Oil Spill Detection Based on Multiscale Multidimensional Residual CNN for Optical Remote Sensing Imagery

Seyd Teymoor Seydi, Mahdi Hasanlou, *Member, IEEE*, Meisam Amani, *Senior Member, IEEE*, and Weimin Huang, *Senior Member, IEEE*

***Abstract*—Oil spill (OS), as one of the main pollutions in the ocean, is a serious threat to the marine environment. Thus, timely and accurate OS detection (OSD) is necessary for ocean management. In this regard, remote sensing (RS) plays a key role due to multiple advantages over large and remote ocean environments. In this study, a new OSD framework based on a deep learning algorithm was developed for optical RS imagery. The proposed method was based on a multiscale multidimensional residual kernel convolution neural network. The proposed method investigated the deep features by the two-dimensional multiscale residual blocks and, then, utilized them at one-dimensional multiscale residual blocks. In this study, Landsat-5 satellite imagery acquired over the Gulf of Mexico was applied to evaluate the performance of the proposed method. The overall accuracy of the proposed method was more than 95%, and the miss detection and false alarm rates were less than 5%, indicating its high potential for OSD. Moreover, it was observed that the proposed method had better performance compared to other OSD algorithms that were investigated in this study.**

***Index Terms*—Convolution neural network (CNN), deep learning, Landsat-5, multiscale kernel convolution, ocean oil spill detection, remote sensing (RS), residual block.**

I. Introduction

EXPLORATION and transportation of oil in the ocean have been considerably increased over the past three decades due to the high oil consumption by different sectors. Oil spill (OS), as one of the main factors polluting the ocean, has negative impacts on coastal and deep-ocean environments [1]. For example, OS results in the contamination of the seafood and the death of marine fauna [2]. Thus, developing accurate and timely OS detection (OSD) algorithms is necessary for preventing the spread of OS and minimizing its negative effects on the ocean and the ecological ecosystem [3]. In this regard, remote sensing (RS) systems provide valuable datasets [4]. For example, satellites collect data with a minimum cost and in a timely manner over large and remote ocean areas [5]. Moreover, the availability of different RS datasets, especially the open-access RS imagery, has considerably facilitated real-time and frequent OSD.

Optical and radar RS datasets are the two resources for OSD, each has its advantages and disadvantages. Optical RS data have usually better spatial and temporal resolutions and contain low noise compared to radar data. On the other hand, radar systems can operate during both daytime and nighttime and in almost any weather condition [6].

Satellite image processing and classification are required in OSD methods. Various machine learning algorithms have so far been developed for image classification using RS data (e.g., [7]–[9]). In this regard, deep learning methods have recently gained more attention due to their multiple advantages over the traditional machine learning algorithms [4]. For example, deep learning algorithms can automatically extract deep features using convolution layers. Moreover, they usually provide higher accuracies compared to other machine learning algorithms [10]–[13].

Like other applications, it was argued that deep learning methods provide high accuracies in OSD [14]. For example, Yekeen *et al.* developed a deep learning framework for OSD using radar imagery. They used the mask-region-based convolutional neural network (Mask R-CNN) for instance segmentation of OS [15]. Moreover, Jiao *et al.* proposed an OSD algorithm based on unmanned aerial vehicles (UAVs) images. First, the high potential OS areas were detected using a deep convolutional neural network (DCNN). Then, the Otsu thresholding algorithm was applied to improve the result of the detection. Finally, the maximally stable extremal regions (MSER) algorithm was utilized for determining the detailed polygon region from the detection box [16]. Additionally, Bianchi *et al.* developed a deep learning architecture called an oil fully convolution network for OSD using radar data. Their proposed algorithm was based on the U-Net, encoder, and decoder [17]. Furthermore, Zhang *et al.* proposed an OSD method based on CNN and the simple linear iterative clustering (SLIC) superpixel method for radar data. They initially extracted several polarimetric features.

Manuscript received December 27, 2020; revised April 25, 2021 and August 31, 2021; accepted October 20, 2021. Date of publication October 27, 2021; date of current version November 10, 2021. This work was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grants under Grant NSERC RGPIN-2017-04508. *(Corresponding author: Mahdi Hasanlou.)*

Seyd Teymoor Seydi and Mahdi Hasanlou are with the School of Surveying and Geospatial Engineering College of Engineering, University of Tehran, Tehran 1439957131, Iran (e-mail: seydi.teymoor@ut.ac.ir; hasanlou@ut.ac.ir).

Meisam Amani is with the Wood Environment & Infrastructure Solutions, Ottawa, ON K2E 7K3, Canada (e-mail: meisam.amani@woodplc.com).

Weimin Huang is with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada (e-mail: weimin@mun.ca).

Digital Object Identifier 10.1109/JSTARS.2021.3123163

Three channels (HH, HV, and VV) were used for calculating SLIC superpixels. The extracted features and superpixels were then considered as inputs of the CNN algorithm. Finally, the result of OS was obtained by the semantic segmentation algorithm using CNN [18]. Krestenitis *et al.* also investigated the performance of different types of deep learning architectures (i.e., U-Net, Link-Net, PSP-Net, DeepLabV2, and DeepLabV3+) for OSD using radar imagery. They concluded that the DeepLabV3+ had the best performance for OSD [19]. Finally, Liu *et al.* proposed a framework for OSD based on a spectral index-based feature selection method and one-dimensional (1-D) CNN using hyperspectral imagery. They compared the efficiency of the proposed method with those of the support vector machine (SVM) and RF classifiers. The result of OSD based on 1-D CNN had a higher accuracy compared to the other two mentioned machine learning methods [20].

As discussed, various OSD methods based on deep learning algorithms have so far been developed for RS data. However, the current algorithms usually have several limitations as discussed below.

1) Most methods have only focused on radar data. These datasets usually contain more noises and have poor temporal resolutions compared to optical imagery. They are also highly dependent on the ocean wind and usually high wind speed causes false alarm (FA) rate or miss detection (MD) in OSD methods.
2) The deep learning methods which are based on encoder and decoder (e.g., U-NET and DeepLabV3+) need a large amount of training dataset and more time for training the network. However, collecting field data over large and remote ocean environments is challenging and expensive.
3) There are several deep learning methods for OSD based on optical UAV imagery. However, these images are not suitable for large ocean areas. They also have limitations in moderate to high wind speeds.
4) Some of the deep learning methods require a feature extraction (i.e., texture and spectral features) step before OSD. The quality of these features is also important and plays a key role in the accuracy of OSD. Thus, an automatic method is required to extract informative features.

Based on the above explanation, optical RS datasets have a high potential of OSD mainly due to their suitable spatial and temporal resolutions, large coverage, and simplicity for interpretation. They also provide more opportunities for OSD due to increasing the availability of different types of optical RS datasets. In this study, a new OSD framework based on a deep learning algorithm was proposed. The proposed method was based on multiscale multidimensional residual CNN architecture. The main objective of this study was to develop a novel OSD framework with the following properties.

1) The proposed framework was an end-to-end system and does not require an additional process for OSD (e.g., feature extraction or feature selection). Additionally, the proposed framework employed multiscale kernel convolution to increase the efficiency of the network in extracting deep features.
2) The proposed framework utilized multidimensional kernel convolution (1-D and 2-D). A combination of these kernels improved the speed of the network and reduced the parameters of the network compared to using only 2-D kernel convolution. Moreover, the proposed method utilized the residual block to provide the accuracy corresponding to architecture with a high number of convolution blocks.

