

Received September 19, 2020, accepted October 2, 2020, date of publication October 12, 2020, date of current version October 22, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3030112

# U-Net Convolutional Networks for Mining Land Cover Classification Based on High-Resolution UAV Imagery

TUAN LINH GIANG<sup>1</sup>, KINH BAC DANG<sup>1</sup> (Member, IEEE),  
QUANG TOAN LE<sup>3</sup>, (Member, IEEE), VU GIANG NGUYEN<sup>3</sup>, (Member, IEEE),  
SI SON TONG<sup>4</sup>, (Member, IEEE), AND VAN-MANH PHAM<sup>1</sup> (Member, IEEE)

<sup>1</sup>Skymap High Technology Company Ltd., Hanoi 100000, Vietnam

<sup>2</sup>Faculty of Geography, VNU University of Science, Hanoi 100000, Vietnam

<sup>3</sup>Space Technology Institute, Vietnam Academy of Science and Technology, Hanoi 100000, Vietnam

<sup>4</sup>Vietnam Academy of Science and Technology, University of Science and Technology of Hanoi, Hanoi 100000, Vietnam

Corresponding author: Kinh Bac Dang (dangkinhbac@hus.edu.vn)

This work was supported in part by the National Program on Space Science and Technology, Ministry of Science and Technology, Vietnam, under Grant VT-UD.06/18-20.

**ABSTRACT** Mining activities are the leading cause of deforestation, land-use changes, and pollution. Land use/cover mapping in Vietnam every five years is not useful to monitor land covers in mining areas, especially in the Central Highland region. It is necessary to equip managers with a better tool to monitor and map land cover using high-resolution images. Therefore, the authors proposed using the U-Net convolutional network for land-cover classification based on multispectral Unmanned aerial vehicle (UAV) image in a mining area of Daknong province, Vietnam. An area of 0.5kmx0.8km was used for training and testing seven U-Net models using seven optimizer function types. The final U-Net model can interpret six land cover types: (1) open-case mining lands, (2) old permanent croplands, (3) young permanent croplands, (4) grasslands, (5) bare soils, (6) water bodies. As a result, two models using Nadam and Adadelta optimizer function can be used to classify six land cover types with accuracy higher than 83%, especially in open-case mining lands and polluted streams flowed out from the mining areas. The trained U-Net models can potentially update new land cover types in other mining areas towards monitoring land cover changes in real-time in the future.

**INDEX TERMS** U-Net convolutional network, unmanned aerial vehicle, deep learning, Daknong, segmentation, permanent cropland, open-cast mining, loss function, optimization.

## I. INTRODUCTION

The Central Highlands of Vietnam has more than 450,000 hectares (ha) of perennial crops, accounting for nearly 90% of the coffee area and 26% of the rubber area [1]. More than 15,792 ha of natural forests in 1990s were destroyed for the construction and mining projects, of which 5,755 ha of forest land has been converted into hydroelectricity till 2015 [2]–[4]. Kissinger *et al.* [5] and Thuy *et al.* [6] emphasized that mining and land cover conversion are two of the six leading causes of deforestation and forest degradation. Although environmental impact assessments have become a mandatory requirement since 1993, illegal mining activities have not been thoughtfully managed [2]. A real-time

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Asif<sup>1</sup>.

land use/cover monitoring machine is needed, replacing the national manual monitoring activities every five years, to quickly assess the environmental quality [7]–[10]. Accordingly, it ensures legal mineral activities in mineral exploitation areas of the Central Highlands.

Deep learning is a branch of artificial intelligence in which computers generate rules based on the raw input data [11], [12]. Based on deep learning, models can improve their performance based on past results or new data sources [13]. Deep learning provides many advantages for humans for predicting natural hazards [14], [15], and helping to make intelligent decisions in real-time without humans intervention [16]. In addition to their applications in social technology, deep learning has been used in various remote sensing analyses, such as detecting ships, turtles, or houses [17]–[19]. Primarily, most of the studies commonly

used satellite images with low to medium resolution images such as MODIS, Landsat, and Sentinel-2 [20]–[22]. However, the application of deep learning to detect land cover from high-resolution UAV images is rare, especially in mining areas.

Recently, the artificial neural network (ANN) has become a useful tool to deal with complex non-linear regression between input data that has not yet been optimized for ancient remote sensing processing techniques such as unsupervised learning, Random Forest, pixel-based and Support Vector Machine [23]–[25]. For example, Pu *et al.* [26] and Dang *et al.* [27] used Convolutional Neural Network (CNN) to classify (1) water quality of inland lakes based on Landsat-8 images; and (2) coastal types in Vietnam based on Sentinel-2 images. Ge *et al.* [28] used ANN models for lithological classification based on various moderate-resolution satellite images. Instead of analyzing satellite images based on pixel-based approaches, the ANN models consider land uses/covers as objects to distinguish their pixels, scale, spatial relations and texture information for generating homogeneous objects [29]. Therefore, developers can optimize the performance and accuracy of land cover classification models in real-time and space from big dataset, reducing the computational costs of traditional physical-based models. In addition, the models can learn or upgrade from previous computation if the developers provide new input data from the real time [30].

The U-Net concept suggested by Ronneberger *et al.* [31] is an updated Fully Convolutional Network model with a symmetrical U-shaped design including a symmetrical contracting and expanding paths. While the contracting path acquires the context, the expanding path predict accurate localization of segmented objects. The U-Net integrates low-level features with comprehensive spatial information (in the contracting path) with high-level features with semantic information (in the expanding path) to boost segmentation precision [32]. Accordingly, successive convolution layers are trained to create a more reliable output based on this knowledge. Based on semantic segmentation approaches, the U-Net models can learn effectively spatial relations between land-cover classes [33]. Therefore, they can better classify objects in whole images, especially in the case of labeled data scarcity, compared to the pixel-based models. Additionally, once the predicted latter are discontinuities due to noises in the mark data and these models implicitly run a regularization process as a post-processing smoothing step [34].

Although the first U-Net models were introduced in 2015, the first models for land-cover classification have been published since 2019, such as land covers in urban [35]–[37], forest [38], coastal areas [39] and sea surface. Most of studies developed the original U-Net with the use of secondary satellite images [40]. For example, Stoian *et al.* [41] have started to use 10m-resolution Sentinel-2 satellite images for land use/cover mapping in many polygons in France. Their input image data is commonly dependent on the satellites, leading to difficulties to monitor land covers in a particular mining

region. An active image-recorder machine such as Unmanned aerial vehicle (UAV) will be an alternative option to classify mining land covers in the real time, especially if this machine can integrate with a deep learning model. In this research, the authors use the U-Net framework to train a mining land cover prediction model based on the multi-spectral UAV image in a particular area of Daknong province, Vietnam. This study will prove the potential of the U-Net model and UAV data in monitoring mining land covers.

The following research questions - relevant to land cover classification – will make this study clearer:

- What are the advantages of integration between deep learning and high-resolution images for monitoring mining areas?
- How are U-Net models for land cover classification on UAV images more effective than benchmark models?
- How do land cover types distribute in a mining area of Daknong province, Vietnam?

In this study, 4-band UAV images were used as input variables for land-cover classification. An area of  $(0.5 \times 0.8)$  km $^2$  was used to develop a U-Net model for land cover classification. Afterward, the trained model was used to interpret another area of  $(2 \times 2)$  km $^2$  in a mining area of Daknong province, Vietnam. Before explaining data collection and model development (from section II.D to II.E), the authors presented a description of land cover classification systems in the research area (section II.A). Notably, all processes to prepare UAV image, preprocess and training models will be explained in detail in section II.B. The trained U-Net models were compared with two benchmark models using Random Forest (RF) and Support Vector Machine (SVM) – presented in section II.G – to check the improvement. Results concerning the classification of mining land cover and model development will be compared with recent studies in section III.

## II. MATERIAL AND METHODS

### A. RESEARCH AREA

To evaluate the ability to monitor minerals exploitation when using the remote sensing technology UAV, the authors selected Tan An quarry in Dak R'moan commune, Gia Nghia City, Dak Lak province as the main research field. The experimental mine is an open-pit mine in operation, and witnessed significant changes in topography due to mining activities. It has an area of 50 ha adjacent to the banks of Dak Tí Hydropower lake and Dak R'Moan inter-commune road. The size of the mining area meets the requirements for close-frequency flight by UAV technology with ideal conditions such as lying on a mountain peak with wide viewing angles and covering the entire field.

Additionally, the hard ground of the mine can ensure that ground control points are immobile during the process of flight control and measurement. There was little dust in the mining object, which provided clear images for building graph and digital model of surface with high resolution. The

flight was conducted on June 28, 2019. On that day, dry weather, moderate wind and little dust mining subjects all contributed to the development of clear aerial photos with optimal image quality. These photos were used to build a digital surface model of high quality later on. To assess the overall process of changes in the status of soil cover in the quarry and surrounding areas, the scope of experimental study expanded to an area of 400 ha, taking the quarry as the center. The mine is an area with the coordinates of  $11^{\circ}59'55''$  -  $12^{\circ}00'14''$  North latitude,  $104^{\circ}06'11''$  -  $104^{\circ}06'41''$  East longitude.

Tan An quarry area is covered with basalt soil with the thickness of 2-5m. Before the exploitation, most of the quarry area was unused hilly land and perennial industrial crops land. During the exploitation process, the quarry area has been gradually expanded and some areas of perennial crop land have also been exploited. Moreover, activities such as dumping and unloading of surface soil, making space for processing and gathering of finished stone products are responsible for changes in other soil types. In 2006, the quarry was just an open-pit exploitation area with a small area of less than 1 ha. The stone is mined into the cliff area. At that time, there was no sign of expansion mining. By 2012, a part of the rock was exploited in the direction of mine expansion, a road leading to the quarry was also formed.

Some of the mine's western perennial industrial land has been converted and removed stones. However, the mining area was not yet put into industrial exploitation. By 2014, after only two years, Tan An quarry has been expanded, with all the corresponding processing and storage areas. From 2016 to 2018, a large area of perennial industrial land was converted to serve the purpose of mine expansion. The extraction process becomes stable. The six types of soil in Tan An quarry can be visually observed through UAV images shown in Figure 1. A part of the quarry area has also been accreted by new layers of soil.

