test set (Section 2.1). The oil-free and look-alike images were combined to form No oil class. The

accuracy of the model during the tests was 95 %, whereas the loss was 0.3487. Fig. S8A presents the confusion matrix, indicating an accuracy of 96 % for the No Oil class and 98 % for Oil class. Additionally, in the Receiver operating characteristic (ROC) curve (Fig. S8B), an Area under curve (AUC) of 99 % was achieved, indicating a robust relationship between sensitivity and specificity.

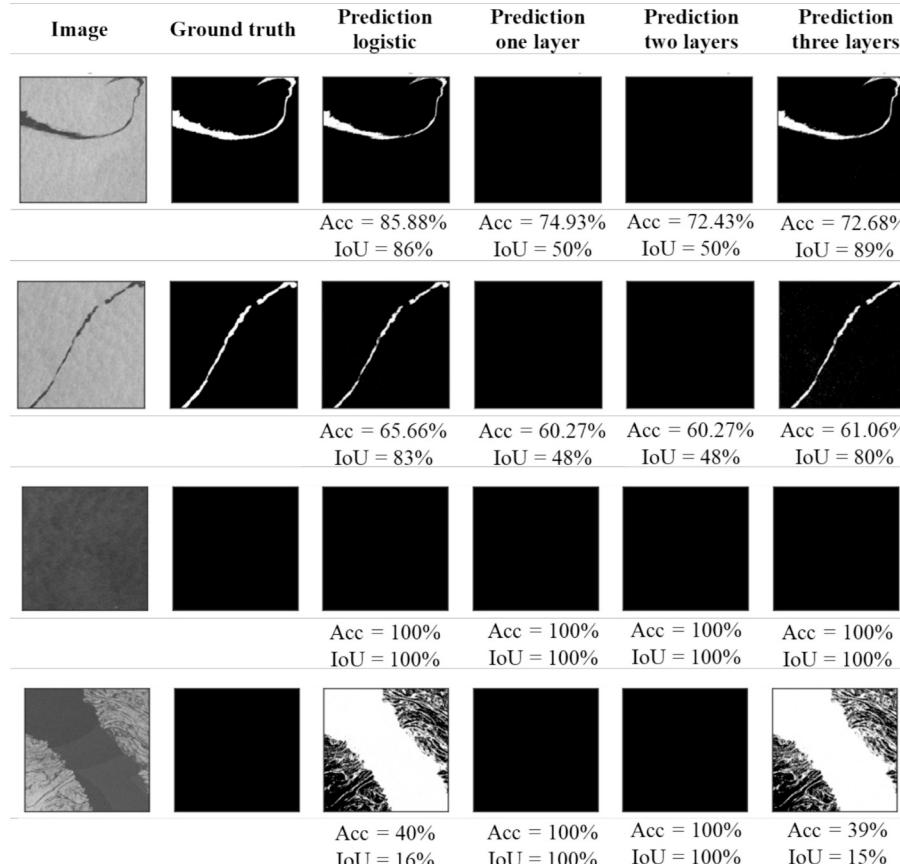
Fig. 2 shows the results of the tests performed on the CNN. “Oil” (True label = 1.00), “No oil” (True label = 0.00), and look-alikes (True label = 0.00) are shown. Probabilities (Prediction) between 0.90 and 1.00 were obtained for the prediction of oil images. For No oil images, the probability values were between 2.2e-28 and 0.46. These results show that the CNN performed well in correctly labelling oil spill images, no oil images, and look-alikes.

### 3.2. Oil spill segmentation models

#### 3.2.1. Multilayer perceptron

Fig. S9 shows the training and validation results of the MLP configurations. The configurations with the best results as a function of the number of hidden layers and neurons were the logistic; one hidden layer and 500 neurons; two hidden layers and 20 neurons in each layer; and three hidden layers and 20 neurons in each layer (Table 2).

Fig. 3 shows the results of the tests conducted using the four configurations presented in Table 2. The output of the MLP generates probability values ranging from 0 to 1. Accordingly, pixels with values  $\geq 0.99$  were classified as oil, while pixels below this threshold were considered as No oil. The first image in the first row corresponds to an oil spill and obtained an accuracy of 85.88 % and an IoU of 86 % for the logistic model, and 72.68 % and 89 % for the three-hidden-layer



**Fig. 3.** Test results for the different configurations. The first column shows the Sentinel-1 SAR image; the second column depicts the gt. Subsequently, the predictions of the different models are presented alongside their respective accuracy and IoU scores. The model outputting probability values necessitated setting the class assignment threshold to 0.99 or greater (binarization). While the logistic model and three-layer hidden model display the best performance, the model segments erroneously in the look-alike image resulting in poor metric scores.

**Table 3**

Different U-net configurations. We started with the original U-net and made successive adjustments, such as changing the loss function to Focal Loss and replacing Conv2DTranspose with Upsampling2D. In addition, filters and convolution depth were varied. The U-net original configuration had the lowest IoU, with scores of 50 % and 40 % for training and validation. However, the configuration that utilized Focal Loss, Upsampling2D, and more convolutional layers and filters obtained the high scores, with a Mean IoU of 96 % in training and 90 % in validation. Highest scores are shown in bold.

Model	Training Mean IoU (%)	Validation Mean IoU (%)
Original U-net	50	40
U-net Modifications:	86	83
- Focal Loss		
U-net Modifications:	91	90
- Focal Loss		
- Upsampling2D		
U-net Modifications:	92	90
- Focal Loss		
- Upsampling2D		
- Filters 16 → 32 → 64 → 128 → 256		
U-net Modifications:	96	90
- Focal Loss		
- Upsampling2D		
- Filters 16 → 32 → 64 → 128 → 256 → 512 → 1024		

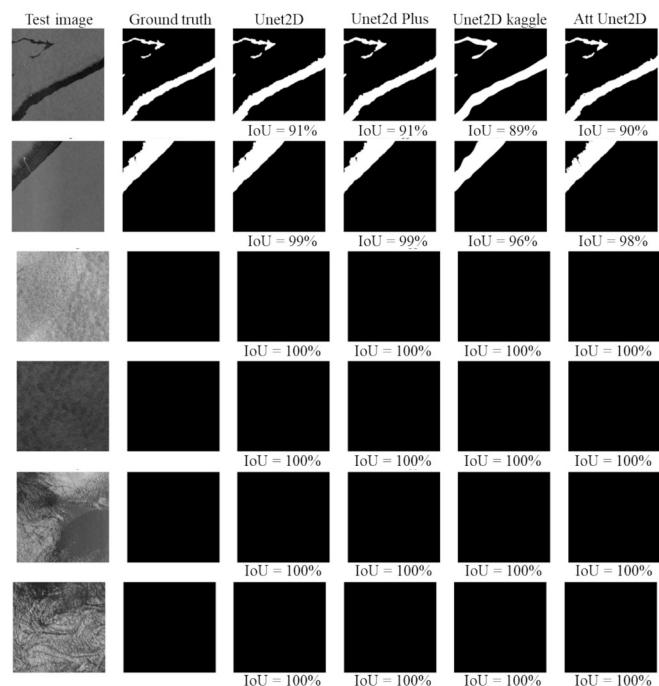
configuration. The one and two hidden layers configurations achieved accuracies of 74.93 % and 72.43 %, respectively, with an IoU of 50 %. The second image, oil spill, gave in an accuracy of 65.66 % and an IoU of 83 % for the logistic, while the three-layers configuration obtained an accuracy of 61.06 % and an IoU of 80 % (Fig. 3).

