Bulletin of Geophysics and Oceanography Vol. XX, n. X, pp. XXX-XXX; Xxxxxx 20XX
DOI 10.4430/bgo00472

Image segmentation for burned area detection from satellite imagery using the U-Net deep learning model

D. Alkan and L. Karasaka

Department of Geomatic Engineering, Konya Technical University, Konya, Turkey

(Received: 4 January 2024; accepted: 22 July 2024; published online: 10 October 2024)

ABSTRACT Fires threaten life all over the world and damage millions of hectares of area every year. Remote sensing provides advantages for damage detection in terms of time and cost. By using satellite imagery, burned areas can be detected without the need to visit the area. Since factors such as image band configuration, optimisation algorithms, and thresholds affect the results, this study aims to observe their impact on burned area detection. Thus, by using Landsat-8 images and U-Net architecture through the Python programming language, various combinations were created and different thresholds were used. According to the results, the combination of 7, 5, 4 bands and the AdaMax algorithm were selected for the final model, and the results were improved by data augmentation. Consequently, accuracy obtained in the final model was 97.76%, which was the highest for a threshold of 0.5. The F1 score obtained for the same threshold was 79.38%.

Key words: burned area, deep learning, Landsat-8, remote sensing, segmentation, U-Net.

1. Introduction

Disasters, which occur due to natural or man-made causes (Martin, 2010), are dangerous events that result in various economic (e.g. damage to livelihoods, infrastructures, etc.) and social (e.g. occurrence of injuries, deaths, psychological effects, etc.) losses (Kalfin *et al*., 2022). Fires, one of the disasters, are caused by a chemical reaction between combustible materials and oxygen in the air; smoke released during a fire is described as a complex chemical mixture (Friedman, 1998). Fires can be caused by uncontrolled combustion or spontaneous natural processes (e.g. lightning, volcanic eruptions, etc.) as well as by controlled burn plans. In addition, human activities (e.g. field work, livestock management, etc.) in fields, forests, or land may also cause fires (Pratama *et al*., 2023). Uncontrolled fires can spread rapidly and cause greater hazards (Chen L. *et al*., 2023). They can cause landslides, erosion, desertification, etc. and limit or prevent living organisms from meeting their needs such as food and accommodation, thus, risking their life. According to the World Fire Statistics report published by CTIF[1], in 2021, nearly 3.1 million fires were reported in the 38 countries surveyed. Of the approximately 1.2 billion people living in these countries, almost 16,800 people lost their lives due to fires (CTIF, 2023). According to

[1] The International Association of Fire and Rescue Service (CTIF) is a worldwide organisation founded in 1900. It was organised to provide safety for firefighters. The CTIF Center of Fire Statistics (CFS) has been publishing a report on world fire statistics every year since 1995, bringing together data (e.g. fire statistics, fire-related deaths, etc.) from European countries and other countries.

GFW[2] data, in the same year, approximately 38% of the world's tree cover was lost due to fires (Tyukavina *et al*., 2022). It is necessary to identify the areas damaged by fires, and even take actions to recover these areas, in order to continue to benefit from the opportunities offered by nature and to maintain the sustainability of life. Remote sensing is one of the techniques that provides monitoring of the Earth's surface, including the detection of burned areas. It performs the survey without any physical contact with the Earth. Remote sensing provides an advantage especially in large areas, such as forests, where observations are very difficult and the possibilities of reaching burned areas are limited with terrestrial methods. Remote sensing satellites collect a large amount of data for investigating changes in the Earth's surface.

Various methods are used to analyse the collected data. Traditional rule-based methods are based on the detection of differences in the spectral response of burned areas, especially in the NIR (Near Infrared) and SWIR (Short Wave Infrared) bands, and the selection of a threshold value (Knopp *et al*., 2020). For the detection of the burned area, pre-fire and post-fire images are generally used. Burned areas can be detected by obtaining spectral indices (e.g. Normalised Difference Vegetation Index, Normalised Burn Ratio, etc.) for each image and taking their differences [e.g. Difference Normalised Difference Vegetation Index (dNDVI), Difference Normalised Burn Ratio (dNBR), etc.]. However, the results obtained may be affected by conditions such as cloud cover. Traditional machine learning based techniques are based on learning the features of burned areas from labelled samples (Knopp *et al*., 2020). Algorithms such as random forests, support vector machines, and linear regression are machine learning algorithms, which use various attributes. Thereby, they maximise inter-class variation and minimise intra-class variation. In this way, burned areas are detected. However, the results are affected by the feature selection process in the pre-step stage, which is difficult and time-consuming (Zhang Q. *et al*., 2021).

Due to such disadvantages of traditional methods, deep learning based methods have been developed. These methods enable feature extraction to be automatically handled, as they do not require any additional input from the user except for the setting of a few hyperparameters (Hu *et al*., 2021). Deep learning is used in different fields, among which classification, segmentation, object detection, and fire detection (Hassaballah and Awad, 2020). Detection of fires and/or burned areas with deep learning methods is a current topic that has frequently been preferred, especially in the last five years (Park and Lee, 2019; Farasin *et al*., 2020; Khryashchev and Larionov, 2020; Knopp *et al*., 2020; Lee *et al*., 2020; Mohla *et al*., 2020; Pinto *et al*., 2020; Rashkovetsky, 2020; Brand and Manandhar, 2021; Knopp, 2021; Rashkovetsky *et al*., 2021; Prabowo *et al*., 2022b). U-Net architecture is generally preferred in studies on this subject. With this architecture, Park and Lee (2019) used Landsat-5 and Landsat-8 data along with spectral indices such as the NDVI, Normalised Difference Water Index, and Fractional Water Index; Knopp *et al*. (2020) and Knopp (2021) used Sentinel-2 data; Lee *et al*. (2020) used Sentinel-1 and Sentinel-2 data; Rashkovetsky (2020) and Rashkovetsky *et al*. (2021) used Sentinel-1, Sentinel-2, Sentinel-3, and Terra and Aqua data; Brand and Manandhar (2021) used MODIS data; Prabowo *et al*. (2022b) used Landsat-8 data to detect fires and/or burned areas. Alternatively, for the detection of burned areas, Farasin *et al*. (2020) used U-Net and CuMedVision architectures along with Sentinel-2 data and indices such as the Burned Area Index, NBR, NBR2, and BAIS2 (Burned Area Index for Sentinel-2). Mohla *et al*. (2020) used Landsat-8 data with a model called AmazonNet based on U-Net architecture. Khryashchev and Larionov (2020) developed the U-ResNet34 model with ResNet34 and U-Net

[2] Global Forest Watch (GFW) is an online platform founded in 1997 by The World Resources Institute (WRI). It aims to monitor changes in forests around the world and make them accessible to everyone.

structures, and performed fire detection using Planet and Resurs-P data. Pinto *et al*. (2020) developed a model, called BA-Net, which used convolutional neural network and long short-term memory architectures and implemented the mapping and dating of burnt regions using VIIRS data. The authors observed that roads (Park and Lee, 2019; Knopp, 2021), clouds (Mohla *et al*., 2020), agricultural areas, and smoke (Knopp *et al*., 2020) may lead to misclassifications. They also included a reference data (mask) in their deep learning architecture in addition to the input data. The reference mask is a binary image consisting only of the colours black and white. As the generation of the reference masks is also a factor that affects the deep learning results, masks are usually generated by experts in the related field.

In this study, the data set established by Prabowo *et al*. (2022b) was used. Prabowo *et al*. (2022b) created a data set consisting of images and reference masks for Indonesia. Indonesia is ranked as one of the countries with the highest rates of natural disasters worldwide according to the Statista[3] Research Department. The risk index for disasters such as landslides, droughts and forest fires is higher than in other countries (Statista Research-Department, 2023b). In 2021 alone, approximately 0.35 million hectares of land burned in Indonesia (Statista Research-Department, 2023a). According to GFW data, approximately 3.4% of Indonesia's tree cover was lost due to fires in the same year (Tyukavina *et al*., 2022). Since Indonesia has large forests which are difficult to monitor with terrestrial methods, Prabowo *et al*. (2022b) used their data set for Indonesia to detect burned areas, but only in a single combination. Conversely, this study aims to examine the effects of different combinations on the results. For this purpose, unlike the works of these authors, various band combinations, optimisation algorithms and threshold values were used, their effects on burned area detection were separately discussed and the generation of the final model was explained and illustrated in detail. In addition, data augmentation for the final model was also examined to verify whether it improves the results.

2. Material and method

2.1. Data set for burned areas

In this study, the data set named "Data set of Deep Learning from Landsat-8 Satellite Images for Estimating Burned Areas in Indonesia" created by Prabowo *et al*. (2022b) was used. This data set consists of burned areas in Indonesia. It contains 227 Landsat-8 images and the reference masks corresponding to these images. The reference masks in the data set were created by experts. The images are in Geo TIFF file format, and are in the World Geodetic System 1984 datum. Some of the images contain regions such as settlements, water bodies, barren lands, clouds, and cloud shadows, while others consist of smoke from still-burning areas. Some of the images contain regions where there are no burned areas (Prabowo *et al*., 2022a). It has been observed that the data set has been updated over time and various versions have been created. In this study, version 2 (file name: fs7mtkg2wk-2) was used. The multispectral images in the data set are in uint16 format, i.e. 16-bit unsigned integer format. They consist of eight bands. These bands are coastal/aerosol, blue, green, red, NIR, SWIR-1, SWIR-2, and cirrus bands. The masks corresponding to the multispectral images in the data set are in uint8 format, i.e. 8-bit unsigned integer format. They consist of a single band in the form of black and white binary images. In these images, burned areas are represented by 1 and unburned areas are represented by 0.

[3] Statista is a worldwide platform founded in 2007. It provides data (e.g. reports, statistics, etc.) on various topics.

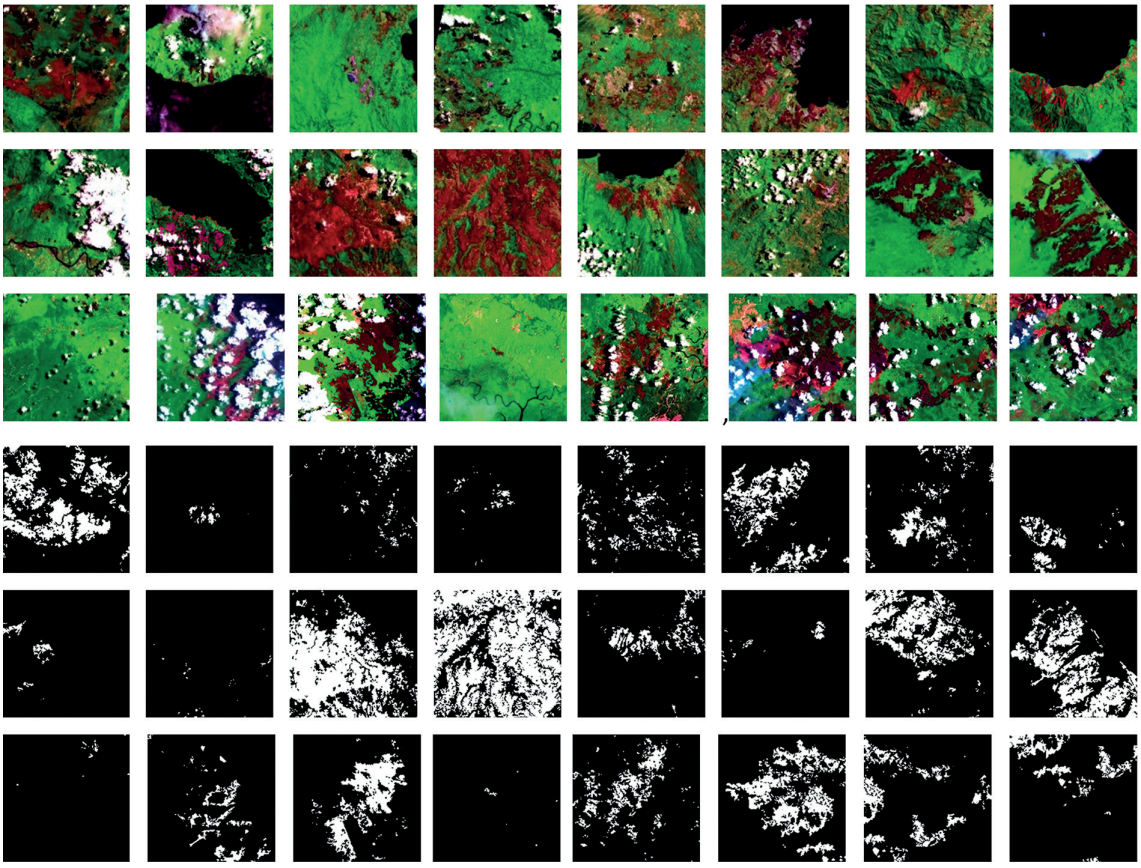

Fig. 1 - Some examples from the data set.

2.2. U-Net deep learning architecture

The U-Net architecture derives its name from its similarity to the letter U. The general structure of this architecture, which includes a total of 23 convolution layers, is shown in Fig. 2. In the figure, each multi-channel feature map is expressed as a blue coloured box. The numbers written above the boxes correspond to the number of channels. The numbers written at the bottom left of the boxes are the x-y size. The boxes shown in white colour are the copied feature maps. The different coloured arrows between the boxes represent different operations. As seen in the figure, the U-Net structure consists of two parts: the downward part on the left side and the upward part on the right side. These two parts are referred to as contracting path and expansive path operations, respectively (Ronneberger *et al.*, 2015). These two terms are also referred to as encoder and decoder in the literature (Zhang J. *et al.*, 2018; Futrega *et al.*, 2021). Down sampling is performed on the contraction path and up sampling is performed on the expansive path.

