

## Article

# Uni-Temporal Multispectral Imagery for Burned Area Mapping with Deep Learning

Xikun Hu, Yifang Ban <sup>\*</sup>  and Andrea Nascetti 

Division of Geoinformatics, KTH Royal Institute of Technology, SE-10044 Stockholm, Sweden; xikun@kth.se (X.H.); nascetti@kth.se (A.N.)

\* Correspondence: yifang@kth.se

**Abstract:** Accurate burned area information is needed to assess the impacts of wildfires on people, communities, and natural ecosystems. Various burned area detection methods have been developed using satellite remote sensing measurements with wide coverage and frequent revisits. Our study aims to expound on the capability of deep learning (DL) models for automatically mapping burned areas from uni-temporal multispectral imagery. Specifically, several semantic segmentation network architectures, i.e., U-Net, HRNet, Fast-SCNN, and DeepLabv3+, and machine learning (ML) algorithms were applied to Sentinel-2 imagery and Landsat-8 imagery in three wildfire sites in two different local climate zones. The validation results show that the DL algorithms outperform the ML methods in two of the three cases with the compact burned scars, while ML methods seem to be more suitable for mapping dispersed burn in boreal forests. Using Sentinel-2 images, U-Net and HRNet exhibit comparatively identical performance with higher kappa (around 0.9) in one heterogeneous Mediterranean fire site in Greece; Fast-SCNN performs better than others with kappa over 0.79 in one compact boreal forest fire with various burn severity in Sweden. Furthermore, directly transferring the trained models to corresponding Landsat-8 data, HRNet dominates in the three test sites among DL models and can preserve the high accuracy. The results demonstrated that DL models can make full use of contextual information and capture spatial details in multiple scales from fire-sensitive spectral bands to map burned areas. Using only a post-fire image, the DL methods not only provide automatic, accurate, and bias-free large-scale mapping option with cross-sensor applicability, but also have potential to be used for onboard processing in the next Earth observation satellites.



**Citation:** Hu, X.; Ban, Y.; Nascetti, A. Uni-Temporal Multispectral Imagery for Burned Area Mapping with Deep Learning. *Remote Sens.* **2021**, *13*, 1509. <https://doi.org/10.3390/rs13081509>

Academic Editor: Alfonso Fernández-Manso

Received: 7 February 2021

Accepted: 5 April 2021

Published: 14 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** Sentinel-2; Landsat-8; burned area mapping; deep learning; semantic segmentation; machine learning

## 1. Introduction

Wildfires are widely recognized as one of the most critical ecosystem disturbances, as they not only result in the significant loss of human lives and properties, but also affect biodiversity and the carbon cycle [1]. Accurate and timely mapping of burned areas is, therefore, needed for the assessment of economic losses caused by the wildfires, managing post-fire hazards such as landslides or mudflows, and planning of remediation and revegetation efforts. Historically, ground-based estimates were used to collect burned area information [2]. With the launch of Earth observation satellites, remote sensing has become a more efficient alternative to monitor wildfire extent due to its timely coverage of fire occurrences regionally and globally [3,4].

Over the past decades, coarse-resolution satellite sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) have been used to identify the burned areas globally based on the thermal emission of burned vegetation [4]. For instance, some MODIS monthly burned area products, such as MCD64A1 [5] at 500 m resolution and FireCCI51 [6] at 250 m resolution, are currently available online, e.g., Google Earth Engine (GEE) platform. However, these burned area products often miss small burned areas that

account for a significant proportion of the total burned area [7]. Therefore, it is desirable to evaluate effective methods for accurate burned area mapping using medium-to-high resolution satellite images.

Open access to Landsat-8 and Sentinel-2 satellite data with a global median average revisit at 2.9 days provides an excellent opportunity for mapping burned areas in near-real-time (NRT) [8]. Various algorithms have been developed using Sentinel-2 and/or Landsat data, but most of the studies require a pre-fire image, dense time-series data, or an empirical threshold [9–11]. To address the main challenge of estimating burn scar from a single post-fire scene, the objective of this study is to investigate DL models for burned area mapping in two climate zones in comparison to other commonly used classical methods.

