Oil Spill Detection using Convolutional Neural Networks and Sentinel-1 SAR Imagery

Eleftheria Kalogirou[1,2], Konstantinos Christofi[1,2], Despoina Makri[1,2], Muhammad Amjad Iqbal[1], Valeria La Pegna[3], Marios Tzouvaras[1,2], Christodoulos Mettas[1,2], Diofantos Hadjimitsis[1,2]

[1] ERATOSTHENES Centre of Excellence, Limassol 3012, Cyprus
[2] Department of Civil Engineering and Geomatics, Cyprus University of Technology, Limassol 3036, Cyprus
[3] Università degli Studi di Roma Tor Vergata, 00133 Roma RM, Italy

Keywords: Oil spills, Remote sensing, AI, Neural Networks, CNN, Marine monitoring, Marine pollution.

Abstract

Oil spills impose significant environmental challenges, leading to critical consequences for marine ecosystems and sea habitant's health. Early delineatin and efficient surveillance are absolutely important to prevent more contamination and support quick hazards reduction. This study focuses on detecting oil spills using satellite imagery and deep learning models, specifically Convolutional Neural Networks (CNN). The dataset used to train the CNN comprised 695 images extracted from Sentinel-1 Synthetic Aperture Radar (SAR) data over the Mediterranean Sea. In particular, 486 images (70%) were allocated for training, 139 images (20%) for validation, and 70 images (10%) for testing. Preprocessing involved a thresholding technique to enhance feature extraction and improve classification precision. The CNN model achieved a high test accuracy of 98.57%, with perfect precision (1.0000), recall of 96.43%, and F1 score of 0.9818, demonstrating strong performance and reliability. These high accuracy levels underscore the model's efficiency in identifying oil spills and its soundness in handling unseen data. The significance of this work is in using satellite-based deep learning models for scalable and automated oil spill detection, therefore providing a reliable and effective substitute for more traditional monitoring systems. The model may be applied over large oceanic areas by using satellite images, thereby supporting marine ecosystem preservation and enhancing environmental risk management connected with oil pollution.

1. Introduction

Oil spills are among the most significant environmental disasters, posing severe threats to marine ecosystems, economic activities, and human health. Between 2010 and 2020, the Eastern Mediterranean experienced over 1,000 oil leak incidents, affecting marine ecosystems and contaminating hundreds of kilometers of shoreline. Reducing the effects of oil spills requires early detection and monitoring. Traditional detection techniques based on optical remote sensing or human observation are severely constrained due to atmospheric weather conditions, limitations in nighttime acquisition, and the large spatial extent that must be monitored. An effective system for automating and improving oil spill monitoring procedures has been developed by combining Artificial Intelligence (AI) techniques with remote sensing techniques, particularly with Sentinel-1 Synthetic Aperture Radar (SAR) data (Cheng et al. 2024). Regardless of sunlight, Sentinel-1, part of the European Space Agency's (ESA) Copernicus mission, can penetrate clouds and provide high-resolution SAR data. SAR data is particularly useful for maritime applications since oil spills suppress surface waves and produce distinctive dark patches in the reflected radar signal. Additionally, Sentinel-1's dual-polarization capability (VV and VH) enables it to distinguish oil spills from similar phenomena, such as low-wind areas, phytoplankton, or wave shadows. These characteristics make Sentinel-1 SAR data highly effective for detecting oil spills under various environmental conditions (Li, Kim et al. 2023).

Several studies have explored the application of deep learning methods with SAR data for efficient oil spill detection in diverse marine environments. Recent advancements have focused on leveraging Sentinel-1 SAR imagery for automated detection and early warning systems. For instance, (Wang et al. 2024) developed a system that processes oil spill detection within approximately 1.5 hours, achieving a false discovery rate of 23.3% and a false negative rate of 24.0% in the Southeastern Mediterranean Sea. This highlights the growing effectiveness of machine learning models in maritime surveillance while emphasizing the need for further improvements in reducing false detections. In another study, an encoder-decoder convolutional neural network (CNN) was trained on verified oil spill incidents in Pakistan's Exclusive Economic Zone (Basit et al. 2024). The model detected 92 previously unreported occurrences by analyzing SAR data from 2017 to 2023. In 2020, the highest number of leaks (26) was recorded, while in 2021 the most affected area (395 km^2) was observed. The findings demonstrated that deep learning significantly enhances oil spill detection and facilitates more effective monitoring techniques.

Considering the challenges posed by SAR data, a novel approach for identifying oil spills in the Suez Canal was proposed by (Cheng et al. 2024). The study introduced a Boosting Random Support Vector Machine (BRSVM) classifier, which achieved an accuracy of 94.72% and an F1-score of 95.33%, demonstrating the effectiveness of this method for large-scale detection. (Kalogirou et al. 2024) examined the environmental challenges posed by oil spills, which contribute significantly to marine pollution and can have long-lasting effects on ecosystems and the economy. The study highlighted that oil spills occur due to accidental discharges or intentional actions such as illegal dumping by cargo ships. The research focused on detecting oil spills in the North-Eastern part of Cyprus using a deep learning model and compared its effectiveness with a conventional Adaptive Thresholding Algorithm. The results indicated that deep learning models achieve higher accuracy, showcasing the potential of advanced technologies in improving oil spill detection. Additionally, leveraging Sentinel-1 data, an autonomous oil spill detection system with early warning capabilities was proposed for the Southeastern Mediterranean (Yang et al.

https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-757-2025 |

2024). This system employed a deep learning-based oil spill detector, achieving a false negative rate of 24.0% and a false discovery rate of 23.3%. The system demonstrated practical applicability, with the complete processing workflow—including SAR image acquisition and spill trajectory modeling—being completed in approximately 1.5 hours.