2.3. Implementation platform

Within the scope of this study, the operations were carried out using the Python programming language. Keras with TensorFlow backend was mainly used. In the implementation, pre-

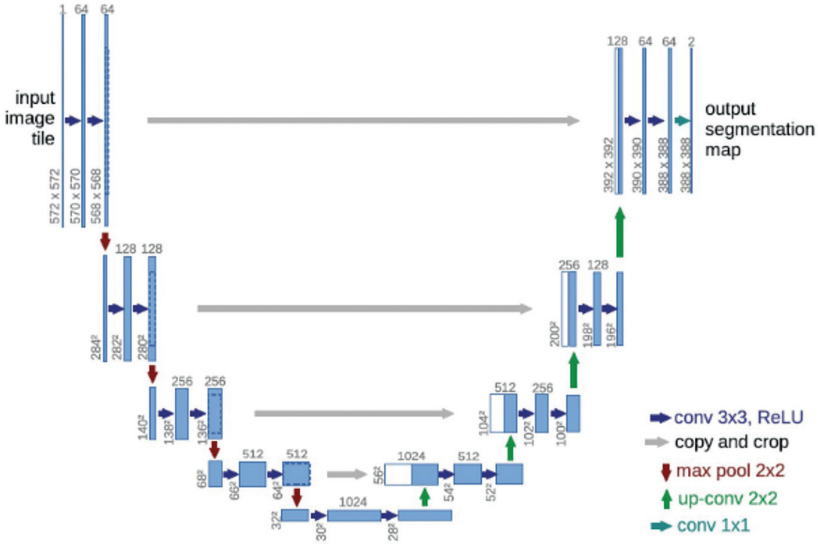

Fig. 2 - The U-Net structure (Ronneberger *et al*., 2015).

processing (generating band combinations, splitting images, separating data set for training, validating, testing, and using data augmentation, etc.) was performed on the computer via the Anaconda Spyder editor. Training, test and accuracy assessments were completed through Google Colaboratory (Colab). The computer model used for the pre-processing is Lenovo Legion Y530 with Intel® Core™ i7-8750H CPU @ 2.20GHz and NVIDIA GeForce GTX 1050 Ti 4 GB graphics card. In Google Colab, the free version providing 12.7 GB of RAM was initially used, and, then, the Google Colab Pro paid version was purchased and resulted to be more user friendly since it offered more time and resources compared to the free version.

2.4. Hyperparameters

While designing the model, hyperparameter should be selected. Hyperparameters are the parameters of the algorithm whose values are determined by the user before training. According to the different hyperparameter selections, different machine learning models are formed. Therefore, it is important to determine the appropriate hyperparameter values (Alkan, 2023). The hyperparameters that must be defined in deep learning networks are described below.

Batch size: the value that defines the amount of data to be selected in each iteration. Choosing a larger batch size results in a faster convergence of the model but requires much more memory (Chen Z. *et al*., 2021). It should be noted that the batch size is limited by the graphical processing unit (GPU) memory (Gao *et al*., 2021). This value is usually chosen as powers of 2, i.e. 4, 8, 16, 32, etc. In this study, batch size 8 was preferred.

Learning rate: a value chosen between 0 and 1 (Lee and Chung, 1970). In order to choose the appropriate learning rate, a number of tests are necessary. Choosing a low learning rate causes the steps taken during the training to be too small (Salim *et al*., 2023) and the model will, consequently, learn slowly. In addition, the model performs a considerable number of epochs and the training time becomes longer. However, the model with a low learning rate is more sensitive to reach the best solution (Wolansky, 2021). Therefore, it is important to find the appropriate learning rate for a good training (Salim *et al*., 2023). In this study, 0.001, 0.0001, and 0.00001 values were used for the learning rate. These values were tried in the relevant network and the value that provided the most appropriate results was preferred as the learning rate in that network.

Number of epochs: the number that defines the number of runs of the algorithm on the entire data set. In this study, the number of epochs used was 100 and early stopping was applied for the final model. With the early stopping, the training process is automatically stopped when the validation accuracy begins to degrade (Zhang T. *et al*., 2021).

Loss function: the function that enables the display of model error. For this purpose, it calculates the difference between the predicted value and the actual value (Chen X. *et al*., 2022). The loss function used in this study is binary cross entropy (BCE), also called log loss (Panwar *et al*., 2017). Eq. 1 shows the mathematical equivalent of the BCE:

$$K = -\frac{1}{N}\sum_{i=0}^{N} y_i \times log(p_i) + (1 - y_i) \times log(1 - p_i) \quad (1)$$

where *N* corresponds to the number of all samples, *y* refers to the label and *p* refers to the probability. Loss functions are necessary to estimate the loss value in each epoch. They are also a part of optimisation (Tan *et al*., 2021).

Optimisation: the parameter concerning the selection of the optimal solution from various choices (Farhan *et al*., 2020). Optimisation algorithms aim to minimise the loss function. Therefore, they update the parameter values in the network throughout the learning process (Salehzadeh *et al*., 2020). Adam (Adaptive Moment Estimation), AdaGrad (Adaptive Gradient) and AdaMax (Maximum Adaptive Moment Estimation) algorithms are some of the examples of optimisation algorithms. Adam is a stochastic gradient descent (SGD) method. It enables the measurement of adaptive learning rates for each parameter (Haji and Abdulazeez, 2021). It was presented by Kingma and Ba (2015). AdaGrad is a method where the learning rate is chosen by the situation. In AdaGrad, a higher parameter gradient will have a lower learning rate, while a lower parameter gradient will have a higher learning rate (Haji and Abdulazeez, 2021). This method was presented by Duchi *et al*. (2011). AdaMax is a method based on an adaptive form of SGD (Haji and Abdulazeez, 2021): it is an extension of the Adam optimisation algorithm (Soydaner, 2020) and was presented by Kingma and Ba (2015).

Regularisation: the technique used to prevent the model from overfitting (Patterson and Gibson, 2017). For this purpose, it reduces the variance in the validation set. This enables the model to better generalise to new samples (Bisong, 2019). In this study, dropout was used for regularisation. The dropout method determines whether each neuron is selected or not. In this way, some of the neurons are included in the process while others are not. This is to avoid overfitting of the network (Srivastava *et al*., 2014).

Activation functions: the functions, also called transfer functions (Krenker *et al*., 2011), used to transmit the output value of neurons, that are in one layer, to the next layers (Ser and Bati, 2019). Through these functions, a task can be learned, understood, and performed (Chieng *et al*., 2018), and complex and non-linear functions can also be modelled (Sureshbabu *et al*., 2023). In this study, the rectified linear unit (ReLU) and sigmoid were used as the activation function. ReLU converts negative values in the input data to zero. Sigmoid, conversely, generates a smaller value when the input data is lower than a certain threshold and generates a higher value when the input data is larger (Lewis-Atwell *et al*., 2022). The ReLU output values are in the range [0, + ∞] while the sigmoid output values are in the range [0, 1]. In Fig. 3, ReLU and sigmoid graphs are shown with their mathematical equivalents.

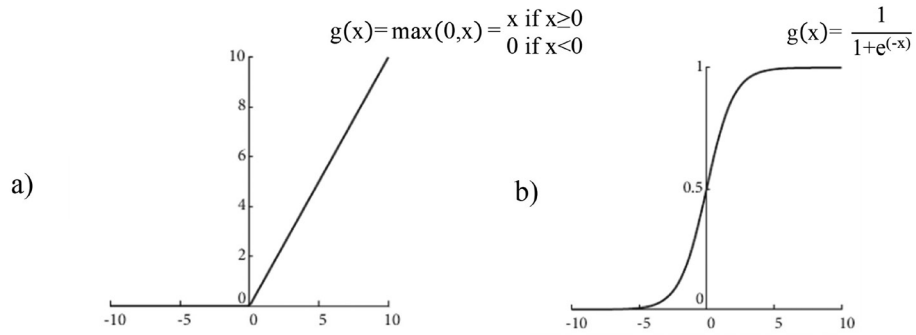

Fig. 3 - Activation function graphs: a) ReLU and b) sigmoid (Jin and Niu, 2021).

2.5. Evaluation metrics

Evaluation metrics are used to evaluate the performance of the model. The evaluation is performed by means of qualitative and quantitative methods. Qualitative methods are performed by comparing visual results with real visual data. Quantitative methods are recall, precision, F_1 score, and Jaccard score metrics (Clarisse, 2021). The calculation of these metrics is generated from the confusion matrix. The confusion matrix, as in Fig. 4, shows the numbers of true positive (*TP*), false positive (*FP*), true negative (*TN*) and false negative (*FN*). For burned areas, the correctly labelled burned samples are represented by *TP*s, the incorrectly labelled unburned samples are represented by *FP*s, the correctly labelled unburned samples are represented by *TN*s, and the incorrectly labelled burned samples are represented by *FN*s (Zhang Q. *et al*., 2021). Table 1 shows the calculation and definitions of the accuracy metrics over the confusion matrix.

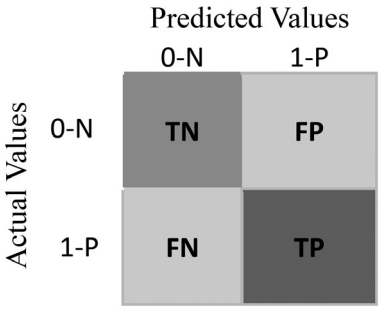

Fig. 4 - Confusion matrix.

2.6. Implementation workflow

In this study, to begin, multispectral images in uint16 format were converted to uint8 format. The obtained images, and their corresponding masks, were divided into three parts to be used in the training, validation, and test process. For this purpose, 10% of the total data set was taken to create the test data set, 10% of the remaining was taken to create the validation data set, and the remaining part was used as the training data set. In this process, the images were randomly separated. The number of images and masks obtained after the separation process are as follows: 184 for training, 20 for validation, and 23 for test. Next, this data set, 512×512×n in size, was subjected to image splitting, thus, reducing the size to 256×256×n due to GPU limitations

Table 1 - Metrics and their definitions.

Metrics	Definitions
$\text{Overall Accuracy}= \frac{TP+TN}{TP+FP+TN+FN}$	Overall accuracy is a metric of the number of correct predictions divided by the total number of predictions.
$\text{Recall}= \frac{TP}{TP+FN}$	Recall indicates how many of the really positive ones are defined as positive in the prediction.
$\text{Precision}= \frac{TP}{TP+FP}$	Precision indicates how many of what is defined as positive in the prediction is also identified as positive in reality.
$F_1= \frac{2\times(\text{Precision}\times\text{Recall})}{\text{Precision}+\text{Recall}}$	The F1 score is a metric that takes into consideration both precision and recall values in predictions.
$\text{Jaccard} = \frac{(A\cap B)}{(A\cup B)}=\frac{TP}{TP+FP+FN}$	The Jaccard score [also called intersection over union (IoU)] is a metric that measures the similarity between the predicted region and the actual region (Huang, 2021).

in memory. The number of final images and masks obtained after the splitting process are as follows: 736 for training, 80 for validation, and 92 for test. Each data set was converted from uint to float format before being used as input in the U-Net architecture and were normalised to be between 0 and 1. Then, the training process was performed with training and validation data sets. The number of epochs was 100, the batch size was 8, and the loss function was BCE. For the learning rate, 0.001, 0.0001, and 0.00001 values were experimented and the most appropriate one was selected. Adam was used as the optimisation algorithm. Subsequently, AdaGrad and AdaMax were also used in the band combination that provided the best and worst results. After the training process, for each combination, the test process was performed with test data sets created in different combinations. For each test image, probability results comprised of values between 0 and 1 were obtained. At that point, in order to generate prediction results consisting of only 0 and 1 values, a threshold value was used. In this study, three different thresholds were used to examine whether the threshold value has an effect on the results. The threshold values chosen were 0.1, 0.5, and 0.9. As a consequence, three prediction results were obtained for each test image. The prediction results were evaluated using the reference masks in the test data set. Overall accuracy, recall, precision, F_1 score, and Jaccard score metrics were used for the evaluation. The flowchart followed in the study is shown in Fig. 5.

2.7. Data augmentation

Data augmentation is the general name given to techniques that increase the number of training data. For this purpose, new images are obtained by applying various transformations to the images in the data set (Alkan, 2023). Within the scope of this study, the results achieved with different combinations were evaluated and the combination that gave the best results was determined. In this combination, data augmentation was performed by applying 90°, 180°, and 270° rotations to the training data set. As a consequence of the data augmentation, 2,208 images and masks were obtained for the training data set. Fig. 6 shows an image from the data set and the new images obtained by applying 90°, 180°, and 270° rotations to this image.