## 2. Related Studies

Sensors like Sentinel-2 and Landsat-8 have multispectral channels in the visible/near-infrared (NIR) and short-wave infrared (SWIR) that are sensitive to fire disturbance. Removal or alteration of leaf structure and canopy cover vegetation caused by wildfire changes the land vegetated surface on the ground even to soil exposure in some extreme events. The amount of radiation reflectance from these changes can be represented as a function of the spectral wavelength [12]. For instance, more radiation in NIR is absorbed by fire-disturbed areas than unburned ones, while burned areas would reflect more radiation in the visible and SWIR bands [13]. Therefore, many well-known spectral indices (covering the visible/NIR, the NIR/short SWIR, and the short SWIR/long SWIR spectral spaces) have been proposed for burned area detection, e.g., MIRBI, EVI, NDVI, BAIM, NBR, NBR2, and BAIS2 [7,14–20]. These indices can be further differentially normalized by the corresponding bitemporal pair (i.e., pre-fire and post-fire scenes) to enhance burned area discrimination (e.g., dNBR as a typical index), which additionally requires the cloud-free pre-fire satellite imagery.

Index thresholding is commonly used for burned area mapping, and the threshold values are often set empirically depending on visual interpretation, biome types, and tree cover percentages [21]. On the other hand, automated methods on indices such as OTSU [22] might have difficulties determining an optimal threshold for indices when the distributions of scene intensity and land cover are complex. Recently, an automatic thresholding chain was proposed for NRT burned area mapping at a national scale using Sentinel-2 data based on the dNDVI and RdNBR with mapping unit 1 hectare (ha) [23]. However, these bitemporal indices increase preprocessing time and limit future applications, especially for the next Earth observation satellite with onboard data processing and limited storage for previous scenes. Some automated methods based on spectral signatures require retraining when applied to diverse landscapes despite tuning thresholds to local conditions [24]. Therefore, the nonparametric machine learning (ML) algorithms have received much attention for their better performance in burned area detection than traditional threshold-based methods using spectral indices [25].

As ML algorithms depend on the distribution of the training data without any assumptions, automated burned area mapping becomes achievable. Various ML methods have been widely used in wildfire science including random forests (RF), Support Vector Machines (SVM), Artificial Neural Networks (ANN), decision trees, and MaxEnt [25]. Furthermore, dense harmonic time-series of Landsat data were used to identify burned areas [26,27]. An automated burned area mapping algorithm using paired Sentinel-2 images was introduced in [28], with an SVM for an initial pixel-based classification and a multiple spectral-spatial classification approach for smoothing the final burned area delineation. Moreover, a global burned area mapping approach was implemented using RF and a seed-growing approach based on time-series Landsat-8 images and GEE [29].

Automated burned area mapping with a uni-temporal post-fire image is promising but relatively challenging because of spectrally similar effects to burned areas caused by unrelated phenomena and disturbances such as shadows, agricultural harvesting, or plowing in the absence of ancillary information [10,30,31]. Various attempts have been



made to map burned areas using a post-fire image. For example, logistic regression (LR) was applied to a single post-fire Landsat-5 Thematic Mapper (TM) image for burned land mapping [30]. Mitrakis et al. [32] compared a variety of ML algorithms including ANN, SVM, and Ada Boost Classifier (AdaBoost) for burned-area mapping in the Mediterranean region with one post-fire Landsat-5 TM image and found that all methods displayed similar accuracy. Likewise, Mallinis and Koutsias [33] compared 10 classification methods and concluded that the variance imposed by the methods is less than the variance imposed by factors differentiated locally in the study sites. Further, Pu and Gong [34] used LR to calculate probabilities of burned scars from a single post-fire Landsat 7 ETM+ image with acceptable overall accuracy. Stroppiana et al. [35] proposed a (semi)-automated multicriteria method for burned area mapping from uni-temporal Landsat TM images in the Mediterranean environment. This soft aggregation approach could reduce omission errors less than 3% but resulted in high commission errors of about 21%.

Recent studies demonstrated that deep learning (DL) algorithms have the capacity to automatically capture object features at multiple scales without the requirement for an extra user input but with a few specific hyperparameters [36]. Semantic segmentation architectures with convolutional neural networks (CNNs) have unique characteristics to extract the contextual information in multiple scales and then label each pixel of an image [37]. It is now playing a significant role in many image analysis tasks ranging from autonomous vehicles, human–computer interaction, robotics, medical research, precision farming, and so on [37–43]. For remote sensing, semantic segmentation algorithms have recently been applied to 2D satellite images and even 3D scenes [44,45]. For instance, automatic extraction of snow cover from high spatial resolution optical images was proposed using DL on a small dataset [46]. A fully convolutional networks model trained on very high resolution (VHR) optical satellite imagery was transferred to Sentinel-2 and SAR data in slum mapping [47]. CloudNet was presented to classify cloud and haze from Sentinel-2 imagery based on deep residual learning, semantic image segmentation, and the concept of atrous convolution [48]. A revised U-Net network structure named DeepUNet was explored for pixel-level sea–land segmentation with images collected from Google Earth and handicraft labeled ground truth images [49]. A similar U-Net architecture with residual units was employed for road area extraction with relatively high accuracy [50]. A Mask Region-Based CNN(R-CNN) model was applied to automatically mapping applications such as ice-wedge polygons [41,42] and archaeological sites [43] with high-resolution or VHR remote sensing imagery.