In summary, the main novelty of the network compared to other state-of-the-art deep learning frameworks are as follows: 1) utilizing multidimensional convolution blocks and combining them with residual blocks to reduce computational complexity and number of the network hyperparameters, residual block allows the gradient to be directly back-propagated to earlier layers to prevent from vanishing or exploding gradient problems; 2) increasing the robustness of the network against the variations of scale by utilizing multiscale blocks. These advantages of the proposed method resulted in better accuracy for OSD compared to other algorithms

II. Case Study and Data

A. Oil Spill Case Study

The largest marine OS in history occurred in the Gulf of Mexico on April 20, 2010 [21]. The oil leaked from the Macondo well for nearly three months and continued to flow until mid-July, 2010. This OS was used as a case study to investigate the potential of the proposed method.

B. Satellite Data

In this study, Landsat-5 imagery was employed to produce OS maps using the proposed method. Landsat-5 carries the thematic mapper (TM) sensor and was launched in March 1984. TM contains seven spectral bands of blue, green, red, near-infrared, shortwave infrared 1, shortwave infrared 2, and thermal.

Two multispectral Landsat-5 images acquired on May 9, 2010 (Dataset #1) and June 26, 2010 (Dataset #2) were utilized in this study [see Fig. 1(a) and (b)]. As clear, Dataset #2 includes clouds to investigate the potential of the proposed method for OSD in more complicated cases.

C. Reference Data

The reference data were generated based on visual interpretation and different reports about the OS in the study area. Figs. 5(f) and 6(f) respectively illustrate the extent of the OS and Non-OS (NOS) areas for Dataset #1 and Dataset #2. Table I also provides the number of generated reference samples for the OS and NOS classes from these two datasets. The generated samples were divided into three groups of training, test, and validation. It should be noted that the clouds were also selected as NOS areas (see Fig. 1).

III. Method

The proposed OSD framework in this study was based on multiscale multidimensional residual CNN (see Fig. 2). After preprocessing satellite imagery, the training and validation

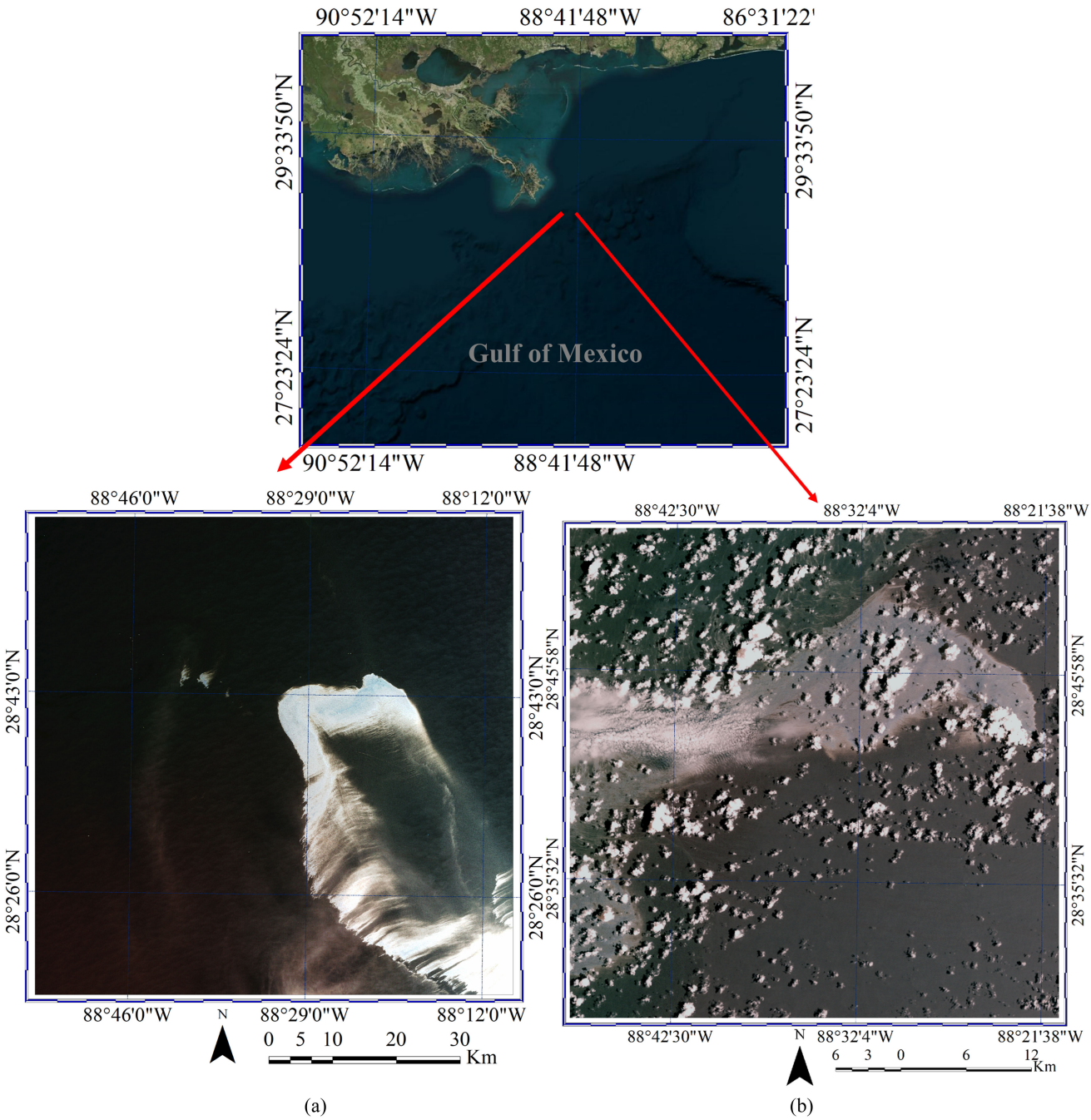


Fig. 1. Study area: the true-color composites of (a) Dataset #1 and (b) Dataset #2, which were acquired on May 9, 2010, and June 26, 2010, respectively.

TABLE I
NUMBER OF REFERENCE SAMPLES FOR THE OS AND NOS CLASSES

Datasets	Class	Number of Samples (%)	Training	Validation	Test
Dataset #1	NOS	25,339	16,470	3,802	5,067
	OS	24,954	16,220	3,743	4,990
Dataset #2	NOS	25,126	16,331	3,770	5,025
	OS	24,499	15,924	3,676	4,899

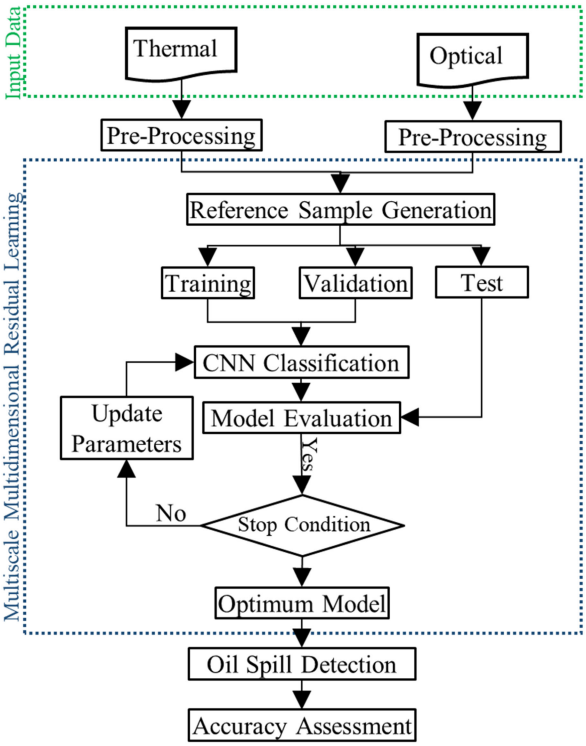


Fig. 2. Overview of the proposed oil spill detection (OSD) framework.

datasets were utilized for training the proposed CNN network and tuning the hyperparameter of CNN, respectively. The test data was also used for the evaluation of the model. CNN's parameters were set by initial values and, then, they were iteratively updated to obtain the optimum values. The training process of the proposed CNN-based framework was repeated until reaching the stop condition (number of iterations). The parameters, which provided the best performance on loss function, were considered as the optimum value of CNN. Thus, the optimum model was obtained based on the optimum values of the hyperparameters. The OS map was finally obtained by predicting the Landsat-5 dataset using the obtained optimum model. More detail about the proposed method is provided in the following subsections.

