remote sensing

MDPI

Article

Deep Multiple Kernel Learning for Active Fire Detection from Landsat 8 Imagery

Amirhossein Rostami [1], Reza Shahosseini [1,*], Shabnam Asgari [1], Arastou Zarei [1] and Saeid Homayouni [1,2]

[1] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran 1417466191, Iran; amirh.rostami7676@ut.ac.ir (A.R.); asgari.shabnam@ut.ac.ir (S.A.); arastou.zarei@ut.ac.ir (A.Z.)

[2] Institut national de la recherche scientifique, Centre Eau Terre Environnement, 490 rue de la Couronne, Quebec City, QC, Canada, G1K 9A9, Canada; saeid.homayouni@inrs.ca (S.H.)

* Correspondence: rshahosseini@ut.ac.ir

Abstract: Active fire detection (AFD) as a hazardous natural disaster in satellite imagery is critical to managing environmental protection policies, supporting decision-making, and enforcing the law. Satellite imagery is essential for detecting active fire due to its Earth's global coverage. However, the classical approaches might not lead to reliable results for the analysis of such big data. Deep learning (DL) methods have recently yielded outstanding results in remote sensing applications. However, a few studies have used these methods due to the lack of a unified dataset to detect and map active fire at a pixel level. This study presented a DL method, namely MultiScale-Net, for segmenting active fire in Landsat 8 images at the pixel level, which differs from simple DL architectures. In the proposed method, several kernels with different sizes were used simultaneously in the network architecture, which resulted in better extraction of fire zones and single fire pixels in an image. Moreover, this research also presents a new innovative Active Fire Index (AFI) index for AFD from Landsat-8 images. AFI index was added to the input data set (SWIR1, SWIR2, Blue) to improve the accuracy of AFD. Therefore, the number of False Negatives (FN) pixels to better detect the extent of the fire compared to using three bands was reduced. Different scenarios were tested, including kernels with different sizes, dilated convolutional layers (DCLs) with variable dilation rates, and input features into the MultiScale-Net. Best quantitative results in precision, sensitivity, F1-score, and IoU criteria were obtained as 95.71%, 93.93%, 91.62%, and 84.54%, respectively, over test samples with test samples of various sizes and shapes of active fire. This level of success in the experimental results shows that MultiScale-Net can be used as an effective method for active fire mapping under various circumstances thanks to its insensitivity to fire size and shape changes.

Keywords: Active Fire; Deep Learning; Active Fire Index; Multi-Size Kernels; Landsat 8 Imagery

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Remote Sens.* **2022**, *14*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

1. Introduction

Fire is one of the most devastating natural disasters causing significant damage to humans and ecosystems [1]. In recent years, forest fires have had deadly irreversible effects around the world, such as the recent fire in Canada (2016) [2], Australia (2019) [3], and California (2020) [4]. According to the Food and Agriculture Organization (FAO) report, in 2015, about 98 million hectares of forest were affected by the fire. These forests were mainly in the tropical regions, where the fire engulfed at least 4% of the total forest area [5]. In 2018, a massive fire in California covered approximately 18,000 km^2, killed 100 people, and caused financial damage of nearly $ 2 billion financial damage [6]. The accident happened again in 2020 in the same state, which damaged approximately 17,000 km^2 of American territory [7]. Such devastating and deadly fires frequently occur around the world. One of the key ways to control and monitor a fire is to pinpoint the exact location of the fire, which can play a significant role in extinguishing a fire. Using images/videos from ground/aerial/satellite sensors and implementing computer vision techniques such as image/video processing can be one of the best ways to map the firing area [8,9].

Ground sensors use optical bands and are the standard tool for early fire detection [10]. However, due to the lack of global coverage of these sensors, one of the best options is to use satellite data to cover almost the entire Earth to monitor active fires. Various satellite sensors have been used for AFD [11]. In the past decade, the Earth observations from three multispectral imaging sensors, namely advanced very-high-resolution radiometer (AVHRR), moderate resolution imaging spectroradiometer (MODIS), and visible, infrared imaging radiometer suite (VIIRS), have been frequently used for AFD [12]. AVHRR instruments are carried by the National Oceanic and Atmospheric Administration (NOAA) family of polar-orbiting operational environmental satellite (POES) and European Meteorological Operational (MetOp) satellites. The instrument provides six channels, three in the visible/near-infrared region and three thermal infrared channels, with 1 km spatial resolution [13]. VIIRS sensor on the Suomi National Polar-orbiting Partnership (S-NPP) satellite incorporates fire-sensitive channels, including a dual-gain high-saturation temperature channel at 4 μm, enabling AFD and characterization [14]. MODIS is widely used and plays an essential role in thermal sensing of land surface [15] and AFD [16] thanks to its daily temporal resolution. Terra and Aqua satellites carry MODIS sensors as part of the National Aeronautics Space Administration (NASA) EO systems. It has a revisit time of 1–2 days and captures data in 36 spectral bands ranging in wavelengths from 0.4 to 14.4 μm and at varying spatial resolutions (2 bands at 250 m, five bands at 500 m, and 29 bands at 1 km) [17]. Giglio et al. [18] validated the Collection 5 and Collection 6 Terra MODIS fire products and deliberated the modifications made to the algorithm used to produce this product compared to its previous version. Parto et al. [19] used a change detection approach to investigate the significant changes in forest features, such as MODIS NDVI and thermal bands for real-time fire detection. He et al. [20] proposed a method to eliminate the solar radiation and thermal path radiance received by the MODIS observations in the Middle Infra-Red (MIR) band and reduce the resulting fire detection errors. A similar method has been implemented on the AVHRR data [21]. Schroeder et al. [22] designed, implemented, and validated an algorithm for middle and thermal infrared bands of VIIRS image data with 375 m of resolution. They showed that VIIRS 375 m data could provide more coherent fire mapping than MODIS 1 km fire data.

Another widely used observation data source for fire monitoring and detection is geostationary satellite data [23]. Geostationary satellites have a high temporal resolution but low spatial resolution. Himawari 8 is a new generation of Japanese geostationary weather satellites operated by the Japan Meteorological Agency. AHI-8 has significantly higher radiometric, spectral, and spatial resolution than its predecessor [24]. Jang et al. [25] proposed a combined three-step forest fire detection algorithm (based on thresholding, machine learning-based modeling, and post-processing) using Himawari-8 geostationary satellite data over South Korea. Xie et al. [26] also presented a spatiotemporal contextual model (STCM) that fully exploits geostationary data's spatial and temporal dimensions based on the data from Himawari-8 Satellite.

One of the essential purposes of fire control is to detect the location and extent of the fire accurately. Despite the widespread use of the satellites mentioned above in fire detection, they have a coarse spatial resolution that causes uncertainty in active fire localization. Thus, spaceborne sensors with a higher spatial resolution, such as Landsat 8, must be used to achieve this goal. The Landsat 8 payload includes the Operational Land Imager (OLI) that has a new 1375 μm band designed to monitor cirrus clouds and the Thermal Infrared Sensor (TIRS) that together provide improved cloud detection capabilities [27,28]. Recently, many algorithms have been developed to detect, segment, and monitor active fires using Landsat 8 images [29]. AFD algorithms for the ASTER sensor were the basis for developing the first fire detection methods in Landsat images [30]. In Landsat data, the probability of saturation of the pixels around the fire pixels is high, especially if the highly reflective surfaces are like buildings, although this problem was partially offset in

Landsat 8 [22]. In the last five years, three algorithms have been developed by Schroeder et al. [31], Murphy et al. [32], and Kumar et al. [30]. Detecting active fire in Landsat 8 images has yielded acceptable results. The first two use thresholds on the reflection of the SWIR1, SWIR2, and NIR bands, and their algorithm is separate for the time of day and night and adopts a time analysis approach. Their thresholds are considered adaptable to the conditions. In contrast, the third algorithm inserts the Red band at the same threshold. It should also be noted that these thresholds are fixed, and with a series of statistical tests on a large number of pixels obtains fire and uses it to detect pixels that have a very high potential for fire.

Machine learning methods, in particular DL, have been successfully applied in solving challenging tasks such as regression and classification problems [33,34], object detection [35], and semantic segmentation [36]. Among the DL method, convolutional neural networks (CNN) are the most widely used tools in computer vision, image processing, and pattern recognition [37]. All previous research in active fire detection has been based on statistical methods and thresholds on fire-sensitive bands from various sensor data. These methods' main problem is determining a threshold that can detect fire effectively in different and complex conditions. One of the main reasons for not using DL networks for AFD in previous studies was the lack of a unified dataset for this purpose. In this paper, thanks to the release of large-scale datasets for AFD by De Almeida et al. [38], we developed a deep ender-decoder network ''MultiScale-Net'' for pixel-level localization of active fire in Landsat 8 images to address this problem. This study developed a high-performance CNN architecture for active fire segmentation to detect individual fire pixels detached from a larger fire zone. In the MultiScale-Net, we used convolutional kernels with various sizes and DCLs of varying rates to solve the challenge of changing the size and shape of the fire; however, the basic U-Net, one of the most widely used and simple CNNs in the segmentation task, was employed in their study, with just a series of minor changes.

This research pursues the following initiatives and goals in two areas of features and network architecture for active fire segmentation:

- Provide a new index for AFD to improve network performance in extracting high-level features.
- For the network architecture, using convolution layers with multiple kernels and different dilation rates is one of the solutions to improve network capability by producing more high-level features but maintaining network accuracy. An efficient deep multiple kernel learning for multiscale AFD from Landsat 8 images is proposed called the MultiScale-Net.

2. Dataset

2.1. Landsat 8 Active Fire Dataset

A new large-scale dataset for active fire segmentation was recently published by De Almeida et al. [38]. This dataset contains image patches of 256×256 pixels, showing the wildfires in several locations around the world, extracted from Landsat 8 images from August to September 2020. The patches are 10-band, 16-bit TIFF images, with channels b1… b7 and b9, b10, b11. The panchromatic channel b8, with 15 meters of spatial resolution, was excluded from the dataset. The three sets of conditions explained in detail in the references [30–32] produced their corresponding ground truth, and we selected the intersection between these sets of conditions as ground truth. As mentioned in [31], bands 7, 6, and 2 (SWIR2, SWIR1, Blue) have been used in this study.

There was a significant class imbalance between fire and non-fire pixels in the dataset, so that approximately 99% of the pixels are non-fire. Most image patches had one or two fire pixels. We chose patches with the highest fire pixels compared to the most imbalanced ones in the dataset. Finally, 144 image patches from the available data were selected and split to 70% for training and validation (100 image patches) and 30% for testing (44 image patches). Test samples were selected for significant challenges of AFD, such as changing the size and shape of the fire in the image scene, the presence of clouds, and single pixels of fire separated from the fire area. Image patches were selected from all five continents (Asia, Europe, Africa, Oceania, and the Americas) to train MultiScale-Net with different geographical, climatic, atmospheric, and other conditions. Also, the test samples were selected to include different shapes and sizes of active fire from large zones to strips and single pixels.

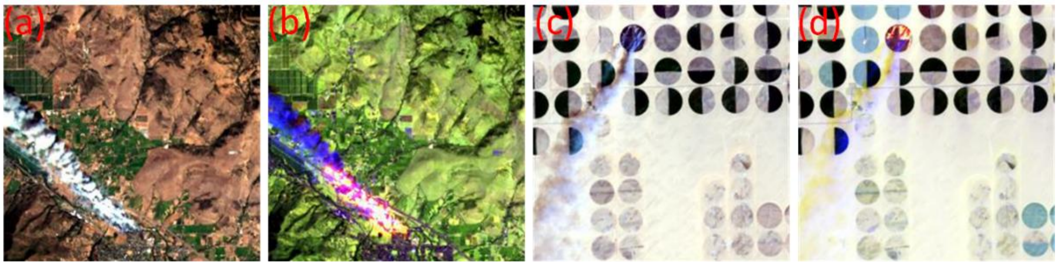


Figure 1. a-c) RGB, b-d) False color (SWIR2, SWIR1, and Blue) of Landsat 8 images from the reference dataset.

2.2. Data Augmentation

Data augmentation is essential for training a model to improve generalization and reduce overfitting, especially when labeled or training data is limited [39]. It is also an excellent way to make the model robust to specific types of noise. There are many ways to boost data, one of the oldest of which is the flipping method. Using this method, the network will become resistant to the changes in direction, azimuth, and rotation. Horizontal flipping is much more common than the vertical one. This augmentation is the easiest method to implement and has proven helpful for datasets like CIFAR-10 and ImageNet [40]. In this study, we used horizontal, vertical, and both horizontal and vertical flipping. As a result, the training data quadrupled to 400 image patches.