In the case of the No oil image in the third row, all models do not detect oil, resulting in accuracy and IoU scores of 1.0 (Fig. 3). Regarding the last image in the fourth row, look-alike, it was noted that the one- and two-layer configurations failed to detect oil (accuracy and IoU scores of 100 %), whereas the logistic and three-layer identified as oil (accuracy 40 % and 39 %, and IoU of 16 % and 15 %) (Fig. 3).

After these comparisons, the one- and two-layer configurations were unsuitable and discarded. Despite having high metric values during training and validation (Table 2), they failed to segment anything in the test (Fig. 3). The logistic, despite its simplicity, provided good approximations, as did the three hidden layers. It is worth noting that the accuracy values did not entirely reflect the configuration performance, in some cases the values were low, between 60 % and 70 %, while the IoU scores were higher, >80 %. This observation led us to conclude that IoU is the optimal metric for evaluating segmentation tasks since the accuracy is susceptible to class imbalance due to the difference in the number of pixels between the background and the foreground. In addition, the models incorrectly segmented the image when a look-alike was present, as the pixel values were like the oil. This highlights the need for context- and shape-aware models.

### 3.2.2. U-Net configurations

Table 3 shows the results of the experiments conducted to determine the optimal configurations and combinations of U-Net parameters for the oil spill segmentation task. The configuration with the highest number of convolutional layers and filters (16 → 32 → 64 → 128 → 256 → 512 → 1024), using Focal Loss as a loss function and Upsampling2D in the expansion part, achieved the highest IoU value of 96 % for training.



**Fig. 4.** Test results with oil, No oil, and look-alikes images of the Att\_Unet2D, Unet2D Kaggle, Unet2D, and Unet2D Plus architectures. The first column shows the images used, the second column displays the gt, and the following columns show the segmentation results of each model. The IoU value obtained for the first two rows is shown at the bottom of each image, these examples corresponding to oil spills. In the results of oil-free (third and fourth rows) and look-alikes image (last two rows) all models obtained scores of IoU = 100 %, indicating that none of the models segmented the surface as oil.

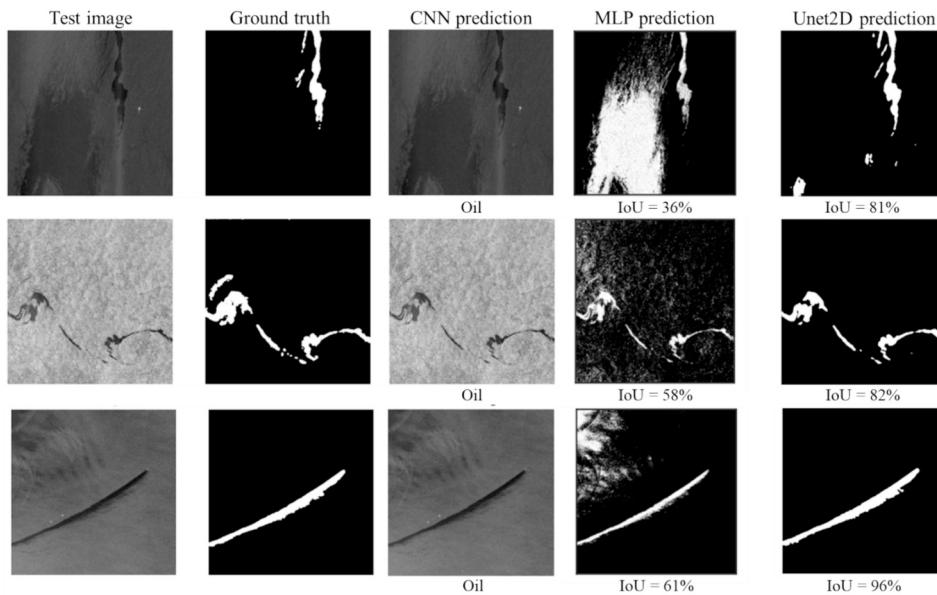
This experiment demonstrates that a configuration with greater depth that extracts a wide variety of complex features can provide a good approximation for oil spill segmentation. Moreover, the experimental results show that the Focal Loss function, which has a focusing parameter, and the Intersection over Union (IoU) are effective alternatives for building oil spill segmentation models. In most images of oil spills, the background tends to dominate, causing saturation of accuracy and loss and high scores even with low model performance.

After identifying the optimal U-Net configuration, tests using variations of the U-Net with the parameters and configuration that yielded high scores were conducted. All architectures achieved a training and validation accuracy of 0.99. Attention U-net (Att\_Unet2D) achieved high scores for Precision (96 %), Recall (95 %), F1 (95%), and IoU (96 %) during training. In validation, Att\_Unet2D, Unet2D Kaggle, and Unet2D showed comparable results with accuracy ranging between 90 % and 91 %, recall between 89 % and 90 %, F1 between 88 % and 89 %, and IoU of 90 %. U-net++ (Unet2D Plus) yielded inferior values across most metrics, except for accuracy, where it showed similar outcomes to the other architectures (Fig. S10).

The comparisons indicate that Att\_Unet2D with the configuration and parameters selected performs better than the rest of the models in metrics, followed by the Unet2D Kaggle model. High scores in training or validation metrics do not always reflect reliable performance in real-world tests. To obtain an accurate assessment of model performance, it is necessary to evaluate the models in more realistic test situations.

### 3.2.3. Testing U-net variations

The U-Net variations were tested with oil spill images, all models performed well in approximating oil spill segmentation (Fig. 4). Achieved IoU values ranging from 89 % to 99 %, the Unet2D kaggle model had the lowest scores, achieving IoU values between 89 % in the first example and 96 % in the second (for more examples see Figs. S11 and



**Fig. 5.** Comparison of segmentation strategies after oil identification in images. The first column contains the test image, the second column contains the gt, and the third column contains the prediction made by the CNN model, which detected the presence of oil in the used examples. The fourth column shows the prediction made by the MLP model with three hidden layers, which achieved IoU values between 36 % and 61 %. The last row displays the prediction made by the Unet2D model, achieving IoU values of 81 %, 82 %, and 96 % for the first, second, and third rows, respectively.

S12).

In addition, oil-free and look-alikes images were used for test (Fig. 4). None of the models were found to segment oil within the scenes, and all models exhibited an IoU value of 100 %, signifying an ideal match between the gt and the prediction (For more examples see Figs. S11 and S12).