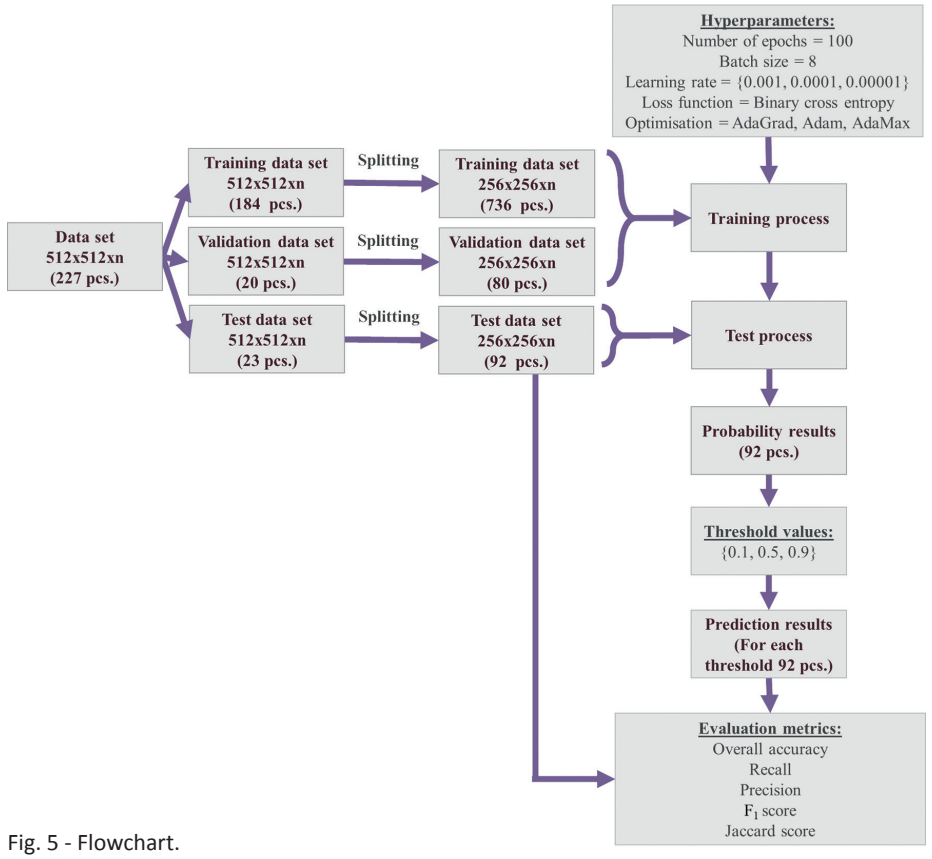

Fig. 5 - Flowchart.

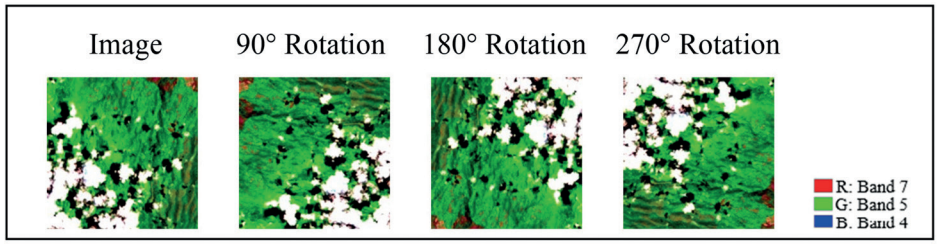

Fig. 6 - Original image and its 90°, 180°, 270° rotations.

3. Results and discussion

This study was carried out in three stages which are discussed in the following sections: band combinations, optimisation algorithms, and final model. Fig. 7 summarises the implementation stages.

3.1. Band combinations

The first implementation stage aimed at investigating the effect of different band combinations on the results. For this purpose, six different combinations were created as 7, 5, 4; 5, 3, 7; 5, 4, 3; 4, 3, 2; 4, 3, 2, 5 and 2, 3, 4, 5, 6, 7 bands. Each combination was separately used with the Adam

Fig. 7 - Implementation stages.

optimisation algorithm. The training process was carried out with the training and validation data sets created in the relevant combination. Table 2 shows the learning rates, training times, training accuracy, and training loss results for each combination.

After the training process, the test process was performed using the test data sets, and was separately carried out for each combination in Table 2. As a conclusion, for each test image, one probability and three prediction results were obtained. Visual results are presented from Figs. 8 to 10. Here, three out of the 92 test images (image 9, image 13, and image 92), the reference masks corresponding to these images, the probability and prediction results obtained in the application are shown. Evaluation metrics were used to evaluate the results. Accuracy, recall, precision, F_1 score, and Jaccard score metrics were calculated for each test image. The metric results for image 9, image 13, and image 92 are presented in Tables 3 to 5. In these tables, the abbreviation "Th." is used to represent the threshold value and "img." is used to represent the image. In the related tables, the results for unburned and burned areas are represented by the initialisms UA and BA, respectively.

Fig. 8 shows the visual results for the combinations ID = 1 and ID = 2. The metric results obtained in these combinations are presented in Table 3. In image 9 (Fig. 8), croplands (blue rectangle) are misclassified in the combinations ID = 1 and ID = 2. The results are worse in the combination ID = 1. Table 3 shows that for image 9, the accuracy and F_1 score values in the burned areas are lower in the combination ID = 1 than for the combination ID = 2. In image 13 (Fig. 8), the road (red rectangle) is misclassified in both combinations. Table 3 shows that for image 13, the highest value

Table 2 - Training information for the combinations ID = 1, ..., 6.

ID	Bands	Optim.	Learning Rate	Training Time	Training Accuracy	Training Loss
1	7, 5, 4	Adam	0.001	1 hour and 36 minutes	0.9886	0.0272
2	5, 3, 7	Adam	0.001	1 hour and 37 minutes	0.9909	0.0236
3	5, 4, 3	Adam	0.0001	1 hour and 53 minutes	0.9916	0.0204
4	4, 3, 2	Adam	0.00001	1 hour and 50 minutes	0.9776	0.0562
5	4, 3, 2, 5	Adam	0.001	1 hour and 45 minutes	0.9806	0.0501
6	2, 3, 4, 5, 6, 7	Adam	0.00001	1 hour and 54 minutes	0.9849	0.0368

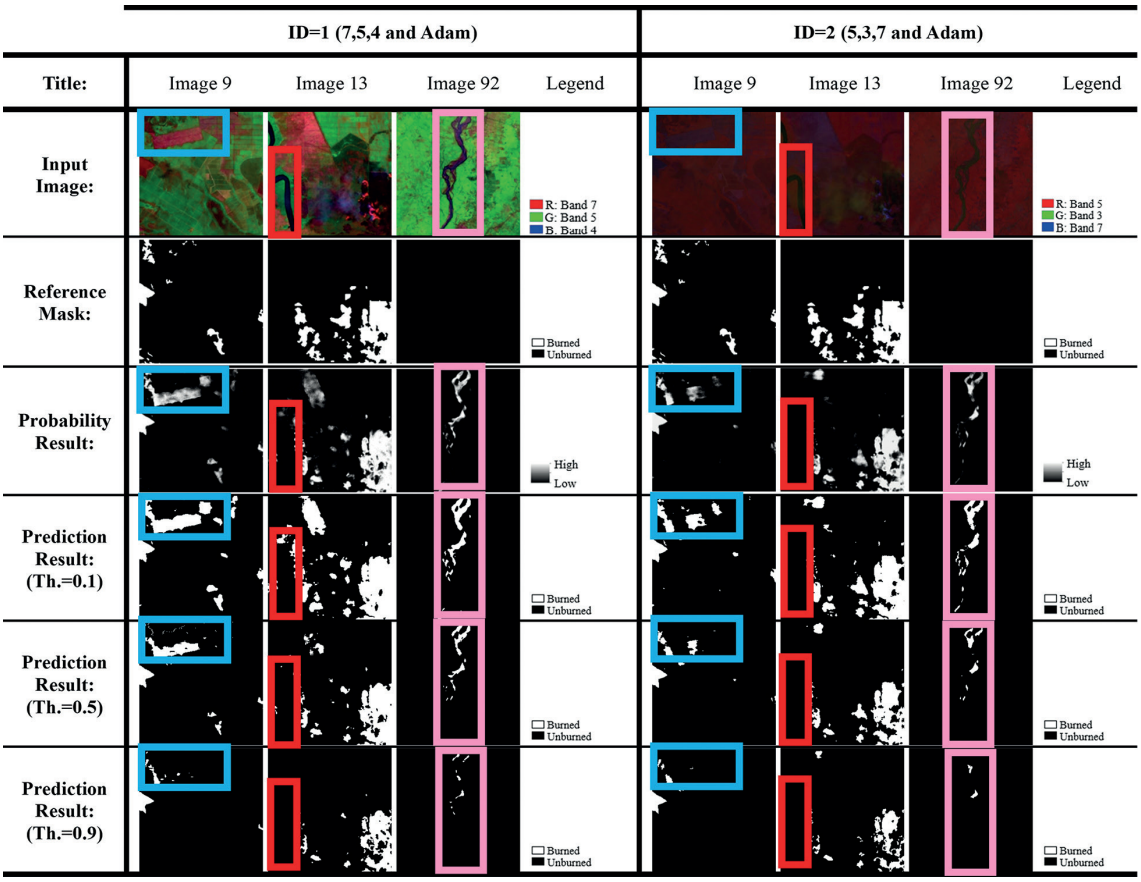

Fig. 8 - Visual results for the combinations ID = 1 and ID = 2, each with three test images, the corresponding reference image, and the results.

among the Jaccard scores is 64.10% for the combination ID = 1 and threshold = 0.5. This value indicates that this prediction result is more similar to the reference data than the other results. In image 92 (Fig. 8), the road (pink rectangle) is misclassified in two combinations although there are no burned areas in the reference mask. The results are worse in the combination ID = 1. Table 3 shows that for image 92, the accuracy values in combination ID = 1 are lower than in combination ID = 2. The F_1 score values in burned areas are calculated as 0% in both combinations. This is due to the fact that there are no burned areas in the reference mask but there are burned areas in the results. Therefore, in view of recall, the question may concern how much of the actually burned areas were also burned in the prediction. The answer to this question is 0% according to Table 3, as there are no burned areas in the reference mask. Similarly, in view of precision, the question on how much of the areas that were burned in the prediction were actually burned may be posed. The answer to this question is also 0% according to Table 3. Again, this is due to no burned areas in the reference mask. As a consequence, the recall, precision, and F_1 score values were calculated as 0% in burned areas. In view of recall in unburned areas, the question may be on how much of the areas that are actually unburned are also unburned in the prediction. The answer to this question is more than 95% according to Table 3. The answer is not 100% as there are burned areas in the results. Likewise, in view of precision in unburned areas, the question may be on how much of the areas that are unburned in the prediction are actually unburned. The answer to

Table 3 - Metric results of the three test images for the combinations ID = 1 and ID = 2.

Th.	Img.	ID = 1 (7, 5, 4 and Adam)								ID = 2 (5, 3, 7 and Adam)							
		Accuracy	Recall		Precision		F_1 Score		Jaccard	Accuracy	Recall		Precision		F_1 Score		Jaccard
			UA	BA	UA	BA	UA	BA			UA	BA	UA	BA	UA	BA	
0.1	9	0.9324	0.9359	0.8465	0.9935	0.3450	0.9638	0.4902	0.3247	0.9553	0.9689	0.6147	0.9844	0.4409	0.9765	0.5135	0.3455
	13	0.8849	0.8919	0.8463	0.9699	0.5851	0.9292	0.6919	0.5289	0.9111	0.9395	0.7533	0.9548	0.6918	0.9471	0.7212	0.5640
	92	0.9681	0.9681	0.0000	1.0000	0.0000	0.9838	0.0000	0.0000	0.9728	0.9728	0.0000	1.0000	0.0000	0.9862	0.0000	0.0000
0.5	9	0.9505	0.9648	0.5909	0.9833	0.4014	0.9740	0.4780	0.3141	0.9660	0.9871	0.4370	0.9777	0.5754	0.9824	0.4967	0.3304
	13	0.9374	0.9744	0.7320	0.9528	0.8375	0.9635	0.7812	0.6410	0.9206	0.9749	0.6194	0.9343	0.8165	0.9542	0.7044	0.5437
	92	0.9821	0.9821	0.0000	1.0000	0.0000	0.9910	0.0000	0.0000	0.9889	0.9889	0.0000	1.0000	0.0000	0.9944	0.0000	0.0000
0.9	9	0.9700	0.9953	0.3364	0.9741	0.7408	0.9846	0.4627	0.3010	0.9708	0.9965	0.3256	0.9737	0.7897	0.9850	0.4611	0.2997
	13	0.9301	0.9972	0.5576	0.9260	0.9726	0.9602	0.7088	0.5490	0.9031	0.9944	0.3963	0.9014	0.9271	0.9456	0.5553	0.3843
	92	0.9958	0.9958	0.0000	1.0000	0.0000	0.9979	0.0000	0.0000	0.9970	0.9970	0.0000	1.0000	0.0000	0.9985	0.0000	0.0000

this question is 100% according to Table 3. This is due to the fact that the unburned areas in the prediction correspond to all the unburned areas in the reference mask.