A. Preprocessing

Preprocessing was the first step and played a key role in the result of OSD. The preprocessing of the reflective spectral bands (i.e., blue, green, red, near-infrared, shortwave infrared 1, and shortwave infrared 2 bands) was started by converting digital number (DN) values to radiance values using (1)

$$L_\lambda = \left(\frac{L\mathrm{max}_\lambda - L\mathrm{min}_\lambda}{\mathcal{Q}\mathrm{cal}_{\max}}\right) \mathcal{Q}\mathrm{cal} + L\mathrm{min}_\lambda \tag{1}$$

where is spectral radiance at the sensor's aperture; Qcal is the quantized calibrated pixel value in DN; $\mathrm{Qcal}_{\min}$ and $\mathrm{Qcal}_{\max}$ are the minimum and maximum quantized calibrated pixel values corresponding to Lmin_λ and Lmax_λ, respectively; and Lmin_λ and Lmax_λ are the spectral at-sensor radiance that were scaled to $\mathrm{Qcal}_{\min}$ and i $\mathrm{Qcal}_{\max}$, respectively. Finally, the atmospheric correction was applied using the fast line-of-sight atmospheric analysis of spectral hypercubes (FLAASH) module in the ENVI software.

For the thermal band, the DN values were converted to brightness temperature (BT). To this end, the spectral radiance was converted to brightness temperature based on (2) [22]

$$T = \frac{K_2}{\ln\left(\frac{K_1}{L_\lambda} + 1\right)} \tag{2}$$

where T is the effective at-satellite temperature in Kelvin; K_1 and K_2 are the calibration constants; and L_λ is the spectral radiance at the sensor's aperture. Finally, the spatial resolution of the thermal band, which was originally 120 m, was resampled to 30 m to conform to the reflective spectral bands.

B. Proposed Architecture

The proposed CNN architecture was based on the multiscale multidimensional residual kernel convolution. Fig. 3 illustrates the type of convolutional blocks utilized in the proposed CNN architecture. These blocks had two main differences which were related to the existence of the residual block and the dimensions of the convolution layers (i.e., 2-D or 1-D).

First, the proposed architecture extracted shallow deep features based on the 2-D-multiscale convolution block. This block contained three convolution layers with different kernel sizes of 5×5, 3×3, and 1×1. The extracted deep features were concatenated and, then, were transformed into another block. The second block was a 2-D-multiscale residual block that had four convolution layers. In this block, the deep features were extracted using three convolution layers with different kernel sizes, which were later concatenated. Subsequently, the output of the convolution block was added to the input deep features of this layer. The proposed architecture extracted the deep features using five 2-D-convolution blocks. In fact, this process was repeated using five 2-D-multiscale residual blocks to extract deeper features.

Then, the deep features were extracted using three 1-D convolution blocks. The main advantages of 1-D convolution blocks were increasing the speed of the training and extracting the deep features [4].

The last layer was a discriminative and prediction layer that had a fully connected layer and a soft-max layer. The fully connected layer investigated the extracted deep features from the previous layer and changed them into a single vector layer [12], [23]. Then, the dropout layer was utilized to randomly remove multiple neurons. This process helped to prevent overfitting [24]. Finally, the output of the dropout layer was transformed to decide the label of the input layer (i.e., OS or NOS).

Fig. 4(a) demonstrates the structure of the shallow deep feature extraction block. In this block, the residual block was not used**.** Fig. 4(b) and (c) also present the mechanism of the 2-D-multiscale residual convolution block and the structure of the 1-D-multiscale residual convolution block, respectively.

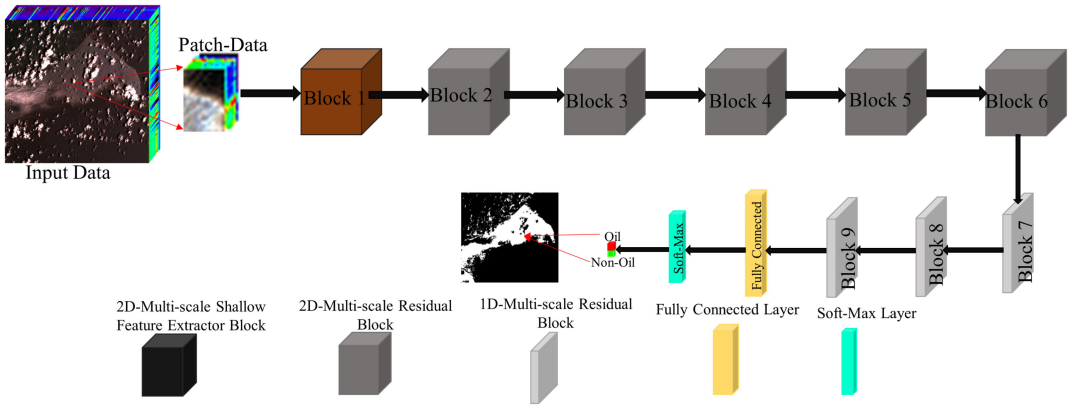


Fig. 3. Overview of the proposed multiscale multidimensional residual CNN architecture for OSD.

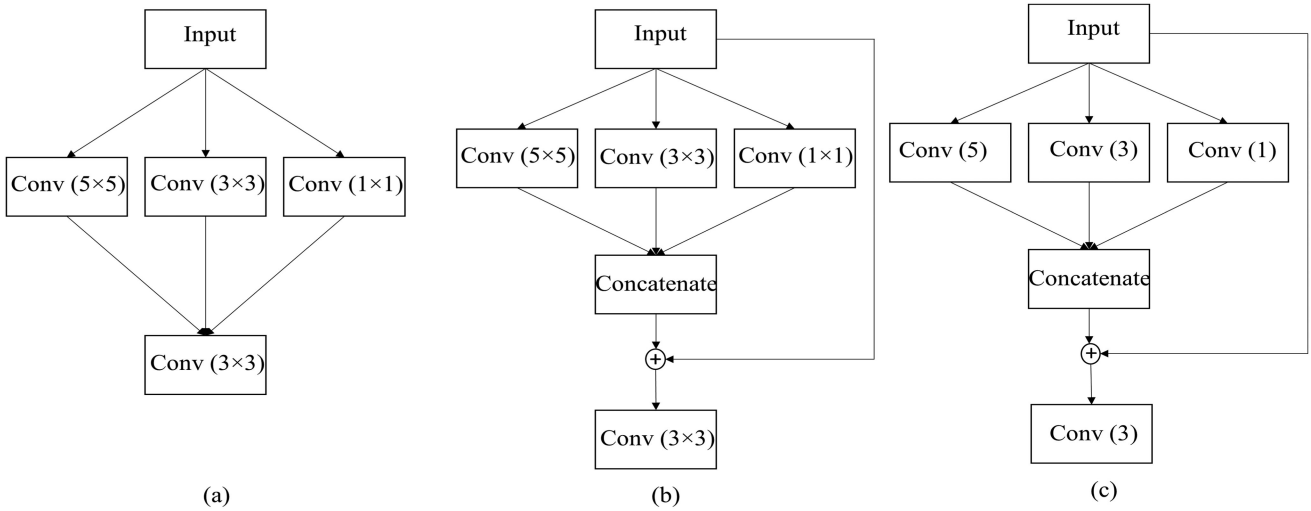


Fig. 4. Structure of the multiscale convolution block. (a) 2-D multiscale deep feature extractor convolution block, (b) 2-D multiscale residual convolution block, and (c) 1-D multiscale residual convolution block.