3. Proposed Methodology

In this study, the proposed architecture is an efficient network for spectral-spatial relationships extraction of remote sensing images with high-level features. Also, in the network input features part, an innovative index has been used to extract better the features related to active fire in the network. Simple data augmentation techniques have also been used to compensate for the lack of training samples. Several configurations were considered based on the network structure, as shown in Table1. According to this table, all possible configurations were considered for AFD. For example, B3K35D1 represents a case where we used three bands, two kernels with sizes 3 and 5, and a dilation rate of 1. In total, we defined 27 different configurations. Figure 2 shows the scenarios experimented on within this study.

184

Figure 2. Different scenarios were used in this study based on the number of input features, different 185
sized kernels, and the DCLs with different rates in MultiScale-Net for AFD. 186

Table 1. Different architecture parameters to configure MultiScale-Net.

Configuration parameters	Input values	Abbreviation
Input feature(s)	SWIR2 + SWIR1 + Blue + AFI	B4
	SWIR2 + SWIR1 + Blue	B3
	AFI	B1
Kernel size	(3, 3)	K3
	(3, 3) + (5, 5)	K35
	(3, 3) + (5, 5) + (7, 7)	K357
Dilation rate	(1, 1) + (1, 1)	D1
	(1, 1) + (2, 2)	D2
	(1, 1) + (3, 3)	D3

3.1. Active Fire Index

Fire detection methods are mainly based on the processing and analysis of thermal and reflective bands. Previous research [31] shows that Band 7 of the Landsat 8 sensor (i.e., SWIR2) is sensitive to fire radiation. Also, the indices provided to date for fire detection in satellite images often require multi-stage processes or the help of vegetation or drought indices [41]. This study proposed an index for detecting active fire and named it the Active Fire Index (AFI), obtained based on Equation (1).

$$\mathrm{AFI} = \frac{\rho_7}{\rho_2} \quad (1)$$

That ρ_7 and ρ_2 represent a value of SWIR2 and the Blue band of Landsat 8 images, respectively. Although the proposed index has a straightforward computational approach (i.e., dividing two bands), it has an excellent ability to separate active fire from the background (see Figure 3). Although active fire has a high reflectivity in the SWIR2 band of the Landsat 8 image, distracting influences such as bright objects and smoke hamper accurate fire identification and segmentation. The AFI removed fire-like objects in brightness and spectral reflection by dividing the SWIR2 by blue, strongly discriminating between fire and non-target pixels. The AFI's performance, particularly in eliminating smoke nuisance, frequently linked with active fire, led to its adoption as an input feature to the MultiScale-Net. The presence of clouds in the image adds to the difficulty of recognizing active fire in remote sensing images [42]. Excessive cloud covering in the image might lead to a misidentification of the current fire. One typical method for lessening the impact of the cloud has been to mask it. Cloud masking accuracy is also very effective in the end outcome of fire detection [43]. Because of the significant reflectance of the cloud in both bands employed in the AFI [44], dividing these two bands resulted in the cloud being nearly eliminated. Simultaneously, the AFI distinguishes between these pixels and the background quite well because of the high reflection of the target (fire) pixels in the SWIR2 spectral region and the low reflection in the blue band range (Figure 3). This discrimination is made even when the number of target pixels is small; however, it does not always perform adequately. Light objects having a high SWIR2 band value and a low reflectance in the blue band, for example, raise the AFI value while not firing (Figure 4, second row).

3.2. Network Architecture

The proposed method architecture is a fully convolutional framework in which the output is a pixel-by-pixel map with the same width and length as the input image. The proposed network differs from previous fire detection architectures [38], which used a simple U-NET for fire detection and mapping. We use an asymmetrical encoder-decoder architecture that, in the decoder branch, in addition to the transpose convolution layers, also uses straight convolution layers to extract high-level features that can capture all the features of small and large fires and their neighbor's pixels.

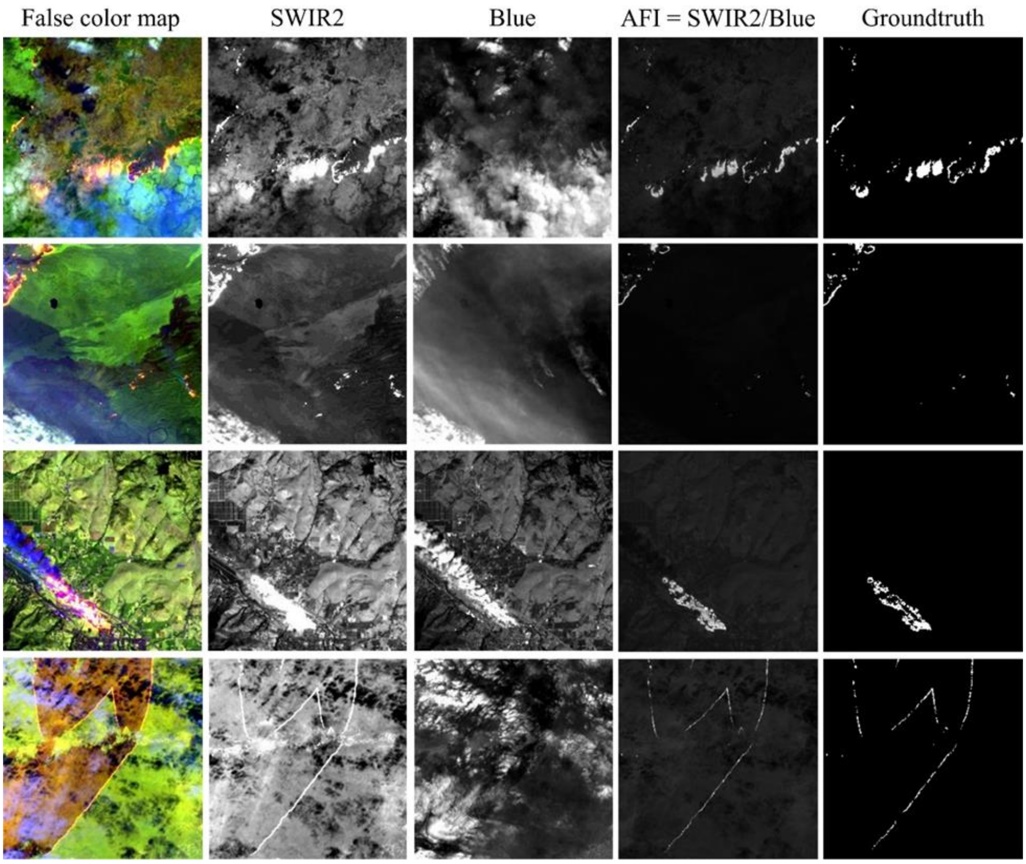


Figure 3. AFI index maps from several Landsat 8 images.

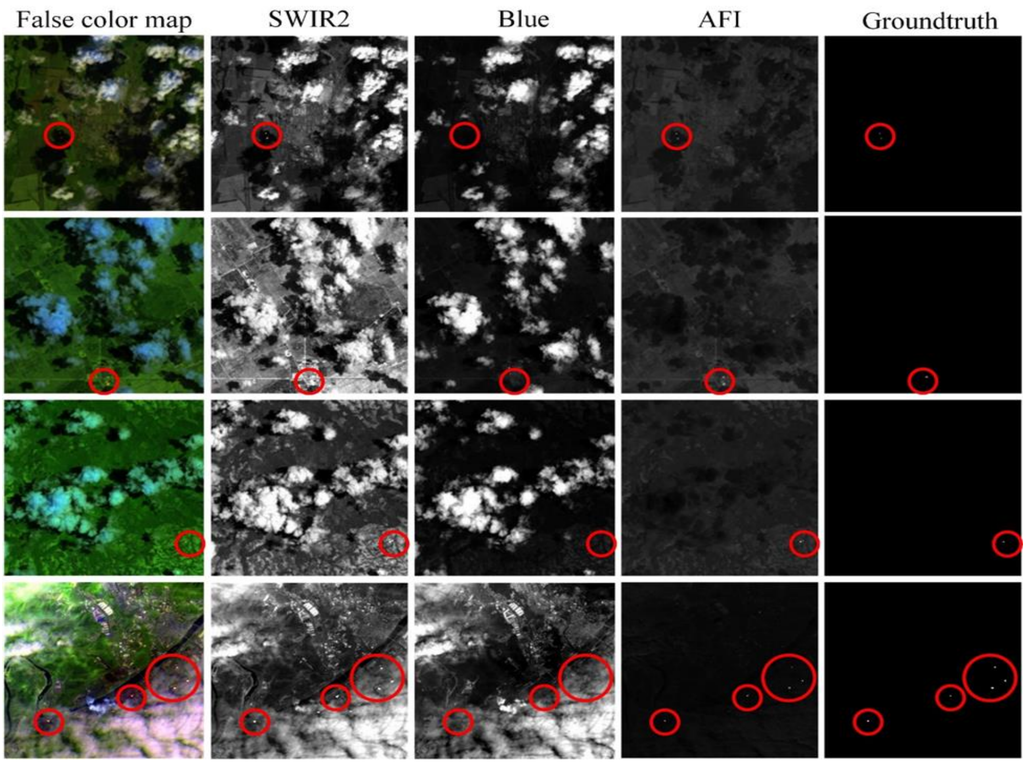


Figure 3. The performance of the AFI for AFD in cloudy scenes. The red circles indicate the approximate firing range.

Also, in the desired architecture, the combination of convolution features with different sizes of filters has been designed and used to take full advantage of the convolution feature of different scales and further improve the accuracy of fire localization [45]. The idea is to use features extracted at various scales from filters with different sizes to provide local and general properties. The feature map of the initial encoder layers retains more spatial details, leading to more corrected boundaries but low-level features. Higher-level spatial features are extracted from the input image in several scales by convolutional layers, where it uses pooling layers to sample feature maps. The decoder branch samples and reconstructs feature maps by inverse convolution layers, which also use convolution layers to extract high-level semantic features. Concatenation links connect the feature maps in equal sizes in the encoder and corresponding decoder layers. Based on that, more concise and highly differentiated features are produced. Such connections can help preserve more spatial detail in decoded feature maps from the decoder branch and enable the training of profound networks. In addition to kernels with different sizes, DCLs have been employed to extract multiscale fires in conditions with variable sizes in the image scene. Dilated convolutions add a new parameter to convolutional layers known as the dilation rate. This parameter provides the distance between values in a kernel. With only nine parameters, a 3×3 kernel with a dilation rate of 2 will have the same field of vision as a 5×5 kernel; consider removing every second column and row from a 5×5 kernel. These provide a wider field of vision at a fixed computational effort. In this study, dilation rates of 1, 2, and 3 in the second layer of convolution were tested in each block (the dilation rate in the first layer of each block was considered constant 1). Our proposed method has the following characteristics:

1. The proposed DL architecture has a network depth of 5 (network length).
2. Take advantage of a new approach that includes convolutional filters with different sizes for the training process.
3. DCLs with different dilation rates.
4. Batch normalization and dropout have been used, which play an essential role in preventing over-fitting [46].
5. Using a binary cross-entropy loss function improves the network's performance. Also, the "glorot_uniform" method initializes the network [47].

Figure 5 shows one of the configurations used in MultisScale-Net, the most complex architecture used in this study. Different configurations based on using DCLs with different dilation rates in the second layer, simultaneous employment of filters with different sizes, and different scenarios of network input features were examined (see Figure 2).