Fig. 9 shows the visual results for the combinations ID = 3 and ID = 4. The metric results obtained

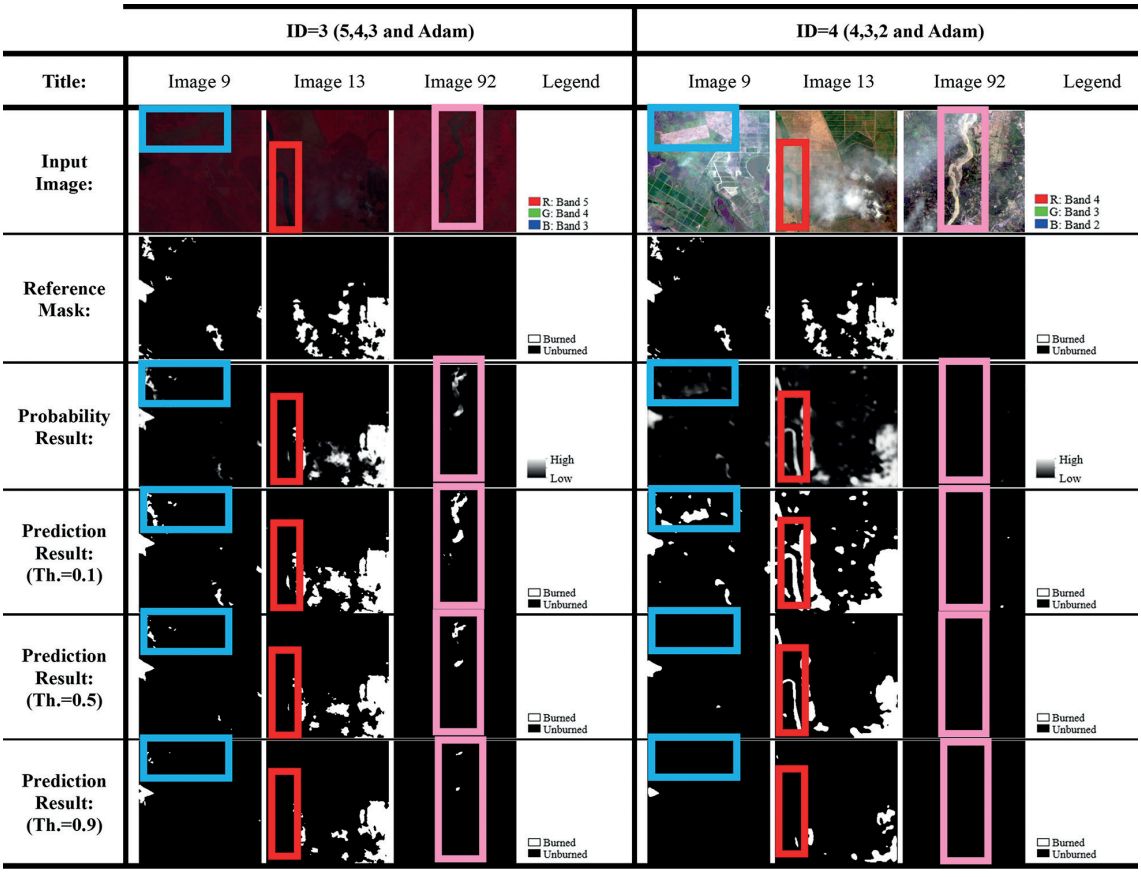

Fig. 9 - Visual results for the combinations ID = 3 and ID = 4, each with three test images, the corresponding reference image, and the results.

in these combinations are presented in Table 4. In image 9 (Fig. 9), the cropland (blue rectangle) is not misclassified in the combination ID = 3 although there are misclassifications in the combination ID = 4 for threshold = 0.1. Table 4 shows that for image 9, the accuracy values and the F_1 score values in the burned areas in the combination ID = 4 are less than in the combination ID = 3. In image 13 (Fig. 9), the road (red rectangle) is misclassified in the combination of ID = 4. However, this misclassification did not occur in the combination ID = 3. Table 4 shows that for image 13, the accuracy values and F_1 score values in the burned areas are higher in the combination ID = 3 than in the combination ID = 4. In image 92 (Fig. 9), the road (pink rectangle) is not misclassified in the combination ID = 4. There is some misclassification in ID = 3, but this is less than in the combinations ID = 1 and ID = 2 in Table 3. Table 4 shows that for image 92, the accuracy values in the combination ID = 3 are higher than in the combinations ID = 1 and ID = 2 in Table 3. However, recall, precision, and F_1 score values in burned areas are the same. Table 4 shows that for image 92, the accuracy for the combination ID = 4 is 99.94% for threshold = 0.1 and 100% for other thresholds. This is due to the fact that burned areas are detected in the prediction result obtained at threshold = 0.1. In the other threshold results, the prediction results consist completely of unburned areas. Therefore, the recall, precision, and F_1 score values obtained at threshold = 0.1 are similar to the results of the previous combinations. However, in the combination ID = 4, the recall, precision, and F_1 score values at the other thresholds were not calculated for burned areas and resulted as 100% for unburned areas. This is due to the fact that, in the prediction results at thresholds of 0.5 and 0.9, the actual and predicted results do not include burned areas and consist completely of unburned areas.

Fig. 10 shows the visual results for the combinations ID = 5 and ID = 6. The metric results obtained for these combinations are presented in Table 5. In Fig. 10, the cropland in image 9 (blue rectangle) and the road in image 13 (red rectangle) are not misclassified in the combinations ID = 5 and ID = 6. However, in the combination ID = 5, for threshold value = 0.1 in image 13, the smoke from the fire can be claimed to be misclassified. In this combination, the result for threshold = 0.5 is more similar to the reference mask. Table 5 shows that for image 13, the accuracy values and F_1 score values in the burned areas are higher for threshold = 0.5 than the other thresholds. In image 92 (Fig. 10), the road (pink rectangle) is not misclassified in the combination ID = 6. However, it is misclassified in the combination ID = 5 for threshold = 0.1. In Table 5 the metrics obtained for image 92 can be explained in a similar way as in the combination ID = 4.

Table 4 - Metric results of the three test images for the combinations ID = 3 and ID = 4.
*x represents the cases where the metric calculation was not performed.

Th.	Img.	ID = 3 (5, 4, 3 and Adam)								ID = 4 (4, 3, 2 and Adam)							
		Accuracy	Recall		Precision		F_1 Score		Jaccard	Accuracy	Recall		Precision		F_1 Score		Jaccard
			UA	BA	UA	BA	UA	BA			UA	BA	UA	BA	UA	BA	
0.1	9	0.9812	0.9933	0.6775	0.9872	0.8008	0.9902	0.7340	0.5798	0.9556	0.9697	0.6000	0.9838	0.4420	0.9767	0.5090	0.3414
	13	0.9166	0.9384	0.7954	0.9622	0.6996	0.9501	0.7444	0.5929	0.8139	0.8276	0.7376	0.9459	0.4354	0.8828	0.5476	0.3770
	92	0.9817	0.9817	0.0000	1.0000	0.0000	0.9907	0.0000	0.0000	0.9994	0.9994	0.0000	1.0000	0.0000	0.9996	0.0000	0.0000
0.5	9	0.9772	0.9989	0.4346	0.9779	0.9398	0.9882	0.5943	0.4228	0.9717	0.9999	0.2624	0.9714	0.9970	0.9855	0.4155	0.2622
	13	0.9277	0.9729	0.6772	0.9436	0.8182	0.9580	0.7411	0.5886	0.8850	0.9390	0.5851	0.9262	0.6336	0.9326	0.6084	0.4372
	92	0.9925	0.9925	0.0000	1.0000	0.0000	0.9962	0.0000	0.0000	1.0000	1.0000	x*	1.0000	x*	1.0000	x*	0.0000
0.9	9	0.9724	0.9999	0.2823	0.9722	0.9986	0.9859	0.4402	0.2822	0.9661	1.0000	0.1169	0.9660	1.0000	0.9827	0.2093	0.1169
	13	0.9285	0.9901	0.5867	0.9300	0.9142	0.9591	0.7147	0.5561	0.8809	0.9913	0.2680	0.8826	0.8479	0.9338	0.4073	0.2557
	92	0.9985	0.9985	0.0000	1.0000	0.0000	0.9992	0.0000	0.0000	1.0000	1.0000	x*	1.0000	x*	1.0000	x*	0.0000

Table 5 - Metric results of the three test images for the combinations ID = 5 and ID = 6
*x represents the cases where the metric calculation was not performed..

Th.	Img.	ID = 5 (4, 3, 2, 5 and Adam)								ID = 6 (2, 3, 4, 5, 6, 7 and Adam)							
		Accuracy	Recall		Precision		F_1 Score		Jaccard	Accuracy	Recall		Precision		F_1 Score		Jaccard
			UA	BA	UA	BA	UA	BA			UA	BA	UA	BA	UA	BA	
0.1	9	0.9815	0.9843	0.9121	0.9964	0.6988	0.9903	0.7913	0.6547	0.9743	0.9988	0.3582	0.9750	0.9279	0.9868	0.5169	0.3485
	13	0.8403	0.8182	0.9642	0.9922	0.4887	0.8968	0.6486	0.4800	0.9360	0.9858	0.6592	0.9413	0.8936	0.9631	0.7588	0.6113
	92	0.9912	0.9912	0.0000	1.0000	0.0000	0.9956	0.0000	0.0000	0.9999	0.9999	0.0000	1.0000	0.0000	0.9999	0.0000	0.0000
0.5	9	0.9781	0.9988	0.4580	0.9788	0.9404	0.9887	0.6160	0.4451	0.9724	1.0000	0.2807	0.9721	1.0000	0.9858	0.4384	0.2807
	13	0.9362	0.9642	0.7810	0.9607	0.7971	0.9624	0.7890	0.6516	0.9161	0.9979	0.4620	0.9114	0.9759	0.9527	0.6271	0.4568
	92	0.9998	0.9998	0.0000	1.0000	0.0000	0.9999	0.0000	0.0000	1.0000	1.0000	x*	1.0000	0.0000	1.0000	x*	0.0000
0.9	9	0.9698	1.0000	0.2131	0.9695	1.0000	0.9845	0.3513	0.2131	0.9679	1.0000	0.1638	0.9677	1.0000	0.9836	0.2815	0.1638
	13	0.9257	0.9987	0.5210	0.9204	0.9860	0.9580	0.6818	0.5172	0.8703	0.9999	0.1509	0.8673	0.9980	0.9289	0.2621	0.1508
	92	1.0000	1.0000	x*	1.0000	0.0000	1.0000	x*	0.0000	1.0000	1.0000	x*	1.0000	0.0000	1.0000	x*	0.0000

Table 6 shows the means of the metric results separately calculated for 92 test images for each combination. Examination of these combinations shows that the best results are given by the 7, 5, 4 bands in the combination ID = 1 and the Adam algorithm. The worst results were

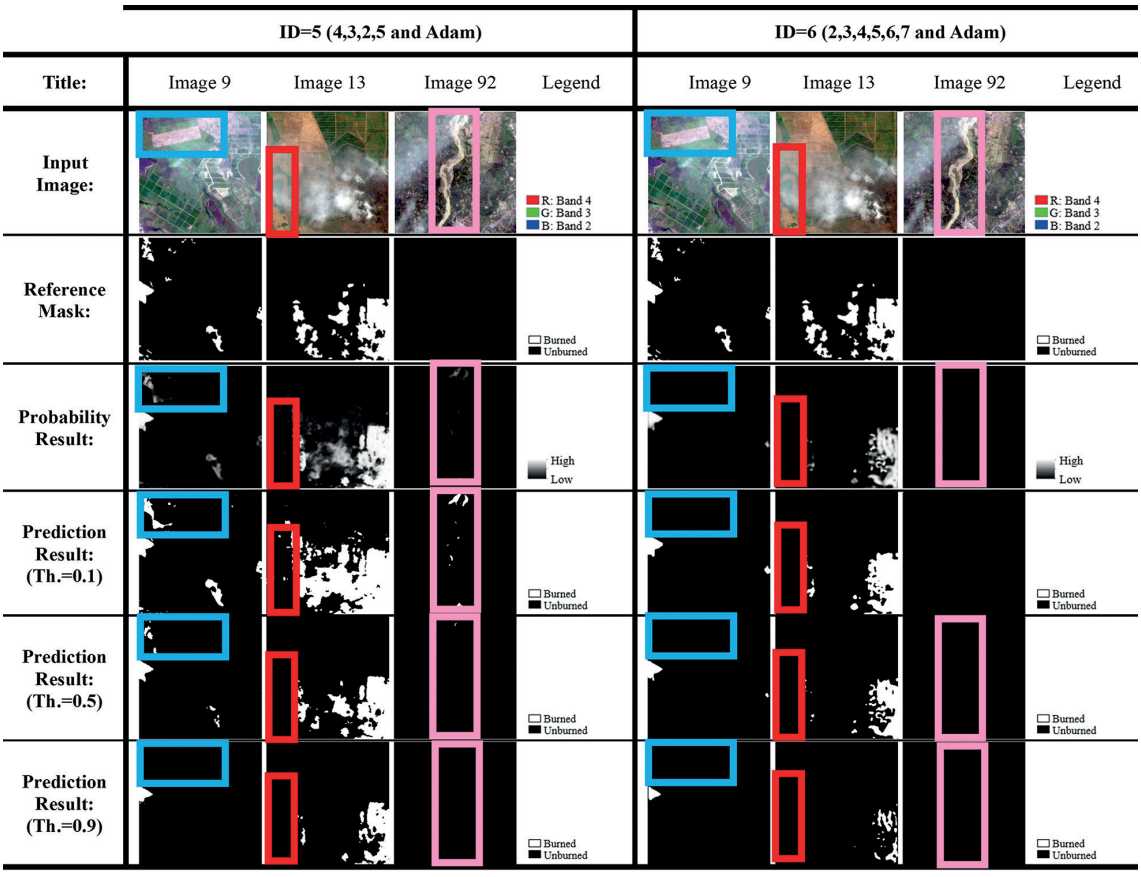

Fig. 10 - Visual results for the combinations ID = 5 and ID = 6, each with three test images, the corresponding reference image, and the results.