For a convolutional layer in the *l*th layer, the computation is expressed according to (3) [25]

$$y^l = g\left(w^l x^{l-1}\right) + b^l \tag{3}$$

where x is the neuron input from the previous layer, $l-1$; g is the activation function; w is the weighted template; and b is the bias vector of the current layer, l. In 2-D convolution, the output of the jth feature map (v) in the ith layer at the spatial location of (x, y) can be computed using (4) [25]

$$v_{i,j}^{xy} = g\left(b_{i,j} + \sum_{m}\sum_{r=0}^{R_i-1}\sum_{s=0}^{S_i-1} W_{i,j}^{r,s} v_{i-1,m}^{(x+r)(y+s)}\right) \tag{4}$$

where m is the feature cube connected to the current feature cube in the $(i-1)$th layer; W is the (r, s)th value of the kernel connected to the mth feature cube in the preceding layer; and R and S are the length and width of the convolution kernel size, respectively. In 1-D convolution, the activation value at spatial position x in the jth feature map of the ith layer is generated using (5) [4]

$$v_{i,j}^{x} = g\left(b_{i,j} + \sum_{m}\sum_{s=0}^{S_i-1} W_{i,j,m}^{r} V_{i-1,m}^{(x+r)}\right). \tag{5}$$

Although some of the CNN architectures are based on only 2-D kernel convolution, the multidimensional convolution kernels were employed in this study to improve the efficiency of CNN. The convolution layer contained the batch normalization and activation function. The activation function was Sigmoid [(6)] in this study

$$f\ (x) = \frac{e^x}{1+e^x}. \tag{6}$$

The Adam optimizer [26], which is an adaptive learning rate optimization algorithm, was used as the optimization algorithm of the proposed CNN's hyperparameter in this study [26]. Moreover, the Kullback-Leibler (KL) divergence was used as a cost function to measure the relative distance between two probability distributions [y and $p(y)$] [10]. This cost function is

TABLE II
CONFUSION MATRIX FOR THE OS AND NOS CLASSES

		Predicted OS	Predicted NOS
Actual	OS	TP	FN
	NOS	FP	TN

defined as follows:

$$D_{\mathrm{KL}}\ (y \parallel p\,(y)) = \sum_{i=1}^{N} y\log\left(\frac{y}{p\,(y)}\right) \tag{7}$$

where y is related to the probability distribution of the original output; $p(y)$ is the predicted probability observation, and N is the number of classes.

C. Accuracy Assessment and Model Comparison

In this study, the confusion matrix was used for statistical accuracy assessment of the OS maps produced using the proposed method. As illustrated in Table II, a confusion matrix has four components of the true positive (TP), which was related to OS pixels, true negative (TN), which was related to NOS pixels, false positive (FP), and false negative (FN). Various accuracy matrices were generated based on these four components and were used to report the accuracy levels. As demonstrated in Table III, these matrices were overall accuracy (OA), F1-score (FS), FA, kappa coefficient (KC), precision (PCC), recall, MD, and specificity.

(See Table II for the components of each equation)

Besides the statistical accuracy assessment of the proposed CNN algorithm, it was also compared with two nondeep-learning methods of the RF [27] and XGboost [32]. Furthermore, the results were compared with those of the two advanced deep learning methods, including 2-D CNN and 2-D Residual-CNN (2D-Res-CNN).

IV. RESULTS

A. Optimum Tuning Parameters Selection

The proposed method and other classification methods have several tuning parameters, the optimum values of which should be determined to obtain the highest possible classification accuracies. In this study, a grid search (GS) algorithm was employed for this purpose [22]. Table IV provides the required tuning parameters for each classifier along with their optimum values obtained using the GS algorithm.

B. OS Map for Dataset #1

The results of OSD using the proposed method and other machine learning algorithms for Dataset #1 are illustrated in Fig. 5. Most of the OS pixels were correctly detected by all classifiers. However, the main difference was in areas where OS was mixed with water. As clear, more errors in the OS maps are related to the FP pixels that are indicated by the red color.

The results of the statistical accuracy assessments of different methods for Dataset #1 are also provided in Table VI. Based on the results, all methods had OAs of more than 98%. Most algorithms had high Recall rates compared to the specificity, PCC, and F1-core rates. Therefore, the performance of most methods in detecting OS pixels was better than detecting NOS pixels. Additionally, the incorrectly classified pixels were related to FP pixels, where the FA rates were high (more than 1.83%). Although the 2D-CNN had a better performance in terms of PCC, FA, and specificity indices compared to the proposed method, it had a high MD rate and a low performance regarding the other accuracy indices, such as OA, and KC. Overall, the proposed method had better accuracy compared to other methods.

C. OS Map for Dataset #2

Fig. 6 shows the results of the OSD for Dataset #2 using different machine learning algorithms. As clear, the results were significantly different than those of Dataset #1. The main challenge was related to the clouds in this data. The spectral signature of clouds is similar to that of the OS class and, thus, the OSD algorithms had considerable errors. For example, there are many FN pixels in the results of all methods. However, this error was minimum in the proposed CNN method [Fig. 6(e)].

The results of statistical accuracy assessments for Dataset #2 are provided in Table VI. The proposed method had the best performance considering most of the accuracy indices. The MD rates of all methods were considerably high except for the proposed method. The BA for all methods was around 91%, however, it was very high (95%) for the proposed CNN method. Although the 2-D Res-CNN provided better performance in the detection of the TN pixels, it had lower accuracy in the detection of the TP pixels.

V. DISCUSSION

A. Sensitivity Analysis of the Parameters

The CNN methods have hyperparameters that need to be optimized. Among these parameters, the patch-size has a key role in the performance of the CNN method [34]. To this end, the performance of the proposed CNN method was investigated using different patch-sizes. As an example, Table VII shows the effects of the patch-size (11×11, 13×13, 15×15, and 17×17) in the result of OSD for Dataset #2. Based on the results, increasing the patch-size generally improved the result of OSD and reduced the MD and FA rates. However, after the patch-size of 15×15, the performance of the CNN algorithm decreased. Based on the results, the patch-size affected the performance of CNN by 1.18% in OA.

The initializer of the hyperparameters also plays a key role in the speed of convergence of the network. In this study, two methods of random and Xavir Initializers were investigated. It was observed that the Xavir Initializer had a better performance in increasing the speed of the training of the network.