The proposed network consists of an encoder part and a decoder part. The encoder part carries out five convolutional blocks. It consists of repeated applications of two 3 × 3, 5 × 5, 7 × 7 convolutional layers with two paddings. Each convolution follows a batch normalization (BN) layer and Rectified Linear Units (ReLU) activation function that in this study, due to the fast computational process and normalization of the output of each layer in the range [0, +∞), it has been used which the effects of noise on the gradient of this function are minimal and a. Moreover, each convolutional block is followed by 2 × 2 max-pooling for down-sampling with the stride of 2. The number of feature channels is doubled after each block. The decoder part corresponds to the encoder part and consists of four transposed convolution layers. Every block in the decoder part consists of a 3 × 3 deconvolution with the stride of 2, a concatenation with the corresponding feature maps from the encoder part, and two 3 × 3, 5 × 5, 7 × 7 convolutions followed by ReLU activation and BN layer. The number of feature channels is halved after each up-sampling process. A 1 × 1 convolution with a Softmax activation function is employed as a classifier at the final layer, followed by a Binary Cross-Entropy (BCE) loss function. The BCE loss function compares each predicted probability with the actual class's output, 0 or 1 (in binary classification). The calculated score then determines the probabilities based on the distance from the expected value. This score shows how the predicted values are close to or far from the actual value, which, to minimize it in the training process, tries to update the network parameters with high accuracy [48]. For binary segmentation, the BCE is used, which is defined as follows:

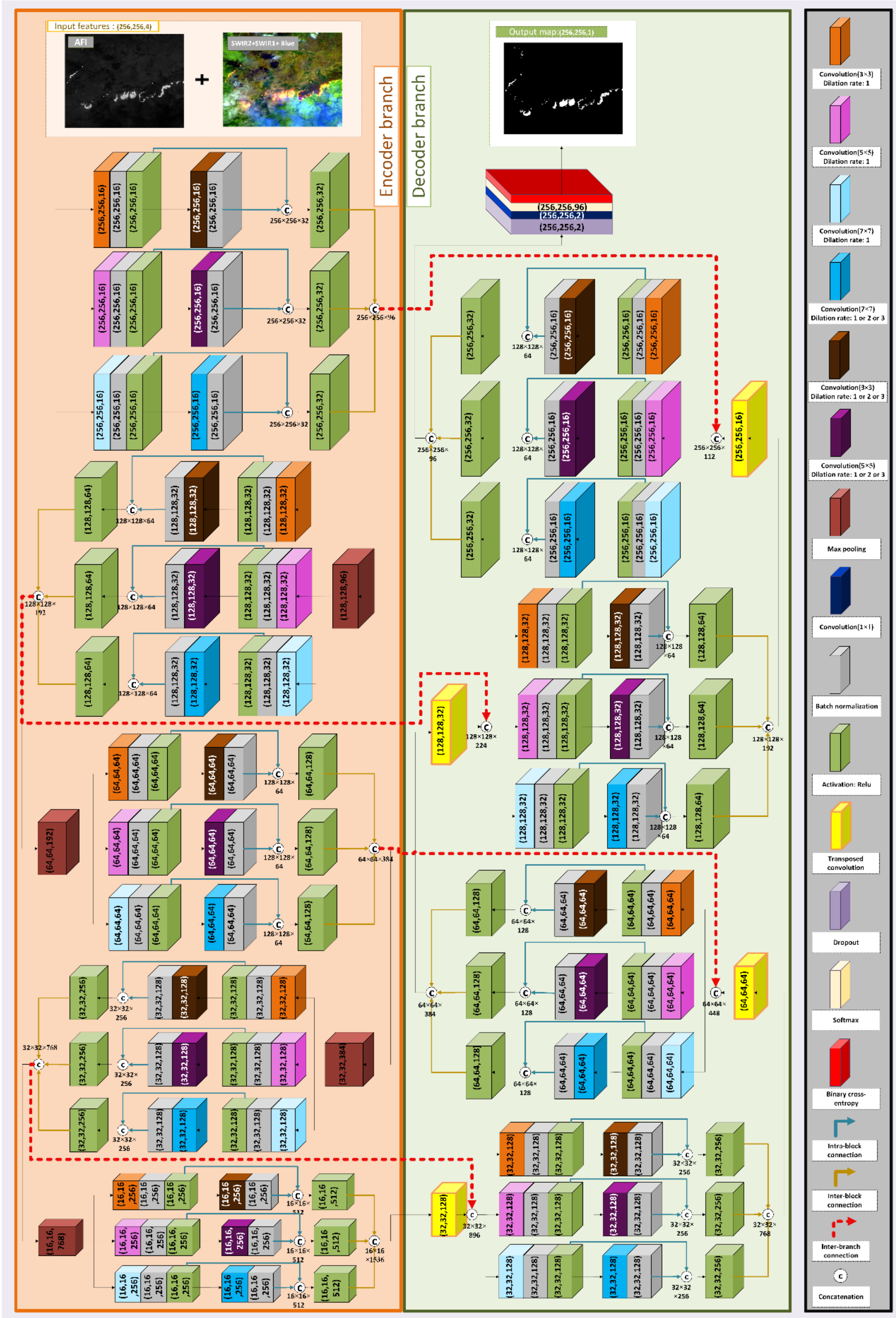


Figure 5. The proposed network for active fire segmentation. "Conv2d (k × k)" shows the convolutional kernel with the size of k × k; "Batch Norm." denotes batch normalization; "ReLU" denotes the rectified linear units.

$$BCE(y,\hat{y}) = -\frac{1}{n}\sum_{i=1}^{n}(y_i \log(\hat{y}) + (1-y_i)\log(1-\hat{y})) \quad (2)$$

that $\hat{\boldsymbol{y}}$ is the probability of fire class, and $\mathbf{1}-\hat{\boldsymbol{y}}$ is the probability of non-fire class. The n denotes the number of pixels, and $\boldsymbol{y_i}$ is the pixel label.

All scenarios were designed and trained using the Keras library on the Google Colab platform, utilizing the TensorFlow processing unit and a Tesla T4 GPU with a batch size of 15 for 200 epochs. Weights and biases values must be initialized before network training. Neural networks are significantly sensitive to initial values of weights, and therefore initial weighting is essential. The initial weighting of weights is done in CNN to counteract gradient vanishing and gradient exploding. The problem of gradient fading often occurs in networks where the values of the weights are small, and the problem of gradient explosion occurs when large amounts of weights are selected. The *glorot* initialization algorithm is one of the known methods in CNN. There are two versions of the *glorot* initialization: "uniform_glorot" (used in this study) and "normal_glorot". Both versions are pretty similar, except that random values are derived from the normal distribution in the second method instead of a uniform distribution. The *glorot* initialization technique works better than uniform random initialization and eliminates the need to guess the appropriate constant limit values [49]. Also, the adaptive moment estimation (Adam) method was used to optimize the trainable network parameters. The learning rate, beta-1, beta-2, and epsilon were chosen as 10^{-4}, 0.9, 0.999, and 10^{-8} respectively in this method.

Tables 2 show the output size, kernel sizes, activation functions, and different types of concatenations in the encoder and decoder part of MultiScale-Net, respectively. The MultiScale-Net employs three types of connections and concatenations:

- Intra-block Concatenation: combine feature maps from the first and second convolution layers using the same filter size in each block.
- Inter-block concatenation merges feature maps using convolutional layers with varying filter sizes from distinct blocks.
- Inter-branch concatenations: merge feature maps from the encoder and the decoder branches' transposed convolution blocks. In the Multiscale-Net, this sequence combines features with highly abstraction ability.

Table 2. MultiScale-Net configuration.

Encoder		Decoder	
Layer	Output	Layer	Output
Conv. 3 × 3 + BN + ReLU Conv. 3 × 3 + BN Intra-block Concatenation + ReLU	256 * 256 * 16 256 * 256 * 16 256 * 256 * 32	Transposed Conv. 3 × 3 Inter-branch Concatenation	32 * 32 * 128 32 * 32 * 896
Conv. 5 × 5 + BN + ReLU Conv. 5 × 5 + BN Intra-block Concatenation + ReLU	256 * 256 * 16 256 * 256 * 16 256 * 256 * 32	Conv. 3 × 3 + BN + ReLU Conv. 3 × 3 + BN Intra-block Concatenation + ReLU	32 * 32 * 128 32 * 32 * 128 32 * 32 * 256
Conv. 7 × 7 + BN + ReLU Conv. 7 × 7 + BN Intra-block Concatenation + ReLU	256 * 256 * 16 256 * 256 * 16 256 * 256 * 32	Conv. 5 × 5 + BN + ReLU Conv. 5 × 5 + BN Intra-block Concatenation + ReLU	32 * 32 * 128 32 * 32 * 128 32 * 32 * 256
Inter-block Concatenation Max-pooling. 2 × 2	256 * 256 * 96 128 * 128 * 96	Conv. 7 × 7 + BN + ReLU Conv. 7 × 7 + BN	32 * 32 * 128 32 * 32 * 128 32 * 32 * 256

		Intra-block Concatenation + ReLU	
		Inter-block Concatenation	32 * 32 * 768
Conv. 3 × 3 + BN + ReLU Conv. 3 × 3 + BN Intra-block Concatenation + ReLU	128 * 128 * 32 128 * 128 * 32 128 * 128 * 64	Transposed Conv. 3 × 3 Inter-branch Concatenation	64 * 64 * 64 64 * 64 * 448
Conv. 5 × 5 + BN + ReLU Conv. 5 × 5 + BN Intra-block Concatenation + ReLU	128 * 128 * 32 128 * 128 * 32 128 * 128 * 64	Conv. 3 × 3 + BN + ReLU Conv. 3 × 3 + BN Intra-block Concatenation + ReLU	64 * 64 * 64 64 * 64 * 64 64 * 64 * 128
Conv. 7 × 7 + BN + ReLU Conv. 7 × 7 + BN Intra-block Concatenation + ReLU	128 * 128 * 32 128 * 128 * 32 128 * 128 * 64	Conv. 5 × 5 + BN + ReLU Conv. 5 × 5 + BN Intra-block Concatenation + ReLU	64 * 64 * 64 64 * 64 * 64 64 * 64 * 128
Inter-block Concatenation Max-pooling. 2 × 2	128 * 128 * 192 64 * 64 * 192	Conv. 7 × 7 + BN + ReLU Conv. 7 × 7 + BN Intra-block Concatenation + ReLU	64 * 64 * 64 64 * 64 * 64 64 * 64 * 128
		Inter-block Concatenation	64 * 64 * 384
Conv. 3 × 3 + BN + ReLU Conv. 3 × 3 + BN Intra-block Concatenation + ReLU	64 * 64 * 64 64 * 64 * 64 64 * 64 * 128	Transposed Conv. 3 × 3 Inter-branch Concatenation	128 * 128 * 32 128 * 128 * 224
Conv. 5 × 5 + BN + ReLU Conv. 5 × 5 + BN Intra-block Concatenation + ReLU	64 * 64 * 64 64 * 64 * 64 64 * 64 * 128	Conv. 3 × 3 + BN + ReLU Conv. 3 × 3 + BN Intra-block Concatenation + ReLU	128 * 128 * 32 128 * 128 * 32 128 * 128 * 64
Conv. 7 × 7 + BN + ReLU Conv. 7 × 7 + BN Intra-block Concatenation + ReLU	64 * 64 * 64 64 * 64 * 64 64 * 64 * 128	Conv. 5 × 5 + BN + ReLU Conv. 5 × 5 + BN Intra-block Concatenation + ReLU	128 * 128 * 32 128 * 128 * 32 128 * 128 * 64
Inter-block Concatenation Max-pooling. 2 × 2	64 * 64 * 384 32 * 32 * 384	Conv. 7 × 7 + BN + ReLU Conv. 7 × 7 + BN Intra-block Concatenation + ReLU	128 * 128 * 32 128 * 128 * 32 128 * 128 * 64
		Inter-block Concatenation	128 * 128 * 192
Conv. 3 × 3 + BN + ReLU Conv. 3 × 3 + BN Intra-block Concatenation + ReLU	32 * 32 * 128 32 * 32 * 128 32 * 32 * 256	Transposed Conv. 3 × 3 Inter-branch Concatenation	256 * 256 * 16 256 * 256 * 112
Conv. 5 × 5 + BN + ReLU Conv. 5 × 5 + BN Intra-block Concatenation + ReLU	32 * 32 * 128 32 * 32 * 128 32 * 32 * 256	Conv. 3 × 3 + BN + ReLU Conv. 3 × 3 + BN Intra-block Concatenation + ReLU	256 * 256 *16 256 * 256 *16 256 * 256 * 32
Conv. 7 × 7 + BN + ReLU Conv. 7 × 7 + BN Intra-block Concatenation + ReLU	32 * 32 * 128 32 * 32 * 128 32 * 32 * 256	Conv. 5 × 5 + BN + ReLU Conv. 5 × 5 + BN Intra-block Concatenation + ReLU	256 * 256 * 16 256 * 256 * 16 256 * 256 * 32
Inter-block Concatenation	32 * 32 * 768	Conv. 7 × 7 + BN + ReLU	256 * 256 * 16