## B. PRE-PROCESSING OF UAV IMAGE AND SAMPLE COLLECTION

UAV technological advancement is most beneficial when applied to monitoring mining area. Due to the fact that mines are densely located over small areas, it is very efficient to collect data for UAV processing. For the purpose of monitoring the mining progress, the experimental mine must be an open-pit one which is still under operation and experiences topography changes caused by exploitation. In this study, image data is taken from UAV database from senseFly of Tan An quarry in Dak Nong province, Vietnam. The data is collected by a Phantom 4 RTK and underwent a 6-step process of OrthoEngine Tool in PCI Geomatica Banff Service Pack 1 (Fully Functional Trial). There are 768 JPEG images used for input data. The critical steps in the pre-processing of UAV data will be introduced briefly as follows:

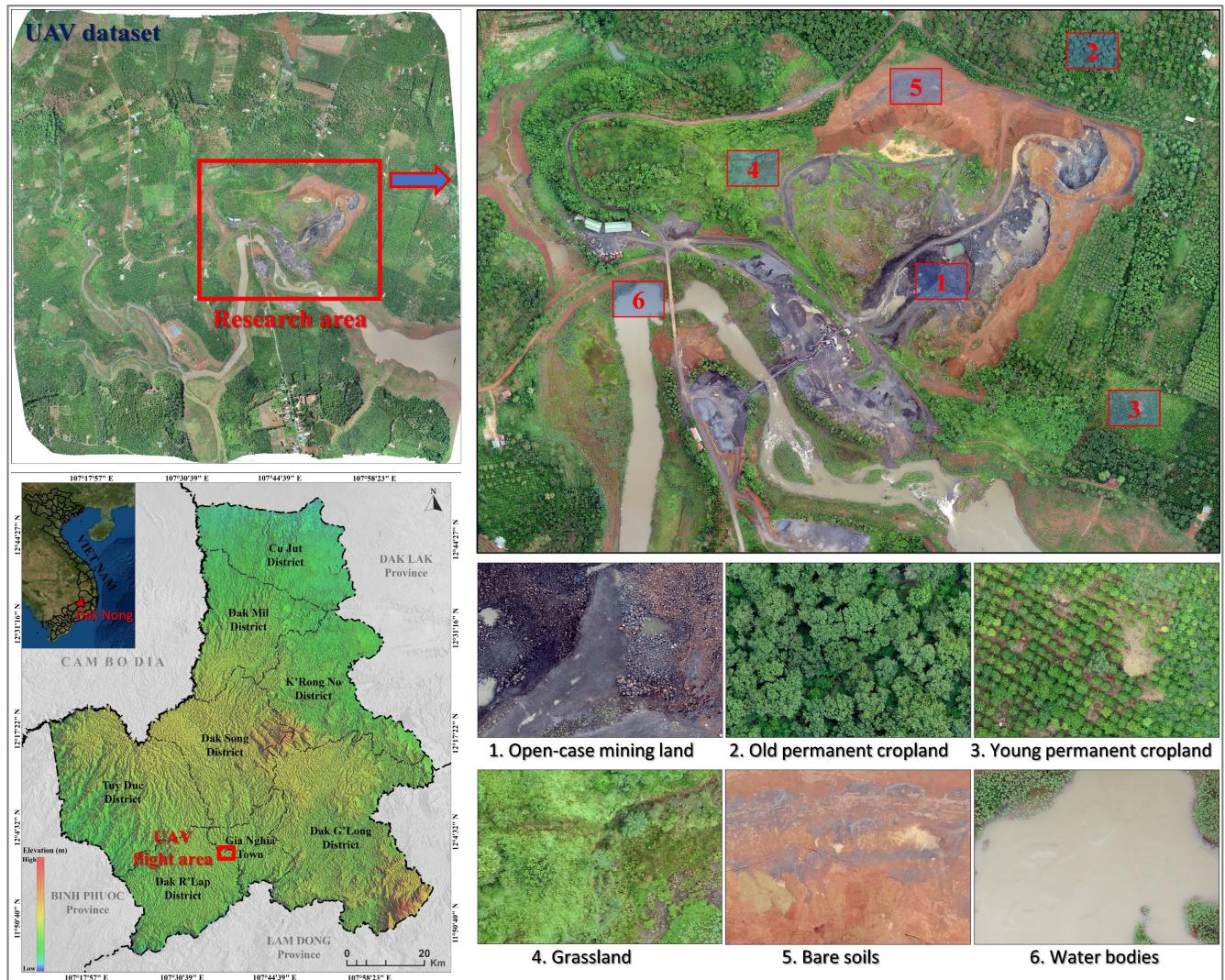
- Step 1 (Set camera calibration): This step helps to identify and correct the distortions of the image which occurred due to curvature of the lens, the focal length,

and the perspective effects. In this step, parameters (Focal length; Principal-point offset; Chip Width and Height; Y-scale factor; Radial-Lens coefficients; Decentering coefficients) are entered manually.

- Step 2 (Collect tie points): Tie point (TPs), a specific location, can be recognized visually in the overlap area among two or more images. By applying image-correlation technique a total of 23,040 optimal tie points were found in 768 images. This result has an important role to play in combining all images together.
- Step 3 (Point refinement): in this step, 23,040 tie points found in step 2 were refined to eliminate the ones with high values of Root mean square (RMS: the residual values of the x and y per point). Tie points with RMS higher than 0.5 were removed from the calculation model.
- Step 4 (GCPs correction): 12 GCP (Ground Control Point) from field with an accuracy of  $\pm 0.5$  pixels were collected to improve georeferencing of the dataset. They are clearly recognized in the raw images with identified ground coordinates. All data were orthorectified with the WGS84/UTM ZONE-48N.
- Step 5 (Generate orthoimages based on DEM): This step generated 768 ortho-images by using raw images, epipolar pair images and a Digital Elevation Model (DEM) of the research area scanned during flying process. This helps to enhance the performance of the correlation process and to reduce the probability of incorrect matches of the raw images. The spatial resolution of the outcome DEM is 0.35 m.
- Step 6 (Mosaicking): In this step, 768 ortho-images were manually edited to form a single and uniform color balancing image. Because 80% of images overlap, all errors related to camera and coordinates were eliminated. The mosaicking process was conducted precisely with the use of Mosaic Tool of the PCI Geomatica Banff software. After the merged image had been generated, the Tan An mining area was presented for the U-Net development that will be explained in section II.D.

## C. PROCESSING DATA

Samples of six LULC types were taken in the area of  $(2 \times 2)$  km<sup>2</sup> around the main mining area. The authors measured in detail the area and the boundary of each LULC in November 2019. Firstly, although the coffee trees in the research area covered a large area, some coffee areas have been mixed with some pepper trees. The density of coffee trees is denser than two other agricultural trees. Therefore, it is challenging to separate coffee and pepper trees in the UAV image. Authors combined coffee and mixed croplands into one land-cover type, so-called permanent croplands. However, they will be separated into old and young permanent croplands depending on how large the canopy of the trees covering the land. The canopy of the old coffee trees covers a larger area than the young one. Therefore, the second land cover type that was separated is the "young permanent croplands". Grassland



**FIGURE 1.** The research area with six samples for six land-cover types in a mining area of Daknong province, Vietnam.

and bare soil respectively, are third and fourth land cover types were separated (Figure 1).

The grasslands in the research area have been left fallow during five years preparing for mining activities in the future, whereas the bare soil areas have been left fallow after finishing mining activities [42]. The whole area used for natural resource mining will be separated into the “open-cast mining” land cover. It includes all mining areas and houses used for exploration activities. Like the last land cover, water bodies were separated and represented small lakes and two small streams that flow out from the mining area to the outside region.

The boundary of all agricultural croplands will be formatted in a polygon file and converted to an integer-8bit raster file (so-called as “mask”). The size of the mask is equal to the size of the input image. Due to the large image size, it is not feasible to read the entire image into memory and then use it for U-Net development. In order to increase the

performance of the training process, the original UAV image was spitted into 12,000 sub-images and resized to sub-images of  $128 \times 128$ -pixel size. Each sub-image was inputted as a sample to train the U-Net models. Afterward, they were separated into two training and testing groups of 80% (or 9,600 sub-images) and 20% (or 2,400 sub-images) for using for the U-Net model development. The ratio of land cover types in the training and testing data is equal.

#### D. BACKGROUND OF MODEL VALIDATION

Once a prediction model is trained, it always needs to be assessed its performance/quality, especially with the deep learning models, avoiding overfitting and underfitting cases [13]. This section will explain all model validation methods used in this study. The validation process was performed throughout the U-Net, Random Forest, and Support Vector Machine model development (see from section II-E to section II-H). Due to the outcomes of all prediction models

are quantitative values, four evaluation metrics were used in this study, including (1) intersection over union value (IoU), (2) class user's accuracy (UA), (3) overall accuracy (OA), and (4) Kappa Coefficient. The IoU value was used during the U-Net development (see section 2-E), whereas three evaluation values, including UA, OA, and Kappa, were used to compare the accuracy between all trained U-Net models and two benchmark models (see section 2-G). The IoU value represents the proportion of the combined area of reference and predicted bounding images (so-called as the intersection region) per the union region [43]. The IoU value is described in the following formula:

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (1)$$

The IoU value has been commonly used in computer vision and deep-learning studies [43]. During the U-Net development process, the IoU value, therefore was calculated continuously for both training and testing image data in each epoch. However, in comparing the prediction-power between the trained U-Net models with two benchmark models, the UA, OA, and Kappa are an alternative option to check all trained models' accuracy at pixel scale. Accordingly, these values can be calculated based on the following formulas:

$$UA_i = \frac{P_{ij}}{M_j} \quad (2)$$

$$OA = \frac{\sum_{i=1}^m P_i}{N} \quad (3)$$

$$Kappa = \frac{N \sum_{i=1; j=1}^m P_{i,j} - \sum_{i=1; j=1}^m R_i M_j}{N^2 - \sum_{i=1; j=1}^m R_i M_j} \quad (4)$$

where  $P_{ij}$  is the number of correctly classified samples in the land-cover class "i" compared with ground truth/input class "j", "N" is the total number of samples, "m" is the number of classes,  $M_j$  is the total number of sample in the class "j" recorded in mask data, and  $R_i$  is the total number of sample in the class "i" recorded in prediction data.

The UA, OA, and Kappa values were only used once all models that were trained. The UA value was used to assess the accuracy of each land cover type predicted from trained models. It explains the positive proportion of the correct classified samples to the predicted samples [43] in a particular land-cover type in this study. In addition to the overall classification accuracy (OA), the Kappa value accounts for all true and false data of the model validation process based on a confusion matrix [44]. Therefore, the use of both OA and Kappa values can make the selection of the best model for mining land-cover classification more accurate. The input samples will be explained in detail in section 2-G.

#### E. U-NET ARCHITECTURE FOR LAND-COVER DETECTION

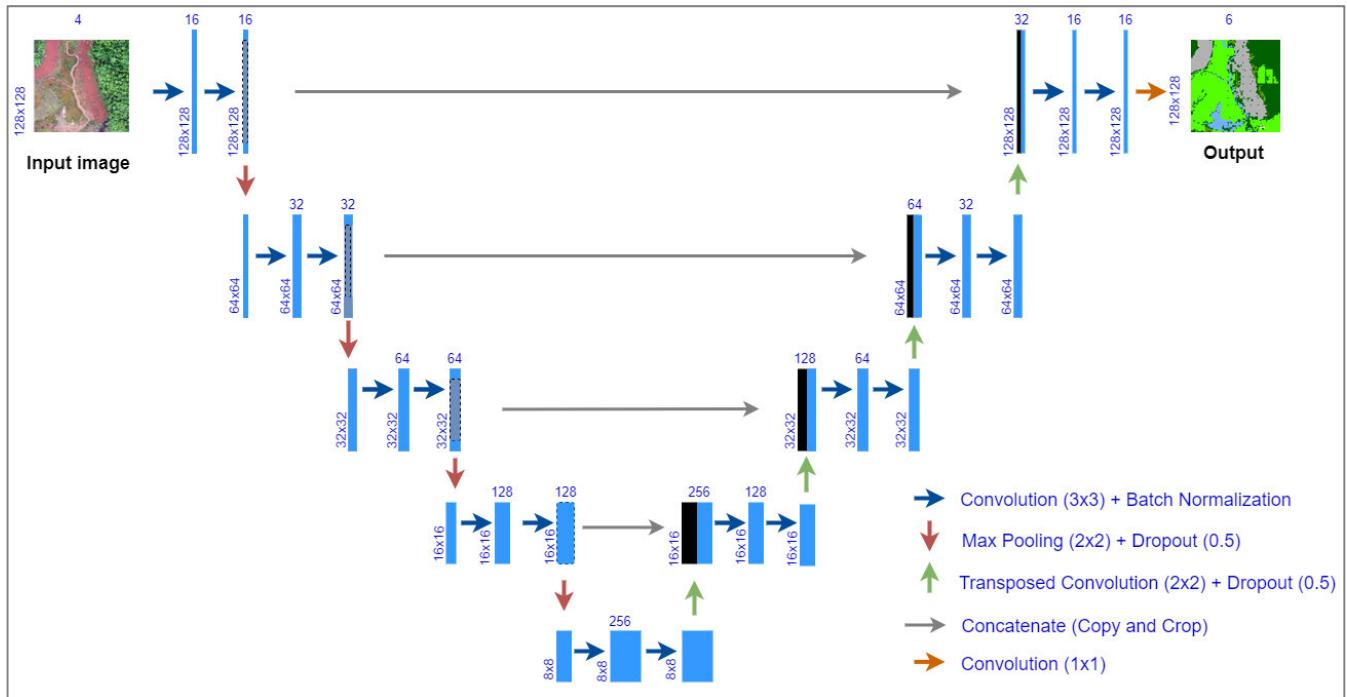
The U-Net model was first developed for segmentation on medical and geological images [45], [46]. It includes a contracting path (in the left side) and an expansive path

(in the right side). The contracting path can be considered as a typical extractor like conventional convolution neural network (CNN) models. The right-hand side performs up-sampling steps, or convert prediction values back to the original image size. The U-Net architecture is shown in Figure 2. Like a Convolutional Neural Network (CNN), the layers of a U-Net as a sequence of different layers consist of 3-dimensional neurons: width, length, and depth [41]. An image's depth is the number of input bands or variables. For example, the image's depth parameter in this study is the number of UAV bands – red, green, blue, and near-infrared spectral bands. Therefore, three dimensions of input sub-images are  $128 \times 128 \times 4$  (width, height, depth, respectively). The neurons in a layer were implemented with small matrices of sub-images, instead of implementing them with entire sub-images at once [37].