TABLE III
METRICS THAT WERE USED FOR THE ACCURACY ASSESSMENT OF THE OS MAPS

Accuracy index	Equation
Overall Accuracy (OA)	$\frac{(\mathrm{TN}+\mathrm{TP})}{(\mathrm{TN}+\mathrm{TP}+\mathrm{FP}+\mathrm{FN})}$
F1-Score (FS)	$\frac{2\times \mathrm{TP}}{(2\times \mathrm{TP}+\mathrm{FP}+\mathrm{FN})}$
False Alarm (FA)	$\frac{(\mathrm{FP})}{(\mathrm{TN}+\mathrm{FP})}$
Precision (PCC)	$\frac{(\mathrm{TN})}{(\mathrm{TN}+\mathrm{FP})}$
Kappa Coefficient (KC)	$\frac{\mathrm{PCC}-\frac{(TP+FP)\times(TP+FN)+(FN+TN)\times(FP+TN)}{(\mathrm{TN}+\mathrm{TP}+\mathrm{FP}+\mathrm{FN})^2}}{1-\frac{(TP+FP)\times(TP+FN)+(FN+TN)\times(FP+TN)}{(\mathrm{TN}+\mathrm{TP}+\mathrm{FP}+\mathrm{FN})^2}}$
Recall	$\frac{(\mathrm{TP})}{(\mathrm{TP}+\mathrm{FN})}$
Miss-Detection (MD)	$\frac{(\mathrm{FN})}{(\mathrm{TP}+\mathrm{FN})}$
Specificity	$\frac{(\mathrm{TN})}{(\mathrm{TN}+\mathrm{FP})}$

TABLE IV
OPTIMUM VALUES OF THE TUNING PARAMETERS OF DIFFERENT CLASSIFIERS

Classifier	Optimum Value
RF	Number Of Trees=105, Number of Features to Split Each Node=3
XGBoost	Nrounds = 500, Max-Depth = 5, Eta = 0.03, Gamma = 0, Min-Child-Weight = 1, Subsample = 0.8, Colsample-Bytree = 0.8.
Proposed CNN	Patch-Size= 15×15×7, Epochs= 10^3, Weight Initializer= Xavir Initializer [33], Dropout-Rate=0.4, Number of Neurons At Full Connected Layer=1500, Initial Learning= 10^{-4}, Epsilon-Value =10^{-9}, Mini-Batch-Size=550, Loss-Function= Kullback-Leibler Divergence, and Optimizers= Adam.

TABLE V
ACCURACY MEASURES OF DIFFERENT METHODS FOR OSD IN DATASET #1

Metric	RF	XGB	2D-CNN	2D- Res-CNN	Proposed CNN
OA (%)	96.37	96.61	97.78	98.07	98.23
PCC (%)	77.95	79.04	98.92	87.19	88.12
MD (%)	2.40	2.03	17.42	1.46	1.32
FA (%)	3.80	3.58	0.12	1.99	1.83
FS (%)	86.67	87.49	90.02	92.52	93.10
BA (%)	96.90	97.20	91.23	98.27	98.43
Recall (%)	97.60	97.97	82.58	98.54	98.68
Specificity (%)	96.20	96.42	99.88	98.01	98.17
KC (%)	0.846	0.855	0.892	0.910	92.10

(OA: overall accuracy, PCC: precision, MD: miss-detection, FA: False Alarm, FS: F1-Score, BA: balanced accuracy, KC: kappa coefficient, RF: random forest, XGB: XGBOOST, CNN: convolution neural network, RES: residual, 2-D: two-dimensional)

TABLE VI
ACCURACY MEASURES OF DIFFERENT METHODS FOR OSD IN DATASET #2

Metric	RF	XGB	2D-CNN	2D-Res-CNN	Proposed CNN
OA (%)	85.16	88.08	94.39	94.68	95.12
PCC (%)	58.27	64.38	82.10	87.03	82.99
MD (%)	14.50	12.30	8.71	14.39	5.55
FA (%)	14.93	11.83	4.85	3.11	4.72
F1-Score (%)	69.30	74.25	86.45	86.31	88.35
BA (%)	85.28	87.94	93.22	91.25	94.87
Recall (%)	85.50	87.70	91.29	85.61	94.45
Specificity (%)	85.07	88.17	95.15	96.89	95.28
KC (%)	0.599	0.667	0.829	0.831	0.852

(OA: overall accuracy, PCC: precision, MD: miss-detection, FA: False Alarm, FS: F1-Score, BA: balanced accuracy, KC: kappa coefficient, KNN: K-Nearest Neighbors, MLP: Multilayer Perceptron, RF: random forest, SVM: support vector machine, XGB: XGBOOST, CNN: convolution neural network).

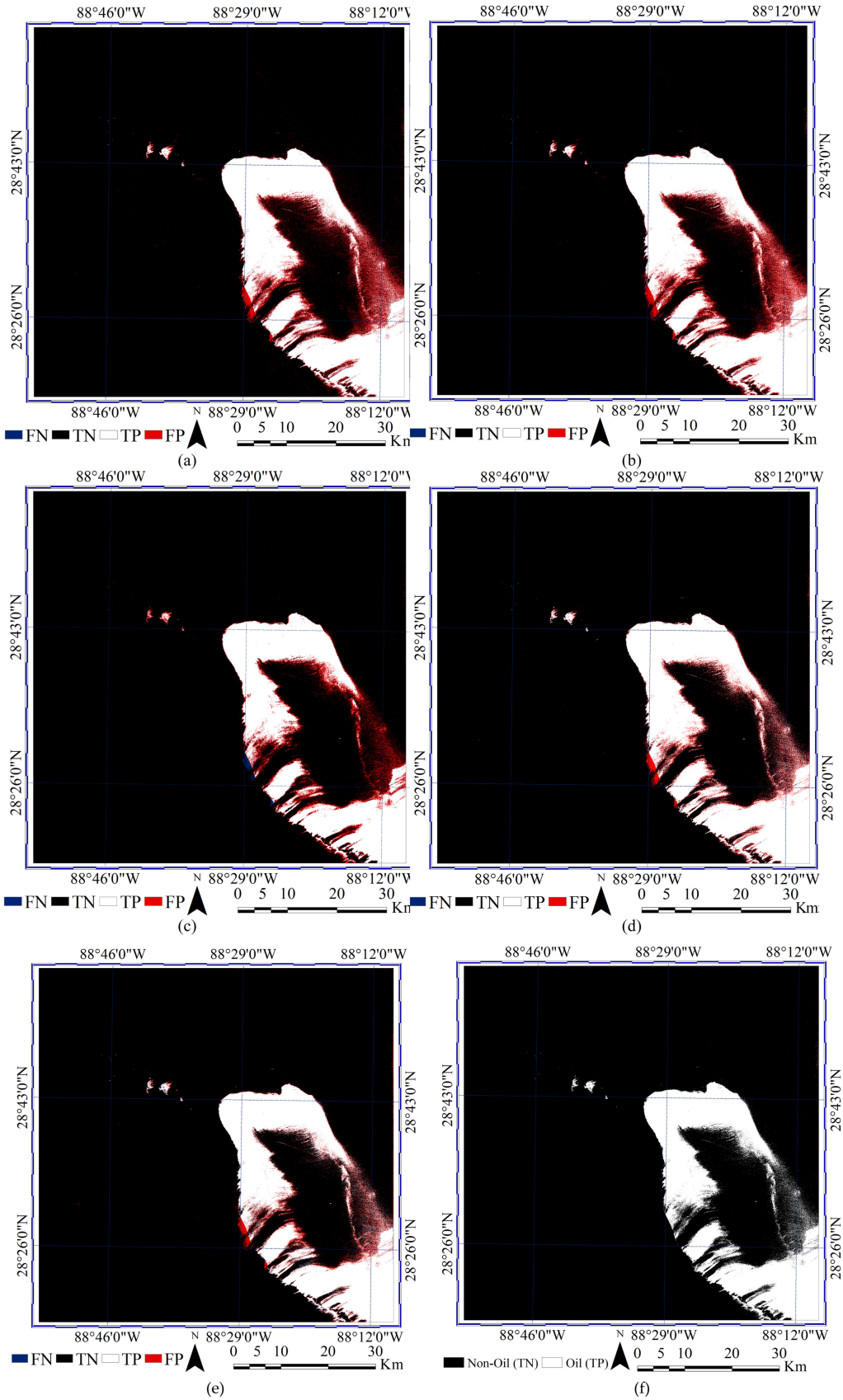


Fig. 5. OS maps for Dataset #1 using (a) RF, (b) GXGB, (c) 2-D CNN, (d) 2-D ResCNN, (e) proposed method, and (f) ground truth. White (TP) and black (TN) colors show the OS and NOS classes, respectively.