Max-pooling. 2 × 2	16 * 16 * 768	Conv. 7 × 7 + BN Intra-block Concatenation + ReLU	256 * 256 * 16 256 * 256 * 32
		Inter-block Concatenation Dropout	256 * 256 * 96 256 * 256 *96
Conv. 3 × 3 + BN + ReLU Conv. 3 × 3 + BN Intra-block Concatenation + ReLU	16 * 16 * 256 16 * 16 * 256 16 * 16 * 512	Conv. 1 × 1 + Softmax	256 * 256 * 2
Conv. 5 × 5 + BN + ReLU Conv. 5 × 5 + BN Intra-block Concatenation + ReLU	16 * 16 * 256 16 * 16 * 256 16 * 16 * 512		
Conv. 7 × 7 + BN + ReLU Conv. 7 × 7 + BN Intra-block Concatenation + ReLU	16 * 16 * 256 16 * 16 * 256 16 * 16 * 512		
Inter-block Concatenation	16 * 16 * 1536		
Total number of trainable parameters		~ 45 M	

3.1. Accuracy Metrics

In the field of AFD, due to the destructive nature of fire, determining the exact extent of the fire and determining the fire pixels in the images of the Landsat 8 satellite is a critical issue. Therefore, the criteria used to assess accuracy are very important. In this research, four accuracy criteria, Precision, Sensitivity, F1-score, and Intersection over Union (IoU), have been considered. The precision also called the correctness in remote sensing literature, is the ratio of predicted fire pixels that are actual fire. The sensitivity, also called completeness, is the ratio of actual fire pixels correctly detected. It is often the case that changing some parameters in the forecasting phase affects accuracy and sensitivity. Depending on the research issue, it should be decided which criteria are essential, and the parameters should be set accordingly. If we optimize our classifier to increase one of the parameters of accuracy or sensitivity and ignore the other, the harmonic mean of those two parameters decreases rapidly. The F1 score parameter is defined as the harmonic mean of the two parameters' precision and sensitivity, the optimality of which indicates the balance for the two parameters' precision and sensitivity [50]. However, due to the possibility of imbalance between fire and non-fire classes, additional criteria need to be calculated to understand the model's actual performance, as precision in unbalanced data sets can be misleading. For this reason, we will also calculate and use the IoU, also known as Jaccard Index (JI). The IoU metric is a standard criterion for semantic segmentation. Takes the pixels that overlap between the prediction and ground truth data and divides it by all the pixels into the prediction or ground truth data [51]. The criteria described above can be obtained from the following equations:

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - score = 2 \times \frac{(Precision \times Sensitivity)}{(Precision + Sensitivity)} \quad (5)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (6)$$

where TP represents the number of fire pixels in the fire class, FP represents the number of fire pixels in the non-fire class, and FN indicates the number of non-fire pixels in the fire class.

4. Results and Discussions

In this study, because many different configurations were used to detect active fire, the network accuracy criteria in different cases are first examined. Then the visual outputs of the best models are displayed.

4.1. Accuracy Assessment of AFD

The accuracy criteria used in this study can well show the strengths and weaknesses of different deep neural network configurations. For this purpose, based on the bands used from Landsat 8 images mentioned in Table 1, first, we evaluate the accuracy criteria in the case where three bands (B3) were used. Except when the dilation rate is 1, Table 3 demonstrates the best precision when three kernels of varying sizes (K357) are used. In addition, raising the dilation rate in the D3 scenario improves precision. However, in the other two scenarios (i.e., K3 and K35), the precision does not rise and fluctuates, so that in K3, raising the dilation rate from D2 to D3, first decreases and then raises it but in K35, the reverse has occurred. In general, the best precision is associated with the K3D1 model, which, of course, differs very little from the K357D3 approach. Except for the D2 scenario, using two kernels (K35) produces the highest sensitivity. In general, increasing the dilation rate does not have a constant effect on changing the sensitivity. By raising the dilation rate from 2 to 3, the K3 scenario initially boosts and then reduces the sensitivity criteria. In the case of the K35, however, the reverse is true. Increasing the dilation rate in the K357 model, on the other hand, wholly reduced the sensitivity. K35D3 model has the maximum sensitivity. Except for scenario D2, the maximum value of the F1-score for K35 reached the sensitivity criteria. The pattern of changes in the F1-score relative to dilation rate was similar to the sensitivity criteria, with minor changes. In the K35D1 model, the greatest F1-score is likewise reached. Again, the highest values achieved for the IoU criteria were the same as for the previous two criteria (sensitivity, F1-score), except for the K35D2 model.

As another scenario, the AFI was added to the other three bands as MultiScale-Net input. According to Table 3, except for the dilation rate of 3, the highest precision is obtained in the case of K3. The precision rose initially and then reduced in all three cases, K3, K35, and K357, as the dilation rate increased while the changes were minor. In general, K3D2 provided the best precision. In terms of the sensitivity criterion, using three kernels simultaneously have the maximum sensitivity in two cases (D2 and D3). In the case of D1, the best outcome is obtained by utilizing K35, which differs slightly from the implementation of K357. Increasing the dilation rate reduces the sensitivity in the case of K3. In general, K35D1 demonstrated the highest level of sensitivity. The employment of more kernels of varying sizes and greater dilation rates (i.e., K357 and D3 scenario, respectively) has boosted the F1-score value. In K357D2, the F1-score criteria yield the best results.

Furthermore, employing three kernels of varying sizes in each convolution layer improves the IoU criteria, and the most significant IoU values in further dilations are associated with the K357 scenario. On the other hand, increasing the dilation rate has the opposite effect and reduces the IoU in K3 and K35 scenarios. In K357, it initially boosts and then decreases the IoU value. The most remarkable results are associated with the K357D2 model.

As another MultiScale-Net training approach, single-feature AFI was used as input, and output accuracy was evaluated. Table 3 indicates that the K35 scenario has the highest precision at dilation rates. Furthermore, increasing the dilation in all three scenarios, K3, K35, and K357, increases precision followed by a drop. K35D1 was the best model. Except for a dilation rate equal to 2, the maximum sensitivity value is associated with the K3 scenario, while the best sensitivity in the D2 scenario is associated with K357.

Increasing the dilation rate has the same effect as increasing the precision measure (first increase, then decrease). The sensitivity criteria for the K357D2 mode were the highest. The maximum F1-score value, except when dilation rate 3 is utilized, is associated with the K357 scenario, and the best F1-score in the D3 is about K3. In K35 and K357, an increase in the dilation rate still resulted in rapid growth and a second reduction in the F1-score criterion, while in K3, this increase resulted in a modest decline followed by a rise in the F1-score. The K357D2 model received the highest F1 score. The highest IoU, as in the F1-score, occurred in the K357 scenario. Regarding the influence of the dilation rate, it can be seen that increasing the dilation rate causes a rise and subsequently a drop in the value of IoU in both K35 and K357 scenarios. In contrast, the K3 scenario went through the opposite trend. The K357D2 model won the competition for the most outstanding accuracy level again, this time in IoU.

After evaluating active fire detection accuracy criteria, some architectures were selected based on the maximum value of metrics used in each considered scenario. Numerical results show that the proposed method extracts and integrates related high-level features from the input images and thus produces fire masks that are very similar to ground truth.

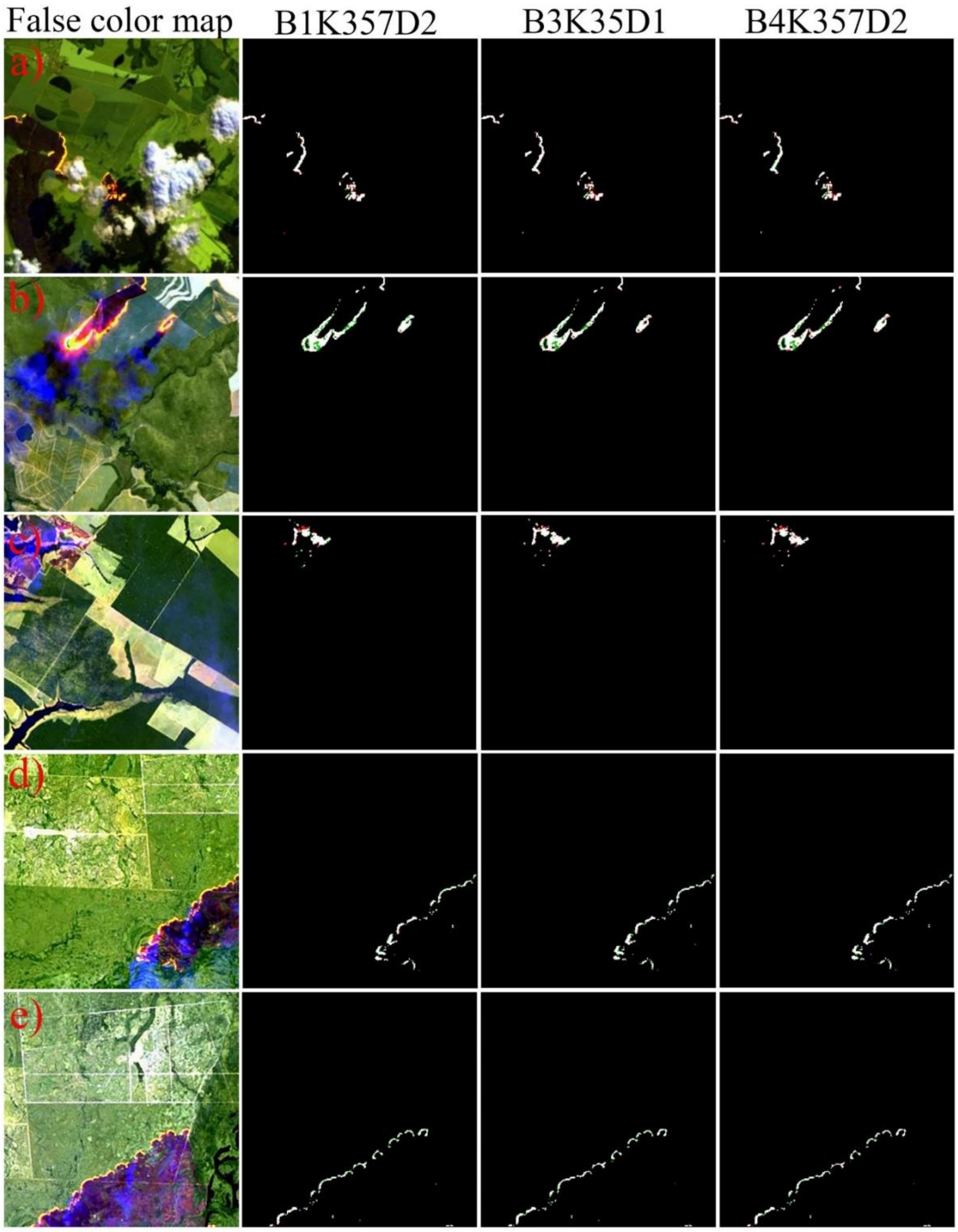
Table 3. Quantitative results of AFD in all scenarios by MultiScale-Net over test samples (P: Precision, S: Sensitivity, F: F1-score, and I: IoU).