Every layer of the U-Net converts original data to new states based on a chosen function. Six consequential types of layers commonly applied to build U-Net architectures include (1) INPUT Layer, (2) Convolutional Layer (CONV), (3) Batch Normalization Layer, (4) Pooling Layer (POOL), and (5) Concatenate Layer, and (6) Dropout Layer. These six-layer types were integrated to form a full U-Net architecture as following:

- INPUT layer is used to insert the raw pixel values of all sub-images ( $128 \times 128 \times 4$ ) to the training model; in this case, the sub-images dimension has 128-pixel wide, 128-pixel height, and four spectral bands.
- CONV layers calculate the outcomes of neurons through a set of filters. Standard CONV layers will calculate new outcomes with the same size with the input, whereas a transposed convolution matrix (abbreviated as "ConvTrans") will be used to up-sample a smaller matrix into a larger one. The filter weight and length have to be smaller than those of the input sub-images. The filter slid across the sub-images, connects with local regions of input sub-images. With the input, new pixel values are determined based on the activation functions selected for the filters (more detailed in section II.E). In this research, as suggested by Li *et al.* [39] and Stoian *et al.* [41], the authors chose 19 CONV layers for the U-Net construction. To minimize the training and validation time, 16, 32, 64, 128, and 256 filters chosen for the 19 CONV layers. The width and length of each filter are at the scale of  $3 \times 3$ , respectively.
- BATCH NORMALIZATION layer is used to normalize the output from the CONV layer to the same scale, before coming to a new calculation. This layer can minimize the distribution changes of the activation values during the training process. It has commonly known as an internal covariate shift problem [47]. Each input layer is normalized by using the mean ( $\beta$ ) and standard deviation (or variance -  $\gamma$ ) parameter of the values in the current batch based on the following formula:

$$y_i = \gamma \hat{x}_i + \beta \quad (5)$$



**FIGURE 2.** The architecture of a U-Net network for mining land cover classification.

whereas the  $\beta$  and  $\gamma$  are trainable parameters,  $\hat{x}_i$  can be calculated by using mean ( $\mu_B$ ) and variance ( $\sigma_B^2$ ) of mini-batch  $B = \{x_1 \dots m\}$  as following formula:

$$\mu_B \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad (6)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (7)$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (8)$$

In total, the batch normalization includes four parameters that will be trained after all CONV layers.

- POOL layer downscale operation to  $2 \times 2$  spatial matrices, like width and height, respectively. This layer also uses specific downscaling process activation functions, which will be explained in section II.E.
- CONCATENATE layer is used to concatenate image information from contracting path to expansive path. Due to the simplification of the original data to new data during the contracting process, the U-Net models combine the data from the previous layers to achieve a more accurate forecast.
- DROPOUT layer is used to randomly deactivates the neurons at each training step, rather than training the data on the initial network. During the iterative process, unnecessary neurons can be deactivated to reduce overfitting and generalization errors. A dropout value is commonly used as a probability of 0.5 to keep the output of every node in a hidden layer. In contrast, a value close to 1.0 is commonly used to keep inputs from the visible layers [48].

The blocks in blue color shown in Figure 2 are the input and output of each calculation layer, whereas the processing layers are presented in arrows. Excepting the last convolutional layer, the other 18 CONV layers are always processed before the batch normalization layers (combined and shown in 18 blue arrows). In the contracting path, the POOL layers run before the Dropout layers to downscale data. Meanwhile, the transposed convolution matrixes integrated with the Dropout layers are used for the upscaling process in the expansive layers. They are represented in four green arrows. Lastly, the concatenate layers are represented in brown arrows.

Table 1 shows the U-Net architecture image processing procedure, with 76 layers. The output of the other layer is the input data of the following layer. The first 32 layers perform the contracting path, whereas the last 36 layers perform the expansive path. The number of filters in the following layer in the U-Net development is double the preceding layer in the contracting path and one half of the preceding layer in the expansive path (Figure 2), compared to CNN models. In contrast, the width and height of the preceding layer is one half of the preceding layer in the contracting path and double of the preceding layer in the expansive one.

The number of parameters of the first Conv2D layer is calculated as following:

$$N_{Conv2D} = (H * W * D + 1) * N_{Filter} \quad (9)$$

where ‘H’ is height, ‘W’ is width, ‘D’ is depth and ‘N Filter’ is number of filters. Each filter has one extra parameter to store the bias value. For example, the first Conv2D-1 layer has  $(3 \times 3 \times 4 + 1) * 16 = 592$  parameters.

**TABLE 1.** Mathematical structure of the developed U-Net for land cover classification.

No.	Layer	Output shape	Parameter
1	InputLayer	128, 128, 4	0
2	Conv2D_1	128, 128, 16	592
3	BatchNormalization_1	128, 128, 16	64
4	Activation_1	128, 128, 16	0
5	Conv2D_2	128, 128, 16	2,320
6	BatchNormalization_2	128, 128, 16	64
7	Activation_2	128, 128, 16	0
8	MaxPooling2D_1	64, 64, 16	0
9	Dropout_1	64, 64, 16	0
10	Conv2D_3	64, 64, 32	4,640
11	BatchNormalization_3	64, 64, 32	128
12	Activation_3	64, 64, 32	0
13	Conv2D_4	64, 64, 32	9,248
14	BatchNormalization_4	64, 64, 32	128
15	Activation_4	64, 64, 32	0
16	MaxPooling2D_2	32, 32, 32	0
17	Dropout_2	32, 32, 32	0
18	Conv2D_5	32, 32, 64	18,496
19	BatchNormalization_5	32, 32, 64	256
20	Activation_5	32, 32, 64	0
21	Conv2D_6	32, 32, 64	36,928
22	BatchNormalization_6	32, 32, 64	256
23	Activation_6	32, 32, 64	0
24	MaxPooling2D_3	16, 16, 64	0
25	Dropout_3	16, 16, 64	0
26	Conv2D_7	16, 16, 128	73,856
27	BatchNormalization_7	16, 16, 128	512
28	Activation_7	16, 16, 128	0
29	Conv2D_8	16, 16, 128	147,584
30	BatchNormalization_8	16, 16, 128	512
31	Activation_8	16, 16, 128	0
32	MaxPooling2D_4	8, 8, 128	0
33	Dropout_4	8, 8, 128	0
34	Conv2D_9	8, 8, 256	295,168
35	BatchNormalization_9	8, 8, 256	1,024
36	Activation_9	8, 8, 256	0
37	Conv2D_10	8, 8, 256	590,080
38	BatchNormalization_10	8, 8, 256	1,024
39	Activation_10	8, 8, 256	0
40	Conv2DTrans_1	16, 16, 128	295,040
41	Concatenate_1	16, 16, 256	0
42	Dropout_5	16, 16, 256	0
43	Conv2D_11	16, 16, 128	295,040
44	BatchNormalization_11	16, 16, 128	512
45	Activation_11	16, 16, 128	0
46	Conv2D_12	16, 16, 128	147,584
47	BatchNormalization_12	16, 16, 128	512
48	Activation_12	16, 16, 128	0
49	Conv2DTrans_2	32, 32, 64	73,792
50	Concatenate_2	32, 32, 128	0
51	Dropout_6	32, 32, 128	0
52	Conv2D_13	32, 32, 64	73,792
53	BatchNormalization_13	32, 32, 64	256
54	Activation_13	32, 32, 64	0
55	Conv2D_14	32, 32, 64	36,928
56	BatchNormalization_14	32, 32, 64	256
57	Activation_14	32, 32, 64	0
58	Conv2DTrans_3	64, 64, 32	18,464
59	Concatenate_3	64, 64, 64	0
60	Dropout_7	64, 64, 64	0
61	Conv2D_15	64, 64, 32	18,464
62	BatchNormalization_15	64, 64, 32	128
63	Activation_15	64, 64, 32	0

**TABLE 1. (Continued.)** Mathematical structure of the developed U-Net for land cover classification.

No.	Layer	Output shape	Parameter
64	Conv2D_16	64, 64, 32	9,248
65	BatchNormalization_16	64, 64, 32	128
66	Activation_16	64, 64, 32	0
67	Conv2DTrans_4	128, 128, 16	4,624
68	Concatenate_4	128, 128, 32	0
69	Dropout_8	128, 128, 32	0
70	Conv2D_17	128, 128, 16	4,624
71	BatchNormalization_17	128, 128, 16	64
72	Activation_17	128, 128, 16	0
73	Conv2D_18	128, 128, 16	2,320
74	BatchNormalization_18	128, 128, 16	64
75	Activation_18	128, 128, 16	0
76	Conv2D_19	128, 128, 6	102
Total parameters			2,164,822

Due to the batch normalization generate four parameters for each CONV layer, the number of parameters in the batch normalization layer is calculated as following:

$$N_{parameter} = 4 * N_i \quad (10)$$

where  $N_i$  is the depth in the input CONV layer. For example, the BatchNormalization\_18 has  $4 * 16 = 64$  parameter. The final Conv2D layer's output is a vector with six values, corresponding to 6 land cover types. Based on 76 layers (1xINPUT, 19xCONV, 4xConv2DTrans, 4xPOOL, 18xBatch-Normalization, 4xConcatenate, 18x Activation, and 8x Dropout layers), the trained U-Net transformed the initial pixel values in 12,000 UAV sub-images to the land cover classes. 23 CONV and 18 Batch Normalization layers contain parameters that can be optimized to improve U-Net model's performance and accuracy. The parameters in the CONV and Batch Normalization layers will be changed with alternative choices of activation and optimizer functions. It will be explained in detail in section II.E.

The accuracy of both training and testing data was checked during U-Net development to avoid overfitting and underfitting issues. The best U-Net will be chosen if it is land cover prediction is compatible with the labels assigned in the input image from the training and testing data. Therefore, the intersection over union value (IoU) was used to assess the accuracy of all trained models. During the training process, the min and max IoU values of training and testing data were recorded to observe the fluctuation of this value in 100 epochs. The IoU value using the training data will be definitely optimized in the trained models, therefore the IoU value using the testing data will be observed to check the accuracy of outcome models. In combination with the IoU, to choose the best model for classifying mining ecosystem types, the trained model need to have the lowest values of all loss functions. The modification of the loss functions will be explained in section 2-F.

The U-Net model is developed based on Keras in Python language, as an API designed for easy manipulation with

Tensorflow developed by Google [38], [49]. The parameters observed during the training process include accuracy and loss of testing and validation data. The U-Net training process is limited to 100 loops (epoch), but the process can be stopped if the coefficient on the set of testing data converges. In other words, all values do not improve after ten epochs.