The development and comparison of machine learning models for oil spill detection in the Persian Gulf were investigated by (Najafizadegan et al. 2023). The study compared Support Vector Machine (SVM), Random Forest (RF), and CNN models. The results showed that CNN achieved the highest accuracy (95.8%), outperforming RF (86.0%) and SVM (78.9%). Furthermore, the study underscored the importance of tracking oil spills using both ascending and descending satellite passes. The accuracy and reliability of oil spill detection systems can be enhanced by continuously training models with newly acquired SAR data. A study by (Li, Park et al. 2023) introduced a self-evolving oil spill detection algorithm, which was tested using eight SAR images. The results demonstrated promising potential for real-time maritime pollution monitoring.

CNN models are at the forefront of this technological revolution, driving advances in image analysis and facilitating applications in pollution detection, marine surveillance, and environmental monitoring. Deep learning has shown great promise in the critical field of oil spill detection, a pressing environmental issue that threatens fisheries, coastal populations, and marine ecosystems. The integration of satellite remote sensing and artificial intelligence provides a robust solution for rapid and accurate monitoring, which is critical for mitigating environmental damage. By leveraging deep learning techniques and satellite imagery, this study aims to develop an automated oil spill detection framework. Automating the detection process enhances monitoring efficiency, allowing for rapid response and damage mitigation. The integration of AI with satellite remote sensing represents a major breakthrough in environmental monitoring, paving the way for advancements in pollution control and marine ecosystem preservation (Gao et al. 2024). The primary objective of this study is to employ satellite imagery and deep learning models to detect oil spills. The research investigates how CNN can be utilized for large-scale, automated oil spill detection, offering a reliable alternative to conventional monitoring techniques while enhancing environmental risk management.

The remainder of this paper is structured as follows:. Section 2 discusses the dataset, including sources, coverage, and preprocessing operations, with subsections on data training and validation, radar satellite images (RSI), and ground truth masks (GTM) processing. Section 3 describes the CNN model architecture, while Section 4 presents the oil spill detection algorithm, including training, optimization, and evaluation methods. Section 5 assesses the model's performance using various metrics such as model accuracy, model loss, training loss, validation loss, and accuracy and loss plots during training and validation. Section 6 presents the results and comparisons with ground truth data. Finally, Section 7 concludes the study, summarizing the findings and future research directions.

2. Dataset and Methodology

This section briefly describes the preprocessing of SAR data, CNN architecture utilized, and model evaluation. The flow graph as shown in Figure 1., illustrates the pipeline for detecting oil spills using Sentinel-1 SAR images and a CNN model. The dataset consists of 50 confirmed oil spill incidents (2014–2019) validated through Ground Truth Measurements (GTM) and Radar Satellite Images (RSI). Preprocessing steps, including noise reduction and radiometric corrections, enhance image quality before feeding the data into a CNN architecture comprising convolutional, pooling, and fully connected layers with a sigmoid output for binary classification (oil spill/no oil spill). The dataset is split into 70% training, 20% validation, and 10% testing, with the model trained using the Adam optimizer and binary cross-entropy loss over 10 epochs. Model evaluation focuses on accuracy and loss metrics, with predictions compared against ground truth data to assess classification performance.

2.1 Dataset

The collection of a significant number of oil spill observations to create a sufficiently large training dataset for deep learning models is a major challenge for various reasons. The dataset used was openly accessible to the public via: `https://drive.google.com/file/d/15WYzzFZvAHmqSIWOPXXRTp_YVd_86818/view?usp=sharing`. A significant challenge was obtaining a balanced dataset, particularly in terms of acquiring a sufficient number of confirmed oil spill-like areas to match the number of oil spill samples. To address this issue and maintain the dataset's validity, images from confirmed oil spill incidents from previous studies were used. The 116 incidents of oil spills recorded between 2014 and 2019 were reviewed, and 50 incidents were selected based on the variety of oil spill sizes and the availability of images with different contrasts.