Fig. S13 shows the performance of the models in the different test scenarios and datasets (Section 2.2.3). The Unet2D model achieved the highest scores, with IoU values of 90 %, 90 %, and 85 % across all datasets. On the other hand, the Unet2D Plus model obtained the lowest scores, with IoU values of 86 % and 87 % for the first two and third datasets, respectively.

This result highlights the need for look-alike filtering prior to segmentation, otherwise false positives may be generated. These tests show that the U-net model, as a standalone solution, cannot completely distinguish look-alikes from oil spills, as evidenced by its scores. This emphasizes the value of employing a CNN model for this task.

#### 4. Discussion

##### 4.1. Importance of context and form in oil discrimination

To classify and discriminate images with oil, a CNN model was developed, achieving high oil detection accuracy with values up to 99 %. This result is due to the fact that these models can extract features and contextual information from the images (LeCun et al., 2015; Yamashita et al., 2018). When examining oil spills images (Fig. S14), there is a progression from simple features, such as edges and intensity, in the initial convolutions to more complex features as the model goes deeper. Generally, the model has learned a regular linear pattern (Fig. S14), commonly observed in ship spills where the continuous movement prevents oil accumulation in one specific point (Solberg, 2012).

Moreover, this linear pattern may also manifest in platform spills. Despite being fixed sources, wind and water movements can cause the oil to adopt a linear shape, aligning with the direction of the wind and currents (Solberg, 2012). SAR images reveal various aspects, including ships and platforms (Jiang et al., 2023). In the fourth and sixth convolutions, an isolated point is visible, which corresponds to the initiation

point of the spill (Fig. S14).

In the case of oil-free images, the features learned by the model are irregular and somewhat chaotic patterns that represent the nature of water movement and roughness (Kim et al., 2010; Li et al., 2018) (Fig. S14). However, distinguishing water from oil is relatively straightforward, as noted in Figs. S13 and S14, where the patterns and shapes are different, with the spill exhibiting a regular pattern. The challenge lies in distinguishing oil from look-alikes (Alpers et al., 2019; Fingas and Brown, 2018).

Observing the features learned by the model in each of the convolutions for the look-alike image (Fig. S16), the patterns at the starting point are based on the intensity and edge aspects. However, the patterns in the fifth convolution appear chaotic and differ from the spill feature (Fig. S15). The slick shape is important in distinguishing oil spills from look-alikes, and the homogeneity and surroundings of the slick are also useful. Isolated slicks in homogeneous environments are more likely to be oil than the same slicks in heterogeneous environments (Solberg, 2012).

When analyzing this behavior, it was found that although look-alikes and spills may have similar characteristics, they can be distinguished by presenting different visual patterns, in addition to the fact that the spills are mostly isolated compared to look-alikes.

##### 4.2. Segmentation models comparison

Using only the pixel values to train a model, as in the case of MLP models, results in false positives, as shown in Fig. 3. This problem arises because oil spills and natural films cause similar attenuation of the surrounding sea noise, ranging from 0.6 to 13 dB and 0.8 to 11.3 dB, respectively (Solberg et al., 2007). These errors can be reduced by using the CNN model to perform an initial filter and avoid segmentation of look-alike images. However, in some situations, there may be cases where both oil and look-alikes are present in the same image. Therefore, MLP and U-net were compared by conducting experiments with three cases involving the presence of oil and look-alike in the same image (Fig. 5).

The CNN model classified these images into the oil class. Therefore, the next step is to segment the oil spill. The MLP model achieved IoU

**Table 4**

Several studies have been conducted to segment oil spills using SAR imagery. The authors present their proposed model and the metrics employed in their study.

Model	Accuracy (%)	IoU (%)	F1 (%)	Author(s)
ANN	91.6			Singha et al., 2013
Mask R-CNN	96.6			Temitope Yeeken et al., 2020
U-Net with attention	91.2	83.4		Zhu et al., 2022
U-Net	85.61	81.46		Zhu et al., 2022
Conditional Adversarial Networks	99			Li et al., 2021
AlexNet	97–98			Wang et al., 2021
Faster -CNN	89.23			Huang et al., 2022
U-Net		53.79		Krestenitis et al., 2019
SVM-ANN	98.13			Dasari et al., 2022
DeepLabV3+		75.08		Dehghani-Dehcheshmeh et al., 2023
FC-DenseNet		73.94		Dehghani-Dehcheshmeh et al., 2023
U-Net		60.85		Dehghani-Dehcheshmeh et al., 2023
CNN y ViT		78.48		Dehghani-Dehcheshmeh et al., 2023
DeepLabv3+ VisionTransformers	94.45	64.59		Dehghani-Dehcheshmeh et al., 2023
U-Net	94.17	43.73		Dehghani-Dehcheshmeh et al., 2023
Feature Merging Network		46.49		Fan et al., 2021
Dual Attention Encoding Network		85.3		Zhai et al., 2022
Ensemble model con UNET		71.12		Roussou et al., 2022
Two-stages CNN		66.67		Nieto-Hidalgo et al., 2018
U-Net Efficient-net-B3		62.19		de Moura et al., 2022
DeepLabv3 Bilateral Segmentation Network (BiSeNetV2)		96.08		Ma et al., 2023
Multichannel Deep neural network	98.56	97.17		Hasimoto-Beltran et al., 2023
Our work	99	96	95	
Depth U-net with Focal Loss				

values of 36 %, 58 %, and 61 % for the first, second, and third cases, respectively. In contrast, the Unet2D model obtained higher values of 81 %, 82 %, and 96 %.

This comparison emphasizes the importance of using contextual information in oil spill segmentation tasks. The Unet2D model achieves highly accurate oil segmentation, whereas the MLP model tends to produce false positives by mistakenly segmenting similar-looking areas as oil. This behavior is critical because real case images may contain both oil and look-alike substances even when a pre-filtering scheme is applied. Evaluating the model's performance in these scenarios enhances its reliability. This is crucial in a warning or monitoring system.

#### 4.3. U-Net with Focal Loss for enhanced oil spill segmentation

The results show that the Unet2D model (Fig. S17) achieved the

highest scores (Fig. 4). This model incorporates Focal Loss, which penalizes easy-to-classify examples that correspond to the background and prevents them from overwhelming the training detector (Lin et al., 2017). This approach enables the model to focus on challenging or isolated examples, such as images of oil spills, where the background is typically the most prominent class within image (Basit et al., 2022; Lin et al., 2017).

Some studies employed CrossEntropy as a loss function and evaluated the model using accuracy (Table 4). However, these two metrics can present bias and errors in segmentation tasks by exclusively focusing on the majority class to generate a score (Basit et al., 2022; Lin et al., 2017). This is evident from those studies that report high accuracy values ranging from 85 % to 94 %. However, when segmenting objects and assessing them using IoU, the scores reduce and vary from 43 % to 89 % (Table 4).

In contrast, the model performed in this study presents consistent metrics with an accuracy of 99 % and an IoU of 96 %. This consistency can be attributed to the use of Focal Loss as a loss function. This choice is particularly useful for problems with significant class imbalance, like oil spills, as this loss function has been shown to outperform other loss functions in these types of cases (Basit et al., 2022; Lin et al., 2017).