Table 6 - The means of the metric results for the combinations ID = 1, ..., 6.

ID	Accuracy			Recall BA			Precision BA			F1-Score BA			Jaccard Score		
	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9
1	0.9741	0.9743	0.9609	0.8364	0.7203	0.5268	0.7047	0.8411	0.9366	0.7509	0.7601	0.6382	0.6399	0.6399	0.4788
2	0.9700	0.9722	0.9608	0.8096	0.718	0.5259	0.6779	0.8247	0.8998	0.726	0.7539	0.6333	0.6065	0.6243	0.4935
3	0.9679	0.9673	0.9577	0.7393	0.5962	0.4399	0.7145	0.8114	0.861	0.7136	0.6688	0.5496	0.5704	0.5243	0.4029
4	0.9436	0.9518	0.9302	0.6297	0.4249	0.1997	0.5177	0.6962	0.6904	0.5396	0.4882	0.2782	0.4233	0.365	0.1786
5	0.9509	0.9677	0.949	0.8131	0.6391	0.3471	0.5996	0.7548	0.852	0.6721	0.6764	0.4513	0.555	0.5564	0.3215
6	0.9609	0.9525	0.9331	0.6071	0.4637	0.2564	0.7434	0.8601	0.8018	0.5208	0.4088	0.2307	0.5208	0.4088	0.2307

obtained with the 4, 3, 2 bands in the combination ID = 4 and the Adam algorithm. Therefore, in the second part of the application, these two band combinations were used with AdaGrad and AdaMax optimisation algorithms to create new combinations.

3.2. Optimisation algorithms

In the second implementation stage, the combinations identified as the best and worst in the first phase were used with AdaGrad and AdaMax optimisation algorithms, with the purpose of examining the effect of the optimisation algorithms on the results. Therefore, the training and validation data sets with combinations 7, 5, 4 and 4, 3, 2 were separately trained with AdaGrad and AdaMax algorithms. Table 7 shows the learning rates, training times, training accuracy and training loss results for each combination.

After the training process, test data sets were created with the same combinations, and the test process was carried out for each combination. The procedures for generating and presenting the test results were followed in the same way as in the first part of the implementation.

Fig. 11 shows the visual results for the combinations ID = 7 and ID = 8. The metric results obtained for these combinations are presented in Table 8. In Fig. 11, the cropland in image 9 (blue rectangle) and the road in image 13 (red rectangle) are misclassified in the combinations ID = 7 and ID = 8. However, the road in image 13 (red rectangle) is not misclassified in both combinations.

Fig. 12 shows the visual results for the combinations ID = 9 and ID = 10. The metric results obtained for these combinations are presented in Table 9. The worst results were observed in these combinations.

The means of the metric results for 92 test images are shown in Table 10 for 7, 5, 4 bands and Table 11 for 4, 3, 2 bands. According to Table 10, the best results are given by the AdaMax algorithm using 7, 5, 4 bands (ID = 8). Similarly, according to Table 11, the best results are given by the Adam algorithm using 4, 3, 2 bands (ID = 4). Evaluating the results, the optimisation algorithms are found to have an effect on the results.

Table 7 - Training information for the combinations ID = 7, ..., 10.

ID	Bands	Optim.	Learning Rate	Training Time	Training Accuracy	Training Loss
7	7, 5, 4	AdaGrad	0.001	1 hour and 47 minutes	0.9827	0.0431
8	7, 5, 4	AdaMax	0.001	1 hour and 46 minutes	0.9907	0.0226
9	4, 3, 2	AdaGrad	0.00001	1 hour and 56 minutes	0.9286	0.2821
10	4, 3, 2	AdaMax	0.00001	1 hour and 57 minutes	0.9556	0.1199

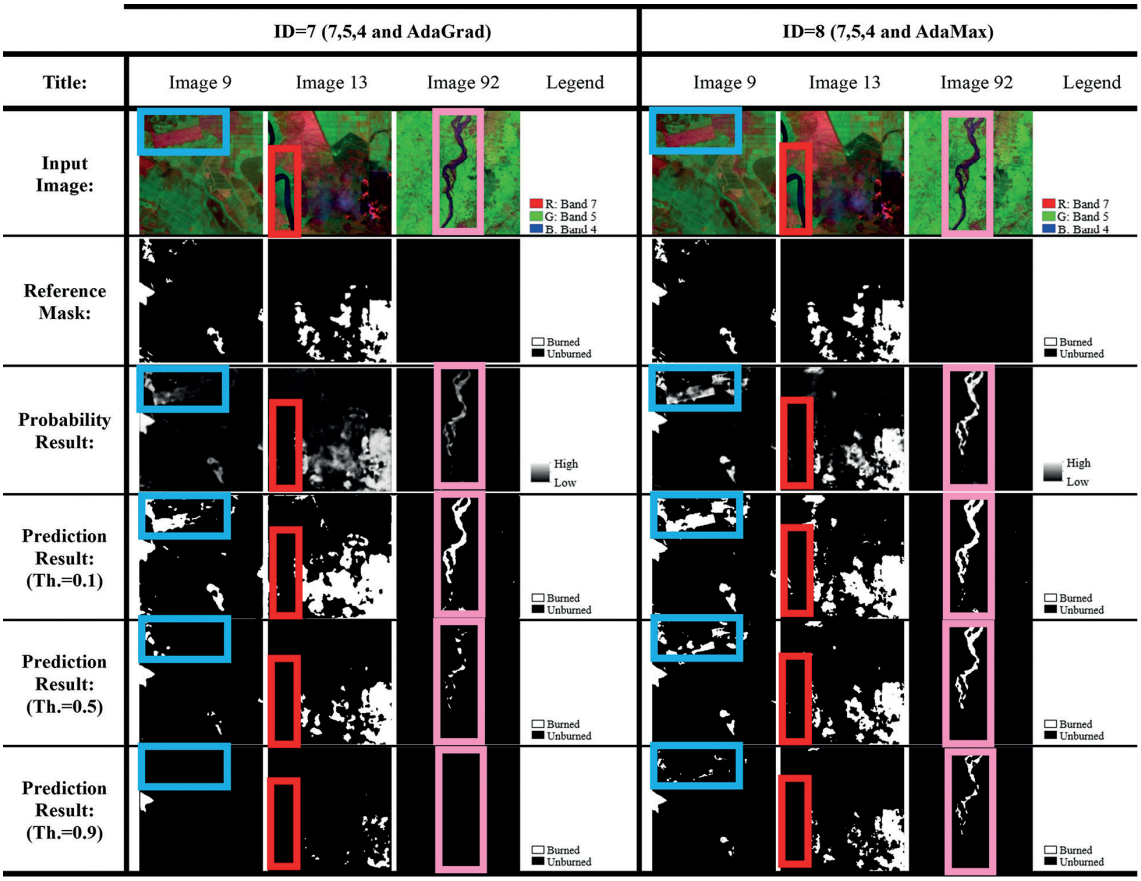

Fig. 11 - Visual results for the combinations ID = 7 and ID = 8, each with three test images, the corresponding reference image, and the results.

Table 8 - Metric results of the three test images for the combinations ID = 7 and ID = 8.
*x represents the cases where the metric calculation was not performed.

Th.	Img.	ID = 7 (7, 5, 4 and AdaGrad)								ID = 8 (7, 5, 4 and AdaMax)							
		Accuracy	Recall		Precision		F_1 Score		Jaccard	Accuracy	Recall		Precision		F_1 Score		Jaccard
			UA	BA	UA	BA	UA	BA			UA	BA	UA	BA	UA	BA	
0.1	9	0.9444	0.9441	0.9523	0.9980	0.4046	0.9703	0.5679	0.3965	0.9258	0.9284	0.8584	0.9939	0.3237	0.9600	0.4701	0.3073
	13	0.8609	0.8413	0.9696	0.9935	0.5241	0.9111	0.6804	0.5157	0.8984	0.9025	0.8753	0.9757	0.6181	0.9377	0.7245	0.5681
	92	0.9602	0.9602	0.0000	1.0000	0.0000	0.9797	0.0000	0.0000	0.9581	0.9581	0.0000	1.0000	0.0000	0.9786	0.0000	0.0000
0.5	9	0.9786	0.9967	0.5240	0.9812	0.8637	0.9889	0.6523	0.4840	0.9568	0.9678	0.6795	0.9870	0.4574	0.9773	0.5468	0.3763
	13	0.9395	0.9767	0.7332	0.9531	0.8500	0.9647	0.7873	0.649277	0.9298	0.9635	0.7432	0.9542	0.7857	0.9588	0.7638	0.6179
	92	0.9927	0.9927	0.0000	1.0000	0.0000	0.9964	0.0000	0.0000	0.9708	0.9708	0.0000	1.0000	0.0000	0.9852	0.0000	0.0000
0.9	9	0.9707	1.0000	0.2366	0.9704	1.0000	0.9850	0.3826	0.2367	0.9739	0.9953	0.4374	0.9779	0.7868	0.9865	0.5622	0.3910
	13	0.8692	0.9995	0.1461	0.8666	0.9812	0.9283	0.2543	0.145690	0.9268	0.9901	0.5755	0.9283	0.9128	0.9582	0.7059	0.5455
	92	1.0000	1.0000	x*	1.0000	x*	1.0000	x*	0.0000	0.9851	0.9851	0.0000	1.0000	0.0000	0.9925	0.0000	0.0000

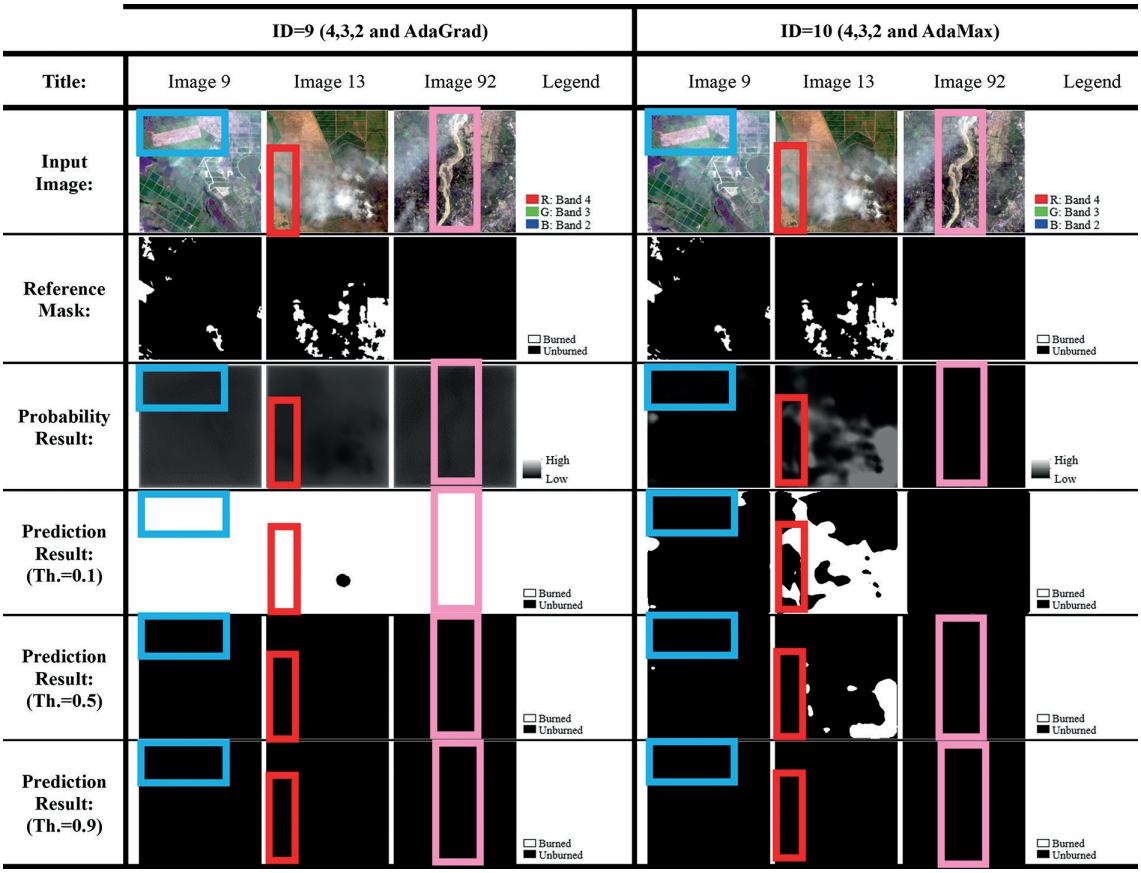

Fig. 12 - Visual results for the combinations ID = 9 and ID = 10, each with three test images, the corresponding reference image, and the results.

Table 9 - Metric results of the three test images for the combinations ID = 9 and ID = 10.
*x represents the cases where the metric calculation was not performed.