As such, there exist opportunities to employ DL-based models in burned-area detection, particularly in cases involving large multivariate datasets [25]. Recently, a DL approach achieved competitive results with low spatial resolution observations ( $0.01^\circ$  spatial resolution grid) for mapping and dating of burned areas [51]. Langford et al. [52] took a weight selection strategy to tackle the imbalanced classification in deep neural networks training for binary wildfire classification across Alaska with MODIS variables. Concerning the medium- and high-resolution satellite images, an implicit Radar Convolutional Burn Index was proposed based on multitemporal Sentinel-1 SAR data and InSAR technique for mapping the burned areas under a convolutional network-based classification framework [53]. Sentinel-1 SAR backscatter was further proved effective to detect burned areas with a CNN-based DL framework [54]. Bermudez et al. [55] used a conditional Generative Adversarial Network to synthesize missing remote sensed optical data from Sentinel-1 SAR data for a region with the presence of burned area. Recently, de Bem et al. [56] analyzed the performance of deep convolutional autoencoders (U-Net and ResUnet) using bitemporal image pair of the Landsat scenes and recommended the sampling window size of 256 by 256 pixels in DL model training.

Using uni-temporal Sentinel-2 imagery, Knopp et al. [57] proposed an automatic processing chain for burned area segmentation based on U-Net. It successfully reaches high overall accuracy, but lacks the test on transferability of the Sentinel-2 trained model to other sensor data. Due to the large proportion of coarse perimeters as reference data, the network

in [57] attempts to create homogeneous burned area delineation with false positives therein. In addition, more DL-based models need to be involved in comprehensive evaluation along with other ML approaches in different landscapes. Considering the spectral consistency between the multispectral Sentinel-2 and Landsat-8 data [58], it implies the potential to make use of transferability and generalization of DL-based models to map burned areas with cross-sensor multispectral data. In our study, several semantic segmentation networks in different classes [37] are compared, including the widely used U-Net [59], region proposal-based CNN architectures like Fast-SCNN [60], increase the resolution of the feature-based DeepLabv3+ [61], and state-of-art enhancement of feature-based high-resolution networks (HRNet) [62].

The overall objective of this study is to specifically explore the capabilities of DL-based models for mapping burned areas in various landscapes including the Mediterranean regions and the boreal forests in Sweden. The specific aims are to address the following limitations of the previous studies by using state-of-the-art DL techniques:

- (1) A poor generalization of spectral indices in heterogeneous regions.
- (2) Additional pre-fire image acquisitions for bitemporal indices.
- (3) Omission errors caused by uni-temporal indices.
- (4) Lack of a more detailed quantitative comparison between different kinds of algorithms.
- (5) Lack of a further investigation about the cross-sensor dataset.