Among the notable works that achieved high accuracy (98.56 %) and IoU (97.17 %) scores, the study conducted by Hasimoto-Beltran et al., 2023 (Table 4) stands out. In their study, IoU was used to monitor training, which is considered to be an effective alternative to semantic segmentation models (Ni Wattanakul et al., 2013; Rahman and Wang, 2016). In the present study, we also employed IoU as a metric to train our segmentation model, which supports its effectiveness in oil spill segmentation.

Hasimoto-Beltran et al. (2023) used 16 images from a single event in their approach. This may introduce bias and limit the generalizability of the model to other events, as demonstrated by their comparative analysis with images from the study of Krestenitis et al. (2019), which resulted in IoU between 54.27 % and 57.13 %.

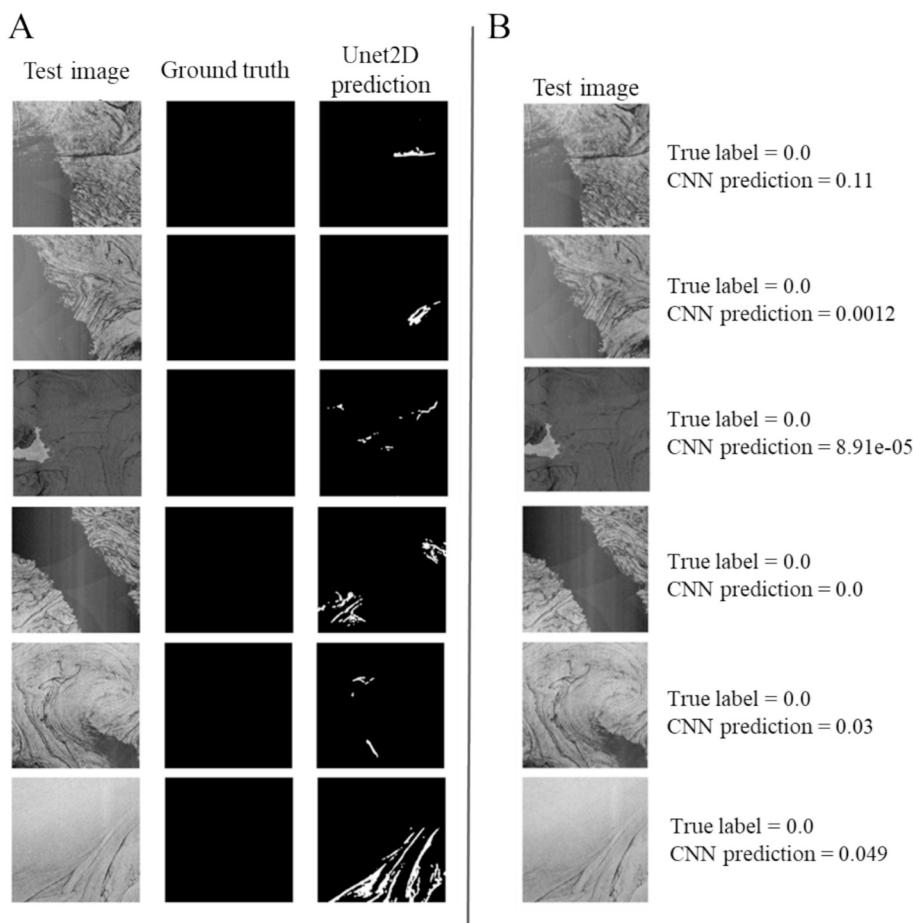
In contrast, a comprehensive and diverse image selection approach was used to Unet2D model (Fig. S1). This strategy captured the wide variability in different types of oil and spill patterns influenced by various physical, chemical, and environmental factors at each site. This approach mitigated the risk of model bias by preventing the model from learning exclusively from one site, thus contributing to enhanced generalizability across diverse oil spill scenarios. In addition, it helped to minimize or eliminate data leakage, which is a common concern in complex datasets (Kaufman et al., 2012). The proposed approach produced consistent metrics and reliable spill segmentation outcomes.

Dehghani-Dehcheshmeh et al. (2023), achieved an accuracy of 94.45 % with 1112 spill images, but only an IoU of 64.59 %. In contrast, our approach, which combines fine-tuning with the use of Focal Loss and IoU, improves the model performance in the oil spill segmentation task compared with other studies. Therefore, these results provide a starting point for combining different strategies to improve the accuracy of oil spill segmentation models.

#### 4.4. Classification and segmentation scheme for oil spills

The Unet2D model developed for oil spill segmentation demonstrates reliable performance in training (IoU = 95 %), validation (IoU = 90 %), and testing (IoU = 0.90 %). However, using only the U-Net model for segmentation would present challenges in look-alikes images, which would affect its performance (Fig. S13). In other words, the model would generate segmentation errors in the presence of look-alikes (Fig. 6), which could lead to false positives.

Testing the CNN model on the same images (Fig. 6A) showed that the resulting prediction probabilities classified them as No Oil images (Fig. 6B). This observation highlights the importance of employing the two-stage scheme for oil spill detection and segmentation. This approach minimizes the possibility of generating false positives.



**Fig. 6.** Examples of images with look-alikes. A) Results for Unet2D prediction, the corresponding mask of each image is observed, together with the prediction, which displays segmentation errors. B) Results from the CNN model. The true label for these images is 0, indicating they belong to the No Oil class. The predicted values are 0 or close to 0, correctly classifying the images in the No Oil class.

Combining both models - CNN for initial detection and Unet2D for subsequent segmentation - creates a more robust strategy for accurate oil spills monitoring.

Based on these findings, a method is proposed for classification, segmentation, area estimation, and location of oil spills using Sentinel-1 imagery is proposed (Fig. S18). The proposed method comprises the following steps: 1) Acquisition and pre-processing Sentinel-1 images; 2) Images splitting; 3) Detection using the proposed CNN model; 4) Segmenting with the proposed U-net model (note that segmentation is performed only if oil is present in the image). If oil is present in multiple sub-images, 5) a mosaic of all sub-images containing oil is created, 6) the spill area is estimated, and finally 7) the spill location is determined.

Previous studies have tested a two-stage mechanism with positive outcomes but reported segmentation errors (Nieto-Hidalgo et al., 2018; Shaban et al., 2021). However, our model has led to considerable improvements in segmentation through fine-tunning and applying the Focal Loss function. This has contributed to a more accurate identification of oil spills. The approach presented in this study provides a more efficient technique for detecting and measuring oil spills in monitoring and response systems.

## 5. Considerations

The development and implementation of ML and DL models oil spill detection and segmentation using SAR imagery has shown significant potential. However, several factors to be considered to enhance the accuracy of the models. First, it is important to increase the amount and global distribution of data to enhance the variability that the model can

learn.