Areas of interest containing confirmed oil spills were extracted from the raw SAR images, followed by radiometric corrections and calibration, which are critical steps to minimize radiometric distortions and ensure that the pixel values correspond precisely to the surface backscatter coefficient (sigma naught measured in dB). Noise filtering is critical for oil spill detection in SAR images, as noise can significantly hinder the extraction of information. The Lee filter was implemented as an effective method for noise reduction while preserving the clarity of edges and key features. Given the large variability in pixel values in radar images, converting pixel values from a linear scale to dB enhances contrast. As an adaptive speckle noise reduction technique for synthetic aperture radar (SAR) images, the Lee filter calculates the local average and variance in a moving window, adjusting the smoothing level based on local statistics. It preserves details in areas of high variance while effectively suppressing noise in homogeneous areas. Since speckle noise in SAR images is multiplicative, applying the filter in the logarithmic domain (dB) converts the noise to an additive form, facilitating suppression

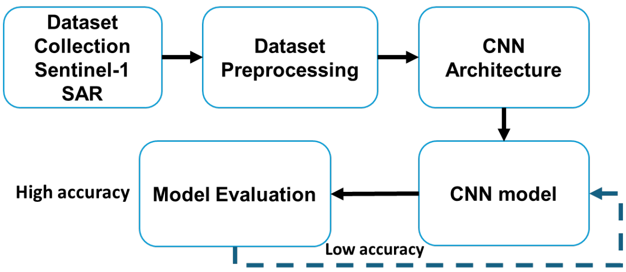

Figure 1. The flowchart of the proposed method for oil spill detection using Sentinel-1 SAR and CNN.

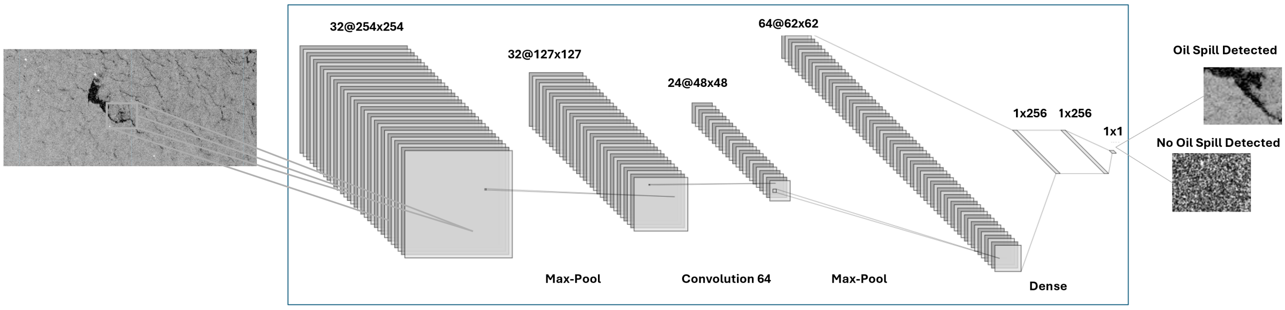

Figure 2. The modeled Convolutional Neural Network (CNN) architecture and detailed parameters for detecting oil spills from SAR imagery.

and improving the overall image sharpness. Research analyzing multi-spot noise reduction techniques for SAR data concluded that Lee filters are among the most effective methods for improving image quality and suppressing noise (Varshini R et al. 2024). After completing the preprocessing steps, the corresponding ground truth mask was created for each image. The accurate oil spill masks are the most critical step to ensure the reliable performance of the model. The image masks were geometrically corrected using the Universal Transverse Mercator (UTM) projection, with preprocessing carried out via the Sentinel Application Platform (SNAP). Therefore, the images were selected to contain only two labels: oil spills and water surfaces. The labeled images were split into smaller patches of size $256 \times 256 \times 1$ using the Patchify library (Cheng et al. 2024; Ahmed et al. 2023).

Figure 2 illustrates the architecture of a CNN designed for oil spill detection using Sentinel-1 SAR satellite images. Initially, a square patch is extracted from the satellite image, focusing on a dark region, indicating an oil spill's presence. The CNN processes this input through multiple convolutional layers, where the first stage generates 32 feature maps (32@ 254x254), which are reduced to 32@127x127 via Max-Pooling. A second convolutional layer with 64 filters produces 64@62x 62, followed by another layer with 24 feature maps (24@48x48). The extracted features are flattened into a one-dimensional array (flatten) and passed through fully connected (dense) layers with 256 neurons, leading to a final output layer with a single neuron (1x1) performing binary classification. The CNN output is either "Oil Spill Detected," if an oil spill is present, or "No Oil Spill Detected," if no indication of an oil spill is found. This model leverages texture differences and radar wave reflection properties to identify oil spill patterns in marine environments effectively.

2.2 Data Training and Validation

A data allocation ratio of 70%–20%–10% was used to train, validate, and test the model, ensuring a balance between performance assessment and training. Using binary cross-entropy as the loss function and the Adam optimization technique, the model was trained for 10 epochs. During training, a separate validation dataset was used to monitor accuracy and prevent overfitting, while the test dataset was employed to evaluate the model's generalization capabilities.

The model was trained using the Adam optimizer, which dynamically adjusts the learning rate by combining the benefits of Momentum and Stochastic Gradient Descent (MSGD) approaches. Since this is a binary classification problem (oil spill or not), binary cross-entropy was selected as the loss function. Activation functions such as sigmoid enabled the network to learn and recognize intricate patterns in the images, and the final layer of the network employs a sigmoid activation function to classify images into two categories: "oil spill" and "no oil spill." The model's fully connected layer generates a numerical output that determines the final classification.