#### F. ALTERNATIVE OPTIONS TO DEVELOP THE U-NET

According to the architecture of the U-Net for the land-cover classification, to develop the U-Net, three types of functions, namely activation function, loss function, and optimizer method can be selected. These functions serve as optimal parameters for filters contained in hidden layers. Which function and method to be used depends on the kind of input data and output labels, and the accuracy/ loss of trained models. These will be discussed in the following section.

##### 1) ACTIVATION FUNCTIONS

To optimize the convergence speed of the U-Net in the CONV layer, an activation function should be chosen. Binary Step or Linear Activation Functions were not selected because of the differences in the appearance of six land-cover types in four UAV spectral bands. Therefore, one in five non-linear activations, which are TanH/Hyperbolic Tangent, ReLU (or Rectified Linear Unit), leaky ReLU, Parametric ReLU, and Swish types will be chosen as the most suitable for optimizing U-Net models [50], [51]. The ReLU function possesses certain advantages over the others. Using max (0, x) - thresholding at zero, the ReLU function – can keep the considerable size of the images ( $128 \times 128 \times 4$ ) and speed-up the convergence process of U-Net models. The other functions fail to achieve such optimal advantages. Furthermore, this function also enables backpropagation in the training process to happen [52]. Besides, the reliability of results from the U-Net applying ReLU function is also better than which of the other functions. For these above reasons, the ReLU function was selected for four U-Net layers.

Two activation functions namely Sigmoid/Logistic and Softmax can be applied for the POOL layer [53]. However, the latter function was chosen rather than the former. Because Softmax function is commonly used to normalize the outputs to classes between 0 and 1 and provides prediction probability in a specific type [54]; while the Sigmoid/Logistic function generates uncertainties related to vanishing gradient problems if the maximum and minimum values of input data are too high. In this study, all POOL layers, therefore, used Softmax function as the best option.

##### 2) LOSS FUNCTIONS

To narrow the gap between the predicted and actual outputs for the trained U-Net, a minimized Cost function (C), or Loss function is needed as convex functions based on choosing optimized values for weights [55]. Due to the dependence of the loss function on weights, input images, and output labels, the weights of the trained networks can help to minimize the loss function, which then contributes to a better

prediction for new input data. This optimization process was explained in detail by Redmon *et al.* [56]. The average loss values were computed using entire training sub-image data and represented by the following function:

$$\mathcal{J} = \frac{1}{n} \sum_{x=1}^n \mathcal{L}^{(x)} \quad (11)$$

with ‘n’ is the size of the training data set, and  $\mathcal{L}^{(x)}$  is the loss value of a single training sub-image during the training process.

Which loss function (e.g., regression, binary classification, and Multi-Class Classification Loss Functions) to be chosen depends on the type of training U-Net [57]. In this study, to classify six land-cover types, authors chose one in three types of Multi-Class Classification Loss Functions which will be introduced as below:

- **Multi-Class Cross-Entropy Loss:** is the standard method to quantify the loss function values in the case of the target values formatted in the set  $[0, 1, 2, \dots, n]$  [17], [58]. In this study, each land-cover type is illustrated by a different integer value. Mathematically, this method uses the framework of maximum likelihood to calculate the loss. A score representing the average probability gaps between the actual and predicted outputs is calculated for all land-cover types. The last score has to be minimized to 0 as the perfect cross-entropy value. This method is demonstrated in the Keras coded in python with a “categorical\_crossentropy” function when compiling the model [59]. The effectiveness of the Multi-Class Cross-Entropy Loss’s performance seems to be reduced, particularly in the encoding process of the network with a large number of labels and a significant memory.
- **Sparse Multi-class Cross-Entropy Loss:** has been designed to enhance the performance of Multi-class Cross-Entropy Loss function in categorizing a considerable number of labels during the training process. Both Sparse Multi-class Cross-Entropy Loss and Multi-Class Cross-Entropy Loss functions can perform the cross-entropy calculation of error; however, there is no need for the encoded target variables during the training process in the former one. This method was developed in the Keras by using “sparse\_categorical\_crossentropy” function when compiling the model [59].
- **Kullback Leibler Divergence Loss:** has been built to measure the differences between the probability distribution of the outcomes and a baseline distribution. The distributions are fitted when the difference value is 0. This method seems to be very similar to the cross-entropy methods; however, it can calculate the amount of lost information when the predicted probability and desired target probability distribution are approximately equal. Ahuja [60] and Galas *et al.* [61] proposed the Kullback Leibler Divergence Loss function is more suitable for complex tasks than simple multi-class

classification. This method has been built in Keras, Python, by the use of the “kullback\_leibler\_divergence” function when compiling the model [59].

### 3) OPTIMIZER FUNCTIONS

Optimization methods to develop neural networks based on a stochastic gradient descent algorithm is commonly employed to eliminate the cost functions. Updating weights in the negative gradient direction, this method increases the accuracy of trained neural networks and minimizes the loss. During the optimization process, the errors of the trained models (or the loss function) must be calculated repeatedly. A cycle of passing data forward and backward via the U-Net model is one epoch [62]. After each epoch, it is necessary to update weights of filters based on optimizer functions. The new parameters will learn the gradient descent of the previous parameters and the weight of filters to find out the optimal location for new ones. Accordingly, the new weight of filters will be calculated respectively with the new parameters. The new filters using updated weights can separate nearly-similar objects on satellite images to reduce the loss value for the next evaluation. During the U-Net development process, the parameters and the weight of filters affect directly to the 19 CONV layers. Seven optimization algorithms sequentially changed comprise traditional SGD (Stochastic Gradient Descent algorithm), Adagrad (Adaptive Gradient Algorithm), Adadelta, Adamax, RMSProp (Root Mean Square Propagation), Adam (Adaptive Moment Estimation) and Nadam (Nesterov-accelerated Adaptive Moment Estimation). The description of these optimization algorithms was provided in Table 2. All in all, the best optimizer method should generate the highest accuracy and lowest loss function values.

## G. EXPERIMENTAL SETUP

In this section, the authors compared the performance of other traditional classifiers with the trained U-Net models. Two classifiers, including random forests (RF) and support vector machines (SVM), were chosen. After completing mining land-cover maps in the research area based on the RF, SVM and seven U-Net models, the prediction results will be compared with the mask. Due to the models using the Random Forest and SVM methods were not trained in epochs, the training process of these two models provide one accuracy value of each model. The cross-validation process used 1000 random points in the testing group (in 20% of the testing data) in the U-Net development to compare with the mask. Three evaluation metrics that were chosen for this step include the UA, OA, and Kappa coefficient. The optimal model will achieve higher OA and Kappa values. The results of this step will be presented in section III.C. The development of these classifiers can be explained as follows:

### 1) RANDOM FOREST (RF)

RF was developed in 2001 by Breiman [66] and became more and more popular used deep learning ensembles, especially remote sensing. A set of decision trees as a forest is trained

**TABLE 2. The seven optimization algorithms to train parameters of the U-Net model in land cover classification, adapted from [59], [62]–[65].**

For-mula No.	Optimizer algorithm	Formula
12	Adam	$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} \hat{m}_t$
13	Adamax	$\theta_{t+1} = \theta_t - \frac{\eta}{u_t} \hat{m}_t$
14	Adagrad	$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{g}_t + \epsilon}} g_t$
15	Adadelta	$\Delta\theta = -\frac{RMS[\Delta\theta]_{t-1}}{RMS[g]_t} g_t$ and $\theta_{t+1} = \theta_t + \Delta\theta$
16	Nadam	$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t + \epsilon}} (\beta_1 \hat{m}_t + \frac{(1-\beta_1)g_t}{1-\beta_1^2})$ $E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2$
17	RMSprop	and $\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g]^2 + \epsilon}} g_t$
18	SGD	$\theta_{t+1} = \theta_t - \eta_t \nabla_\theta Q(\theta_t; x^{(i)}, y^{(i)})$

where  $\theta$  is parameter value;  $\eta$  is the learning rates;  $t$  is time step;  $\epsilon = 10^{-8}$ ;  $g_t$  is the gradient;  $E[g]$  — moving average of squared gradients;  $m$ ,  $v$  are estimates of first and second moments;  $u_t$  - the max operation;  $\beta$  - moving average parameter (good default value — 0.9);  $\eta$  — step size.

randomly from training data. Then, the forest, which contains many decision trees, is developed as a composite classifier model [23]. Once the forest is created, users can use the RF model to predict across each tree with new input data. Three parameters can be modified to optimize a RF model including the number of trees in the forests, the number of features (calculated as the square root of the reflectance bands in remote sensing imagery, corresponding to 4 UAV bands in this study) and the number of samples taken at each leaf (was set at “1”) [67]. In this study, the number of trees was tested with 10, 100, 500, and 1000, and finally with 100 trees the results obtained a relatively higher accuracy compared to three other numbers.

### 2) SUPPORT VECTOR MACHINE (SVM)

SVM has known as a supervised learning model to analyze and classify image data, especially with small datasets [29]. In two-dimensional data, a hyperplane was identified based on the SVM model to separate effect the data into two categories. In multi-dimensional data, as in this study, the data is converted to corresponding dimensional space and a plane is identified to separate data to categories [68]. Two parameters can be adjusted to optimize the SVM models, including Kernel coefficient “gamma” and error term “C” values. The higher gamma value can make the SVM models fitted with the training dataset. Although the error is reduced, it can make over-fitting issues. Meanwhile, the higher “C” value can help the SVM models to use more training datasets for the

supporting process but it can be less accurate if the selected training dataset is too large [69]. Therefore, the “gamma” and “C” values were selected based on the OA and Kappa values. The optimal values were found for the “gamma” value at 0.25 and “C” value at 100.

#### H. APPLICATION OF TRAINED U-NET MODELS FOR NEW LAND COVER CLASSIFICATION

Once the best U-Net for the land cover classification from high-resolution images was developed, its most important function is to interpret new UAV images. In this study, the authors focused on land cover types in rock mining in Daknong province, Vietnam. Therefore, a new UAV image in a mining area of  $(2 \times 2)$  km<sup>2</sup> was taken in 2019. The image data collection and pre-processing were also done, as explained in section II.B and II.C. Once the new image was inputted to the trained U-Net, the model accesses the trained parameters to transform the new images into specific spatial matrices, generate intermediate matrixes in 76 layers, and interpret the final land cover classes for every pixel within the new image. All these prediction processes do not require any new training samples.

### III. RESULTS

#### A. U-NET MODEL PERFORMANCE

Figure 3 describes the loss function values of seven models using seven optimization methods. The indicators used for comparison are the “loss value” of “training and validation data” over epochs. In general, validation and training loss values were optimized after about 10 to 15 epochs in all methods. The U-Net model’s training process using the Adamax was stopped after ten epochs, and the best model was found in the first epoch. It yields IoU results with a min of 81.5, a mean of 82.3, and a max of 82.5. Its standard deviation of 0.3 is better than other ones (Table 3). After about 30 epochs, the models using Adam and RMSprop only have an accuracy of about 81%, whereas the models using the Nadam and Adadelta achieved the higher accuracy values of nearly 84%.

The model using the SGD optimizer has the accuracy value of 68%, the training and validation loss values are too different. The prediction works in detail at the local scale – looks like the outcome using the pixel-based methods. It can be easy to observe in the location of young permanent croplands. The model using the SGD optimizer did not identify spatial distribution of young crop lands and bare soils among the croplands. It is different with the large region of the bare soils in brown color (Figure 4 and 5). The distribution of each land cover will be explained as following sections.