The continuous degradation of oil must be considered from the moment of its release. As the oil undergoes decomposition and weathering, which makes it difficult to characterize and over time it can appear more like surfaces of a biological origin (Solberg, 2012). This degradation is important because it allows the models to capture characteristics that are present at the edges of the spill zone, where the oil can vary in concentration, volume, and even combine to form emulsions. Moreover, all of these variations depend on the source and location of the spill (Gade et al., 1998; Kim et al., 2010; Solberg, 2012).

It is necessary to consider the location of the main sources, oil facilities like wells, platforms, and vessels. A monitoring system should be developed and focused on these facilities. While this might limit the possibility of exploring areas with potential natural spills, it can increase the precision and effectiveness of the monitoring system in place (Solberg, 2012; Solberg et al., 2007; Topouzelis et al., 2015).

The temporal resolution of satellites is an important aspect to consider when developing a monitoring mechanism because of the need for frequent data availability. However, the Sentinel-1 SAR revisit time is five days, which limits real-time monitoring. Nevertheless, it benefits from its global availability.

Finally, when developing the model, it is essential to consider the advantages and disadvantages of both ML and DL and conduct extensive parameter, metric, and depth searches in the architectures. These searches may be useful for solving complex and specific problems that sometimes require the development of a unique model or perspective.

## 6. Conclusions

The search for the combination of parameters and architecture (fine-tuning) of the models allowed us to find the combination that produced the best approximations in the classification and segmentation of oil spills. The CNN model with the selected parameters and configuration provided good approximations for oil spill classification and proved to be a useful first filter to eliminate false positives.

For the segmentation task, models trained on pixel values alone are confused when presented with look-alike images because this type of model does not consider the spatial and contextual information that is important to differentiate oil from other surfaces. Because of the unbalanced class problem, conventional metrics, and loss functions, such as accuracy and cross-entropy, proved to be inadequate for oil spill segmentation. A fine-tuned U-net model with Focal Loss and IoU increases the certainty of spill segmentation. A comparison of U-net variations provides an overview of their advantages and disadvantages in oil spill segmentation. This will allow us to determine the method that gives the best results for the oil spill segmentation task.

The use of a two-stage approach employing DL techniques for oil spill detection and segmentation provides a good approximation. It also provides reliability for oil spill monitoring. This is because the possibility of a false positive in the presence of a look-alike is reduced.

Through the search of parameters and configurations of DL models, this work generates a perspective and gives an idea of the possibilities of combinations that help to generate a perspective with greater certainty for the classification and segmentation of oil spills. It also contributes to the proposal of a monitoring mechanism using SAR images and DL. This tool will facilitate informed decision-making in the case of such disasters.

## CRediT authorship contribution statement

**Rubicel Trujillo-Acatitla:** Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Investigation, Formal analysis, Data curation. **José Tuxpan-Vargas:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Investigation, Conceptualization. **Cesaré Ovando-Vázquez:** Writing – review & editing, Writing – original draft, Supervision, Software, Project administration, Methodology, Investigation, Formal analysis. **Erandi Monterrubio-Martínez:** Writing – review & editing, Visualization, Resources, Methodology, Data curation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgements

This work was supported by the Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT, <https://conahcyt.mx/>), through grant CVU: 699765. We thank the División de Geociencias Aplicadas ([https://ipicyt.edu.mx/Geociencias\\_Aplicadas/areas\\_geociencias\\_aplicadas.php](https://ipicyt.edu.mx/Geociencias_Aplicadas/areas_geociencias_aplicadas.php)) and the Centro Nacional de Supercómputo (<https://cns.ipicyt.edu.mx/>) of the Instituto Potosino de Investigación Científica y Tecnológica A.C. (<https://www.ipicyt.edu.mx/index.php>) for the computational grant TKII-R2018-COV1.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.marpolbul.2024.116549>.