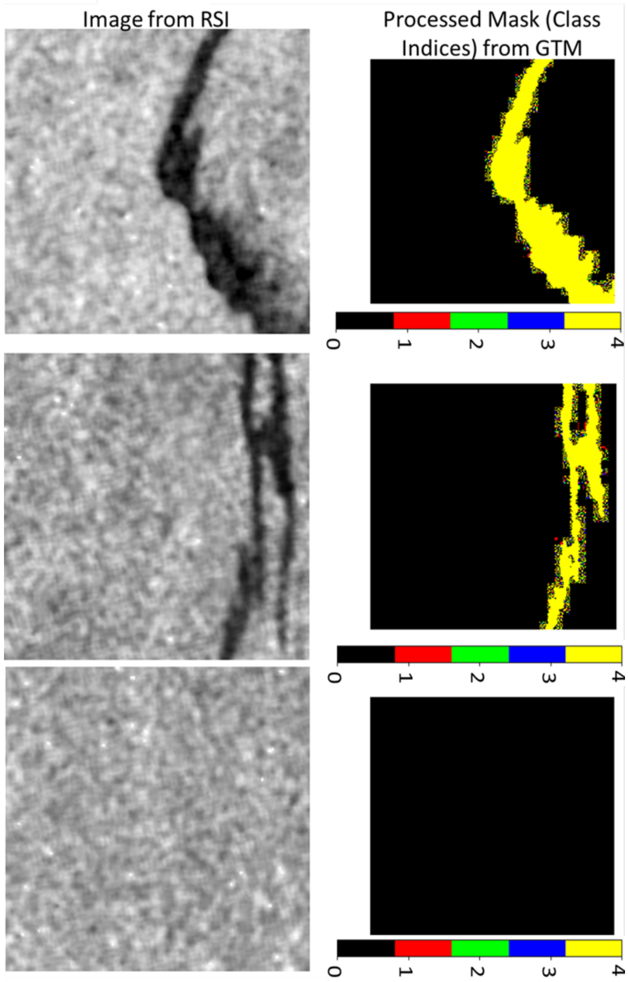

Figure 3. Sample SAR images (left column) and corresponding ground truth mask (GTM) image (right column); yellow areas are the oil spills.

2.3 Data Preparation: RSI Images and GTM Processing Masks

Before the model training, the data preparation process was carried out, which involved RSI and GTM. Proper data preparation is crucial for effective model training, as it ensures that the input data is normalized and correctly formatted. Initially, the RSI images were loaded and normalized to the [0,1] scale using the `image_to_array` method, which converts the images into numerical arrays. This normalization is essential, as images with pixel values in this range help the model learn the patterns in the data more effectively. Subsequently, the GTM masks were loaded in grayscale, representing categories (e.g., oil spill or not) and were normalized using a predefined color map (`COLOR_MAP`).

A four-class oil spill classification is given by (Hassani et al. 2020). As shown in Figure 3, the oil spill detection mask is primarily displayed in a yellow hue, representing the main body of the oil, while the water is depicted in black to indicate areas without any detected contamination. Minor appearances in red, blue, and green are due to small variations or noise during data processing, causing some pixels at the boundaries between classes to be classified slightly differently. These minor deviations reflect the sensitivity of the classification method and do not affect the overall representation of the oil spill; they simply highlight small anomalies that may occur during the image conversion and processing stages. After loading the images and masks, proper image-mask matching was performed, ensuring that each RSI image matched the corresponding GTM mask. This process was crucial to ensure that the model would be trained with the correct labels for each pixel. Proper matching of images and masks is necessary for the model to learn the correct association between the image features and their corresponding mask labels. The goal of this process was to prepare the data in the right way so that the model could be trained effectively and achieve good generalization on unseen data.

2.4 The Basic Structure of CNN

CNNs are powerful deep learning models specialized in image processing and analysis. They are widely used in various applications, including environmental monitoring, where they help detect oil spills in satellite imagery. A CNN consists of three main components: the input layer, which receives and processes images; the feature extraction layers, which use filters to identify important patterns; and the classification layers, which categorize the image based on extracted features. The convolutional layers play a pivotal role in feature extraction by applying filters that detect edges, textures, and shapes within an image. These filters help identify key characteristics of an oil spill, such as texture variations and boundary contours. To retain spatial information, zero padding is applied at the edges, while pooling layers help reduce dimensionality, preserving the most relevant features while lowering computational costs.

The final stage of a CNN includes fully connected layers, which take the extracted features and classify the image into predefined categories. In the case of oil spill detection, the model determines whether an oil spill is present or not. By stacking multiple layers, CNNs effectively learn hierarchical representations of image data, making them well-suited for complex visual recognition tasks (Dehghani-Dehcheshmeh et al. 2023).

2.5 Proposed Oil Spill Detection Algorithm

The proposed oil spill detection algorithm is based on a CNN designed to analyze satellite images and identify oil spills with high accuracy. Despite the presence of noise and distortions in satellite data, CNNs have a strong ability to detect patterns and recognize key features in images. The detection process follows three key stages.