#### B. LAND-COVER INTERPRETATION IN THE MINING AREA OF DAKNONG PROVINCE

Figure 4 shows the land-cover map based on manual image interpretation and segmentation methods. This map was used as a mask to train all U-Net models. According to this result, the area of old permanent croplands (16 hectares) is

**TABLE 3. The IoU values of seven models using seven optimizer algorithms.**

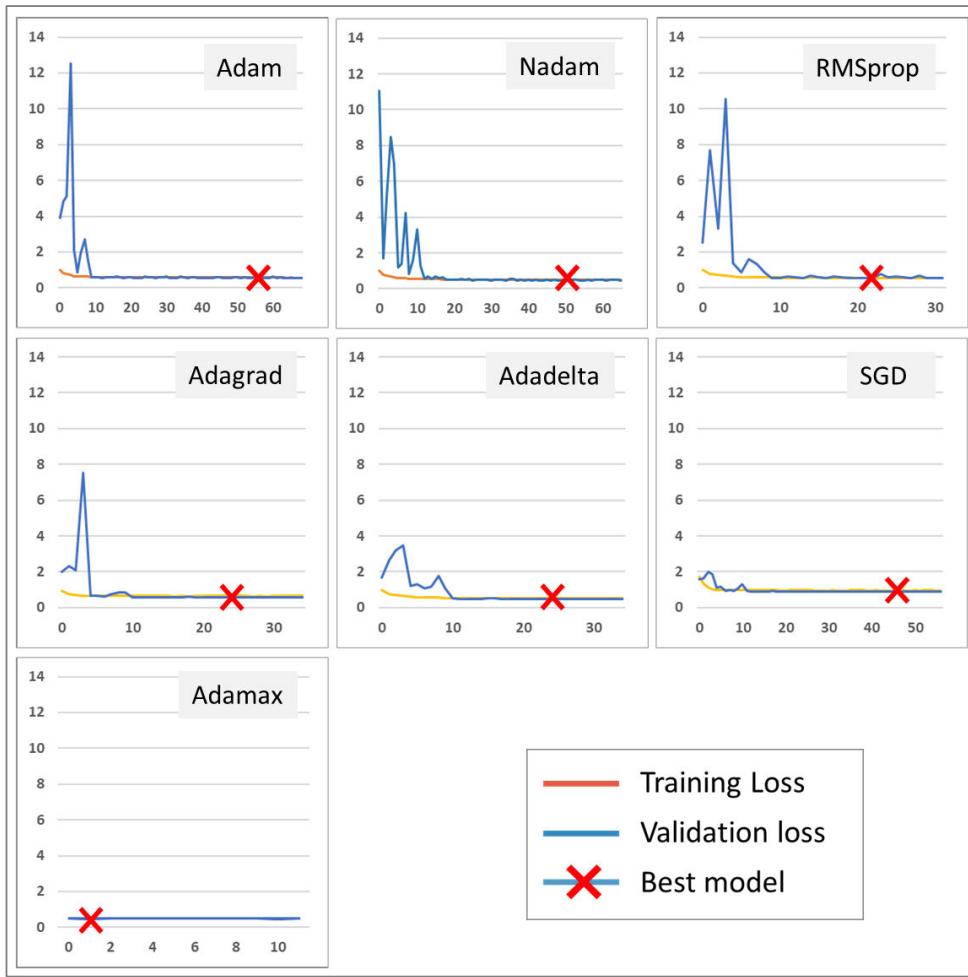
No.	Model using optimizer algorithm	IoU (%)			
		Min	Mean	Max	Standard deviation
1	Adam	18.3	74.6	81.2	14.3
2	Adamax	81.5	<b>82.3</b>	82.5	0.3
3	Nadam	18.4	74.2	<b>83.2</b>	16.9
4	Adagrad	24.8	72.3	79.0	15.5
5	Adadelta	31.7	73.5	<b>83.6</b>	16.7
6	SGD	27.9	63.4	67.5	10.0
7	RMSprop	26.2	70.6	81.1	16.4

four times higher than those of the young permanent croplands. The area of grasslands is about 14 hectares, whereas the open-case mining areas covered more than 9 hectares. As shown in Figure 4, the open mine is surrounded by the permanent croplands with various species of composing trees such as cashew, coffee and durian. These species are not separately cultivated in different land parcels but they are mixed between cashew, durian and coffee tree. The interposing cultivation makes the image classification confused.

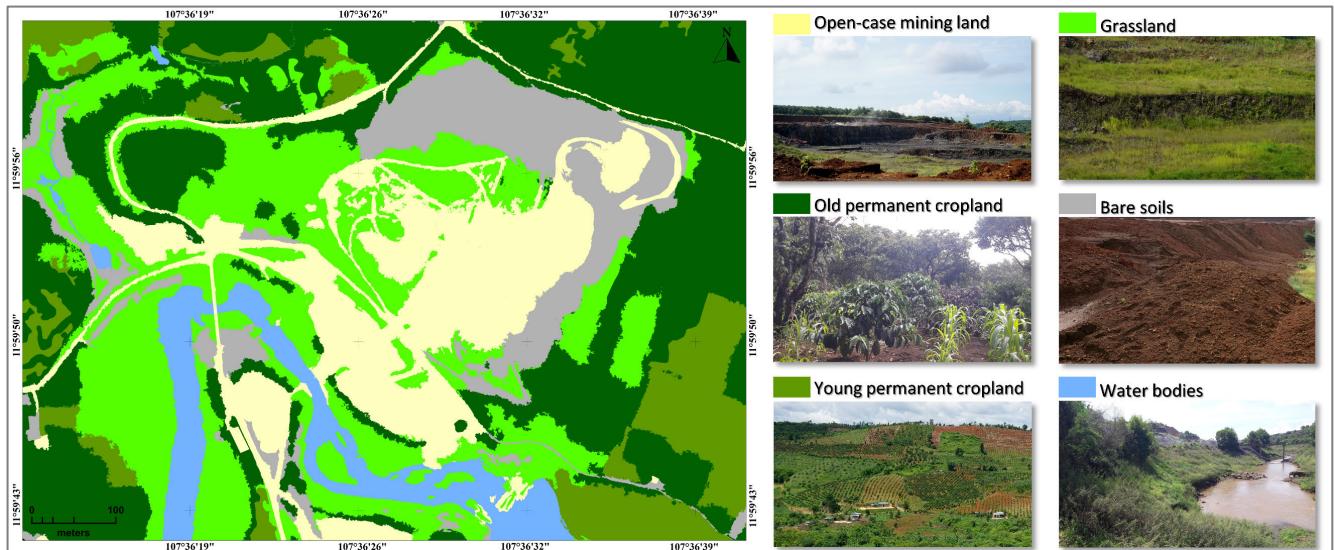
The area of the open rock mine is about 17.7 ha counted in June 2020, characterized by two discrete areas including a exploiting core and a taking-off soil region. The exploiting area has a rocky surface in a gray cover, and the taking off soil layer is in yellow-red due to the concentration of volcanic material. In addition, service areas taking parts of the mine, occupy 4.8 ha for the storage, product processing, and transport. The service areas are clearly observed on satellite image with bright gray of concrete surface.

The south and southwest side of the mine is the bank of the Dak Tik lake which is submerge in a short time during the rainy season. Thus, the bank is covered by grass most of the rest of year. Grass cover presents as a smooth, bright green region, and closed water surface in satellite images.

The diagrams in Figure 5 describe the image interpretation results based on seven U-Net models using seven different optimizer options. In general, all models can effectively detect the old permanent croplands, open-cast mining areas, water bodies, bare soils, and grasslands. However, three models using the Adam, Adagrad, and RMSprop optimizers could not detect the vast area of the young permanent croplands in the eastern side of the research area. The prediction process was done after 120 seconds. In terms of accuracy, the models using the Nadam and Adadelta optimizers are still the best ones. Using the Adam optimizer, the U-Net model can detect the differences between the young permanent croplands with



**FIGURE 3.** The loss function values of seven models using seven optimization methods.



**FIGURE 4.** A land cover classification in the mining area of Daknong province, Vietnam based on manual image interpretation and segmentation method.

the bare soils that distribute among crops. Whereas, the model using the Adadelta learn thoroughly from the mask and then

combine the narrow area of bare soils among the young permanent croplands with the crops into one object.

**TABLE 4.** The Cross-validation of nine models for mining land cover classification.

Sample Distribution			Class accuracy (%)								
Class	Samples		Adam	Adamax	Nadam	Adagrad	Adadelta	SGD	RMSprop	RF	SVM
Open-case mining land	141		80.6	80.9	<b>85.1</b>	79.5	82.3	76.7	74.5	81.8	80.5
Old permanent cropland	252		<b>89.7</b>	83.3	<b>88.9</b>	<b>92.5</b>	<b>90.1</b>	<b>89.3</b>	<b>93.6</b>	<b>90.9</b>	82.9
Young permanent cropland	164		69.5	82.6	<b>85.8</b>	72.7	<b>89.4</b>	20.5	80.3	76.9	79.5
Grassland	208		70.7	72.1	75.5	73.6	75.9	71.8	74.6	71.2	72.7
Bare soils	123		83.4	<b>85.4</b>	<b>85.4</b>	81.3	82.9	73.1	78.9	77.2	80.5
Water bodies	112		<b>87.5</b>	<b>89.3</b>	83.1	84.8	<b>88.4</b>	81.8	<b>92.8</b>	82.1	<b>89.3</b>
Overall accuracy (%)											
TOTAL	1000		80.2	82.3	<b>84.0</b>	80.7	<b>84.8</b>	68.7	82.4	80.0	81.0
Kappa Coefficient											
			0.78	0.80	<b>0.82</b>	0.79	<b>0.83</b>	0.66	0.82	0.78	0.79

### C. ACCURACY COMPARISON AMONG THE TRAINED MODELS

The accuracy differences between nine trained models (including seven U-Net models and two benchmark models) are presented in Table 4. The prediction maps based on the RF and SVM models can be seen in Figure 5. Due to the testing samples are a part of validation dataset, the OA values of seven models are nearly similar to the results shown in Table 3. As shown in Figure 5, the OA and Kappa values the model using the SGD optimizer have the lowest values, compared to other models. The results predicted from the RF and SVM models have the OA and Kappa values significantly lower than most of U-Net models. Compare with the manual interpretation mask, the model using the Adadelta optimizer provides the best prediction.