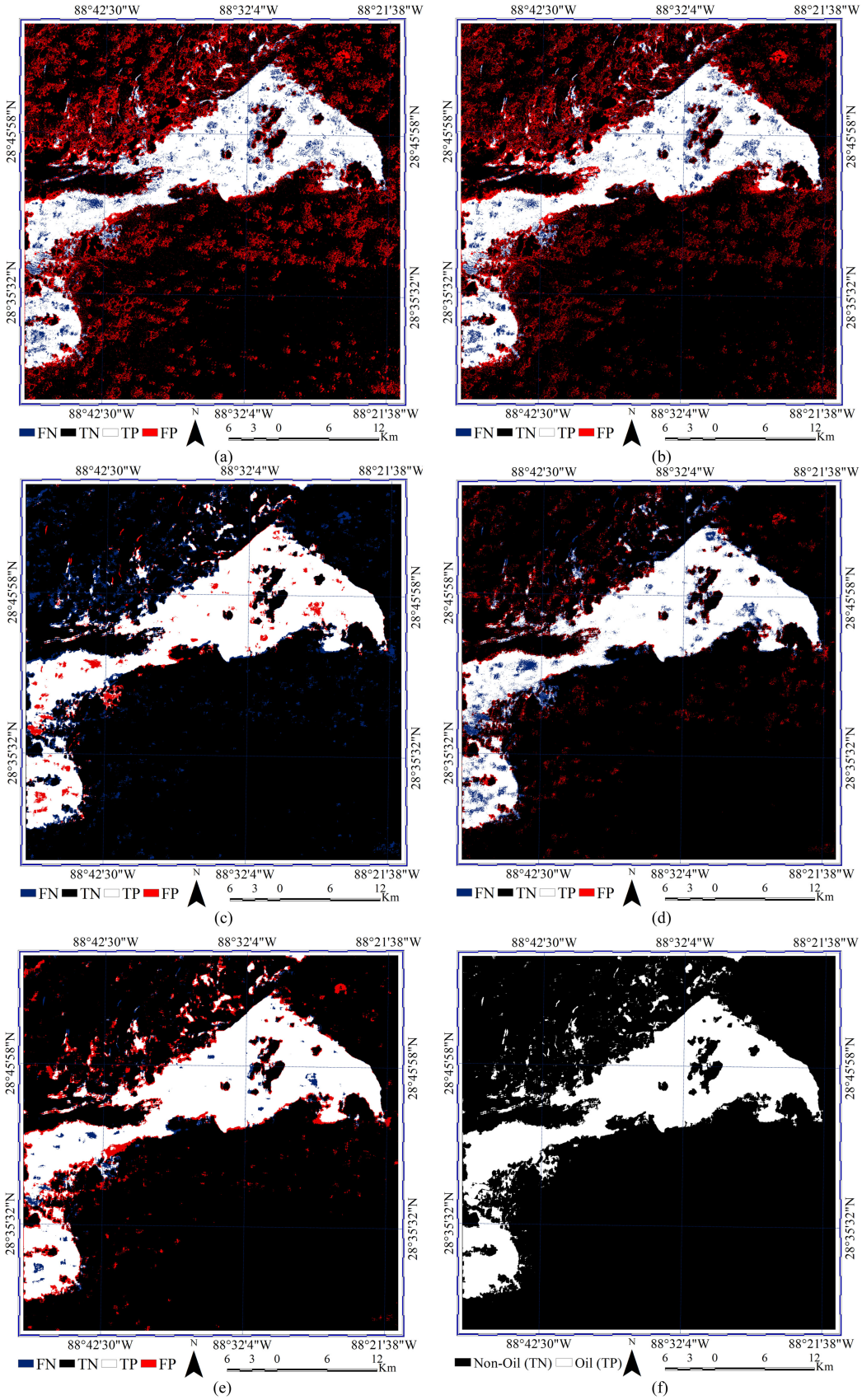


Fig. 6. OS maps for Dataset #2 using (a) RF, (b) GXGB, (c) 2-D CNN, (d) 2-D ResCNN, (e) proposed method, and (f) ground truth. White (TP) and black (TN) colors show the OS and NOS classes, respectively.

TABLE VII
EFFECTS OF INCREASING THE PATCH-SIZE ON THE RESULT OF OSD USING THE PROPOSED METHOD APPLIED TO DATASET #2

Patch-Size	OA (%)	KC (%)
Patch-Size 11×11	93.93	0.819
Patch-Size 13×13	94.43	0.833
Patch-Size 15×15	95.11	0.852
Patch-Size 17×17	94.89	0.846

TABLE VIII
EFFECTS OF INCREASING THE NUMBER OF SPECTRAL BANDS ON THE RESULT OF THE PROPOSED OSD METHOD APPLIED TO DATASET #2

Spectral Band	OA	KC
R,G,B	92.71	0.786
R,G, B, NIR	93.95	0.820
R,G, B, NIR,SWIR-1	94.39	0.830
R,G, B, NIR,SWIR-1, SWIR-2	94.58	0.838
R,G, B, NIR,SWIR-1, SWIR-2,Thermal	95.12	0.852

It was also observed that including more convolution layers in the proposed network increased the performance of the network and, consequently, the MD and FA rates were decreased. However, it also increased training time due to increasing the number of hyperparameters.

B. Spectral Bands

The number of spectral bands was another factor that could affect the result of OSD using the proposed method. As an example, Table VIII presents the accuracy levels using different patch-sizes and the effects of increasing the spectral bands on the accuracy obtained for Dataset #2. It was observed that increasing the spectral bands improved the performance of the proposed method using different patch-sizes. The lowest accuracy was obtained using three spectral bands (red, blue, and green) in the patch-size of 11×11 (OA $= 92.28\%$), while the highest accuracy was obtained for the patch-size of 15×15 with all spectral bands.

C. Cloud Cover and Noise Condition

Clouds and their shadows are one of the main challenges in OSD using optical RS datasets. Dataset #1 contained fewer cloudy regions compared to Dataset #2. Thus, most classifiers had difficulties in correctly delineating clouds from OS areas in Dataset #2. Although the thermal band played a key role in the detection of clouds, many OS pixels were wrongly identified due to the spectral similarity of these two targets. However, as demonstrated in Fig. 6 and Table VI, the proposed CNN algorithm was relatively robust in delineating clouds from OS areas.

Based on the results presented in Figs. 5 and 6, there were many noisy pixels in the results of the XGB and RF classifiers. This was less serious for the deep learning-based methods especially the proposed algorithm. Therefore, it was concluded that the proposed method was more robust to noise and environmental conditions.