Table 1. Training and Validation Performance Metrics of the CNN Model

Epoch	Accuracy	Loss	Validation Accuracy	Validation Loss
1/10	0.5040	2.4372	0.7612	0.5681
2/10	0.7623	0.4970	0.8504	0.3470
3/10	0.8854	0.2991	0.9281	0.2719
4/10	0.9173	0.2539	0.9353	0.2450
5/10	0.9270	0.2204	0.9525	0.1568
6/10	0.9580	0.1741	0.9496	0.1340
7/10	0.9706	0.1085	0.9741	0.1278
8/10	0.9686	0.1184	0.9612	0.1204
9/10	0.9667	0.0995	0.9784	0.0913
10/10	0.9639	0.1139	0.9755	0.0805

- First, convolutional layers extract essential features such as edges, textures, and shapes by applying specialized filters. These layers process raw image data, identifying the fundamental visual properties of an oil spill.
- Next, pooling layers reduce the spatial dimensions of the extracted features while retaining the most significant information. Max pooling is commonly used to select the most prominent features, enhancing model efficiency and minimizing computational load. This step also helps prevent overfitting, improving the model's generalization to new images.
- Finally, fully connected layers analyze the collected features and classify the image into one of two categories: "oil spill" or "no oil spill." These layers integrate all extracted patterns to produce a final classification decision.

By leveraging the hierarchical feature extraction capabilities of CNNs, the proposed model can detect oil spills in satellite images with a high degree of accuracy, even in challenging environmental conditions (Lee et al. 2017).

3. CNN Model Evaluation

This section discusses the model accuracy and loss during training and validation of CNN for oil spill detection.

3.1 Model Accuracy and Loss

As shown in Table 1, the model performed well during training, achieving a training accuracy of 96.39% and training loss of 0.1139. The high accuracy suggests that the model is able to make correct predictions most of the time, while the relatively low loss shows that it is minimizing errors in its predictions. Additionally, the model's performance on the validation set was also successful, with a validation accuracy of 97.55% and a validation loss of 0.0805. This indicates that the model is able to generalize well to unseen data, as the validation accuracy is almost as high as the training accuracy. The low validation loss further supports this, suggesting that the model is not overfitting and is maintaining good performance on data that it hasn't encountered before.

Overall, these metrics demonstrate that the model has been trained effectively. It can make accurate predictions on both the training and validation datasets, and its ability to generalize to new data suggests that it will perform well when deployed in real-world applications.

https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-757-2025

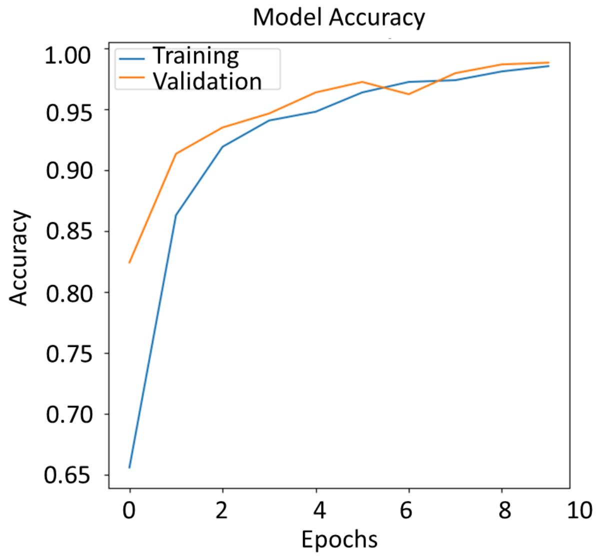

Figure 4. Model Training and Validation Accuracy Over Epochs.

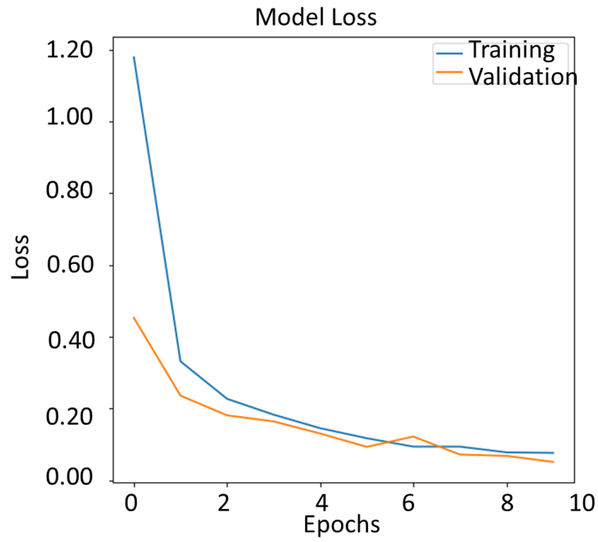

Figure 5. Model Training and Validation Loss Over Epochs.

3.2 Training and Validation Loss

The training loss of 0.1139 indicates that the model is learning well from the training data. Loss is a metric that measures the difference between the predicted values and the actual ground truth. A smaller loss means that the model’s predictions are closer to the true values, which is the goal during training. The steady decrease in training loss over the epochs suggests that the model is gradually improving its predictions for the training data. This predicted behavior shows that the model is gradually learning and reducing the inaccuracy.

The validation loss of 0.0805 is equally low, which is also a positive outcome. Validation loss is calculated on a separate dataset that the model has not seen before, helping to assess how well it generalizes to new, unseen data. The fact that the validation loss is not much higher than the training loss suggests that the model is not overfitting. If the validation loss were significantly higher, it would indicate that the model failed to generalize and was too specialized to the training data. Therefore, the model demonstrates good generalization ability, as it performs well on both the training and validation datasets.