Th.	Img.	ID = 9 (5,4,3 and AdaGrad)								ID = 10 (4,3,2 and AdaMax)							
		Accuracy	Recall		Precision		F_1 Score		Jaccard	Accuracy	Recall		Precision		F_1 Score		Jaccard
			UA	BA	UA	BA	UA	BA			UA	BA	UA	BA	UA	BA	
0.1	9	0.0384	0	1	0	0.0384	0	0.07392	0.0384	0.9544	0.9825	0.2501	0.9704	0.3632	0.9764	0.2962	0.1739
	13	0.1581	0.0088	0.9862	0.7803	0.1520	0.0175	0.2635	0.1517	0.6209	0.5619	0.9482	0.9837	0.2806	0.7152	0.4330	0.2764
	92	0	0	0	0	0	0	0	0	0.9993	0.9993	0	1	0	0.9997	0	0
0.5	9	0.9616	1	0	0.9616	0	0.9804	0	0	0.9622	1	0.0155	0.9622	1	0.9807	0.0305	0.0155
	13	0.8473	1	0	0.8473	0	0.9173	0	0	0.8862	0.9612	0.4702	0.9096	0.6858	0.9347	0.5579	0.3869
	92	1	1	x*	1	x*	1	x*	0	1	1	x*	1	x*	1	x*	0
0.9	9	0.9616	1	0	0.9616	0	0.9804	0	0	0.9616	1	0	0.9616	0	0.9804	0	0
	13	0.8473	1	0	0.8473	0	0.9173	0	0	0.8473	1	0	0.8473	0	0.9173	0	0
	92	1	1	x*	1	x*	1	x*	0	1	1	x*	1	x*	1	x*	0

Table 10 - The means of the metric results for the combinations ID = 1, ID = 7 and ID = 8.
The highest values in the columns are shown in bold.

ID	Optim.	Accuracy			Recall BA			Precision BA			F_1-Score BA			Jaccard Score		
		Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9
1	Adam	**0.9741**	0.9743	0.9609	0.8364	0.7203	0.5268	**0.7047**	**0.8411**	**0.9366**	**0.7509**	0.7601	0.6382	**0.6399**	0.6399	0.4788
7	AdaGrad	0.9589	0.97	0.9427	0.8301	0.6376	0.3759	0.5831	0.6477	0.7938	0.6613	0.6799	0.4816	0.5489	0.5692	0.3395
8	AdaMax	0.9707	**0.9758**	**0.9682**	**0.8459**	**0.7775**	**0.6199**	0.6477	0.7983	0.9014	0.7181	**0.7764**	**0.7147**	0.6094	**0.6513**	**0.5588**

Table 11 - The means of the metric results for the combinations ID = 4, ID = 9 and ID = 10.
The highest values in the columns are shown in bold.

ID	Optim.	Accuracy			Recall BA			Precision BA			F_1-Score BA			Jaccard Score		
		Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9	Th. =0.1	Th. =0.5	Th. =0.9
4	Adam	**0.9436**	**0.9518**	**0.9302**	0.6297	**0.4249**	**0.1997**	**0.5177**	**0.6962**	**0.6904**	**0.5396**	**0.4882**	**0.2782**	**0.4233**	**0.365**	**0.1786**
9	AdaGrad	0.3156	0.895	0.895	**0.7598**	0	0	0.1126	0	0	0.1621	0	0	0.1041	0	0
10	AdaMax	0.9227	0.9351	0.895	0.3938	0.2019	0	0.4956	0.4311	0	0.3877	0.2433	0	0.301	0.1745	0

3.3. Final model

In the last implementation stage, all combinations (ID = 1 … 10) generated in the first two parts were compared and the combination that gave the best results was determined. Following data augmentation, by applying 90°, 180°, and 270° rotations to the training data set in the determined combination, training and test processes were performed. Thus, the results of the final model were obtained.

Table 12 summarises the results obtained in the first two parts of the implementation. According to this table, the training accuracy is more than 90%. Among the test results, the

Table 12 - Summary of training and test results for the combinations ID = 1 … 10.
The highest values in the columns are shown in bold.

ID	Bands	Optim.	TRAIN	TEST					
			Accuracy	Accuracy			F_1 Score BA		
				Th.=0.1	Th.=0.5	Th.=0.9	Th.=0.1	Th.=0.5	Th.=0.9
1	7,5,4	Adam	0.9886	**0.9741**	0.9743	0.9609	**0.7509**	0.7601	0.6382
2	5,3,7	Adam	0.9909	0.97	0.9722	0.9608	0.726	0.7539	0.6333
3	5,4,3	Adam	0.9916	0.9679	0.9673	0.9577	0.7136	0.6688	0.5496
4	4,3,2	Adam	0.9776	0.9436	0.9518	0.9302	0.5396	0.4882	0.2782
5	4,3,2,5	Adam	0.9806	0.9509	0.9677	0.949	0.6721	0.6764	0.4513
6	2,3,4,5,6,7	Adam	0.9849	0.9609	0.9525	0.9331	0.5208	0.4088	0.2307
7	7,5,4	AdaGrad	0.9827	0.9589	0.97	0.9427	0.6613	0.6799	0.4816
8	7,5,4	AdaMax	0.9907	0.9707	**0.9758**	**0.9682**	0.7181	**0.7764**	**0.7147**
9	4,3,2	AdaGrad	0.9286	0.3156	0.895	0.895	0.1621	0	0
10	4,3,2	AdaMax	0.9556	0.9227	0.9351	0.895	0.3877	0.2433	0

highest accuracy and F_1score values were obtained for the combination ID = 8 with a threshold of 0.5. These values are 97.58% and 77.64%, respectively.

As a result of the implementation, the combination ID = 8 shows higher results than the other combinations. In this combination, 7, 5, 4 bands and the AdaMax optimisation algorithm were used. Therefore, the 7, 5, 4 bands were used for the final model and the training data set in this band combination was rotated by 90°, 180°, and 270°. By creating a new training data set in this way, the training process was performed using the AdaMax optimisation algorithm. In the training process, the maximum number of epochs was chosen as 100 and early stopping was used. The training was automatically early-stopped at epoch 49. Information about the process is summarised in Table 13.

Table 13 - Training information for the final model.

ID	Bands	Optim.	Learning Rate	Training Time	Training Accuracy	Training Loss
11	7, 5, 4	AdaMax	0.001	2 hours and 29 minutes	0.9911	0.0236

After the training process, the test process was performed for each combination with test data sets in the same combination. Data augmentation was not applied to the test data set. The procedures for generating and presenting the test results were followed in the same way as in the first two implementation phases. In this way, the final model results were obtained.

Fig. 13 shows the visual results for the combination ID = 11. The metric results obtained for this combination are presented in Table 14. In Fig. 13 the croplands (blue rectangle) in image 9 and a small part of the road (pink rectangle) in image 92 are misclassified. In image 13, the road (red rectangle) is not misclassified.

Together with the final model, the confusion matrices calculated for image 9, image 13, and image 92 are presented in Tables 15 to 17 for all the combinations performed in this study (ID = 1, ... , 11).

Table 18 shows the metric results calculated for 92 test images for the final model in the combination ID = 11. Unlike the previous two parts of the implementation, in this part, in addition

Table 14 - Metric results of the three test images for the final model.

Th.	Img.	ID = 11 (7,5,4 and AdaMax)							
		Accuracy	Recall		Precision		F_1 Score		Jaccard
			UA	BA	UA	BA	UA	BA	
0.1	9	0.9427	0.9525	0.6982	0.9875	0.3695	0.9698	0.4833	0.3186
	13	0.8907	0.8914	0.8870	0.9777	0.5956	0.9326	0.7127	0.5536
	92	0.9919	0.9919	0	1	0	0.9960	0	0
0.5	9	0.9488	0.9635	0.5805	0.9829	0.3885	0.9731	0.4655	0.3033
	13	0.9221	0.9465	0.7870	0.9610	0.7262	0.9537	0.7554	0.6070
	92	0.9957	0.9957	0	1	0	0.9979	0	0
0.9	9	0.9508	0.9714	0.4346	0.9773	0.3778	0.9743	0.4042	0.25330
	13	0.9314	0.9808	0.6568	0.9407	0.8608	0.9604	0.7451	0.5938
	92	0.9987	0.9987	0	1	0	0.9994	0	0

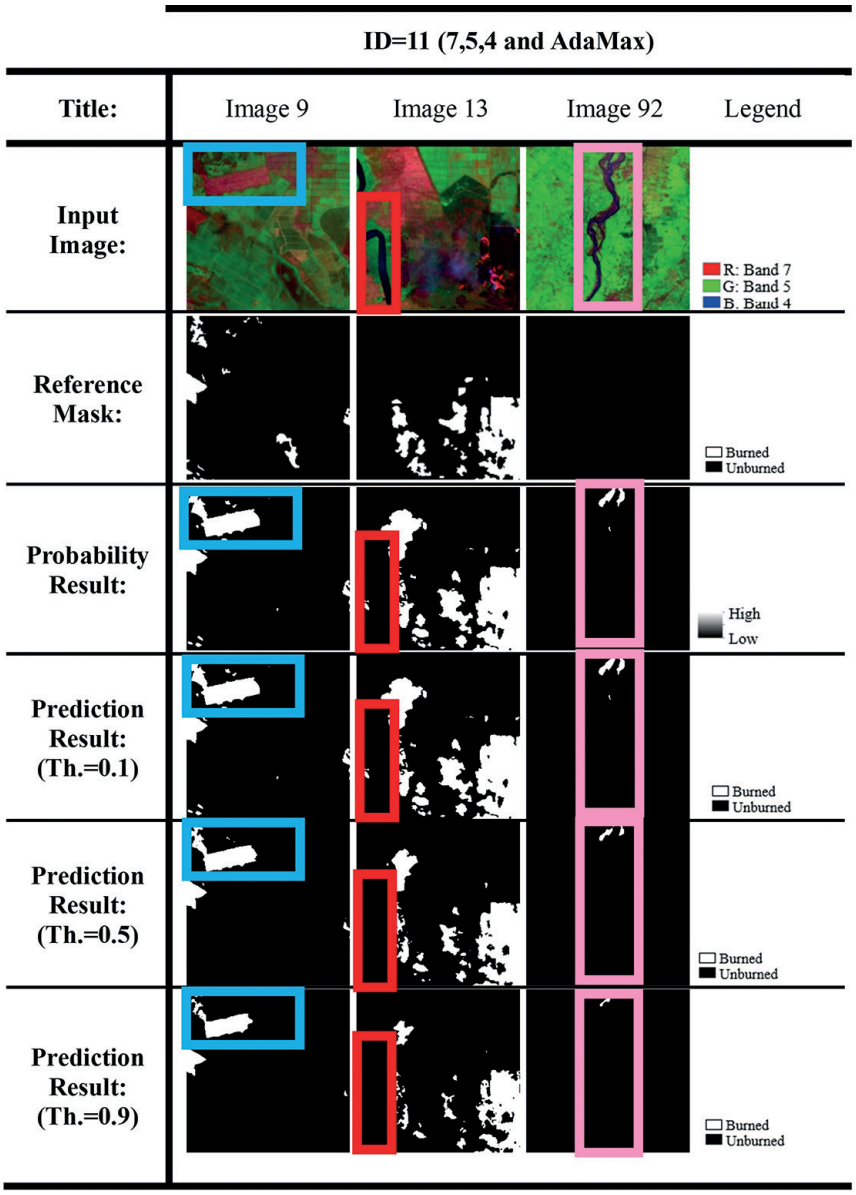

Fig. 13 - Visual results for the combination ID = 11 with three test images, the corresponding reference image, and the results.

Table 15 - Confusion matrices calculated for ID = 1 ... 11 combinations with the threshold value = 0.1.

Image No.	Confusion Matrix (Th.=0.1)										
	ID = 1	ID = 2	ID = 3	ID = 4	ID = 5	ID = 6	ID = 7	ID = 8	ID = 9	ID = 10	ID = 11
9	58979 4042 386 2129	61061 1960 969 1546	62597 424 811 1704	61116 1905 1006 1509	62032 989 221 2294	62951 70 1614 901	59496 3525 120 2395	58511 4510 356 2159	0 63021 0 2515	61918 1103 1886 629	60025 2996 759 1756
13	49524 6005 1538 8469	52171 3358 2469 7538	52111 3418 2047 7960	45958 9571 2626 7381	45434 10095 358 9649	54744 785 3410 6597	46719 8810 304 9703	50117 5412 1248 8759	490 55039 138 9869	31200 24329 518 9489	49501 6028 1130 8877
92	63448 2088 0 0	63754 1782 0 0	64335 1201 0 0	65495 41 0 0	64960 576 0 0	65532 4 0 0	62928 2608 0 0	62790 2746 0 0	0 65536 0 0	65492 44 0 0	65008 528 0 0

Table 16 - Confusion matrices calculated for ID = 1 ... 11 combinations with the threshold value = 0.5.