## References

- (ESA), E.S.A. 2018. SNAP - ESA Sentinel Application Platform.
- Abadi, Martín, Agarwal, Ashish, Barham, Paul, Brevdo, Eugene, Chen, Zhifeng, Citro, Craig, Corrado, Greg S., Davis, Andy, Dean, Jeffrey, Devin, Matthieu, Ghemawat, Sanjay, Goodfellow, Ian, Harp, Andrew, Irving, Geoffrey, Isard, Michael, Jozefowicz, Rafal, Jia, Yangqing, Kaiser, Lukasz, Kudlur, Manjunath, Levenberg, Josh, Mané, Dan, Schuster, Mike, Monga, Rajat, Moore, Sherry, Murray, Derek, Olah, Chris, Shlens, Jonathon, Steiner, Benoit, Sutskever, Ilya, Talwar, Kunal, Tucker, Paul, Vanhoucke, Vincent, Vasudevan, Vijay, Viégas, Fernanda, Vinyals, Oriol, Warden, Pete, Wattenberg, Martin, Wicke, Martin, Yu, Yuan, Zheng, Xiaoqiang, 2022. TensorFlow: Large-scale Machine Learning on Heterogeneous Systems.
- Abou Samra, R.M., Ali, R.R., 2022. Monitoring of oil spill in the offshore zone of the Nile Delta using Sentinel data. Mar. Pollut. Bull. 179, 113718 <https://doi.org/10.1016/j.marpolbul.2022.113718>.
- Abou Samra, R.M., Ali, R.R., 2024. Tracking the behavior of an accidental oil spill and its impacts on the marine environment in the Eastern Mediterranean. Mar. Pollut. Bull. 198, 115887 <https://doi.org/10.1016/j.marpolbul.2023.115887>.
- Abou Samra, R.M., El-Gammal, M., Eissa, R., 2021. Oceanographic factors of oil pollution dispersion offshore the Nile Delta (Egypt) using GIS. Environ. Sci. Pollut. Res. 28, 25830–25843. <https://doi.org/10.1007/s11356-021-12570-0>.
- Albawi, S., Mohammed, T.A., Al-Zawi, S., 2017. Understanding of a convolutional neural network. In: 2017 International Conference on Engineering and Technology (ICET), pp. 1–6. <https://doi.org/10.1109/ICEngTechnol.2017.8308186>.
- Alpers, W., Liu, A.K., Wu, S.Y., 2019. Satellite remote sensing SAR. In: Encyclopedia of Ocean Sciences. Academic Press, pp. 429–442. <https://doi.org/10.1016/B978-0-12-409548-9.11615-X>.
- Al-Ruzouq, R., Gibril, M.B.A., Shanableh, A., Kais, A., Hamed, O., Al-Mansoori, S., Khalil, M.A., 2020. Sensors, features, and machine learning for oil spill detection and monitoring: a review. Remote Sens. <https://doi.org/10.3390/rs12203338>.
- Alzubaidi, L., Zhang, J., Humaidi, A.J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M.A., Al-Amidie, M., Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J. Big Data 8, 53. <https://doi.org/10.1186/s40537-021-00444-8>.
- Arzt, M., Deschamps, J., Schmid, C., Pietzsch, T., Schmidt, D., Tomancak, P., Haase, R., Jug, F., 2022. LABKIT: labeling and segmentation toolkit for big image data. Front. Comput. Sci. 4 <https://doi.org/10.3389/fcomp.2022.777728>.
- Babagolimatiokalei, J., 2022. Monitoring of oil slicks in the Persian Gulf using Sentinel 1 images. J. Ocean Eng. Sci. <https://doi.org/10.1016/j.joes.2022.05.029>.
- Basit, A., Siddique, M.A., Bhatti, M.K., Sarfraz, M.S., 2022. Comparison of CNNs and vision transformers-based hybrid models using gradient profile loss for classification of oil spills in SAR images. Remote Sens. 14 <https://doi.org/10.3390/rs14092085>.
- Brekke, C., Solberg, A.H.S., 2005. Oil spill detection by satellite remote sensing. Remote Sens. Environ. 95, 1–13. <https://doi.org/10.1016/j.rse.2004.11.015>.
- Briggs, I.L., Briggs, B.C., 2018. Petroleum industry activities and human health. In: The Political Ecology of Oil and Gas Activities in the Nigerian Aquatic Ecosystem. Academic Press, pp. 143–147. <https://doi.org/10.1016/B978-0-12-809399-3.00010-0>.
- Burgherr, P., 2007. In-depth analysis of accidental oil spills from tankers in the context of global spill trends from all sources. J. Hazard. Mater. 140, 245–256. <https://doi.org/10.1016/j.jhazmat.2006.07.030>.
- Cantorna, D., Dafonte, C., Iglesias, A., Arcay, B., 2019. Oil spill segmentation in SAR images using convolutional neural networks. A comparative analysis with clustering and logistic regression algorithms. Appl. Soft Comput. 84, 105716 <https://doi.org/10.1016/j.asoc.2019.105716>.
- Chaturvedi, S.K., Banerjee, S., Lele, S., 2020. An assessment of oil spill detection using Sentinel 1 SAR-C images. J. Ocean Eng. Sci. 5, 116–135. <https://doi.org/10.1016/j.joes.2019.09.004>.
- Chen, Y., Sun, Y., Yu, W., Liu, Y., Hu, H., 2022. A novel lightweight bilateral segmentation network for detecting oil spills on the sea surface. Mar. Pollut. Bull. 175, 113343 <https://doi.org/10.1016/j.marpolbul.2022.113343>.
- Chollet, F., et al., 2015. Keras.
- Daly, K.L., Passow, U., Chanton, J., Hollander, D., 2016. Assessing the impacts of oil-associated marine snow formation and sedimentation during and after the Deepwater Horizon oil spill. Anthropocene 13, 18–33. <https://doi.org/10.1016/j.ancene.2016.01.006>.
- Dasari, K., Anjaneyulu, L., Nadimikeri, J., 2022. Application of C-band sentinel-1A SAR data as proxies for detecting oil spills of Chennai, East Coast of India. Mar. Pollut. Bull. 174, 113182 <https://doi.org/10.1016/j.marpolbul.2021.113182>.
- de Moura, N.V.A., de Carvalho, O.L.F., Gomes, R.A.T., Guimarães, R.F., de Carvalho Júnior, O.A., 2022. Deep-water oil-spill monitoring and recurrence analysis in the Brazilian territory using Sentinel-1 time series and deep learning. Int. J. Appl. Earth Obs. Geoinf. 107, 102695 <https://doi.org/10.1016/j.jag.2022.102695>.
- Dehgani-Dehcheshmeh, S., Akoondzadeh, M., Homayouni, S., 2023. Oil spills detection from SAR Earth observations based on a hybrid CNN transformer networks. Mar. Pollut. Bull. 190, 114834 <https://doi.org/10.1016/j.marpolbul.2023.114834>.