Configuration scenarios				**Accuracy metrics (%)**			
Input	Kernel size	Dilation rate		P	S	F	I
B3	K3	D1		**95.71**	86.16	90.68	82.96
		D2		89.48	92.18	90.81	83.17
		D3		93.42	86.97	90.08	81.95
	K35	D1		91.06	91.83	**91.45**	**84.24**
		D2		93.65	86.89	90.14	82.06
		D3		88.02	**93.93**	90.87	83.28
	K357	D1		92.26	89.78	91	83.49
		D2		94.01	87.08	90.41	82.51
		D3		95.29	84.05	89.32	80.7
			Average	<u>**92.54**</u>	88.76	90.53	82.71
			Best Model	K3D1	K35D3	K35D1	K35D1
B1	K3	D1		93.25	85.35	89.12	80.38
		D2		87.8	90.04	88.91	80.04
		D3		90.58	88.53	89.54	81.07
	K35	D1		**94.5**	83.42	88.61	79.56
		D2		91.77	89.25	90.49	82.64
		D3		93.63	84.54	88.85	79.95
	K357	D1		93.59	85.22	89.21	80.52
		D2		89.76	**92.49**	**91.11**	**83.67**
		D3		92.63	80.21	85.97	75.4
			Average	91.94	86.56	89.09	80.35
			Best Model	K35D1	K357D2	K357D2	K357D2

B4	K3	D1		91	89.56	90.27	82.27
		D2		**92.51**	87.96	90.18	82.12
		D3		92.05	84.15	87.92	78.54
	K35	D1		89.95	**92.52**	91.22	83.86
		D2		92.49	88.92	90.67	82.93
		D3		92.18	89.05	90.59	82.79
	K357	D1		90.51	92.39	91.44	84.23
		D2		91.72	91.52	**91.62**	**84.54**
		D3		91.63	90.91	91.27	83.94
			Average	91.56	**89.66**	**90.58**	**82.79**
			Best Model	K3D2	K35D1	K357D2	K357D2

4.1. Qualitative Evaluation of MultiScale-Net

According to Equation (6), the IoU is equal to 1 if both FP and FN are equal to zero. In other words, the DL network should not have any misdiagnosis. For this reason, the IoU criterion is a strict metric, and for visual outputs, the models with the highest IoU value were selected (i.e., B1K357D2, B3K35D1, and B4K357D2). The binary output maps of the selected architectures are shown in Figure 6. All three models have detected and segmented active fire locations with relatively good accuracy. There is no significant difference between their outputs. The slight differences are frequently seen in locations where the extent of the fire ranges from a wide area to a single pixel. In such circumstances, the B4K357D2 outperformed other models, and the amount of FNs in this model's outputs is reduced, particularly in the extraction of single-pixel fires disconnected from the fire zone (e.g., samples (a), (i), and (j) in Figure 6). Furthermore, all three models chosen from the MultiScale-Net perform better in the situation of fire beams than single pixels isolated from the fire zones (e.g., samples (d), (e), and (h) in Figure 6). However, the MultiScale-Net is robust enough to alter the size and shape of the fire in the image scene.



8

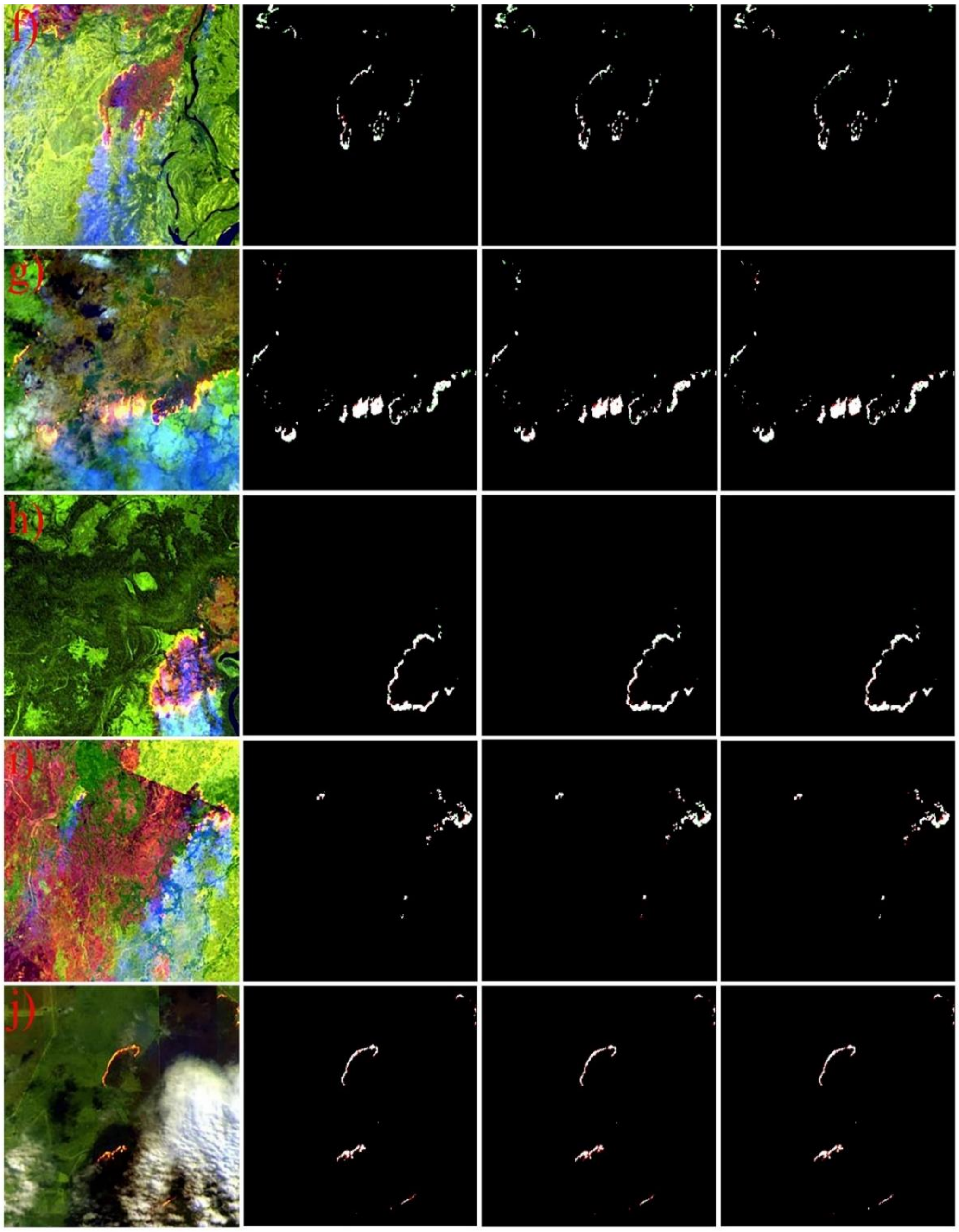


Figure 6. The results of AFD using the three best models of MultiScale-Net. White and black pixels indicate the TP and TN, respectively. Red pixels indicate the FN, and green pixels demonstrate FP.

The sample (j) (Figure 6) has more visual distinction between fire and backdrop than the others, with more minor fire-like characteristics. The number of FP pixels in the output map of B1K357D2, B3K35D1, and B4K357D2 for this sample is 3, 2, and 0, respectively. In contrast, all three models' output maps contain many FN pixels. As a result, it can be argued that, while decreasing the complexity of image conditions for detecting active fire has resulted in fewer false alarms (i.e., FP pixels), it has also resulted in the MultiScale-Net being negligent. Many fire pixels, particularly those on the boundary between fire and background, are misclassified as non-fire (i.e., FN pixels) by the MultiScale-Net.

Most segmentation mistakes have happened in the fire zones' peripheral pixels. However, in rare situations, pixels surrounding a large fire region are classified as fire pixels by the MultiScale-Net (i.e., FP pixels) (sample (b) in Figure 6). This misclassification is most likely owing to the extremely high temperature of the surrounding pixels (TP pixels) and its influence on the spectral reflectance of the inside pixels (i.e., FP pixels). In certain instances, the fire zone's center may be incorrectly classified as non-fire (see Figure 7). In some test samples, the pixel saturation issue was seen visually, so it also involved producing ground truth in this dataset (i.e., the conditions developed in [30–32]). In Figure 7, the digital numbers of pixels in the SWIR2 band, seen in green, are minimal, and it is zero in some of these pixels. As mentioned in the introduction, one of the sensitive bands in detecting and segmenting active band fire is SWIR2, which is zero in the green pixels in this image. Visually, and given the zone of fire in this image, maybe that these pixels are saturated. Again, the B4K357D2 model is more similar to the ground truth.

It should be noted that such visual interpretations may be associated with error because the nature of fire, especially its appearance in satellite images with limited spatial resolution and the significant impact of the Earth's atmosphere, is accompanied by uncertainty. Figure 8 shows slight variations between network predictions in different models in another test sample of the dataset. The main difference is where the two fire zones connect. The closest generated map to the ground-truth map is in the B4K357D2 model, but there is a separation between the two fire zones in the other two cases.

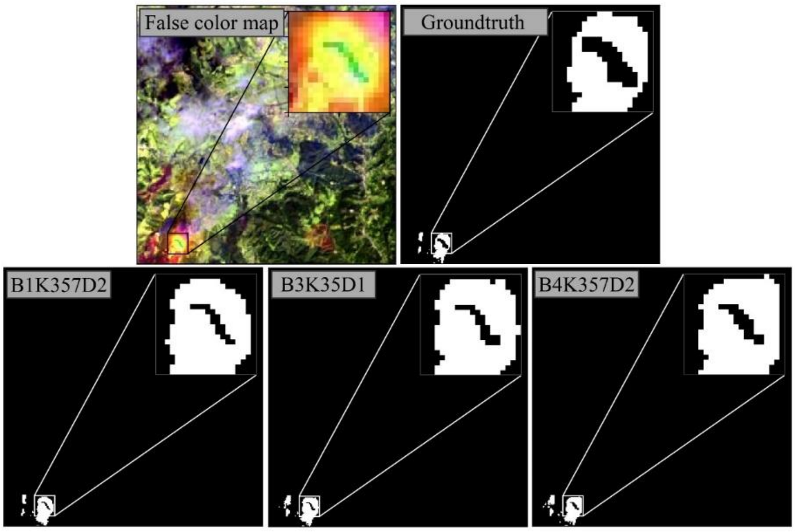


Figure 7. Pixel saturation problem in AFD in Landsat8 images.

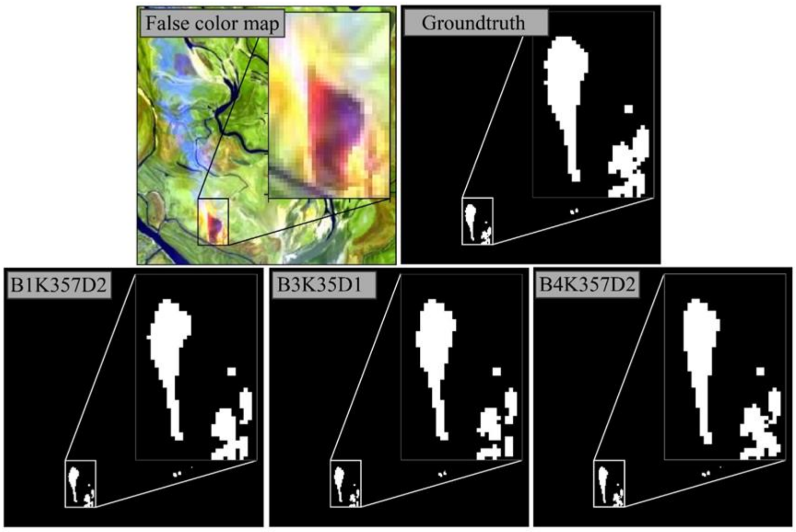


Figure 8. Comparison between three best models for AFD by MultiScale-Net.

4.3. Qualitative Evaluation of MultiScale-Net

One of the significant challenges in detecting and locating active fire is the presence of many clouds in the scene, so that in some studies, first masking the cloud and then locating active fire has been done. Also, the minimal number of target pixels (fire) makes it difficult to identify them in the image. In this study, however, the issue was considered challenging (i.e., heavy cloud conditions in the scene and the low number of fire pixels) to assess the performance of the best-selected models from the previous steps. According to Figure 9, the two models, B4K357D2 and B3K35D1, have the same performance and are different from ground truth, but the B1K357D2 model has a different performance than the previous two models. Again, it can be seen that it is visually difficult to distinguish fire pixels from non-fire pixels, and there is much uncertainty. The proposed DL models may have been correctly identified. However, all three models have false positive (FP) pixels; the pixels are considered fire pixels, but non-target pixels in the ground truth. As shown in Figure 10, both models B3K35D1 and B4K357D2 have been able to detect a single pixel of fire surrounded by many cloud pixels, but model B1K357D2 has been unable to do so. This result shows that the simultaneous use of kernels of different sizes accurately extracts the spectral-spatial properties associated with active fire, allowing even a single pixel of fire surrounded by background pixels to be properly segmented. Also, visually, there is uncertainty about the pixel below the target, which may be fire.