3.3 Qualitative Analysis of CNN Model

The accuracy plots, as shown in Figure 4, provide valuable insights into how well the model is performing. In these plots, the training accuracy (represented by the blue line) and the validation accuracy (represented by the orange line) are almost identical. This alignment is a positive indicator, as it suggests that the model is generalizing well and is not overfitting, moving forward with the training epochs. Overfitting typically manifests as high training accuracy with significantly lower validation accuracy, but that is not the case here. The model appears to be learning effectively from the training data and applying that knowledge to make accurate predictions on unseen validation data.

In the early epochs, the model shows important improvement in accuracy, which is typical during the first stages of training as the model learns the basic patterns in the data. As training progresses, the rate of improvement in accuracy begins to slow down, but this is expected as the model approaches its performance limits on the training data.

3.4 Fluctuations in Validation Loss

Small fluctuations in validation loss are observed after a few epochs, as shown in Figure 5. These fluctuations are expected in any training process. They typically do not pose a problem unless they cause a significant increase in validation loss. They can arise due to changes in the data distribution or small adjustments made by the model during training. Slight variations in validation loss can be influenced by the complexity of the model or hyperparameter choices. Weight updates or changes in the data subsets used in each epoch may also lead to minor fluctuations in performance.

3.5 Predicted Labels

For several important reasons, the process of generating predicted labels in a binary classification model for oil spill detection is vital. Using the sigmoid activation function, which scales the output between 0 and 1, it first automates the classification operation by converting the numerical outputs of the model into probability scores. These probabilities are then converted into binary labels (“oil spill” or “no oil spill”) based on a predefined threshold (typically 0.5), allowing for efficient processing of large data sets without the need for continuous human interaction. In addition, generating expected labels is significant for evaluating the performance of the model. It is possible to calculate critical performance metrics, such as accuracy, sensitivity, and specificity, by comparing these predictions with the ground truth labels. These metrics provide insights into the model’s performance, indicating opportunities for development and ensuring the system’s reliability in real-world scenarios. Another critical element to avoid overfitting is monitoring the predicted labels on separate validation and testing datasets. This continuous evaluation ensures that the model not only understands the features of the training data but also successfully generalizes to new, untested data, which is significant in real-world situations, such as environmental monitoring. In conclusion, automated generation of predicted labels improves the overall performance of the system, allowing the model to be seamlessly integrated into decision-making processes and real-time monitoring systems (Sun et al. 2024).

4. Results

The dataset was prepared from Sentinel-1 Level-1 Ground Range Detected (GRD) images, which is a C-band SAR system and provided by the ESA through the Copernicus Open Access Hub. With a pixel spacing of 10 meters, the SAR sensor provides ground coverage of approximately 250 kilometers. The raw data from the vertical-vertical (VV) polarization were processed to create the SAR dataset.

The model accuracy results indicated in Table 1 that the model is effectively learning from the training data. The performance of the model in both the training and validation phases has been promising. The training accuracy of 96.39% indicates that the model was able to effectively learn from the training data and generalize well to the test set, achieving a high level of precision in predicting the target class. This is supported by a relatively low training loss of 0.1139, which shows that the model's predictions closely align with the true labels in the test set. The validation accuracy, which reached 97.55%, further reinforces the model's generalization capability, demonstrating that it performs well not only on the training data but also on different datasets. The validation loss of 0.0805 further supports this finding, suggesting that the model maintains high predictive accuracy with minimal error when tested on the validation set. The relatively small gap between the test and validation accuracy indicates that the model is neither underfitting nor overfitting. It suggests that the model has effectively captured the underlying patterns in the training data while still maintaining the ability to generalize to unseen examples. The close proximity between the test and validation loss values (training loss = 0.1139, validation loss = 0.0805) further implies that the model is not overfitting to the training data and is therefore robust to variations in new data.

As shown in Figure 6, it can be observed how the model accurately detected oil spills, confirming its reliability and precision in identifying marine pollution. The successful classification indicates that the model has effectively learned to recognize the characteristics of oil spills, making it a valuable tool for environmental monitoring and pollution management. Also, its capacity to generalize to new data implies that it can be used in many geographic locations, supporting the preservation of marine ecosystems and facilitating prompt action in the event of pollution incidents. The following images illustrate the model's prediction outcomes, where the detected oil spills are validated against real-world data. This demonstrates the model's effectiveness in real scenarios, reinforcing the role of artificial intelligence in monitoring and safeguarding marine ecosystems.

The model achieved a high-test accuracy of 0.9857, demonstrating its strong performance in classification. With a perfect precision score of 1.0000, the model exhibits no false positives, ensuring that all predicted positive cases are truly relevant. The recall score of 0.9643 indicates a slight margin of false negatives, meaning that a small portion of actual positive cases were not detected. The F1-score, which balances precision and recall, reached 0.9818, confirming the overall robustness and reliability of the model in accurately identifying the target class.