- Del Frate, F., Petrocchi, A., Lichtenegger, J., Calabresi, G., 2000. Neural networks for oil spill detection using ERS-SAR data. *IEEE Trans. Geosci. Remote Sens.* 38, 2282–2287. <https://doi.org/10.1109/36.868885>.
- Elkan, C., 2012. Evaluating Classifiers. Univ. San Diego, California, pp. 1–11. Retrieved [01-11-2012] from <http://cseweb.ucsd.edu/~elkanB250>.
- Fan, Y., Rui, X., Zhang, G., Yu, T., Xu, X., Poslad, S., 2021. Feature merged network for oil spill detection using SAR images. *Remote Sens.* 13 <https://doi.org/10.3390/rs13163174>.
- Feinauer, D.M., Latif, G., Alenazy, A.M., Tayem, N., Alghazo, J., Alzubaidi, L., 2022. Oil spill identification using deep convolutional neural networks. In: 2022 14th International Conference on Computational Intelligence and Communication Networks (CICN), pp. 240–245. <https://doi.org/10.1109/CICN56167.2022.10008373>.
- Filipponi, F., 2019. Sentinel-1 GRD preprocessing workflow. In: Proceedings 18. <https://doi.org/10.3390/ECRS-3-06201>.
- Fingas, M., Brown, E.C., 2015. Oil spill remote sensing. In: Fingas, M. (Ed.), *Handbook of Oil Spill Science and Technology*. John Wiley & Sons, Inc.
- Fingas, M., Brown, C.E., 2018. A review of oil spill remote sensing. *Sensors*. <https://doi.org/10.3390/s18010091>.
- Fiscella, B., Giacaspri, A., Nirchio, F., Pavese, P., Trivero, P., 2000. Oil spill detection using marine SAR images. *Int. J. Remote Sens.* 21, 3561–3566. <https://doi.org/10.1080/014311600750037589>.
- Gade, M., Alpers, W., Hühnerfuss, H., Masuko, H., Kobayashi, T., 1998. Imaging of biogenic and anthropogenic ocean surface films by the multifrequency/multipolarization SIR-C/X-SAR. *J. Geophys. Res. Ocean.* 103, 18851–18866. <https://doi.org/10.1029/97JC01915>.
- Gauthier, M.-F., Weir, L., Ou, Z., Arkett, M., De Abreu, R., 2007. Integrated satellite tracking of pollution: a new operational program. In: 2007 IEEE International Geoscience and Remote Sensing Symposium, pp. 967–970. <https://doi.org/10.1109/IGARSS.2007.4422960>.
- GDAL/OGR contributors, 2021. {GDAL/OGR} Geospatial Data Abstraction software Library. <https://doi.org/10.5281/zenodo.5884351>.
- Goutte, C., Gaussier, E., 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In: Losada, D.E., Fernández-Luna, J.M. (Eds.), *Advances in Information Retrieval*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 345–359.
- Guo, Y., Zhang, H.Z., 2014. Oil spill detection using synthetic aperture radar images and feature selection in shape space. *Int. J. Appl. Earth Obs. Geoinf.* 30, 146–157. <https://doi.org/10.1016/j.jag.2014.01.011>.
- Hasimoto-Beltran, R., Canul-Ku, M., Díaz Méndez, G.M., Ocampo-Torres, F.J., Esquivel-Trava, B., 2023. Ocean oil spill detection from SAR images based on multi-channel deep learning semantic segmentation. *Mar. Pollut. Bull.* 188, 114651 <https://doi.org/10.1016/j.marpolbul.2023.114651>.
- Huang, X., Zhang, B., Perrie, W., Lu, Y., Wang, C., 2022. A novel deep learning method for marine oil spill detection from satellite synthetic aperture radar imagery. *Mar. Pollut. Bull.* 179, 113666 <https://doi.org/10.1016/j.marpolbul.2022.113666>.
- Huz, R., Lastra, M., López, J., 2019. Other environmental health issues: oil spill. In: Encyclopedia of Environmental Health. Elsevier, pp. 792–796. <https://doi.org/10.1016/B978-0-12-409548-9.11156-X>.
- Indola, S., Goswami, A.K., Mishra, S.P., Asopa, P., 2018. Conceptual understanding of convolutional neural network - a deep learning approach. *Procedia Comput. Sci.* 132, 679–688. <https://doi.org/10.1016/j.procs.2018.05.069>.
- Ivshina, I.B., Kuyukina, M.S., Krivoruchko, A.V., Elkin, A.A., Makarov, S.O., Cunningham, C.J., Peshkur, T.A., Atlas, R.M., Philp, J.C., 2015. Oil spill problems and sustainable response strategies through new technologies. *Environ. Sci. Process Impacts* 17, 1201–1219. <https://doi.org/10.1039/C5EM00070J>.
- Jiang, Q., Ji, M., Wang, J., Sun, P., 2023. Remote sensing methods for striped marine oil spill detection in narrow ship channels. *Ocean Eng.* 289, 116162 <https://doi.org/10.1016/j.oceaneng.2023.116162>.
- Jones, C.E., 2023. An automated algorithm for calculating the ocean contrast in support of oil spill response. *Mar. Pollut. Bull.* 191, 114952 <https://doi.org/10.1016/j.marpolbul.2023.114952>.
- Kaufman, S., Rossel, S., Perlrich, C., Stitelman, O., 2012. Leakage in data mining. *ACM Trans. Knowl. Discov. Data* 6, 1–21. <https://doi.org/10.1145/2382577.2382579>.
- Kim, D., Moon, W.M., Kim, Y.-S., 2010. Application of TerraSAR-X data for emergent oil-spill monitoring. *IEEE Trans. Geosci. Remote Sens.* 48, 852–863. <https://doi.org/10.1109/TGRS.2009.2036253>.
- Kingma, D.P., Ba, J., 2014. Adam: a method for stochastic optimization. In: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. <https://doi.org/10.48550/arXiv.1412.6980>.
- Krestenitis, M., Orfanidis, G., Ioannidis, K., Avgerinakis, K., Vrochidis, S., Kompatsiaris, I., 2019. Oil spill identification from satellite images using deep neural networks. *Remote Sens.* 11 <https://doi.org/10.3390/rs11151762>.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. <https://doi.org/10.1038/nature14539>.
- Leifer, I., Lehr, W.J., Simecek-Beatty, D., Bradley, E., Clark, R., Dennison, P., Hu, Y., Matheson, S., Jones, C.E., Holt, B., Reif, M., Roberts, D.A., Svejkovsky, J., Swayze, G., Wozencraft, J., 2012. State of the art satellite and airborne marine oil spill remote sensing: application to the BP Deepwater Horizon oil spill. *Remote Sens. Environ.* 124, 185–209. <https://doi.org/10.1016/j.rse.2012.03.024>.
- Li, Y., Yang, X., Ye, Y., Cui, L., Jia, B., Jiang, Z., Wang, S., 2018. In: Yuan, H., Geng, J., Liu, C., Bian, F., Surapunt, T. (Eds.), *Detection of Oil Spill Through Fully Convolutional Network BT - Geo-spatial Knowledge and Intelligence*. Springer Singapore, Singapore, pp. 353–362.
- Li, Y., Lyu, X., Frery, A.C., Ren, P., 2021. Oil spill detection with multiscale conditional adversarial networks with small-data training. *Remote Sens.* 13 <https://doi.org/10.3390/rs13122378>.
- Lin, T.-Y., Goyal, P., Girshick, R.B., He, K., Dollár, P., 2017. Focal Loss for Dense Object Detection. CoRR abs/1708.0.
- Ma, Y., Zeng, K., Zhao, C., Ding, X., He, M., 2014. Feature selection and classification of oil spills in SAR image based on statistics and artificial neural network. In: 2014 IEEE Geoscience and Remote Sensing Symposium, pp. 569–571. <https://doi.org/10.1109/IGARSS.2014.6946486>.
- Ma, X., Xu, J., Pan, J., Yang, J., Wu, P., Meng, X., 2023. Detection of marine oil spills from radar satellite images for the coastal ecological risk assessment. *J. Environ. Manag.* 325, 116637 <https://doi.org/10.1016/j.jenvman.2022.116637>.
- Mehlig, B., 2019. Artificial Neural Networks. CoRR abs/1901.0.
- Mera, D., Bolon-Canedo, V., Cotos, J.M., Alonso-Betanzos, A., 2017. On the use of feature selection to improve the detection of sea oil spills in SAR images. *Comput. Geosci.* 100, 166–178. <https://doi.org/10.1016/j.cageo.2016.12.013>.
- Mo, Y., Wu, Y., Yang, X., Liu, F., Liao, Y., 2022. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing* 493, 626–646. <https://doi.org/10.1016/j.neucom.2022.01.005>.
- Nieto-Hidalgo, M., Gallego, A.-J., Gil, P., Pertusa, A., 2018. Two-stage convolutional neural network for ship and spill detection using SLAR images. *IEEE Trans. Geosci. Remote Sens.* 56, 5217–5230. <https://doi.org/10.1109/TGRS.2018.2812619>.
- Niwattanakul, S., Singthongchai, J., Naenudorn, E., Wanapu, S., 2013. Using of Jaccard coefficient for keywords similarity. In: Proceedings of the International Multiconference of Engineers and Computer Scientists, pp. 380–384.
- Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M.C.H., Heinrich, M.P., Misawa, K., Mori, K., McDonagh, S.G., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D., 2018. Attention U-Net: Learning Where to Look for the Pancreas. CoRR abs/1804.0.
- O'Shea, K., Nash, R., 2015. An Introduction to Convolutional Neural Networks. Corr abs/1511.0.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Rahman, M.A., Wang, Y., 2016. Optimizing intersection-over-union in deep neural networks for image segmentation. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Porikli, F., Skaff, S., Entezari, A., Min, J., Iwai, D., Sadagic, A., Scheidegger, C., Isenberg, T. (Eds.), *Advances in Visual Computing*. Springer International Publishing, Cham, pp. 234–244.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), *Medical Image Computing and Computer-assisted Intervention — MICCAI 2015*. Springer International Publishing, Cham, pp. 234–241.
- Roussou, R., Katz, N., Sharon, G., Glizerin, Y., Kosman, E., Shuster, A., 2022. Automatic recognition of oil spills using neural networks and classic image processing. *Water* 14 <https://doi.org/10.3390/w14071127>.
- Ruby, U., Yendapalli, V., 2020. Binary cross entropy with deep learning technique for image classification. *Int. J. Adv. Trends Comput. Sci. Eng.* 9 <https://doi.org/10.30534/ijatcse/2020/175942020>.
- Shaban, M., Salim, R., Abu Khalifeh, H., Khelifi, A., Shalaby, A., El-Mashad, S., Mahmoud, A., Ghazal, M., El-Baz, A., 2021. A deep-learning framework for the detection of oil spills from SAR data. *Sensors* 21. <https://doi.org/10.3390/s21072351>.
- Singha, S., Bellerby, T.J., Trieschmann, O., 2012. Detection and classification of oil spill and look-alike spots from SAR imagery using an Artificial Neural Network. In: 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 5630–5633. <https://doi.org/10.1109/IGARSS.2012.6352042>.
- Singha, S., Bellerby, T.J., Trieschmann, O., 2013. Satellite oil spill detection using artificial neural networks. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 6, 2355–2363. <https://doi.org/10.1109/JSTARS.2013.2251864>.
- Solberg, A.H.S., 2012. Remote sensing of ocean oil-spill pollution. *Proc. IEEE* 100, 2931–2945. <https://doi.org/10.1109/JPROC.2012.2196250>.
- Solberg, A.H.S., Brekke, C., Husøy, P.O., 2007. Oil spill detection in Radarsat and Envisat SAR images. *IEEE Trans. Geosci. Remote Sens.* 45, 746–755. <https://doi.org/10.1109/TGRS.2006.887019>.
- Taravat, A., Del Frate, F., 2012. Development of band ratioing algorithms and neural networks to detection of oil spills using Landsat ETM+ data. *EURASIP J. Adv. Signal Process.* 2012, 107. <https://doi.org/10.1186/1687-6180-2012-107>.
- Taye, M.M., 2023. Theoretical understanding of convolutional neural network: concepts, architectures, applications, future directions. *Computation* 11. <https://doi.org/10.3390/computation11030052>.
- Temitope Yekeen, S., Balogun, A., Wan Yusof, K.B., 2020. A novel deep learning instance segmentation model for automated marine oil spill detection. *ISPRS J. Photogramm. Remote Sens.* 167, 190–200. <https://doi.org/10.1016/j.isprsjprs.2020.07.011>.
- Topouzelis, K., Psyllos, A., 2012. Oil spill feature selection and classification using decision tree forest on SAR image data. *ISPRS J. Photogramm. Remote Sens.* 68, 135–143. <https://doi.org/10.1016/j.isprsjprs.2012.01.005>.
- Topouzelis, K., Tarchi, D., Vespe, M., Posada, M., Muellenhoff, O., Ferraro, G., 2015. Detection, tracking, and remote sensing: satellites and image processing (spaceborne oil spill detection). In: Fingas, M. (Ed.), *Handbook of Oil Spill Science and Technology*. John Wiley & Sons, Inc.
- Trujillo-Acatitla, R., Tuxpan-Vargas, J., Ovando-Vázquez, C., 2022. Oil spills: detection and concentration estimation in satellite imagery, a machine learning approach. *Mar. Pollut. Bull.* 184, 114132 <https://doi.org/10.1016/j.marpolbul.2022.114132>.
- Van Rossum, G., Drake, F.L., 2009. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA.