## IV. DISCUSSION

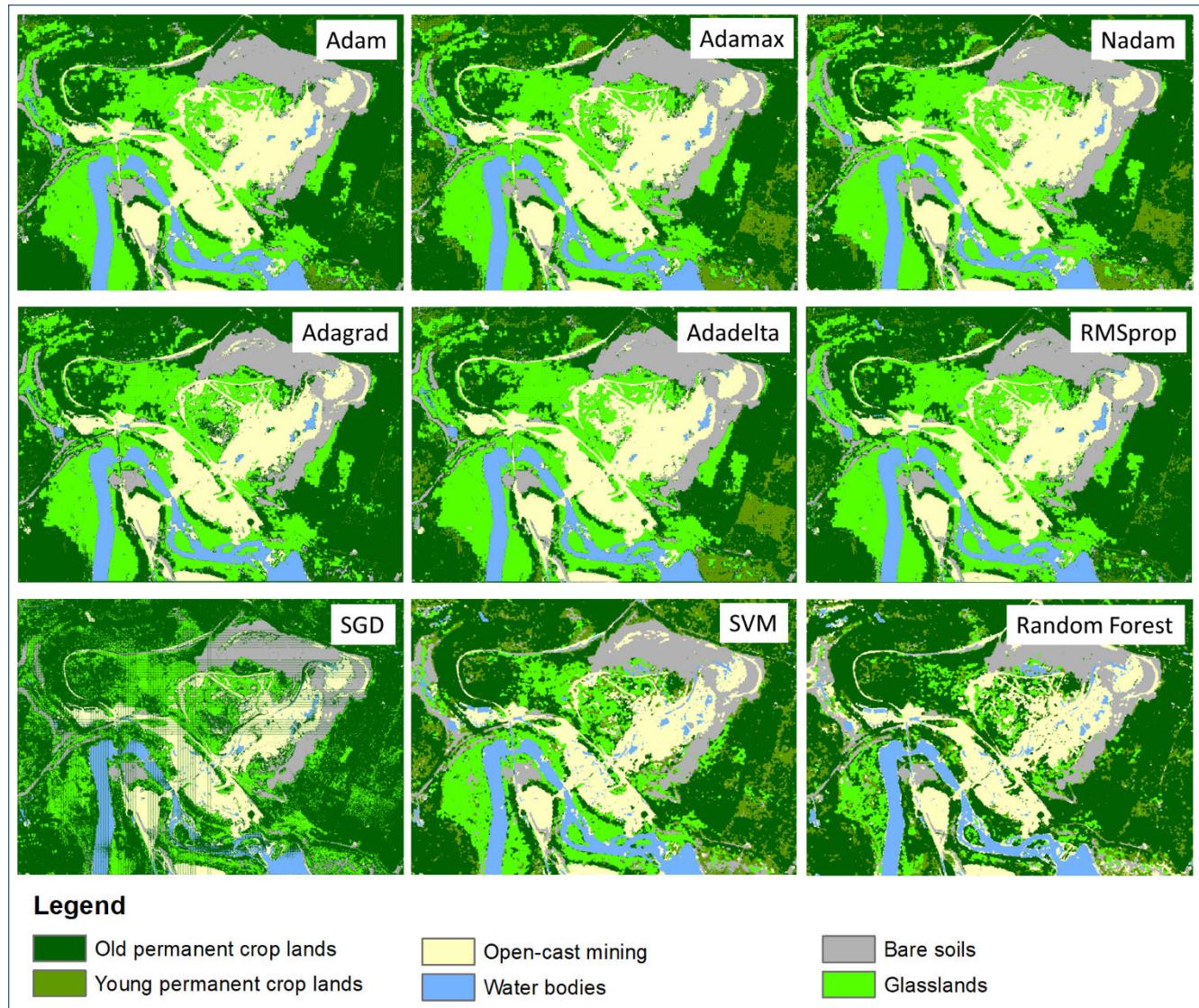
### A. COMPARISON WITH FORMAL LAND COVER CLASSIFICATION SYSTEMS

Classification of land cover is closely linked to ecological conditions on the Earth's surface and partly represents the global ecosystem health, quality of the water, and useful for sustainable land management. Most land-cover classification studies typically use low/medium-spatial resolution satellite images as done by Yao *et al.* [35], Patra *et al.* [70], Mansaray *et al.* [71], and Hadeel *et al.* [72]. Commonly, they are lack of spatial information; making difficulties in identifying land covers in high heterogeneous regions accurately. This study took full advantage of features of the high-resolution images taken from UAV images such as rich texture, shape, and spatial distribution information to the train land-use prediction models. The image segmentation step was done to generate a sample map or a mask to training models [73]–[75]. To make a mask for training the U-Net models, in this study, the authors also used image segmentation. However, the results from the image segmentation step were ineffective to use directly as a mask. An object generalization was made as a

mask to train the model (shown in Figure 4). Once the U-Net models were trained, this step can be eliminated. According to our results, it can be argued that the modified U-Net architecture is capable of producing land cover maps that improve the benchmark classification produced by former Random Forest and SVM models. The land-cover classification models can be attached directly to the UAV vehicle to detect the land covers quickly without post data analysis.

Importantly, the time to predict land-use types from digital images based on the trained U-Net models is also improved significantly, compared to the traditional methods. Commonly, scientists have to take various ground control points as a mask to interpret land-use types from UAV and other satellite images. Whole interpretation process requires a hard work both indoor and outdoor. Sometimes, it can take a month to have a land-use map for a particular time. The ground control points cannot be used for the new data or new areas. With the use of deep learning methods applied for trained U-Net models, now users can generate land-use map, particularly in the mining areas in a short time (about 120s in this study). Additionally, the trained models are not only used for the research area, but it can also be used for different areas where are monitored by the UAV technology.

Compared to the land-cover classification in urban and agricultural low-lands using deep learning from former studies [35], [37], [76]–[78], this study chose a typical mining area on basalt highland area of Vietnam. Except for the agricultural lands, two land cover types were not trained in former studies, including the open-cast mining areas and polluted water bodies. During the rock mining activities, basalt soils was extracted and left on the top soil surfaces. It leads to the specific texture and homogeneous spatial distribution of the open-case mining areas with other land cover types. A stream in the research area were polluted by the mining activities, leading to the high reflection of sunlight; whereas the freshwater stream will absorb energy from sunlight. Therefore, the pollution level of stream flowing from the open-case mining



**FIGURE 5.** Land cover interpretation based on Random Forest, Support Vector Machine and the seven U-Net models using seven optimizer functions in the mining area of Daknong province.

areas also can monitor if the UAV images are taken for a long time.

#### B. IMPROVEMENT OF LAND COVER CLASSIFICATION

Four trained U-Net models using Adam, Adagrad, RMSprop, and SGD optimizer functions (Group 1) cannot separate old and young permanent croplands due to the spectral similarity these two land covers. Whereas, the three models using Adamax, Nadam and Adadelta (Group 2) can separate these two agricultural lands based on the differences in their texture and shape. The differences between the two types of crops are their canopy. In the same area of croplands, the old permanent crops have a more massive canopy than the young ones. Small bare-soil regions that are distributed among the young permanent croplands can be eliminated by using the models in Group 1. Meanwhile, the models in Group 2 can detect both the small bare-soil regions and young permanent crops

as an object that is different from an object covered by full old permanent crops. It explains why the models in Group 2 have higher accuracy values than those in Group 1. It is crucial if developers in the future want to detect agricultural lands in different growing periods. New samples of agricultural land cover types can be automatically updated in next UAV flights to increase the accuracy of the trained U-Net model. It also can update the sample from high-resolution image from Google earth data sources. It should be done in further studies.

The U-Net development for land cover classification requires a cost and time-consuming dedication of developers, especially of scientists. To monitor mining land covers in real-time, the managers have to be equipped with a high-performance machine to (1) transmit data directly from the UAV machine to server with a high-speed connection, (2) pre-process UAV images, and (3) interpret the land cov-

ers based on the trained deep-learning model. In this study, the authors used a CPU Intel(R) Core™ i7-9750H CPU @ 2.60GHz CPU with 16GB RAM and GPU NVIDIA GeForce RTX 2060. The average time per epoch is more than 45 seconds. It needs to be improved with a supercomputer, especially in the case of the rapid land-cover classification based on the UAV data. Additionally, the trained model has to be updated from the new data. Instead of using seven optimizers to improve the U-Net models, various optimization techniques such as evolutionary or swarm intelligence algorithm can also be used in future work. It would be a potential method to train new knowledge for the trained U-Net models from new UAV data. Therefore, the proposed method does not only require a high-performed machine, but it also requires the development of 5G technology in the future.

## V. CONCLUSION

Based on the use of a U-Net model to classify land covers in a mining area of Daknong province, Vietnam, the individual research questions mentioned in the introduction section can be answered as follows:

- **What are the advantages of integration between deep learning and high-resolution images for monitoring mining areas?** The integration can reduce the time and cost to take samples when users want to interpret land cover types from a new UAV image. Once a U-Net model is trained, it can update new samples and be used to interpret quickly new data.
- **How are U-Net models for land cover classification on UAV images more effective than benchmark models?** The trained U-Net model in this study with the accuracy of 83% was used successfully to interpret six land cover types from UAV image taken in a mining area of Vietnam. The final model using Adadelta provided the more accurate results than two benchmark models (RF and SVM). This model can become a useful tool for managers in mining land cover management in the future.
- **How do land cover types distribute in a mining area of Daknong province, Vietnam?** Six land cover types were identified in the research area. Although the permanent croplands take 40%, open-cast mining take account for nearly 20%. Only one stream flowing out from the mining area has been affected by mining activities. Therefore, its high potential to monitor the land cover changes of this mining area based on the use of the trained U-Net models in the future.

However, the classification of land covers in mining areas requires a faster process to better support decision-makers during the monitoring activities. Although this research explained in detail scientific meaning and steps to train a model for mining land covers, it is challenging to complete all proposed processes on the field; apply it to the real-time monitoring. Further studies can use the trained models to set up on a digital chip to record and process UAV images directly.

## ACKNOWLEDGMENT

The authors also want to thank Pham Thi Xuan Quynh for language correction. The authors are grateful for the time and efforts from the editors and the anonymous reviewers on improving our manuscript.

## CONFLICT DECLARATION

Authors have no conflict of interest

## REFERENCES

- [1] M. Anthony, *Diversification by smallholder farmers: Viet Nam Robusta Coffee*. Rome, Italy: FAO, 2007.
- [2] P. T. Xuan, T. T. Anh, D. T. T. Tra, H. T. T. Nga, P. T. Dang, N. T. Lien, and N. V. Pho, "Environmental issues of mining activities in tay nguyen," *Vietnam J. Earth Sci.*, vol. 37, no. 2, pp. 139–147, Nov. 2015.
- [3] T. Chiramba, S. Mogoi, I. Martinez, and T. Jones, "Payment for forest ecosystem services (PFES): Pilot implementation in Lam Dong Province, Vietnam," in *Proc. UN-Water Int. Conf.*, Zaragoza, Spain, Oct. 2011, ch. 3, pp. 122–129.
- [4] N. Hieu, D. Van Bao, V. Van Vinh, T. P. N. Pham, D. T. Phuong, and P. H. T. Thi, "Geomorphological resources for tourism development in central Highlands and South Central coast (in Vietnamese)," *Vietnam J. Earth Sci.*, vol. 1S, no. 31, pp. 35–47, 2015.
- [5] G. Kissinger, *Drivers of Deforestation and Forest Degradation, or REDD+ Policymakers*. Vancouver, BC, Canada: Lexeme Consulting, 2012.
- [6] P. T. Thuy, M. Moeliono, N. T. Hien, N. H. Tho, and V. T. Hien, *The context of REDD + in Vietnam Drivers, Agents and Institutions*, CIFOR. Bogor, Indonesia: Center for International Forestry Research, 2012.
- [7] P. Meyfroidt and E. F. Lambin, "Forest transition in Vietnam and displacement of deforestation abroad," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 38, pp. 16139–16144, 2009.
- [8] D. Muller and M. Zeller, "Land use dynamics in the central highlands of vietnam: A spatial model combining village survey data with satellite imagery interpretation," *Agricul. Econ.*, vol. 27, no. 3, pp. 333–354, Nov. 2002, doi: [10.1111/j.1574-0862.2002.tb00124.x](https://doi.org/10.1111/j.1574-0862.2002.tb00124.x).
- [9] T.-H.-D. Phan, R. Brouwer, L. P. Hoang, and M. D. Davidson, "A comparative study of transaction costs of payments for forest ecosystem services in vietnam," *Forest Policy Econ.*, vol. 80, pp. 141–149, Jul. 2017, doi: [10.1016/j.forpol.2017.03.017](https://doi.org/10.1016/j.forpol.2017.03.017).
- [10] I. Jadin, V. Vanacker, and H. T. T. Hoang, "Drivers of forest cover dynamics in smallholder farming systems: The case of northwestern vietnam," *Ambio*, vol. 42, no. 3, pp. 344–356, Apr. 2013, doi: [10.1007/s13280-012-0348-4](https://doi.org/10.1007/s13280-012-0348-4).
- [11] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geosci. Frontiers*, vol. 7, no. 1, pp. 3–10, Jan. 2016, doi: [10.1016/j.gsf.2015.07.003](https://doi.org/10.1016/j.gsf.2015.07.003).
- [12] X. Ma, "Fully convolutional network for rice seedling and weed image segmentation at the seedling stage in paddy fields," *PLoS ONE*, vol. 14, no. 4, 2019, Art. no. e0215676, doi: [10.1371/journal.pone.0215676](https://doi.org/10.1371/journal.pone.0215676).
- [13] K. B. Dang, B. Burkhard, W. Windhorst, and F. Müller, "Application of a hybrid neural-fuzzy inference system for mapping crop suitability areas and predicting rice yields," *Environ. Model. Softw.*, vol. 114, pp. 166–180, Apr. 2019, doi: [10.1016/j.envsoft.2019.01.015](https://doi.org/10.1016/j.envsoft.2019.01.015).
- [14] A. Gebrehiwot, L. Hashemi-Beni, G. Thompson, P. Kordjamshidi, and T. Langari, "Deep convolutional neural network for flood extent mapping using unmanned aerial vehicles data," *Sensors*, vol. 19, no. 7, p. 1486, Mar. 2019, doi: [10.3390/s19071486](https://doi.org/10.3390/s19071486).
- [15] G. Tsagkatakis, A. Aidini, K. Fotiadou, M. Giannopoulos, A. Pentari, and P. Tsakalides, "Survey of deep-learning approaches for remote sensing observation enhancement," *Sensors*, vol. 19, no. 18, p. 3929, 2019, doi: [10.3390/s19183929](https://doi.org/10.3390/s19183929).
- [16] K. B. Dang, W. Windhorst, B. Burkhard, and F. Müller, "A Bayesian Belief Network-Based approach to link ecosystem functions with rice provisioning ecosystem services," *Ecol. Indicators*, vol. 100, pp. 30–44, May 2018, doi: [10.1016/j.ecolind.2018.04.055](https://doi.org/10.1016/j.ecolind.2018.04.055).
- [17] Q. Shi, W. Li, R. Tao, X. Sun, and L. Gao, "Ship classification based on multifeature ensemble with convolutional neural network," *Remote Sens.*, vol. 11, no. 4, p. 419, Feb. 2019, doi: [10.3390/rs11040419](https://doi.org/10.3390/rs11040419).
- [18] P. C. Gray, A. B. Fleishman, D. J. Klein, M. W. McKown, V. S. Bézy, K. J. Lohmann, and D. W. Johnston, "A convolutional neural network for detecting sea turtles in drone imagery," *Methods Ecol. Evol.*, vol. 10, no. 3, pp. 345–355, Dec. 2018, doi: [10.1111/2041-210X.13132](https://doi.org/10.1111/2041-210X.13132).