Image No.	Confusion Matrix (Th.=0.5)										
	ID = 1	ID = 2	ID = 3	ID = 4	ID = 5	ID = 6	ID = 7	ID = 8	ID = 9	ID = 10	ID = 11
9	60805 2216 1029 1486	62210 811 1416 1099	62951 70 1422 1093	63019 2 1855 660	62948 73 1363 1152	63021 0 1809 706	62813 208 1197 1318	60994 2027 806 1709	63021 0 2515 0	63021 0 2476 39	60723 2298 1055 1460
13	54108 1421 2682 7325	54136 1393 3809 6198	54023 1506 3230 6777	52144 3385 4152 5855	53540 1989 2191 7816	55415 114 5384 4623	54234 1295 2670 7337	53500 2029 2570 7437	55529 0 10007 0	53373 2156 5302 4705	52560 2969 2131 7876
92	64363 1173 0 0	64806 730 0 0	65044 492 0 0	65536	65525 11 0 0	65536	65060 476 0 0	63624 1912 0 0	65536	65536	65255 281 0 0

Table 17 - Confusion matrices calculated for ID = 1, ..., 11 combinations with the threshold value = 0.9.

Image No.	Confusion Matrix (Th.=0.9)										
	ID = 1	ID = 2	ID = 3	ID = 4	ID = 5	ID = 6	ID = 7	ID = 8	ID = 9	ID = 10	ID = 11
9	62725 296 1669 846	62803 218 1696 819	63020 1 1805 710	63021 0 2221 294	63021 0 1979 536	63021 0 2103 412	63021 0 1920 595	62723 298 1415 1100	63021 0 2515 0	63021 0 2515 0	61221 1800 1422 1093
13	55372 157 4427 5580	55217 31 6041 3966	54978 551 4136 5871	55048 481 7325 2682	55455 74 4793 5214	55526 3 8497 1510	55501 28 8545 1462	54979 550 4248 5759	55529 0 10007 0	55529 0 10007 0	54466 1063 3434 6573
92	65261 275 0 0	65337 199 0 0	65439 97 0 0	65536	65536	65536	65536	64559 977 0 0	65536	65536	65453 83 0 0

Table 18 - Minimum, maximum, mean, and standard deviation values of metric results for the final model.

Statistic	Accuracy			Recall BA			Precision BA			F_1 Score BA			Jaccard Score		
	Th.=0.1	Th.=0.5	Th.=0.9	Th.=0.1	Th.=0.5	Th.=0.9	Th.=0.1	Th.=0.5	Th.=0.9	Th.=0.1	Th.=0.5	Th.=0.9	Th.=0.1	Th.=0.5	Th.=0.9
Min.	0.7728	0.7402	0.6727	0	0	0	0	0	0	0	0	0	0	0	0
Max.	1	1	1	1	0.9894	0.9385	0.9568	1	1	0.9749	0.9809	0.9658	0.9510	0.9625	0.9340
Mean	0.9717	0.9776	0.9684	0.8590	0.7755	0.5968	0.6891	0.8342	0.9180	0.7554	0.7938	0.6974	0.6221	0.6632	0.5378
Std.	0.0371	0.0378	0.0513	0.2549	0.2486	0.2578	0.2411	0.2313	0.2208	0.2375	0.2293	0.2463	0.2660	0.2796	0.2898

to the mean values of the metric results, the minimum, maximum, and standard deviation values are also given. When the mean values in Table 18 are examined, the highest accuracy was achieved with 97.76% for a threshold value of 0.5. The highest mean recall value for the burned areas was 85.90% for a threshold value of 0.1, the highest mean precision value was 91.80% for a threshold value of 0.9, and the highest mean F_1 score value was 79.38% for a threshold value of 0.5. The mean Jaccard score was the highest with a value of 66.32% for a threshold value of 0.5. For a threshold of 0.5, the minimum accuracy was 77.28%, the maximum 100% and the standard deviation 3.71%. The maximum accuracy is 100% due to the presence of images with no burned areas in the reference mask and in the results. The lowest standard deviation implies that there are no major differences between the results obtained for the 92 test images.

Table 19 shows the calculated metric results for unburned areas for 92 test images in the combination of ID = 11. According to Table 19, the mean recall value for unburned areas is the highest with 99.71% for a threshold value of 0.9, the mean precision value is the highest with 98.88% for a threshold value of 0.1, and the mean F_1 score value is the highest with 98.36% for a threshold value of 0.5.

Table 19 - Minimum, maximum, mean, and standard deviation values of metric results for the final model.

Statistic	Recall UA			Precision UA			F_1-Score UA		
	Th.=0.1	Th.=0.5	Th.=0.9	Th.=0.1	Th.=0.5	Th.=0.9	Th.=0.1	Th.=0.5	Th.=0.9
Min.	0.6819	0.8903	0.9714	0.7399	0.6803	0.5525	0.7751	0.7967	0.7049
Max.	1	1	1	1	1	1	1	1	1
Mean	0.9696	0.9888	0.9971	0.9888	0.9795	0.9619	0.9786	0.9836	0.9777
Std.	0.0479	0.0190	0.0058	0.0348	0.04969	0.0724	0.0376	0.0336	0.0451

Fig. 14 shows the loss, accuracy, recall, precision, and Jaccard score graphs obtained by training the final model. When Fig. 14 is analysed, the loss values are observed to be approaching 0 for the training and validation data sets. Accuracy, sensitivity, precision, and Jaccard score values are approaching to 1. According to Fig. 14a, the loss values obtained for the training and validation data sets are 0.02 and 0.08, respectively. In the study by Prabowo *et al*. (2022b), the loss value was reported as 0.07. In addition, according to Fig. 14e, the Jaccard score values obtained for the training and validation data sets were 0.94 and 0.84, respectively. In the study by Prabowo *et al*. (2022b), the Jaccard score value was reported as 0.93. The values obtained as a consequence of the training of the final model can be stated to be consistent with Prabowo *et al*. (2022b).

To summarise, when all results were examined, it was observed that different band combinations, optimisation algorithms, and threshold values affected the results. Better results

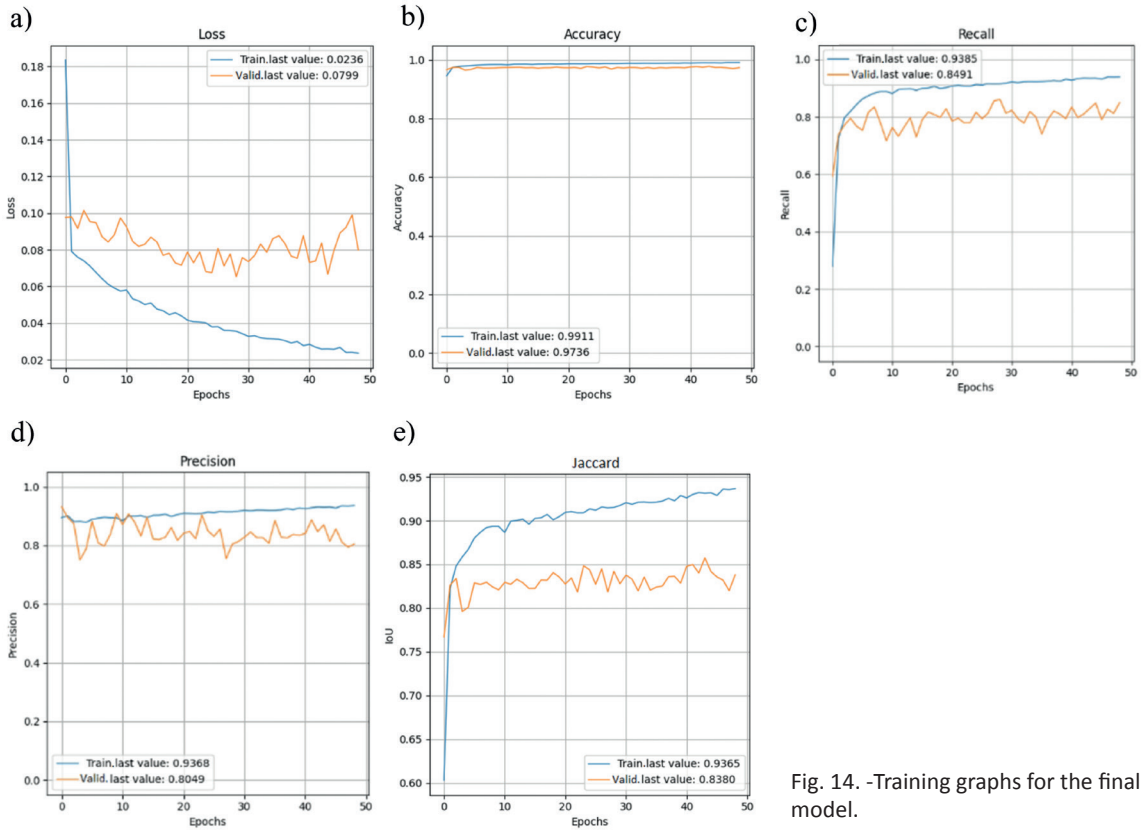

Fig. 14. -Training graphs for the final model.

were obtained with a threshold value of 0.1 and 0.5, while worse results were obtained with a threshold value of 0.9. Also, roads and croplands were misclassified in some combinations. Therefore, the use of different combinations influences the results. However, the misclassification of a road in a test image does not mean that the road is also misclassified in another test image for the same combination. For example, in the test images shared by Alkan and Karasaka (2023), road and agricultural areas were not misclassified in most combinations, while they were misclassified mostly in this study. Therefore, the conditions in the region (smoke, etc.) affect the test results as much as the band combinations. However, clouds were not misclassified. By comparing the combinations performed in this study, the best results were obtained by using 7, 5, 4 bands and the AdaMax optimisation algorithm, and the results were improved by adding a data augmentation process to the training data set. Table 20 summarises the results for the ID = 8 combination (without data augmentation) and the ID = 11 combination (with data augmentation) in order to make it easier for the reader to examine the effect of data augmentation. According to this table, the accuracy and F_1 score results obtained in the ID = 11 combination, which is the final model, are mostly higher than in the ID = 8 combination. Therefore, it can be said that the application of 90°, 180°, and 270° rotations to the training data set has a positive effect on the results.

Table 20 - Summary of training and test results for the combinations ID = 8 and ID = 11.
The highest values in the columns are shown in bold.

ID	Bands	Optim.	Data Augmentation	TRAINING	TEST					
				Accuracy	Accuracy			F_1-Score BA		
					Th.=0.1	Th.=0.5	Th.=0.9	Th.=0.1	Th.=0.5	Th.=0.9
8	7,5,4	AdaMax	None	0.9907	0.9707	0.9758	0.9682	0.7181	0.7764	**0.7147**
11	7,5,4	AdaMax	90°, 180° and 270° rotations	0.9911	**0.9717**	**0.9776**	**0.9684**	**0.7554**	**0.7938**	0.6974

4. Conclusions

According to this study, better results were obtained for threshold values of 0.1 and 0.5. Visual results showed that roads and agricultural areas could be misclassified. In addition, the smoke in the fire area made the detection of burned areas difficult and caused misclassifications in some combinations. However, clouds were not misclassified in general, but the possibility of burned areas under the clouds could not be detected. When comparing the metric results, the best results were obtained by using the 7, 5, 4 (SWIR 2, NIR, red) bands with the AdaMax optimisation algorithm. Therefore, in this combination, the final model was generated by data augmentation of the training data via 90°, 180°, and 270° rotations. In the training of the final model, the cluster size, activation function and number of epochs were the same as in the previous combination, but with an early stopping in the final model to prevent the network from overfitting. As a conclusion, data augmentation, performed by applying 90°, 180°, and 270° rotations to the training data set, was observed to have a positive effect on the results. In the final model, generated within the scope of the study, the highest values were obtained at a threshold value of 0.5, where the mean accuracy was 97.76%, the mean Jaccard score was 66.32%, the mean recall for burned areas was 75.55%, the mean precision was 83.42%, and the mean F_1 score value was 79.38%.

REFERENCES

Alkan D.; 2023: *Detection of burned areas from satellite images using u-net deep learning architecture.* Dissertation thesis, Konya Technical University, 124 pp., in Turkish, unpublished, <https://hdl.handle.net/20.500.13091/4686>.

Alkan D. and Karasaka L.; 2023: *Segmentation of LANDSAT-8 images for burned area detection with deep learning.* Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-M-1-2023, 455-461, doi: 10.5194/isprs-archives-XLVIII-M-1-2023-455-2023.

Bisong E.; 2019: *Regularization for deep learning.* In: Bisong E. (ed), Building machine learning and deep learning models on Google cloud platform: a comprehensive guide for beginners, Apress, Berkeley, CA, USA, pp. 415-421, doi: 10.1007/978-1-4842-4470-8_34.

Brand A.K. and Manandhar A.; 2021: *Semantic segmentation of burned areas in satellite images using a U-Net-based convolutional neural network.* Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B3-2021, pp. 47-53, doi: 10.5194/isprs-archives-XLIII-B3-2021-47-2021.

Chen L., Wang H., Ouyang Y., Zhou Y., Wang N. and Li Q.; 2023: *FSLens: a visual analytics approach to evaluating and optimizing the spatial layout of fire stations.* arXiv:2307.12227v2, doi: 10.48550/arXiv.2307.12227.

Chen X., Yu R., Ullah S., Wu D., Li Z., Li Q., Qi H., Liu J., Liu M. and Zhang Y.; 2022: *A novel loss function of deep learning in wind speed forecasting.* Energy, 238, Part B, doi: 10.1016/j.energy.2021.121808.

Chen Z., Zhu B., Wang L. and Zhang Y.; 2021: *Detection and counting of hot forming parts in video based on deep learning.* In: Advanced High Strength Steel and Press Hardening, Proceedings of the 5th International Conference (ICHSU2020), Shanghai, China, Nov. 17-18, 2020, pp. 478-484.