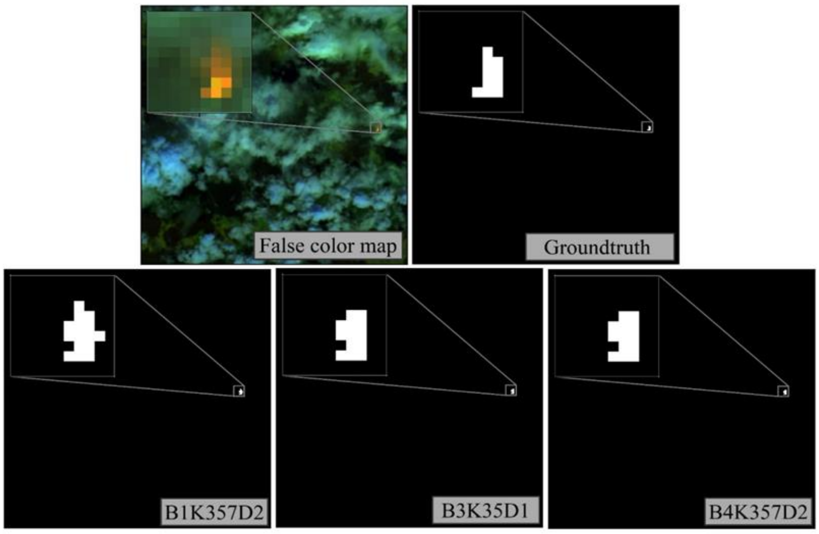
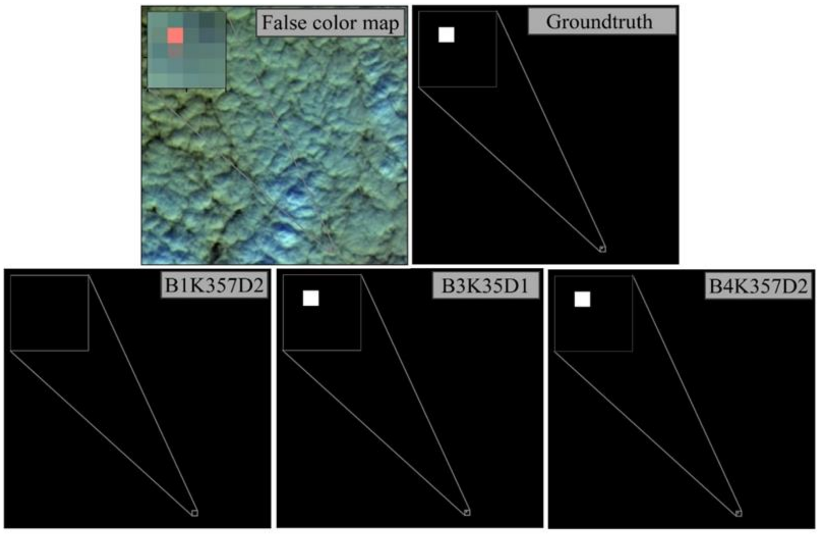


Figure 9. Qualitative comparison between three best models for AFD in severe cloudy scenes and few fire pixels



.

Figure 10. Qualitative comparison between three best models for AFD in severe cloudy scenes and few fire pixels

4.4. Effect of adding AFI to the three Landsat 8 bands

This study presented a new index to detect active fire in Landsat 8images. The best value for both the F1-score and IoU criteria is for a condition obtained from 4 bands, that is, by adding the AFI to the three Landsat 8 bands (B3). Of course, the use of several kernels with different sizes or different dilation rates may have resulted in the obtained accuracy. For this reason, the network configuration was averaged from all possible scenarios in all three scenarios (B1, B3, and B4).

As shown in Table 3, the highest average values of IoU, F1-score, and Sensitivity are related to the B4 scenario. However, in the case of the precision criteria, the highest mean value is related to B3. Even the average precision of the B1 scenario is better than B4. Given the fact that the best precision in the B1 is about 2% higher than in the B4, and the average value may have increased due to this high maximum value, the number of TP, FP, and FN pixels should be analyzed more attentively. Figure 11 shows that the number of FN pixels in B4 mode is less than the other two cases, while the number of FP pixels in B3 mode is less than the others. In the field of AFD, FN pixels are fire pixels. However, the network has identified these pixels as non-fire, which can be very dangerous in terms of fire management and firefighting, but FP pixels, some pixels are not actual fire that the MultiScale-Net recognizes these pixels as fire and is a kind of "false alarm."

On the other hand, the addition of AFI has caused the number of TP pixels to be higher than the other two modes, and in fact, the MultiScale-Net has been able to detect more fire pixels correctly. Therefore, the lower precision in B4 than B3 mode is due to more FP pixels. Nevertheless, in criteria such as sensitivity, where the number of FN pixels becomes more prominent than FP, the B4 mode leaves an outstanding performance. Also, in most of the test samples, the B4 scenario was more in line with ground truth (as in Figures 7 and 8), but in severe cloud conditions and a small number of target pixels in the scene, there was no difference in the results (as in Figure 9 and 10). Therefore, our innovative index visually helps separate active fire from the background by adding three more bands (SWIR2, SWIR1, and Blue) to improve the performance of the DL network in locating active fire. Although using the AFI as the sole input of the network did not lead to better results than the other two scenarios, the network predictions were reasonably good in this scenario, with even fewer FP pixels than in the B4.

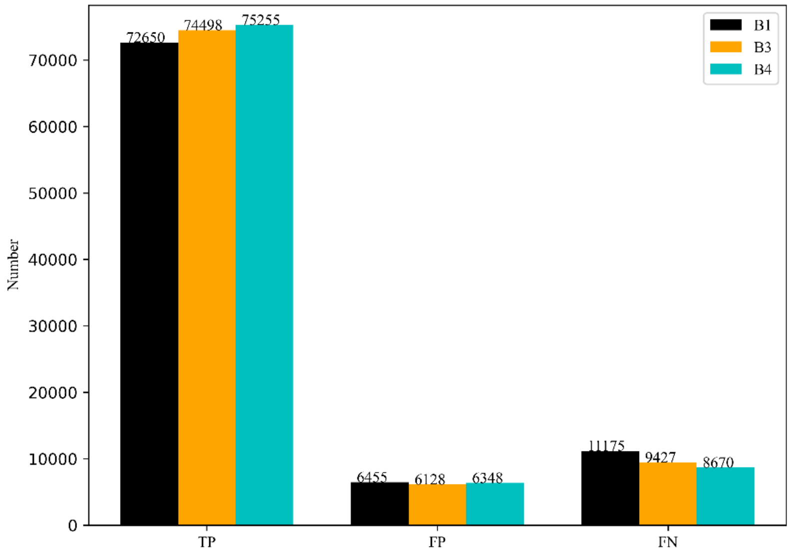


Figure 11. Number of TP, FP, and FN pixels in B1, B3, and B4 scenarios for AFD by MultiScale-Net

4.5. Effect of multi-size kernels and DCLs

Multi-size kernels have been used in building extraction [52], cloud and cloud shadow segmentation [53] and have achieved good results. This study used this technique to extract better the fire's extent and more excellent decision between fire and other objects with high reflectivity in the band range of SWIR2 by the deep encoder-decoder network. According to Table 3, out of the 12 best models, except for two cases, all the rest belong to configurations that use more than one kernel. The two cases that did not include this mode (i.e., the use of multi-size kernels) were for the precision criterion, and in the other criteria, the multi-size kernels were more successful. Out of 10 successes of multi-size kernels, K357 and K35 mode have been repeated five times. One of the test samples was selected to analyze the multi-size AFD ability of the multi-size kernels visually. Considering a specific scenario and a constant dilation rate, the MultiScale-Net output was compared in these cases (see Fig 12). One of the challenges in detecting active fire is to change the scale of the fire in an image scene so that there is a fire zone in one part of the image, and in the other part, there are small fires up to a few pixels. As shown in Figure 12, the K357 model performed better than the other two models and was relatively robust to fire resizing. Of course, there is still a slight difference with the ground truth.

The DCLs were used in deep neural networks to extract targets and features with different dimensions. In this study, DCLs with dilation rates of 1, 2, and 3 were used to locate the active fire better and meet the challenge of changing the target's scale in the image. According to Table 3, only one of the top models used dilation rate 3, which did not show good efficiency in other models. Of the remaining 11 top models, six were rated 2, and five were rated 1 (see Table 3). However, it should be noted that increasing the dilation rate from 1 to 2 improved accuracy in some cases but not always.

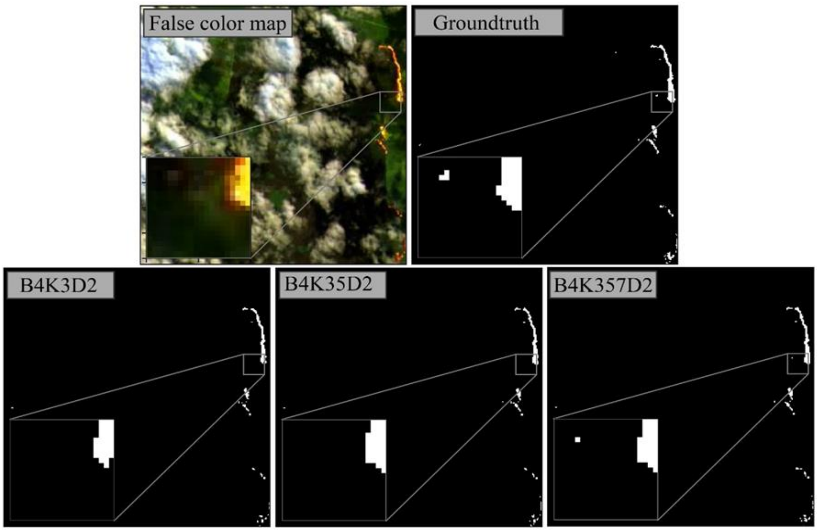


Figure 12. The effect of simultaneous use of several kernels with different sizes for AFD in MultiScale-Net

The number of FP, FN, and TPs in different scenarios has been investigated to evaluate the performance of MultiScale-Net in further detail and the effect of simultaneous use of convolutional kernels with different sizes and DCLs with different dilation rates (Figure 13). The dilation rate 1 (D1) has the lowest FPs in scenario B1 (i.e., the AFI as the only input feature of the MultiScale-Net). Among the numerous multi-size kernel scenarios, the 3 and 5 kernel sizes had the fewest FPs simultaneously (K35). The dilation rate 2 (D2) resulted in the lowest FNs. In contrast, the utilization of three kernels of varying sizes resulted in the lowest number of FNs, except for the dilation rate of 3 (D3).

In summary, in scenario B1, it can be observed that the dilation rates of 1 and 2 and the employment of chisels of various sizes have enabled the multiscale net to perform better in fire detection and have reduced the number of segmentation errors. Examining multiple models with varied kernel sizes and dilation rates in scenario B3 does not yield a relevant result regarding FPs. The lowest number of FPs, on the other hand, is associated with a dilation rate of 3 (D3) and a fixed-size kernel of 3 (K3). At dilation rates of 2 and 3 (D2 and D3, respectively), the lowest number of FPs is associated with the model that employs three kernels of varying sizes. Even though the K35D3 model has the fewest FNs, adjusting kernel sizes and dilation rates do not provide a consistent outcome. The simultaneous use of two kernels with sizes 3 and 5 (K35) resulted in the lowest number of FNs in rates of dilations 1 and 3. In contrast, in dilation rate 2 (D2), the simultaneous use of two kernels with sizes 3 and 5 (K35) resulted in the highest number of FNs.

Dilation rate 2 had the lowest FPs in scenario B4 (i.e., three bands of Landsat 8 + AFI as the MultiScale-Net input features). In contrast, the specific link between FPs and the increasing kernels with various sizes cannot be determined. The lowest number of FPs, on the other hand, is because the MultiScale-Net employs two kernels of sizes 3 and 5. The dilation rate 1 had the lowest FNs in all kernel scenarios of different sizes. However, there was still no consistent trend of changes in the number of convolution blocks with different kernel sizes and the number of FNs.

In contrast, the model with the lowest number of related FNs used two kernels of sizes 3 and 5 (i.e., K35D1 model, see the last row of Figure 16). According to Figure 13, the K35D1, K3D1, and K35D2 have the lowest FPs in the B1, B3, and B4 input feature scenarios, respectively, while the K357D2, K35D3, and K35D1 have the lowest FN. As a result, it can be approximately deduced that using kernels of varying sizes, particularly two kernels of sizes 3 and 5 (K35), produces small mistakes in detecting active fire compared to a kernel of constant size 3. However, varied dilation rates (particularly rate 3) DCLs do not significantly lower AFD inaccuracy. However, in some circumstances, the dilation rate 2 has been successful.