- Vasudevan, R.K., Ziatdinov, M., Vlcek, L., Kalinin, S.V., 2021. Off-the-shelf deep learning is not enough, and requires parsimony, Bayesianity, and causality. *npj Comput. Mater.* 7, 16. <https://doi.org/10.1038/s41524-020-00487-0>.
- Wan, J., Cheng, Y., 2013. Remote sensing monitoring of Gulf of Mexico oil spill using ENVISAT ASAR images. In: 2013 21st International Conference on Geoinformatics, pp. 1–5. <https://doi.org/10.1109/Geoinformatics.2013.6626165>.
- Wang, X., Liu, J., Zhang, S., Deng, Q., Wang, Z., Li, Y., Fan, J., 2021. Detection of oil spill using SAR imagery based on AlexNet model. *Comput. Intell. Neurosci.* 2021, 4812979 <https://doi.org/10.1155/2021/4812979>.
- Wang, D., Wan, J., Liu, S., Chen, Y., Yasir, M., Xu, M., Ren, P., 2022. BO-DRNet: an improved deep learning model for oil spill detection by polarimetric features from SAR images. *Remote Sens.* 14 <https://doi.org/10.3390/rs14020264>.
- Wang, D., Liu, S., Zhang, C., Xu, M., Yang, J., Yasir, M., Wan, J., 2023. An improved semantic segmentation model based on SVM for marine oil spill detection using SAR image. *Mar. Pollut. Bull.* 192, 114981 <https://doi.org/10.1016/j.marpolbul.2023.114981>.
- White, H.K., Hsing, P.-Y., Cho, W., Shank, T.M., Cordes, E.E., Quattrini, A.M., Nelson, R.K., Camilli, R., Demopoulos, A.W.J., German, C.R., Brooks, J.M., Roberts, H.H., Shedd, W., Reddy, C.M., Fisher, C.R., 2012. Impact of the *Deepwater Horizon* oil spill on a deep-water coral community in the Gulf of Mexico. *Proc. Natl. Acad. Sci.* 109, 20303–20308. <https://doi.org/10.1073/pnas.1118029109>.
- Xu, J., Wang, X., Zhu, X., Cui, C., Liu, P., Li, B., 2017. Research on marine radar oil spill network monitoring technology. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1868–1871. <https://doi.org/10.1109/IGARSS.2017.8127341>.
- Yamashita, R., Nishio, M., Do, R.K.G., Togashi, K., 2018. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 9, 611–629. <https://doi.org/10.1007/s13244-018-0639-9>.
- Zeng, K., Wang, Y., 2020. A deep convolutional neural network for oil spill detection from spaceborne SAR images. *Remote Sens.* 12 <https://doi.org/10.3390/rs12061015>.
- Zhai, J., Mu, C., Hou, Y., Wang, J., Wang, Y., Chi, H., 2022. A dual attention encoding network using gradient profile loss for oil spill detection based on SAR images. *Entropy* 24. <https://doi.org/10.3390/e24101453>.
- Zhang, J., Feng, H., Luo, Q., Li, Y., Wei, J., Li, J., 2020. Oil spill detection in quad-polarimetric SAR images using an advanced convolutional neural network based on SuperPixel model. *Remote Sens.* 12 <https://doi.org/10.3390/rs12060944>.
- Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2018. UNet++: {A} Nested U-Net Architecture for Medical Image Segmentation. *Corr abs/1807.1*.
- Zhu, Q., Zhang, Y., Li, Z., Yan, X., Guan, Q., Zhong, Y., Zhang, L., Li, D., 2022. Oil spill contextual and boundary-supervised detection network based on marine SAR images. *IEEE Trans. Geosci. Remote Sens.* 60, 1–10. <https://doi.org/10.1109/TGRS.2021.3115492>.