D. Deep Feature Extraction

Designing suitable architecture in deep learning-based methods is of great importance. The results of OSD from two deep learning methods (2-D CNN, and 2D Res-CNN) were investigated in this study. Although these methods resulted in high FA, PCC, and specificity values, they had a low performance regarding other accuracy measures. For example, these methods provided the FA values under 1% for Dataset#1; however, the MD rate was more than 17%. It is also worth noting that although the sizes of the sample data for both OS and NOS classes were similar (25 000 pixels), the two deep learning methods provided different results. However, it was expected that they had low errors in terms of MD and FA. In contrast, the proposed deep learning method had a higher accuracy in this regard, where it provided low MD and FA rates for both datasets.

The proposed CNN method was able to extract the deep features in the multiscale convolution filters using multidimensional kernel convolution. This method took the advantage of deep features into account to investigate both spatial and spectral features using different convolution kernels. It is worth noting that there are many other feature extraction methods, such as texture features or spectral attribute profiles that manually extract features. However, one of the main challenges of manual feature extraction methods is the fact that feature extraction/selection is a time-consuming process. However, the deep learning methods automatically extract spatial/spectral features in a time-efficient approach.

Additionally, there are many other state-of-the-art deep learning-based architectures, such as U-Net, Deep-LabV3+, and Seg-Net, which need a large amount of training data, the collection of which is expensive and time-consuming for OSD applications. Moreover, the training of these deep learning architectures is time-consuming. However, one of the advantages of the proposed method was no need for a large number of samples, and its training required a lower time compared to the above-mentioned architectures.

Finally, although it took more time to train the proposed network compared to other deep learning methods (i.e., one more hour), the proposed method was selected as the preferred algorithm for OSD due to providing the highest accuracy and more robust results

VI. CONCLUSION

In this study, a new deep learning method was proposed to produce accurate OS maps using optical RS imagery. The main limitations of the current deep learning methods for OSD were initially discussed. Then, a new OSD method was developed to address those issues. Additionally, this research investigated the effect of increasing the number of spectral bands and patch size on accuracy. It was observed that increasing the number of spectral bands reduced FA rates. In summary, the proposed method improved the quality of OSD results in two different datasets. Overall, it was concluded that the proposed deep learning method had the following advantages over the currently available methods: 1) higher accuracy and lower errors for OSD, 2) higher efficiency in delineating OS from clouds in

optical images, 3) lower MD and FA rates, 4) ability to provide excellent performance even using three spectral bands, and 5) taking the advantages of the residual block, multiscale, and multidimensional kernel convolutions.

REFERENCES

[1] D. Song *et al.*, "A novel marine oil spillage identification scheme based on convolution neural network feature extraction from fully polarimetric SAR imagery," *IEEE Access*, vol. 8, pp. 59801–59820, 2020.

[2] B. G. Gautama, N. Longépé, R. Fablet, and G. Mercier, "Assimilative 2-D Lagrangian transport model for the estimation of oil leakage parameters from SAR images: Application to the Montara oil spill," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 11, pp. 4962–4969, Nov. 2016.

[3] Y. Cheng *et al.*, "Monitoring of oil spill trajectories with COSMO-SkyMed X-band SAR images and model simulation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 7, pp. 2895–2901, Jul. 2014.

[4] S. T. Seydi, M. Hasanlou, and M. Amani, "A new end-to-end multi-dimensional CNN framework for land cover/land use change detection in multi-source remote sensing datasets," *Remote Sens.*, vol. 12, no. 12, 2020, Art. no. 2010.

[5] S. Sun and C. Hu, "The challenges of interpreting oil–water spatial and spectral contrasts for the estimation of oil thickness: Examples from satellite and airborne measurements of the deepwater horizon oil spill," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2643–2658, May 2018.

[6] J. Fan, F. Zhang, D. Zhao, and J. Wang, "Oil spill monitoring based on SAR remote sensing imagery," *Aquatic Procedia*, vol. 3, pp. 112–118, 2015.

[7] M. Amani *et al.*, "Application of Google Earth Engine cloud computing platform, sentinel imagery, and neural networks for crop mapping in Canada," *Remote Sens.*, vol. 12, no. 21, 2020, Art. no. 3561.

[8] M. Amani, S. Mahdavi, and O. Berard, "Supervised wetland classification using high spatial resolution optical, SAR, and LiDAR imagery," *J. Appl. Remote Sens.*, vol. 14, no. 2, 2020, Art. no. 024502.

[9] S. Mahdavi, B. Salehi, M. Amani, J. Granger, B. Brisco, and W. Huang, "A dynamic classification scheme for mapping spectrally similar classes: Application to wetland classification," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 83, 2019, Art. no. 101914.

[10] S. Berisha, F. F. Shahraki, D. Mayerich, and S. Prasad, "Deep learning for hyperspectral image analysis, Part I: Theory and algorithms," in *Hyperspectral Image Analysis*. Berlin, Germany: Springer, 2020, pp. 37–68.

[11] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning,* vol. 2, Cambridge, MA, USA: MIT Press, 2016.

[12] V. Liermann, S. Li, and N. Schaudinnus, "Deep learning: An introduction," in *The Impact of Digital Transformation and FinTech on the Finance Professional*. Berlin, Germany: Springer, 2019, pp. 305–340.

[13] Q. Yuan *et al.*, "Deep learning in environmental remote sensing: Achievements and challenges," *Remote Sens. Environ.*, vol. 241, 2020, Art. no. 111716.

[14] X. Yu, H. Zhang, C. Luo, H. Qi, and P. Ren, "Oil spill segmentation via adversarial *f*-divergence learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 4973–4988, Sep. 2018.

[15] S. Yekeen and A. Balogun, "Automated marine oil spill detection using deep learning instance segmentation model," *Int. Arch. Photogrammetry, Remote Sens. Spatial Inf. Sci.*, vol. 43, pp. 1271–1276, 2020.

[16] Z. Jiao, G. Jia, and Y. Cai, "A new approach to oil spill detection that combines deep learning with unmanned aerial vehicles," *Comput. Ind. Eng.*, vol. 135, pp. 1300–1311, 2019.

[17] F. M. Bianchi, M. M. Espeseth, and N. Borch, "Large-scale detection and categorization of oil spills from SAR images with deep learning," *Remote Sens.*, vol. 12, no. 14, 2020, Art. no. 2260.

[18] J. Zhang, H. Feng, Q. Luo, Y. Li, J. Wei, and J. Li, "Oil spill detection in quad-polarimetric SAR images using an advanced convolutional neural network based on superpixel model," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 944.

[19] M. Krestenitis, G. Orfanidis, K. Ioannidis, K. Avgerinakis, S. Vrochidis, and I. Kompatsiaris, "Oil spill identification from satellite images using deep neural networks," *Remote Sens.*, vol. 11, no. 15, 2019, Art. no. 1762.

[20] B. Liu, Y. Li, G. Li, and A. Liu, "A spectral feature based convolutional neural network for classification of sea surface oil spill," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 4, 2019, Art. no. 160.

[21] R. M. Abbriano *et al.*, "Deepwater horizon oil spill: A review of the planktonic response," *Oceanography*, vol. 24, no. 3, pp. 294–301, 2011.

[22] G. Chander, B. L. Markham, and D. L. Helder, "Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors," *Remote Sens. Environ.*, vol. 113, no. 5, pp. 893–903, 2009.

[23] K. Huang, A. Hussain, Q.-F. Wang, and R. Zhang, *Deep Learning: Fundamentals, Theory and Applications*. Berlin, Germany: Springer, 2019.