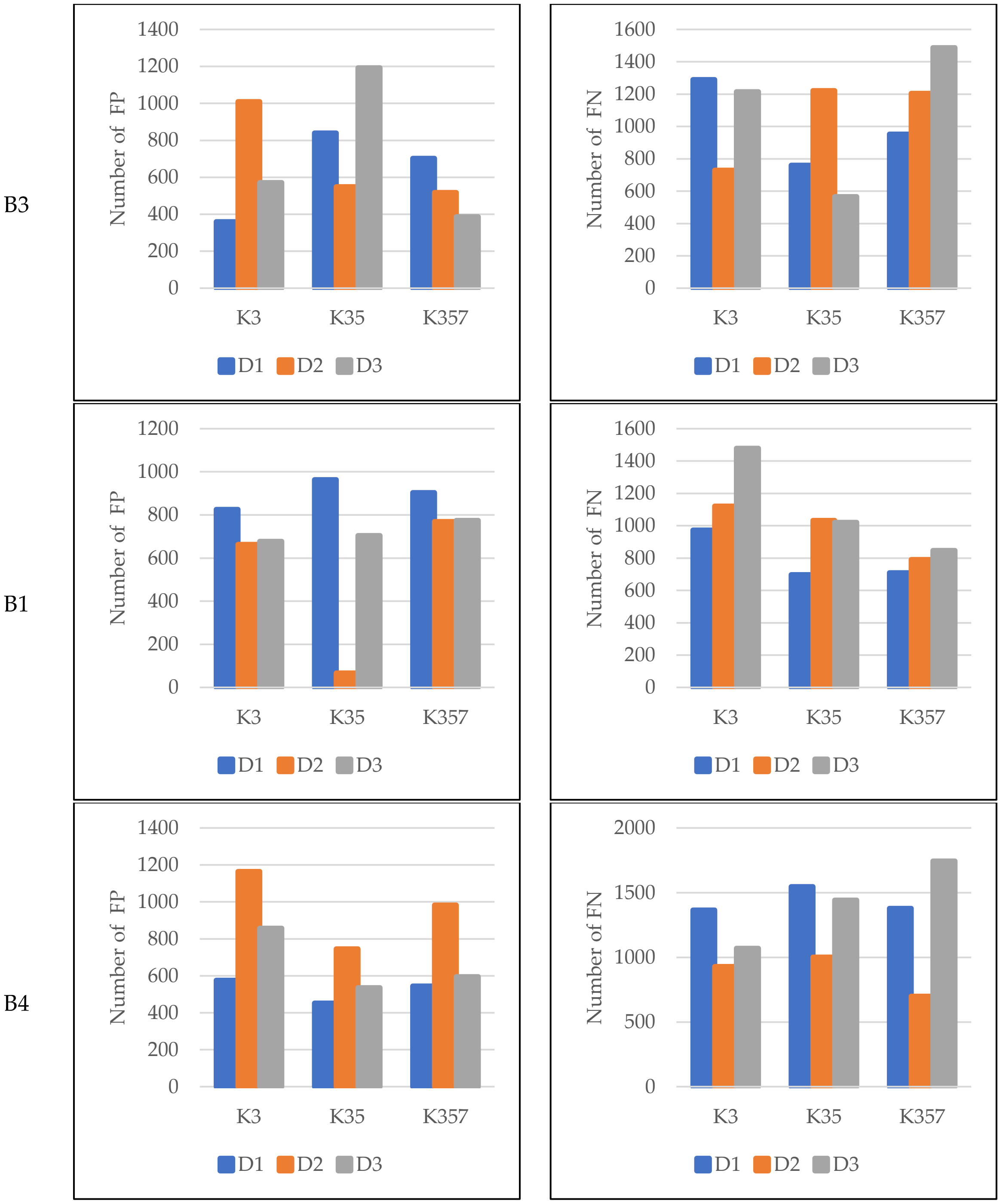


Figure 13. The numbers of FP and FN over all the test samples by different models with various kernel sizes and dilation rates in MultiScale-Net.

4.6. Comparison Analysis

Previous studies in AFD and segmentation have primarily relied on thresholds on certain spectral bands and the analysis of image pixel statistics, with just a few studies employing deep convolution neural networks. The absence of bulk data for training is the main reason for not implementing these approaches. Because of the modeling of complicated nonlinear relationships and the use of multiple convolution kernels and max-pooling layers, these networks contain many parameters. A large amount of labeled samples for the training phase is needed to estimate these parameters. De Almeida et al. [38] provide the first (as far as we know) large-scale dataset for active fire segmentation, encompassing over 150,000 image patches derived from Landsat 8 images collected throughout the world in August and September 2020, including fires in many regions. The primary distinction between our study and the reference study is constructing the CNN architectural structure and network input properties. Our primary objective is to create a high-performance CNN architecture for active fire segmentation to vary the fire's size and shape to make it more robust. It can even detect individual fire pixels separated from a larger fire region.

The main goal of the De Almeida et al. research was to prepare and publish a large-scale dataset for identifying active fire through DL algorithms and demonstrating the potential of these methods for AFD. As a result, their study relied only on the U-Net, which is one of the most widespread and straightforward CNNs available for segmentation and classification tasks in the remote sensing community. The basic U-Net excepts that the max-pooling and dropout layer are embedded after each convolutional layer were used in their study. However, our study employed convolutional kernels with multiple sizes and DCLs with varied rates in the MultiScale-Net. The feature maps of using these multiscale convolutional layers were concatenated into each convolutional block to provide high-level features that allowed the MultiScale-Net to detect active firing with reasonable accuracy.

Furthermore, the network was resistant to active fire resizing due to multi-size kernels and, in some instances, DCLs with varying rates. U-Net, on the other hand, lacks all of these characteristics. Furthermore, we introduced the AFI as a novel indicator for identifying active fire. The inclusion of this index as a MultiScale-Net input to the spectral bands of Landsat 8 images resulted in the extraction of more significant features from the image and enhanced network accuracy in active fire mapping. Two scenarios for U-Net input features were considered; three and ten Landsat 8 spectral bands.

Nevertheless, in our study, three scenarios were considered for network input features; three Landsat 8 spectral bands with AFI, three Landsat 8 bands, and AFI individually. Another notable distinction between the two studies is using different CNN training and testing phases scenarios. The U-Net 10c (ten bands of Landsat 8 image as input), U-Net 3c (three bands of Landsat 8 image as input), and U-Net 3c light (similar to U-Net 3c, but with just a quarter of the number of convolution filters in each layer) were utilized in the reference research. Also, as ground-truth data, binary segmentation masks based on the set of conditions described in Schroeder et al. [31], Murphy et al. [32], and Kumar and Roy [30] individually, voting and intersection of these three types of masks (5 types in total) were employed (a total of 15 different scenarios). The maps were also compared to the manually annotated data during the testing phase. The MultiScale-Net, on the other hand, was evaluated in 27 different scenarios in terms of network architecture (such as the simultaneous use of kernels of different sizes -K3, K35, K357- and DCLs with variable dilation rate –D1, D2, D3-), input features (B1, B3, and B4).

5. Conclusion

Active fires are among the most harmful natural hazards affecting climate change and life worldwide. With the development of artificial intelligence technology, DL methods have made great strides in computer image processing and vision. This research provided a "Multiscale-Net" AFD algorithm to identify visible, active fire pixels using Landsat-8 data. The proposed CNN benefited from two significant differences with the early DL networks: 1) convolutional kernels of different sizes simultaneously in each convolution layer. 2) Use of convolution layers with different dilation rates. The main advantages of the proposed algorithm can be summarized in three aspects. The first is that various observational samples from five continents were used to make the results reliable. The second is that MultiScale-Net showed satisfactory performance in extracting fires of different sizes in the challenging test samples. In another case, this study proposed AFI, a new innovative indicator for AFD, derived from the SWIR2 and blue divisions. Multiscale-Net uses it as an input feature to increase accuracy.

Different models (a total of 27) were examined based on the change of these three scenarios (input features-B1, B3, B4-, multi-size kernels -K3, K35, K357-, and dilation rate-D1, D2, D3-). Most of the best models were the ones that used multi-size kernels (in some cases, changing dilation rates from 1 to 2 improved the quantitative results, but 3 was inefficient) and integrated AFI with three Landsat 8 bands as the input features with the highest Precision, Sensitivity, F1-score, and IoU 95.71, 93.93, 91.62, and 84.54, over 19 test samples, respectively. In addition, scenario-based analyses of quantitative and qualitative experimental results were carried out. Among the input feature scenarios, B4 (AFI + three Landsat-8 spectral bands) had the greatest mean (average of all possible models in each scenario) of Sensitivity, F1-score, and IoU, with values of 89.66, 90.57, and 82.79, respectively. This scenario had the lowest FN pixels among the other scenarios, which is an essential advantage because this inaccuracy ignores the extent of the fire and causes irreversible damage during the firefighting operation. Qualitative studies also revealed that using kernels of different sizes concurrently robustness the MultiScale-Net against changes in patterns and size of the active fire in the image. This multi-size kernel in scenario B4 enhances the MultiScale-Net performance at fire zone linkage and in areas with single fire pixels away from major fire zones. The MultiScale-Net was also tested under cloudy conditions and with a single fire pixel, with good outcomes

This study was a new step in using deep learning techniques in fire detection in satellite images that have received less attention. The findings of this study can be used to manage and control fires and reduce fire damage. In future research, we propose that due to the non-dependence of the proposed method on thermal bands, the potential of Sentinel-2 data be investigated due to its high spatial and temporal resolution. Also, the processing time of the proposed method can be evaluated in cloud environments such as Google Earth Engine (GEE).

Author Contributions: Conceptualization, A.R. and R.S.; methodology, A.R., S.A. and A.Z.; software, A.R., S.A. and A.Z..; formal analysis, S.H.; investigation, A.R. and R.S.; writing—original draft preparation, A.R.; writing—review and editing, A.R., R.S., S.H., A.Z.; visualization, R.S.; supervision, R.S.; project administration, R.S, and S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Data Availability Statement: The authors state that all the data in this study are publicly available at https://github.com/pereira-gha/activefire.

Acknowledgments: The authors would like to present their appreciation to De Almeida et al. for providing and publishing the unique and valuable dataset to the remote sensing community

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Seydi, S.T.; Hasanlou, M.; Chanussot, J. DSMNN-Net: A Deep Siamese Morphological Neural Network Model for Burned Area Mapping Using Multispectral Sentinel-2 and Hyperspectral PRISMA Images. *Remote Sensing* **2021**, *13*, doi:10.3390/rs13245138.
2. Lalani, N.; Drolet, J.L.; McDonald-Harker, C.; Brown, M.R.G.; Brett-MacLean, P.; Agyapong, V.I.O.; Greenshaw, A.J.; Silverstone, P.H. Nurturing Spiritual Resilience to Promote Post-Disaster Community Recovery: The 2016 Alberta Wildfire in Canada. *Front Public Health* **2021**, *9*, 682558–682558, doi:10.3389/fpubh.2021.682558.
3. Seydi, S.T.; Akhoondzadeh, M.; Amani, M.; Mahdavi, S. Wildfire Damage Assessment over Australia Using Sentinel-2 Imagery and MODIS Land Cover Product within the Google Earth Engine Cloud Platform. *Remote Sensing* **2021**, *13*, doi:10.3390/rs13020220.
4. Keeley, J.E.; Syphard, A.D. Large California Wildfires: 2020 Fires in Historical Context. *Fire Ecology* **2021**, *17*, 22, doi:10.1186/s42408-021-00110-7.
5. *FAO. 2020. Global Forest Resources Assessment 2020 – Key Findings. Rome.;*
6. Gin, J.L.; Balut, M.D.; Der-Martirosian, C.; Dobalian, A. Managing the Unexpected: The Role of Homeless Service Providers during the 2017–2018 California Wildfires. *Journal of Community Psychology* **2021**, *49*, 2532–2547, doi:10.1002/jcop.22653.
7. Ball, G.; Regier, P.; González-Pinzón, R.; Reale, J.; Van Horn, D. Wildfires Increasingly Impact Western US Fluvial Networks. *Nature Communications* **2021**, *12*, 2484, doi:10.1038/s41467-021-22747-3.
8. Moumgiakmas, S.S.; Samatas, G.G.; Papakostas, G.A. Computer Vision for Fire Detection on UAVs—From Software to Hardware. *Future Internet* **2021**, *13*, doi:10.3390/fi13080200.
9. Xu, R.; Lin, H.; Lu, K.; Cao, L.; Liu, Y. A Forest Fire Detection System Based on Ensemble Learning. *Forests* **2021**, *12*, doi:10.3390/f12020217.
10. Alkhatib, A.A.A. A Review on Forest Fire Detection Techniques. *International Journal of Distributed Sensor Networks* **2014**, *10*, 597368, doi:10.1155/2014/597368.
11. Wooster, M.J.; Roberts, G.J.; Giglio, L.; Roy, D.P.; Freeborn, P.H.; Boschetti, L.; Justice, C.; Ichoku, C.; Schroeder, W.; Davies, D.; et al. Satellite Remote Sensing of Active Fires: History and Current Status, Applications and Future Requirements. *Remote Sensing of Environment* **2021**, *267*, 112694, doi:10.1016/j.rse.2021.112694.
12. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A Review on Early Forest Fire Detection Systems Using Optical Remote Sensing. *Sensors* **2020**, *20*, doi:10.3390/s20226442.
13. Cihlar, J.; Chen, J.; Li, Z. Seasonal AVHRR Multichannel Data Sets and Products for Studies of Surface-Atmosphere Interactions. *Journal of Geophysical Research: Atmospheres* **1997**, *102*, 29625–29640, doi:10.1029/97JD01195.
14. Csiszar, I.; Schroeder, W.; Giglio, L.; Ellicott, E.; Vadrevu, K.P.; Justice, C.O.; Wind, B. Active Fires from the Suomi NPP Visible Infrared Imaging Radiometer Suite: Product Status and First Evaluation Results. *Journal of Geophysical Research: Atmospheres* **2014**, *119*, 803–816, doi:10.1002/2013JD020453.
15. Zarei, A.; Shah-Hosseini, R.; Ranjbar, S.; Hasanlou, M. Validation of Non-Linear Split Window Algorithm for Land Surface Temperature Estimation Using Sentinel-3 Satellite Imagery: Case Study; Tehran Province, Iran. *Advances in Space Research* **2021**, *67*, 3979–3993, doi:10.1016/j.asr.2021.02.019.
16. Xu, W.; Wooster, M.J.; He, J.; Zhang, T. First Study of Sentinel-3 SLSTR Active Fire Detection and FRP Retrieval: Night-Time Algorithm Enhancements and Global Intercomparison to MODIS and VIIRS AF Products. *Remote Sensing of Environment* **2020**, *248*, 111947, doi:10.1016/j.rse.2020.111947.