- [19] Q. Guo, S. Jin, M. Li, Q. Yang, K. Xu, Y. Ju, J. Zhang, J. Xuan, J. Liu, Y. Su, Q. Xu, and Y. Liu, "Application of deep learning in ecological resource research: Theories, methods, and challenges," *Sci. China Earth Sci.*, vol. 63, no. 10, pp. 1457–1474, Oct. 2020, doi: [10.1007/s11430-019-9584-9](https://doi.org/10.1007/s11430-019-9584-9).
- [20] C. Bacour, F. Baret, D. Béal, M. Weiss, and K. Pavageau, "Neural network estimation of LAI, fAPAR, fCover and LAI $\times$ c<sub>ab</sub>, from top of canopy MERIS reflectance data: Principles and validation," *Remote Sens. Environ.*, vol. 105, no. 4, pp. 313–325, Dec. 2006, doi: [10.1016/j.rse.2006.07.014](https://doi.org/10.1016/j.rse.2006.07.014).
- [21] F. Zambrano, A. Vrieling, A. Nelson, M. Meroni, and T. Tadesse, "Prediction of drought-induced reduction of agricultural productivity in chile from MODIS, rainfall estimates, and climate oscillation indices," *Remote Sens. Environ.*, vol. 219, pp. 15–30, Dec. 2018, doi: [10.1016/j.rse.2018.10.006](https://doi.org/10.1016/j.rse.2018.10.006).
- [22] Q. Feng, J. Yang, D. Zhu, J. Liu, H. Guo, B. Bayartungalag, and B. Li, "Integrating multitemporal sentinel-1/2 data for coastal land cover classification using a multibranch convolutional neural network: A case of the yellow river delta," *Remote Sens.*, vol. 11, no. 9, p. 1006, Apr. 2019, doi: [10.3390/rs11091006](https://doi.org/10.3390/rs11091006).
- [23] X. Chen, "A new application of random forest algorithm to estimate coverage of moss-dominated biological soil crusts in semi-Arid Mu Us Sandy Land, China," *Remote Sens.*, vol. 11, no. 11, p. 1286, 2019.
- [24] T. Liu and A. Abd-Elrahman, "Deep convolutional neural network training enrichment using multi-view object-based analysis of unmanned aerial systems imagery for wetlands classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 139, pp. 154–170, May 2018, doi: [10.1016/j.isprsjprs.2018.03.006](https://doi.org/10.1016/j.isprsjprs.2018.03.006).
- [25] M. R. Alizadeh and M. R. Nikoo, "A fusion-based methodology for meteorological drought estimation using remote sensing data," *Remote Sens. Environ.*, vol. 211, pp. 229–247, Jun. 2018, doi: [10.1016/j.rse.2018.04.001](https://doi.org/10.1016/j.rse.2018.04.001).
- [26] F. Pu, C. Ding, Z. Chao, Y. Yu, and X. Xu, "Water-quality classification of inland lakes using Landsat8 images by convolutional neural networks," *Remote Sens.*, vol. 11, no. 14, p. 1674, Jul. 2019, doi: [10.3390/rs11141674](https://doi.org/10.3390/rs11141674).
- [27] K. B. Dang, V. B. Dang, Q. T. Bui, V. V. Nguyen, T. P. N. Pham, and V. L. Ngo, "A convolutional neural network for coastal classification based on ALOS and NOAA satellite data," *IEEE Access*, vol. 8, pp. 11824–11839, 2020, doi: [10.1109/ACCESS.2020.2965231](https://doi.org/10.1109/ACCESS.2020.2965231).
- [28] W. Ge, Q. Cheng, Y. Tang, L. Jing, and C. Gao, "Lithological classification using sentinel-2A data in the shibanjing ophiolite complex in inner mongolia, China," *Remote Sens.*, vol. 10, no. 4, p. 638, Apr. 2018, doi: [10.3390/rs10040638](https://doi.org/10.3390/rs10040638).
- [29] M. Piragnolo, A. Masiero, and F. Pirotti, "Comparison of Random Forest and Support Vector Machine classifiers using UAV remote sensing imagery," *Geophys. Res. Abstract EGU Gen. Assement*, vol. 19, no. 3, pp. 2017–15692, 2017. [Online]. Available: <http://meetingorganizer.copernicus.org/EGU2017/EGU2017-15692-1.pdf>
- [30] T. D. Phan, J. C. R. Smart, S. J. Capon, W. L. Hadwen, and O. Sahin, "Applications of Bayesian belief networks in water resource management: A systematic review," *Environ. Model. Softw.*, vol. 85, pp. 98–111, Nov. 2016, doi: [10.1016/j.envsoft.2016.08.006](https://doi.org/10.1016/j.envsoft.2016.08.006).
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (Lecture Notes in Computer Science), vol. 9351, N. Navab, J. Hornegger, W. Wells, and A. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241, doi: [10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [32] G. Jakovljevic, M. Govendarica, and F. Alvarez-Taboada, "A deep learning model for automatic plastic mapping using unmanned aerial vehicle (UAV) data," *Remote Sens.*, vol. 12, no. 9, p. 1515, May 2020, doi: [10.3390/RS12091515](https://doi.org/10.3390/RS12091515).
- [33] F. I. Diakogiannis, F. Waldner, P. Caccetta, and C. Wu, "ResUNet-A: A deep learning framework for semantic segmentation of remotely sensed data," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 94–114, Apr. 2020, doi: [10.1016/j.isprsjprs.2020.01.013](https://doi.org/10.1016/j.isprsjprs.2020.01.013).
- [34] M. Carranza-García, J. García-Gutiérrez, and J. Riquelme, "A framework for evaluating land use and land cover classification using convolutional neural networks," *Remote Sens.*, vol. 11, no. 3, p. 274, Jan. 2019, doi: [10.3390/rs11030274](https://doi.org/10.3390/rs11030274).
- [35] X. Yao, H. Yang, Y. Wu, P. Wu, B. Wang, X. Zhou, and S. Wang, "Land use classification of the deep convolutional neural network method reducing the loss of spatial features," *Sensors*, vol. 19, no. 12, p. 2792, Jun. 2019, doi: [10.3390/s19122792](https://doi.org/10.3390/s19122792).
- [36] P. Zhang, Y. Ke, Z. Zhang, M. Wang, P. Li, and S. Zhang, "Urban land use and land cover classification using novel deep learning models based on high spatial resolution satellite imagery," *Sensors*, vol. 18, no. 11, p. 3717, Nov. 2018, doi: [10.3390/s18113717](https://doi.org/10.3390/s18113717).
- [37] L. Garg, P. Shukla, S. Singh, V. Bajpai, and U. Yadav, "Land use land cover classification from satellite imagery using mUnet: A modified Unet architecture," in *Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl.*, vol. 14, 2019, pp. 359–365.
- [38] F. H. Wagner, "Using the U-net convolutional network to map forest types and disturbance in the Atlantic rainforest with very high resolution images," *Remote Sens. Ecol. Conservation*, vol. 5, no. 4, pp. 360–375, 2019, doi: [10.1002/rse2.111](https://doi.org/10.1002/rse2.111).
- [39] R. Li, W. Liu, L. Yang, S. Sun, W. Hu, F. Zhang, and W. Li, "DeepUNet: A deep fully convolutional network for pixel-level sea-land segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3954–3962, Nov. 2018, doi: [10.1109/JSTARS.2018.2833382](https://doi.org/10.1109/JSTARS.2018.2833382).
- [40] 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. Geoinf.*, vol. 8, no. 4, p. 160, Mar. 2019, doi: [10.3390/ijgi8040160](https://doi.org/10.3390/ijgi8040160).
- [41] A. Stoian, V. Poulain, J. Ingla, P. Poughon, and D. Derkzen, "Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems," *Remote Sens.*, vol. 11, no. 17, p. 1986, 2019, doi: [10.3390/rs11171986](https://doi.org/10.3390/rs11171986).
- [42] J. Morris, "The Vietnamese bauxite mining controversy: The emergence of a new oppositional politics," Ph.D. dissertation, Dept. Environ. Sci., Policy Manage., Univ. California, Berkeley, Berkeley, CA, USA, 2013, p. 205.
- [43] A. E. Maxwell and T. A. Warner, "Thematic classification accuracy assessment with inherently uncertain boundaries: An argument for center-weighted accuracy assessment metrics," *Remote Sens.*, vol. 12, no. 12, p. 1905, Jun. 2020, doi: [10.3390/rs12121905](https://doi.org/10.3390/rs12121905).
- [44] A. Bogoliubova and P. Tymkow, "Accuracy assessment of automatic image processing for land cover classification of St. Petersburg protected area," *Acta Scientiarum Polonorum. Geodesia et Descriptio Terrarum*, vol. 13, pp. 5–22, 2014.
- [45] I. Varfolomeev, I. Yakimchuk, and I. Safonov, "An application of deep neural networks for segmentation of microtomographic images of rock samples," *Comput.*, vol. 8, no. 4, p. 72, 2019, doi: [10.3390/computers8040072](https://doi.org/10.3390/computers8040072).
- [46] X. Wu, Y. Shi, S. Fomel, and L. Liang, "Convolutional neural networks for fault interpretation in seismic images," in *Proc. SEG Tech. Program Expanded Abstr.*, Aug. 2018, pp. 1946–1950, doi: [10.1190/segam2018-2995341.1](https://doi.org/10.1190/segam2018-2995341.1).
- [47] A. Garcia-Pedrero, M. Lillo-Saavedra, D. Rodriguez-Esparragon, and C. Gonzalo-Martin, "Deep learning for automatic outlining agricultural parcels: Exploiting the land parcel identification system," *IEEE Access*, vol. 7, pp. 158223–158236, 2019, doi: [10.1109/ACCESS.2019.2950371](https://doi.org/10.1109/ACCESS.2019.2950371).
- [48] G. J. Scott, R. A. Marcum, C. H. Davis, and T. W. Nivin, "Fusion of deep convolutional neural networks for land cover classification of high-resolution imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1638–1642, Sep. 2017, doi: [10.1109/LGRS.2017.2722988](https://doi.org/10.1109/LGRS.2017.2722988).
- [49] A. Gulli, *Deep Learning With Keras—Implement Neural Networks With Keras on Theano and TensorFlow*. Birmingham, U.K.: Packt, 2017.
- [50] A. F. M. Agarap, "Deep learning using rectified linear units (ReLU)," *Neural Evol. Comput. Comput. Sci.*, vol. 1, pp. 2–8, 2019.
- [51] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," 2018, *arXiv:1811.03378*. [Online]. Available: <http://arxiv.org/abs/1811.03378>
- [52] H. A. H. Al-najjar, B. Kalantar, B. Pradhan, and V. Saeidi, "Land cover classification from fused DSM and UAV images using convolutional neural networks," *Remote Sens.*, vol. 11, no. 12, p. 1461, 2019.
- [53] B. Gao and L. Pavel, "On the properties of the softmax function with application in game theory and on the properties of the softmax function with application in game theory and reinforcement learning," *Comput. Sci., Math.*, pp. 1–10, 2018. [Online]. Available: <https://arxiv.org/abs/1704.00805>
- [54] A. F. T. Martins and R. F. Astudillo, "From softmax to sparsemax: A sparse model of attention and multi-label classification," in *Proc. 33rd Int. Conf. Mach. Learn.*, vol. 48, 2016, p. 10.
- [55] H. Perez, J. H. M. Tah, and A. Mosavi, "Deep learning for detecting building defects using convolutional neural networks," *Sensors*, vol. 19, no. 16, p. 3556, Aug. 2019, doi: [10.3390/s19163556](https://doi.org/10.3390/s19163556).
- [56] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).