Chieng H.H., Wahid N., Ong P. and Perla S.R.K.; 2018: *Flatten-T Swish: a thresholded ReLU-Swish-like activation function for deep learning.* International Journal of Advances in Intelligent Informatics, 4, 76-86, doi: 10.26555/ijain.v4i2.249.

Clarisse K.N.I.; 2021: *Using different deep learning methods for Covid-19 classification from CT scans segmented by generative adversarial networks and UNet.* Dissertation thesis, Konya Technical University, 78 pp., in Turkish, unpublished, <https://hdl.handle.net/20.500.13091/393>.

CTIF (International Association of Fire and Rescue Services); 2023: *World Fire Statistics, Report No 28*. <https://www.ctif.org/sites/default/files/2023-06/CTIF_Report28-ESG.pdf>.

Duchi J., Hazan E. and Singer Y.; 2011: *Adaptive subgradient methods for online learning and stochastic optimization.* Journal of Machine Learning Research, 12, 2121-2159.

Farasin A., Colomba L., Palomba G., Nini G. and Rossi C.; 2020: *Supervised burned areas delineation by means of Sentinel-2 imagery and Convolutional Neural Networks.* In: Lee Hughes A., McNeill F. and Zobel C. (eds), Proceedings of the 17th ISCRAM Conference – Blacksburg, VA, USA May 2020, Virginia Tech, Blacksburg, VA, USA, pp. 1060-1071.

Farhan R.I., Maolood A.T. and Hassan N.F.; 2020: *Optimized deep learning with binary PSO for intrusion detection on CSE-CIC-IDS2018 dataset.* Journal of Al-Qadisiyah for computer science and mathematics, 12, 16-27, doi: 10.29304/jqcm.2020.12.3.706.

Friedman R.; 1998: *Principles of fire protection chemistry and physics. 3rd ed.* Jonas and Bartlett Publishers, Sudbury, MA, USA, , 296 pp.

Futrega M., Milesi A., Marcinkiewicz M. and Ribalta P.; 2021: *Optimized U-Net for brain tumor segmentation.* In: Crimi A. and Bakas S. (eds), Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries, BrainLes 2021, Springer, Cham, Switzerland, pp. 15-29, doi: 10.1007/978-3-031-09002-8_2.

Gao L., Zhang Y., Han J. and Callan J.; 2021: *Scaling deep contrastive learning batch size under memory limited setup.* In: Rogers A., Calixto I., Vulić I., Saphra N., Kassner N., Camburu O.-M., Bansal T. and Shwartz V. (eds), Proceedings of the 6th Workshop on Representation Learning for NLP, Association for Computational Linguistics, pp. 316-321, <https://aclanthology.org/2021.repl4nlp-1.31.pdf>.

Haji S.H. and Abdulazeez A.M.; 2021: *Comparison of optimization techniques based on gradient descent algorithm: a review.* PalArch's Journal of Archaeology of Egypt/Egyptology, 18, 2715-2743, ISSN 1567-214x.

Hassaballah M. and Awad A.I. (eds); 2020: *Deep learning in computer vision: principles and applications.* Routledge & CRC Press, Boca Raton, FL, USA, 338 pp., ISBN 9781032242859.

Hu X., Ban Y. and Nascetti A.; 2021: *Uni-temporal multispectral imagery for burned area mapping with deep learning.* Remote Sens., 13, 29 pp., doi:10.3390/rs13081509.

Huang B.; 2021: *Using synthetic satellite imagery from virtual worlds to train deep learning models for object recognition*. PhD thesis, Department of Electrical Computer Engineering, Duke University, Durham, NC, USA, 109 pp., unpublished, <https://hdl.handle.net/10161/22951>.

Jin R. and Niu Q.; 2021: *Automatic fabric defect detection based on an improved YOLOv5.* Mathematical Problems in Engineering, 2021, 13 pp., doi: 10.1155/2021/7321394.

Kalfin Sukono, Supian S. and Mamat M.; 2022: *Insurance as an alternative for sustainable economic recovery after natural disasters: a systematic literature review.* Sustainability, 14, 18 pp., doi: 10.3390/su14074349.

Khryashchev V. and Larionov R.; 2020: *Wildfire segmentation on satellite images using deep learning.* In: 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT), Moscow, Russia, doi: 10.1109/MWENT47943.2020.9067475.

Kingma D.P. and Ba J.; 2015: *Adam: A method for stochastic optimization.* In: International Conference on Learning Representations, San Diego, CA, USA, 15 pp., doi: 10.48550/arXiv.1412.6980.

Knopp L.; 2021: *Development of a burned area processor based on Sentinel-2 data using deep learning.* PFG, 89, 357-358, doi:10.1007/s41064-021-00177-6.

Knopp L., Wieland M., Rättich M. and Martinis S.; 2020: *A deep learning approach for burned area segmentation with Sentinel-2 data*. Remote. Sens., 12, 22 pp., doi: 10.3390/rs12152422.

Krenker A., Bešter J. and Kos A.; 2011: *Introduction to the artificial neural networks.* In: Suzuki K. (ed.), Artificial neural networks - Methodological advances and biomedical applications, InTechOpen, London, UK, pp. 3-18, doi: 10.5772/15751.

Lee C. and Chung C.; 1970: *The effect of learning rate to an unsupervised neural network for inspection process.* WIT Transactions on Information and Communication Technologies, 16, 20 pp., doi: 10.2495/ai960211.

Lee I.K., Trinder J.C. and Sowmya A.; 2020: *Application of U-Net Convolutional Neural Network to bushfire monitoring in Australia with Sentinel-1/-2 data.* Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci., XLIII-B1-2020, 573-578, doi: 10.5194/isprs-archives-XLIII-B1-2020-573-2020.

Lewis-Atwell T., Townsend P.A. and Grayson M.N.; 2022: *Machine learning activation energies of chemical reactions. WIREs* Comput. Mol. Sci., 12, 31 pp., doi: 10.1002/wcms.1593.

Martin M.L.; 2010: *Child participation in disaster risk reduction: the case of flood-affected children in Bangladesh.* Third World Quarterly, 31, 1357-1375, doi: 10.1080/01436597.2010.541086.

Mohla S., Mohla S., Guha A. and Banerjee B.; 2020: *Multimodal noisy segmentation based fragmented burn scars identification in Amazon Rainforest.* In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, pp. 4122-4126, doi: 10.1109/SMC42975.2020.9283432.

Panwar S., Das A., Roopaei M. and Rad P.; 2017: *A deep learning approach for mapping music genres.* In: 2017 12th System of Systems Engineering Conference (SoSE), Waikoloa, HI, USA, 5 pp., doi: 10.1109/SYSOSE.2017.7994970.

Park S.-W. and Lee Y.-W.; 2019: *Semantic segmentation of Landsat Images for detection of wildfire-damaged areas.* In: The 40th Asian Conference on Remote Sensing (ACRS 2019), Daejeon, Korea, Korean Society of Remote Sensing, Seoul, Korea, pp. 2271-2273, ISBN: 978-1-7138-0326-3.

Patterson J. and Gibson A.; 2017: *Deep learning: a practitioner's approach.* O'Reilly Media Inc., Sebastopol, CA, USA, 532 pp., ISBN: 9781491914250.

Pinto M.M., Libonati R., Trigo R.M., Trigo I.F. and Da Camara C.C.; 2020: *A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images.* ISPRS Journal of Photogrammetry and Remote Sensing, 160, 260-274, doi:10.1016/j.isprsjprs.2019.12.014.

Prabowo Y., Dimara Sakti A., Adi Pradono K., Amriyah Q., Ulfa K., Halim Rasyidy F., Bengkulah I., Thufaili Imdad M., Ali S. and Arief R.; 2022a: *Dataset of deep learning from Landsat-8 satellite images for estimating burned areas in Indonesia*. Mendeley Data, V5,, doi: 10.17632/fs7mtkg2wk.5.

Prabowo Y., Sakti A.D., Adi Pradono K., Amriyah Q., Halim Rasyidy F., Bengkulah I., Ulfa K., Surya Candra D., Thufaili Imdad M. and Ali S.; 2022b: *Deep learning dataset for estimating burned areas: case study, Indonesia.* Data, 7, 11 pp., doi: 10.3390/data7060078.

Pratama A.H.I.N., Widodo P. and Saragih H.J.R.; 2023: *Prevention and mitigation of forest and land fires in Riau Province in support of national security.* International Journal of Innovative Science and Research Technology, 8, 4 pp., doi: 10.5281/zenodo.7631287.

Rashkovetsky D.; 2020: *Wildfire detection from multi-sensor satellite imagery using a deep learning approach.* Department of Aerospace and Geodesy, Technical University, Munich, Germany, Dissertation, 103 pp.

Rashkovetsky D., Mauracher F., Langer M. and Schmitt M.; 2021*: Wildfire detection from multisensor satellite imagery using deep semantic segmentation.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 7001-7016, doi: 10.1109/JSTARS.2021.3093625.

Ronneberger O., Fischer P. and Brox T.; 2015: *U-Net: convolutional networks for biomedical image segmentation.* In: Navab N., Hornegger J., Wells W.M. and Frangi A.F. (eds), Medical image computing and computer-assisted intervention – MICCAI 2015, part 3, Springer, Cham, Switzerland, pp. 234-241, , doi: 10.1007/978-3-319-24574-4_28.

Salehzadeh A., Calitz A.P. and Greyling J.; 2020: *Human activity recognition using deep electroencephalography learning.* Biomedical Signal Processing and Control, 62, doi: 10.1016/j.bspc.2020.102094.

Salim S., Jamil M.M.A., Ambar R., Zaki W.S.W. and Mohammad S.; 2023: *Learning rate optimization for enhanced hand gesture recognition using Google teachable machine.* In: 2023 IEEE 13th International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, pp. 332-337, doi: 10.1109/ICCSCE58721.2023.10237148.

Ser G. and Bati C.T.; 2019: *Determining the best model with deep neural networks: Keras application on mushroom data* . YYU Journal of Agricultural Sciences, 29, 406-417, doi: 10.29133/yyutbd.505086, in Turkish.

Soydaner D.; 2020*: A comparison of optimization algorithms for deep learning.* International Journal of Pattern Recognition and Artificial Intelligence, 34, 26 pp., doi: 10.1142/S0218001420520138.

Srivastava N., Hinton G., Krizhevsky A., Sutskever I. and Salakhutdinov R.; 2014: *Dropout: a simple way to prevent neural networks from overfitting.* Journal of machine learning research, 15, 1929-1958, <https://jmlr.csail.mit.edu/papers/volume15/srivastava14a/srivastava14a.pdf>.

Statista Research-Department; 2023a: *Area burnt by forest fires in Indonesia from 2015 to 2021*. <https://www.statista.com/statistics/1083530/indonesia-area-burnt-by-forest-fires/#statisticContainer>.

Statista Research-Department; 2023b: *Natural disasters in Indonesia - statistics & facts*. <https://www.statista.com/topics/8305/natural-disasters-in-indonesia/#topicOverview>.

Sureshbabu S., Tejero F., Sanchez-Moreno F., MacManus D.G. and Sheaf C.; 2023: *Deep-learning methods for non-linear transonic flow-field prediction.* In: AIAA AVIATION 2023 Forum, 12-16 June 2023, San Diego, CA, USA, doi: 10.2514/6.2023-3719.

Tan K.L., Lee C.P. and Lim K.M.; 2021: *FN-Net: a deep convolutional neural network for fake news detection.* In: 2021 9th International Conference on Information and Communication Technology (ICoICT), Yogyakarta, Indonesia, IEEE, pp. 331-336, doi: 10.1109/ICoICT52021.2021.9527500.

Tyukavina A., Potapov P., Hansen M.C., Pickens A.H., Stehman S.V., Turubanova S., Parker D., Zalles V., Lima A., Kommareddy I., Song X.P., Wang L. and Harris N.; 2022: *Global trends of forest loss due to fire from 2001 to 2019.* Frontiers in Remote Sensing, 3, 20 pp., doi: 10.3389/frsen.2022.825190.

Wolansky I.; 2021: *A deep dive into the basics of deep learning.* Medical Sciences, 65, 252-256, doi: 10.25040/ntsh2021.02.23.

Zhang J., Jin Y., Xu J., Xu X. and Zhang Y.; 2018: *MDU-net: Multi-scale densely connected u-net for biomedical image segmentation*. Health Inf. Sci. Syst., 11, 10 pp., doi: 10.1007/s13755-022-00204-9.

Zhang Q., Ge L., Zhang R., Metternicht G.I., Du Z., Kuang J. and Xu M.; 2021: *Deep-learning-based burned area mapping using the synergy of Sentinel-1&2 data.* Remote Sensing of Environment, 264, 10 pp., doi: 10.1016/j.rse.2021.112575.

Zhang T., Zhu T., Gao K., Zhou W. and Yu P. S.; 2021: *Balancing learning model privacy, fairness, and accuracy with early stopping criteria.* IEEE Transactions on Neural Networks and Learning Systems, 34, 5557-5569, doi: 10.1109/TNNLS.2021.3129592. .

Corresponding author: Lütfiye Karasaka
Department of Geomatic Engineering, Konya Technical University
Yeni Istanbul street, Konya, Turkey
Phone: +90 332 205 1641; e-mail: lkarasaka@ktun.edu.tr