[24] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.

[25] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, "HybridSN: Exploring 3-D–2-D CNN feature hierarchy for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Feb. 2020.

[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Accessed: Nov. 04, 2021, Jan., 2017, [Online]. Available: http://arxiv.org/abs/1412.6980.

[27] S. Tong, X. Liu, Q. Chen, Z. Zhang, and G. Xie, "Multi-feature based ocean oil spill detection for polarimetric sar data using random forest and the self-similarity parameter," *Remote Sens.*, vol. 11, no. 4, 2019, Art. no. 451.

[28] T. Chen and S. Lu, "Subcategory-aware feature selection and SVM optimization for automatic aerial image-based oil spill inspection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5264–5273, Sep. 2017.

[29] A. Matkan, M. Hajeb, and Z. Azarakhsh, "Oil spill detection from SAR image using SVM based classification," *Int. Arch. Photogrammetry Remote Sens. Spatial Inf. Sci. SMPR*, vol. 1, pp. 55–60, 2013.

[30] A. Taravat and F. Del Frate, "Development of band ratioing algorithms and neural networks to detection of oil spills using Landsat ETM+ data," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, 2012, Art. no. 107.

[31] Y. Cao, L. Xu, and D. Clausi, "Exploring the potential of active learning for automatic identification of marine oil spills using 10-Year (2004–2013) RADARSAT data," *Remote Sens.*, vol. 9, no. 10, 2017, Art. no. 1041.

[32] A.-B. Salberg and S. Ø. Larsen, "Classification of ocean surface slicks in simulated hybrid-polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 12, pp. 7062–7073, Dec. 2018.

[33] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.

[34] J. Hamwood, D. Alonso-Caneiro, S. A. Read, S. J. Vincent, and M. J. Collins, "Effect of patch size and network architecture on a convolutional neural network approach for automatic segmentation of OCT retinal layers," *Biomed. Opt. Exp.*, vol. 9, no. 7, pp. 3049–3066, 2018.

[35] K. N. Topouzelis, "Oil spill detection by SAR images: Dark formation detection, feature extraction and classification algorithms," *Sensors*, vol. 8, no. 10, pp. 6642–6659, 2008.

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

[37] S. Chehresa, A. Amirkhani, G.-A. Rezairad, and M. R. Mosavi, "Optimum features selection for oil spill detection in SAR image," *J. Indian Soc. Remote Sens.*, vol. 44, no. 5, pp. 775–787, 2016.

[38] S. Liu, M. Chi, Y. Zou, A. Samat, J. A. Benediktsson, and A. Plaza, "Oil spill detection via multitemporal optical remote sensing images: A change detection perspective," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 324–328, Mar. 2017.

[39] M.-S. Lee, K.-A. Park, H.-R. Lee, J.-J. Park, C.-K. Kang, and M. Lee, "Detection and dispersion of thick and film-like oil spills in a coastal bay using satellite optical images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 11, pp. 5139–5150, Nov. 2016.

[40] A. Pisano, "Development of oil spill detection techniques for satellite optical sensors and their application to monitor oil spill discharge in Mediterranean sea," Doctoral Thesis, Alma Mater Studiorum - Universit di Bologna, 2011. doi: 10.6092/unibo/amsdottorato/3520

Seyd Teymoor Seydi received the B.Eng. degree in surveying and geomatics engineering from the University of Shahid Rajaei, Iran in 2015, and the M.Eng. degree in remote sensing from the University of Tehran, Iran in 2018. He is currently working toward the Ph.D. degree in remote sensing with the University of Tehran, Iran.

His research interests include multitemporal multispectral/hyperspectral and SAR remote sensing processing and classification, and machine learning algorithms.

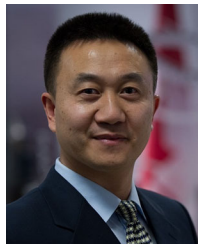
Mahdi Hasanlou (Member, IEEE) received the B.Sc. degree in surveying and geomatics engineering, the M.Sc. degree in remote sensing, and the Ph.D. degree in remote sensing from the University of Tehran, Tehran, Iran, in 2003, 2006, and 2013, respectively.

Since 2013, he has been an Associate Professor with the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, where he is the Head with the Remote Sensing Laboratory and also the Head with the Remote Sensing and Photogrammetry Group. His research interests include hyperspectral, thermal, optical, and SAR remote sensing for urban and agro-environmental applications.

Meisam Amani (Senior Member, IEEE) received the B.Eng. degree in geomatics engineering from the University of Tabriz, Tabriz, Iran, in 2012, the M.Eng. degree in remote sensing engineering from K.N. Toosi University of Technology, Tehran, Iran, in 2014, and the Ph.D. degree in electrical engineering from Memorial University of Newfoundland, St. John's, NL, Canada, in 2018.

He is currently a Senior Remote Sensing Engineer and the Key Specialty Leader of Data Analytics with a global consulting and engineering company, Wood PLC, Ottawa, ON, Canada, where he manages and leads various industrial, governmental, and academic remote sensing projects worldwide. Over the past 11 years, he has worked on different applications of remote sensing, including but not limited to land cover/land use classification, soil moisture estimation, drought monitoring, water quality assessment, watershed management, power/transmission line monitoring, fog detection and nowcasting, and ocean wind estimation. To do these, he has utilized various remote sensing datasets (e.g., UAV, optical, LiDAR, SAR, scatterometer, radiometer, and altimeter) along with different machine learning and big data processing algorithms. A list of his research works, including over 50 peer-reviewed journal and conference papers.

Dr. Amani is an Associate Editor for IEEE JSTARS and the Lead Guest Editor for a special issue in the *Remote Sensing* journal. He also serves as a regular Reviewer in about 15 international remote sensing journals. He was the recipient of the prestigious Professional Engineers and Geoscientists Newfoundland and Labrador Environmental Award in 2020 for his contribution to wetland mapping in Canada using advanced machine learning and big data processing algorithms.



Weimin Huang (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in radio physics from Wuhan University, Wuhan, China, in 1995, 1997, and 2001, respectively, and the M.Eng. degree in electrical engineering from Memorial University of Newfoundland, St. John's, NL, Canada, in 2004.

From 2008 to 2010, he was a Design Engineer with Rutter Technologies, St. John's, NL, Canada. Since 2010, he has been with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, where he is currently a Professor. He has authored more than 250 research papers. His research interests include the mapping of oceanic surface parameters via high-frequency ground wave radar, X-band marine radar, and global navigation satellite systems.

Dr. Huang has been a Technical Program Committee Member. He served as a Technical Program Co-chair of the IEEE Newfoundland Electrical and Computer Engineering Conference, in 2012 and 2013. He is currently an Area Editor for the *IEEE Canadian Journal of Electrical and Computer Engineering*, an Associate Editor for the IEEE ACCESS, and an Editorial Board Member of *Remote Sensing*, and a Guest Editor for the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. He serves as a regular Reviewer for over 60 international journals and a Reviewer for many IEEE international conferences, such as RadarCon, International Conference on Communications, IEEE Global Communications Conference, IEEE Geoscience and Remote Sensing Society, and Oceans. He was the recipient of the Postdoctoral Fellowship from Memorial University of Newfoundland. In 2017, he was the recipient of the Discovery Accelerator Supplements Award from the Natural Sciences and Engineering Research Council of Canada. He was also the recipient of the IEEE Geoscience and Remote Sensing Society 2019 Letters Prize Paper Award.