17. Xiaoxiong Xiong; Emily J. Aldoretta; Amit Angal; Tiejun Chang; Xu Geng; Daniel O. Link; Vincent V. Salomonson; Kevin A. Twedt; Aisheng Wu Terra MODIS: 20 Years of on-Orbit Calibration and Performance. *Journal of Applied Remote Sensing* **2020**, *14*, 1–16, doi:10.1117/1.JRS.14.037501.
18. Giglio, L.; Schroeder, W.; Justice, C.O. The Collection 6 MODIS Active Fire Detection Algorithm and Fire Products. *Remote Sensing of Environment* **2016**, *178*, 31–41, doi:10.1016/j.rse.2016.02.054.
19. Parto, F.; Saradjian, M.; Homayouni, S. MODIS Brightness Temperature Change-Based Forest Fire Monitoring. *Journal of the Indian Society of Remote Sensing* **2020**, *48*, 163–169, doi:10.1007/s12524-019-01071-w.
20. He, L.; Li, Z. Enhancement of a Fire-Detection Algorithm by Eliminating Solar Contamination Effects and Atmospheric Path Radiance: Application to MODIS Data. *null* **2011**, *32*, 6273–6293, doi:10.1080/01431161.2010.508057.
21. He, L.; Li, Z. Enhancement of a Fire Detection Algorithm by Eliminating Solar Reflection in the Mid-IR Band: Application to AVHRR Data. *null* **2012**, *33*, 7047–7059, doi:10.1080/2150704X.2012.699202.
22. Schroeder, W.; Oliva, P.; Giglio, L.; Csiszar, I.A. The New VIIRS 375m Active Fire Detection Data Product: Algorithm Description and Initial Assessment. *Remote Sensing of Environment* **2014**, *143*, 85–96, doi:10.1016/j.rse.2013.12.008.
23. C. B. Engel; S. D. Jones; K. Reinke A Seasonal-Window Ensemble-Based Thresholding Technique Used to Detect Active Fires in Geostationary Remotely Sensed Data. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *59*, 4947–4956, doi:10.1109/TGRS.2020.3018455.
24. Zang, Z.; Li, D.; Guo, Y.; Shi, W.; Yan, X. Superior PM2.5 Estimation by Integrating Aerosol Fine Mode Data from the Himawari-8 Satellite in Deep and Classical Machine Learning Models. *Remote Sensing* **2021**, *13*, doi:10.3390/rs13142779.
25. Jang, E.; Kang, Y.; Im, J.; Lee, D.-W.; Yoon, J.; Kim, S.-K. Detection and Monitoring of Forest Fires Using Himawari-8 Geostationary Satellite Data in South Korea. *Remote Sensing* **2019**, *11*, doi:10.3390/rs11030271.
26. Xie, Z.; Song, W.; Ba, R.; Li, X.; Xia, L. A Spatiotemporal Contextual Model for Forest Fire Detection Using Himawari-8 Satellite Data. *Remote Sensing* **2018**, *10*, doi:10.3390/rs10121992.
27. Roy, D.P.; Kovalskyy, V.; Zhang, H.K.; Vermote, E.F.; Yan, L.; Kumar, S.S.; Egorov, A. Characterization of Landsat-7 to Landsat-8 Reflective Wavelength and Normalized Difference Vegetation Index Continuity. *Remote Sensing of Environment* **2016**, *185*, 57–70, doi:10.1016/j.rse.2015.12.024.
28. Kovalskyy, V.; Roy, D.P. A One Year Landsat 8 Conterminous United States Study of Cirrus and Non-Cirrus Clouds. *Remote Sensing* **2015**, *7*, doi:10.3390/rs70100564.
29. Liu, Z.; K Wu; Jiang, R.; Zhang, H. A SIMPLE ARTIFICIAL NEURAL NETWORK FOR FIRE DETECTION USING LANDSAT-8 DATA. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **2020**, *XLIII-B3-2020*, 447–452, doi:10.5194/isprs-archives-XLIII-B3-2020-447-2020.
30. Kumar, S.S.; Roy, D.P. Global Operational Land Imager Landsat-8 Reflectance-Based Active Fire Detection Algorithm. *null* **2018**, *11*, 154–178, doi:10.1080/17538947.2017.1391341.
31. Schroeder, W.; Oliva, P.; Giglio, L.; Quayle, B.; Lorenz, E.; Morelli, F. Active Fire Detection Using Landsat-8/OLI Data. *Remote Sensing of Environment* **2016**, *185*, 210–220, doi:10.1016/j.rse.2015.08.032.
32. Murphy, S.W.; de Souza Filho, C.R.; Wright, R.; Sabatino, G.; Correa Pabon, R. HOTMAP: Global Hot Target Detection at Moderate Spatial Resolution. *Remote Sensing of Environment* **2016**, *177*, 78–88, doi:10.1016/j.rse.2016.02.027.
33. M. Ansari; S. Homayouni; A. Safari; S. Niazmardi A New Convolutional Kernel Classifier for Hyperspectral Image Classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2021**, *14*, 11240–11256, doi:10.1109/JSTARS.2021.3123087.

34. Sadegh Ranjbar; Arastou Zarei; Mahdi Hasanlou; Mehdi Akhoondzadeh; Jalal Amini; Meisam Amani Machine Learning Inversion Approach for Soil Parameters Estimation over Vegetated Agricultural Areas Using a Combination of Water Cloud Model and Calibrated Integral Equation Model. *Journal of Applied Remote Sensing* **2021**, *15*, 1–17, doi:10.1117/1.JRS.15.018503.
35. Z. -Q. Zhao; P. Zheng; S. -T. Xu; X. Wu Object Detection With Deep Learning: A Review. *IEEE Transactions on Neural Networks and Learning Systems* **2019**, *30*, 3212–3232, doi:10.1109/TNNLS.2018.2876865.
36. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-a: A Deep Learning Framework for Semantic Segmentation of Remotely Sensed Data. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, *162*, 94–114, doi:10.1016/j.isprsjprs.2020.01.013.
37. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *Journal of Big Data* **2021**, *8*, 53, doi:10.1186/s40537-021-00444-8.
38. de Almeida Pereira, G.H.; Fusioka, A.M.; Nassu, B.T.; Minetto, R. Active Fire Detection in Landsat-8 Imagery: A Large-Scale Dataset and a Deep-Learning Study. *ISPRS Journal of Photogrammetry and Remote Sensing* **2021**, *178*, 171–186, doi:10.1016/j.isprsjprs.2021.06.002.
39. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *Journal of Big Data* **2019**, *6*, 60, doi:10.1186/s40537-019-0197-0.
40. Wang, Y.; Pan, X.; Song, S.; Zhang, H.; Huang, G.; Wu, C. Implicit Semantic Data Augmentation for Deep Networks. *Advances in Neural Information Processing Systems* **2019**, *32*, 12635–12644.
41. García-Llamas, P.; Suárez-Seoane, S.; Fernández-Guisuraga, J.M.; Fernández-García, V.; Fernández-Manso, A.; Quintano, C.; Taboada, A.; Marcos, E.; Calvo, L. Evaluation and Comparison of Landsat 8, Sentinel-2 and Deimos-1 Remote Sensing Indices for Assessing Burn Severity in Mediterranean Fire-Prone Ecosystems. *International Journal of Applied Earth Observation and Geoinformation* **2019**, *80*, 137–144, doi:10.1016/j.jag.2019.04.006.
42. Giglio, L.; Descloitres, J.; Justice, C.O.; Kaufman, Y.J. An Enhanced Contextual Fire Detection Algorithm for MODIS. *Remote Sensing of Environment* **2003**, *87*, 273–282, doi:10.1016/S0034-4257(03)00184-6.
43. E. Cadau; G. Laneve Improved MSG-SEVIRI Images Cloud Masking and Evaluation of Its Impact on the Fire Detection Methods. In Proceedings of the IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium; July 7 2008; Vol. 2, p. II–1056.
44. Tang, H.; Yu, K.; Hagolle, O.; Jiang, K.; Geng, X.; Zhao, Y. A Cloud Detection Method Based on a Time Series of MODIS Surface Reflectance Images. *null* **2013**, *6*, 157–171, doi:10.1080/17538947.2013.833313.
45. V. Badrinarayanan; A. Kendall; R. Cipolla SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2017**, *39*, 2481–2495, doi:10.1109/TPAMI.2016.2644615.
46. Garbin, C.; Zhu, X.; Marques, O. Dropout vs. Batch Normalization: An Empirical Study of Their Impact to Deep Learning. *Multimedia Tools and Applications* **2020**, *79*, 12777–12815, doi:10.1007/s11042-019-08453-9.
47. Panicker, R.O.; Kalmady, K.S.; Rajan, J.; Sabu, M.K. Automatic Detection of Tuberculosis Bacilli from Microscopic Sputum Smear Images Using Deep Learning Methods. *Biocybernetics and Biomedical Engineering* **2018**, *38*, 691–699, doi:10.1016/j.bbe.2018.05.007.
48. Y. Ho; S. Wookey The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. *IEEE Access* **2020**, *8*, 4806–4813, doi:10.1109/ACCESS.2019.2962617.

49. Xavier Glorot; Yoshua Bengio Understanding the Difficulty of Training Deep Feedforward Neural Networks. In Proceedings of the Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics; Yee Whye Teh, Mike Titterington, Eds.; PMLR, March 31 2010; Vol. 9, pp. 249–256.
50. Chicco, D.; Jurman, G. The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics* **2020**, *21*, 6, doi:10.1186/s12864-019-6413-7.
51. C. Gonzales; W. Sakla Semantic Segmentation of Clouds in Satellite Imagery Using Deep Pre-Trained U-Nets. In Proceedings of the 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR); October 15 2019; pp. 1–7.
52. Mehdi Khoshboresh-Masouleh; Fatemeh Alidoost; Hossein Arefi Multiscale Building Segmentation Based on Deep Learning for Remote Sensing RGB Images from Different Sensors. *Journal of Applied Remote Sensing* **2020**, *14*, 1–21, doi:10.1117/1.JRS.14.034503.
53. Khoshboresh-Masouleh, M.; Shah-Hosseini, R. A Deep Learning Method for Near-Real-Time Cloud and Cloud Shadow Segmentation from Gaofen-1 Images. *Computational Intelligence and Neuroscience* **2020**, *2020*, 8811630, doi:10.1155/2020/8811630.