- [57] M. Lapin, M. Hein, and B. Schiele, "Analysis and optimization of loss functions for multiclass, top-k, and multilabel classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1533–1554, Jul. 2018, doi: [10.1109/TPAMI.2017.2751607](https://doi.org/10.1109/TPAMI.2017.2751607).
- [58] P. C. Gray, A. B. Fleishman, D. J. Klein, and M. W. McKown, "A convolutional neural network for detecting sea turtles in drone imagery," *Methods Ecol. Evol.*, vol. 10, no. 3, Mar. 2019, doi: [10.1111/2041-210X.13132](https://doi.org/10.1111/2041-210X.13132).
- [59] A. Gulli and S. Pal, *Implement Neural Networks With Keras on Theano and TensorFlow, Deep Learning With Keras*. Birmingham, U.K.: Packt, 2017.
- [60] K. Ahuja, "Estimating kullback-leibler divergence using kernel machines," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 690–696, doi: [10.1109/IEECONF44664.2019.9049082](https://doi.org/10.1109/IEECONF44664.2019.9049082).
- [61] D. J. Galas, G. Dewey, J. Kunert-Graf, and N. A. Sakharenko, "Expansion of the Kullback-Leibler divergence, and a new class of information metrics," *Axioms*, vol. 6, no. 2, p. 8, 2017, doi: [10.3390/axioms6020008](https://doi.org/10.3390/axioms6020008).
- [62] L. Wang, Y. Yang, R. Min, and S. Chakradhar, "Accelerating deep neural network training with inconsistent stochastic gradient descent," *Comput. Sci. Mach. Learn. Univ.*, vol. 3, p. 12, 2017.
- [63] M. Z. Alom, "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, 2019, doi: [10.3390/electronics8030292](https://doi.org/10.3390/electronics8030292).
- [64] V. Iglovikov, S. Mushinskiy, and V. Osin, "Satellite imagery feature detection using deep convolutional neural network: A kaggle competition," 2017, *arXiv:1706.06169*. [Online]. Available: <http://arxiv.org/abs/1706.06169>
- [65] D. Falbel, "R interface to 'Keras,'" R. Package Version 2.2.4.1, Tech. Rep., 2019. Accessed: Oct. 11, 2020. [Online]. Available: <https://CRAN.R-project.org/package=keras>
- [66] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.
- [67] T. Berhane, C. Lane, Q. Wu, B. Autrey, O. Anenkhonov, V. Chepinoga, and H. Liu, "Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory," *Remote Sens.*, vol. 10, no. 4, p. 580, Apr. 2018, doi: [10.3390/rs10040580](https://doi.org/10.3390/rs10040580).
- [68] A. Karatzoglou, D. Meyer, and K. Hornik, "Support Vector Algorithm in R," *J. Stat. Softw.*, vol. 15, no. 9, pp. 1–28, 2006.
- [69] J. Du, Y. Liu, Y. Yu, and W. Yan, "A prediction of precipitation data based on support vector machine and particle swarm optimization (PSO-SVM) algorithms," *Algorithms*, vol. 10, no. 2, p. 57, May 2017, doi: [10.3390/a10020057](https://doi.org/10.3390/a10020057).
- [70] S. Patra, S. Sahoo, P. Mishra, and S. C. Mahapatra, "Impacts of urbanization on land use/cover changes and its probable implications on local climate and groundwater level," *J. Urban Manage.*, vol. 7, no. 2, pp. 70–84, Sep. 2018, doi: [10.1016/j.jum.2018.04.006](https://doi.org/10.1016/j.jum.2018.04.006).
- [71] L. Mansaray, W. Huang, D. Zhang, J. Huang, and J. Li, "Mapping rice fields in urban shanghai, southeast China, using sentinel-1A and landsat 8 datasets," *Remote Sens.*, vol. 9, no. 3, p. 257, Mar. 2017, doi: [10.3390/rs9030257](https://doi.org/10.3390/rs9030257).
- [72] A. Hadeel, M. Jabbar, and X. Chen, "Remote sensing and GIS application in the detection of environmental degradation indicators," *Geo-Spatial Inf. Sci.*, vol. 14, no. 1, pp. 39–47, Jan. 2011, doi: [10.1007/s11806-011-0441-z](https://doi.org/10.1007/s11806-011-0441-z).
- [73] T. M. Berhane et al., "Comparing pixel- and object-based approaches in effectively classifying wetland-dominated landscapes," *Remote Sens.*, vol. 10, no. 1, 2018, doi: [10.3390/rs10010046](https://doi.org/10.3390/rs10010046).
- [74] M. Krestenitis, G. Orfanidis, K. Ioannidis, K. Avgerinakis, S. Vrochidis, and I. Kompatiari, "Oil spill identification from satellite images using deep neural networks," *Remote Sens.*, vol. 11, no. 15, p. 1762, 2019, doi: [10.3390/rs11151762](https://doi.org/10.3390/rs11151762).
- [75] C. Munyati, T. Ratshibvumo, and J. Ogola, "Landsat TM image segmentation for delineating geological zone correlated vegetation stratification in the Kruger National Park, South Africa," *Phys. Chem. Earth, A/B/C*, vols. 55–57, pp. 1–10, 2013, doi: [10.1016/j.pce.2009.11.014](https://doi.org/10.1016/j.pce.2009.11.014).
- [76] Q. Hu, W. Wu, T. Xia, Q. Yu, P. Yang, Z. Li, and Q. Song, "Exploring the use of Google Earth imagery and object-based methods in land Use/Cover mapping," *Remote Sens.*, vol. 5, no. 11, pp. 6026–6042, Nov. 2013, doi: [10.3390/rs5116026](https://doi.org/10.3390/rs5116026).
- [77] H. A. H. Al-Najjar, "Land cover classification from fused DSM and UAV images using convolutional neural networks," *Remote Sens.*, vol. 11, no. 12, p. 1461, 2019, doi: [10.3390/rs11121461](https://doi.org/10.3390/rs11121461).
- [78] C. Zhang, I. Sargent, X. Pan, H. Li, A. Gardiner, J. Hare, and P. M. Atkinson, "An object-based convolutional neural network (OCNN) for urban land use classification," *Remote Sens. Environ.*, vol. 216, pp. 57–70, Oct. 2018, doi: [10.1016/j.rse.2018.06.034](https://doi.org/10.1016/j.rse.2018.06.034).



**TUAN LINH GIANG** received the B.S. degree in cartographic, remote sensing, and GIS from the Faculty of Geography, VNU University of Science, in 2010, and the M.S. degree in environmental bio-agriculture from the Public University of Navarra, Spain, in 2013.

Since 2010, she has been a Research Assistant with different NGO Agricultural Organization, such as CIAT SKYMAP High Technology Company Ltd., and GREENFIELD, Vietnam. She has experience on land use/cover policies. Her research interests include applications of deep learning, remote sensing and GIS in ecological modeling, sustainable agriculture, and land management.



**KINH BAC DANG** (Member, IEEE) received the B.S. degree in geomorphology and natural hazards, in 2010, the M.S. degree in cartographic, remote sensing, and GIS from the Faculty of Geography, VNU University of Science, in 2013, and the Ph.D. degree in agricultural sciences/ecology from the University of Kiel, Kiel, Germany.

From 2016 to 2018, he was a Research Assistant with the Institute for Natural Resource Conservation, University of Kiel. Since 2019, he has been a Lecturer with the Faculty of Geography, VNU University of Science. He is the author of different articles published in *Environmental Modelling and Software*, *Journal of Environmental Management*, and *Ecological Indicators* journals. His research interests include interdisciplinary sciences and applications of deep learning, remote sensing and GIS in ecological modeling, sustainable agriculture, and natural hazards.

D. Dang received awards and honors, include the TOSHIBA Scholarship, Japan, the 911-VIED Fellowships, Vietnam, and the Certificate of Reviewing from *Ecological Indicators* journal. He serves as a Senior Reviewer for journal *Ecological Indicators*.



**QUANG TOAN LE** (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in mapping remote sensing and GIS from the Faculty of Geography, VNU University of Science-Hanoi National University, in 2007, 2011, and 2019, respectively.

He was with the Center for Applied Research in Remote Sensing and GIS, in 2007 and the Space Technology Institute, Vietnam Academy of Science and Technology (VAST), in 2008. He was a Remote Sensing and GIS Specialist with different projects funded by FSDP, the World Bank (WB), the International Union for Conservation of Nature and Natural (IUCN), and the Netherlands Development Organization (SNV). His research interests include using remote sensing and GIS for management of environment and natural resources toward the sustainable development, establishing and updating database for environment and natural resources assessment using remote sensing and GPS data, and organizing and carrying out training transfer program in remote sensing and GIS.



**VU GIANG NGUYEN** (Member, IEEE) is currently pursuing the Ph.D. degree in urbanization, disaster prevention, and nature landscape with KU Leuven University, Belgium.

He was a Technical Consultant with NGOs and NPOs, such as the Asian Development Bank (ADB), the International Union for Conservation of Nature (IUCN), and the Netherlands Development Organization (SNV), where he carried out capacity building training on land use planning and rural livelihood development, creating land use, land cover maps, and forest cover change analysis. He is currently a Researcher and a Deputy Chief Manager with the Department of Remote Sensing Technology, GIS and GPS, Space Technology Institute (STI), Vietnam. His studies focus on application of remote sensing and GIS for investing, monitoring, and managing the use of natural resources. His research interests include modeling of land use/cover change, estimating biomass, and creating emission maps from multisources using remote sensing data.



**VAN-MANH PHAM** (Member, IEEE) received the bachelor's and master's degrees in cartography, remote sensing, and GIS from the VNU University of Science, Hanoi, Vietnam, in 2007 and 2013, respectively, where he is currently pursuing the Ph.D. degree. His research interests include remote sensing, spatial analysis, satellite image analysis, land-cover/land-use, and geoinformatics.

• • •



**SI SON TONG** (Member, IEEE) received the Ph.D. degree in remote sensing applications for the coastal zone from the University of Reims Champagne-Ardenne, France, in 2017. Since 2018, he has been a Lecturer and a Researcher with the Department of Space and Applications, University of Science and Technology of Hanoi. His research interests include using multisource remote sensing data and spatial analysis tools to monitor land use/cover change.