The confusion matrix in Figure 7 shows the performance of the oil spill detection model, demonstrating excellent accuracy

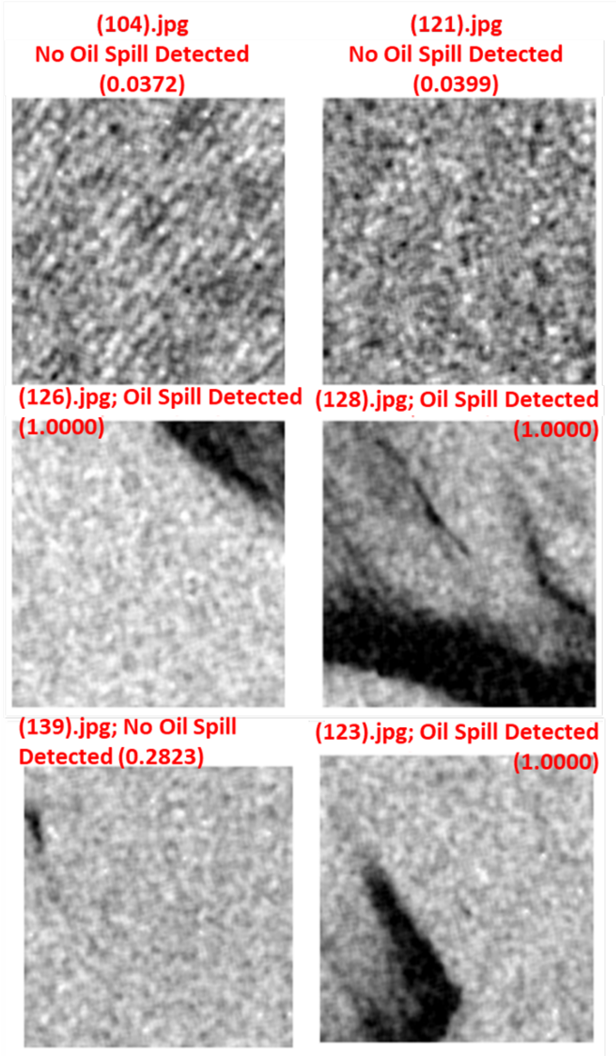

Figure 6. Predicted Labels: detected oil spill and no oil spill.

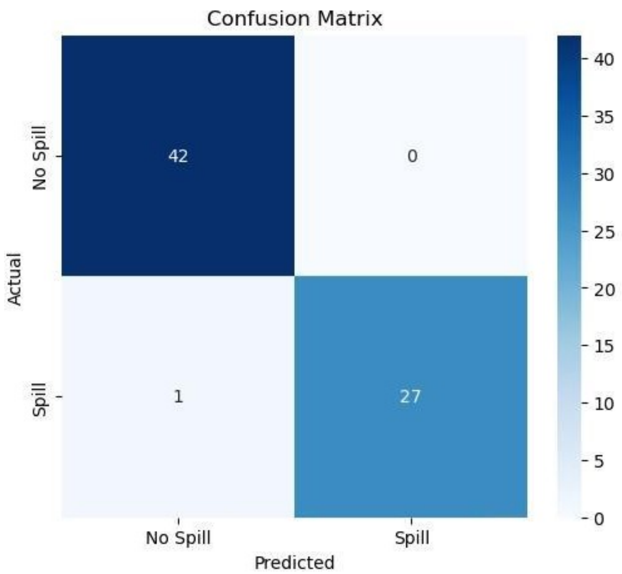

Figure 7. Confusion matrix of the model achieved for the given dataset.

98.57% and perfect precision for oil spill detection 100%. Specifically, the model correctly identified 27 oil spill cases (true positives) and 42 non-spill cases (true negatives), with no false positives, highlighting its reliability. However, there was one false negative, meaning the model missed one actual oil spill, resulting in a recall of 96.43%. While the overall performance is very high, this indicates a small chance that the model might fail to detect some real oil spills, which could be critical in real-world applications.

These results signify that the model is performing well and can make accurate predictions, which is especially important when applied to tasks such as environmental monitoring, where reliable predictions are crucial. In conclusion, a key component of machine learning applications is model performance, which establishes the accuracy and dependability of predictions. While achieving high accuracy is critical, maintaining the model's stability and adaptability to various datasets is just as essential. Models are frequently tested in real-world situations in practical applications, where data distributions may differ and unexpected difficulties may occur. Therefore, ongoing assessment, adjustment, and refinement are required to ensure optimal performance.

5. Conclusions and Future Work

The CNN model achieved a high test accuracy of 98.57%, with perfect precision (1.0000), recall of 96.43% and F1 score of 0.9818, demonstrating strong performance and reliability. The results demonstrate that the model has achieved high performance in both the training and validation phases, with high accuracy and minimal loss. The fact that the model achieved 96.39% training accuracy during training and 97.55% accuracy during validation confirms that it has learned to generalize well from the training data to unseen examples. Furthermore, the relatively low training and validation loss values indicate that the model is capable of making predictions with high accuracy, which is essential for tasks such as detecting oil spills or other environmental hazards in satellite imagery. The model's ability to generalize well to the validation set suggests that it has successfully captured relevant features within the dataset without overfitting. This makes it a promising tool for deployment in real-world scenarios, where unseen data may vary. The good performance on both the training and validation sets highlights that the model is reliable and can be trusted to make accurate predictions in a wide range of situations, providing confidence in its application for environmental monitoring, especially in the detection of phenomena such as oil spills. Given the high performance of the model, it can be confidently stated that the model is ready for integration into operational systems that require accurate and reliable predictions. Furthermore, its success validates the effectiveness of using deep learning techniques for such image classification tasks, particularly in environmental science applications. Although the model has achieved strong results, there are areas that could be deployed in future research and development to improve its performance and increase its applicability to wider areas.

One of the main priorities for the future development of the model is to extend the dataset beyond the Mediterranean Sea. While the current dataset has provided valuable training data, further strengthening the robustness of the model can be achieved by incorporating a greater variety of images from different geographical areas and environments that will allow the model to learn to detect oil spills. We will also address the impacts of environmental hazards under different conditions and ocean current impacts on oil spills over certain time spans (Iqbal et al. 2022). By incorporating these techniques, the model could achieve greater accuracy in detecting various conditions of oil spill pollution. Being an accurate oil spill detection crucial for preventing and minimizing environmental impacts, in this work it has been demonstrated that the automated satellite image analysis allows for faster and more precise detection compared to traditional methods, helping authorities make timely decisions to address pollution and restore the marine environment.

6. Acknowledgements

The present work was carried out in the framework of the AI-OBSERVERTM project (`https://ai-observer.eu/`) titled "Enhancing Earth Observation Capabilities of the Eratosthenes Centre of Excellence on Disaster Risk Reduction through Artificial Intelligence," that has received funding from the European Union's Horizon Europe Framework Programme HORIZON-WIDERA-2021-ACCESS-03 (Twinning) under Grant Agreement No 101079468. The authors also acknowledge the 'EXCELSIOR': ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment H2020 Widespread Teaming project (`www.excelsior2020.eu`) in which the Eratosthenes CoE has been established. The 'EXCELSIOR' project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 857510, from the Government of the Republic of Cyprus through the Directorate General for the European Programmes, Coordination, and Development and the Cyprus University of Technology.

References

Ahmed, Samira et al. (2023). 'Deep neural network for oil spill detection using Sentinel-1 data: application to Egyptian coastal regions'. In: *Geomatics, Natural Hazards and Risk* 14.1, pp. 76–94.

Basit, Abdul et al. (2024). 'Deep Learning-Based Detection of Oil Spills in Pakistan's Exclusive Economic Zone from January 2017 to December 2023.' In: *Remote Sensing* 16.13.

Cheng, Lingxiao et al. (2024). 'A marine oil spill detection framework considering special disturbances using Sentinel-1 data in the Suez Canal'. In: *Marine Pollution Bulletin* 208, p. 117012.

Dehghani-Dehcheshmeh, Saeid et al. (2023). 'Oil spills detection from SAR Earth observations based on a hybrid CNN transformer networks'. In: *Marine Pollution Bulletin* 190, p. 114834.

Gao, Yuhao et al. (2024). 'Relating CNN-Transformer Fusion Network for Remote Sensing Change Detection'. In: *2024 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, pp. 1–6.

Hassani, Behnam et al. (2020). 'Oil spill four-class classification using UAVSAR polarimetric data'. In: *Ocean Science Journal* 55.3, pp. 433–443.

Iqbal, Muhammad Amjad et al. (2022). ‘Doppler Centroid Estimation for Ocean Surface Current Retrieval from Sentinel-1 SAR Data’. In: *2021 18th European Radar Conference (EuRAD)*, pp. 429–432.

Kalogirou, Eleftheria et al. (2024). ‘Empirical Analysis of Oil Spill Detection Methods’. In: *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 3673–3677.

Lee, Yeakang et al. (2017). ‘Making a More Reliable Classifier via Random Crop Pooling’. In: *Robot Intelligence Technology and Applications 4: Results from the 4th International Conference on Robot Intelligence Technology and Applications*. Springer, pp. 309–318.

Li, Chenglei, Duk-jin Kim et al. (2023). ‘A self-evolving deep learning algorithm for automatic oil spill detection in Sentinel-1 SAR images’. In: *Remote Sensing of Environment* 299, p. 113872.

Li, Chenglei, Soyeon Park et al. (2023). ‘A Deep Learning Based Self-Evolving Oil Spill Detection Algorithm Using Sentinel-1 SAR Images’. In: *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 1289–1292.

Najafizadegan, Sahand et al. (2023). ‘Variable-complexity machine learning models for large-scale oil spill detection: The case of Persian Gulf’. In: *Marine Pollution Bulletin* 195, p. 115459.

Sun, Zhen et al. (2024). ‘Utilizing deep learning algorithms for automated oil spill detection in medium resolution optical imagery’. In: *Marine Pollution Bulletin* 206, p. 116777.

Varshini R, Sanjjushri et al. (2024). ‘Speckle Noise Analysis for Synthetic Aperture Radar (SAR) Space Data’. In: *arXiv e-prints*, arXiv–2408.

Wang, Dawei et al. (2024). ‘Marine oil spill detection using improved polarimetric feature based on polarization SAR image’. In: *International Journal of Remote Sensing* 45.3, pp. 911–929.

Yang, Yi-Jie et al. (2024). ‘A near real-time automated oil spill detection and early warning system using Sentinel-1 SAR imagery for the Southeastern Mediterranean Sea’. In: *International Journal of Remote Sensing* 45.6, pp. 1997–2027.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-757-2025