### 3. Study Areas and Data Characteristics

#### 3.1. Study Areas

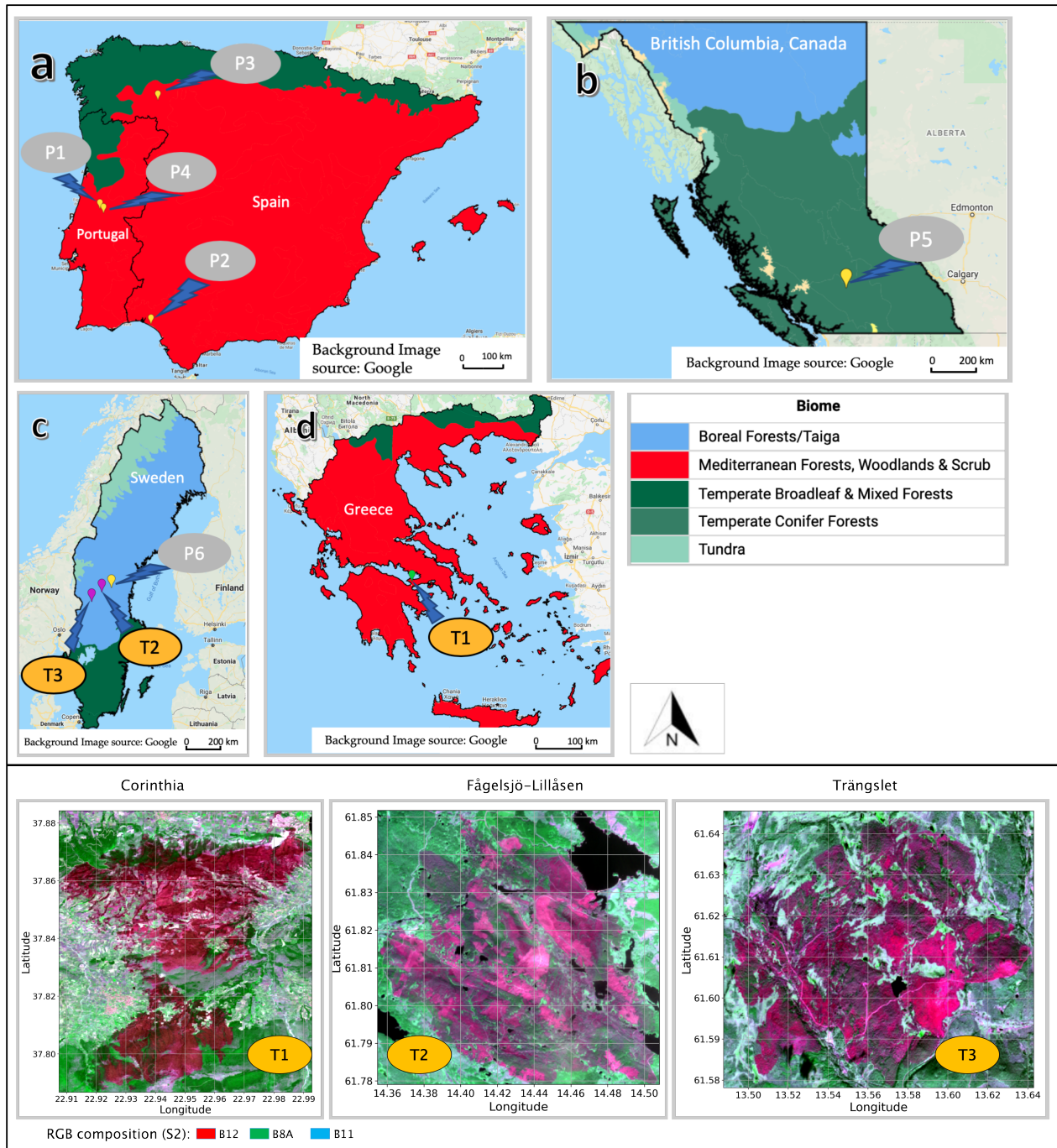
Figure 1 provides an overview of the training and testing study sites highlighting the different biome types. The Mediterranean forests, woodlands, and scrubs dominates the vegetation in Portugal, Spain (a) and Greece (d). In Figure 1a, four large wildfire events—two in Portugal (P1 in Leiria District and P4 in Castelo Branco) and two in Spain (P2 in Donana and P3 in Encinedo)—are selected as training sites for models (see Table 1 for details). Historically, the Mediterranean Basin is extremely vulnerable to wildfires, frequently happening in the summer period (between June and September). For instance, the Castelo Branco fire covers approximately 9646 ha near two villages with 1471 people affected among the total of 15,596 living nearby.

**Table 1.** Description of the study sites and the training (P1–P6) and testing (T1–T3) sites. S2: Sentinel-2; L8: Landsat-8. REF: reference. POST Date: Post-fire acquisition dates. res.: Resolution.

	Country	Sites	Event Date	End Date	Burned Area (ha)	REF Date	POST Date (S2)	POST Date (L8)	Width × Height in 20 m Res.
P1	Portugal	Leiria District	2017-06-17	2017-06-24	45,135	2017-06-24	2017-07-04	*	2240 × 2022
P2	Spain	Donana	2017-06-24	2017-06-30	8446	2017-08-08	2017-07-01	*	1120 × 1037
P3	Spain	Encinedo	2017-08-22	2017-09-01	9934	2017-10-10	2017-09-02	*	1247 × 550
P4	Portugal	Castelo Branco	2019-07-20	2019-07-23	9646	2019-07-24	2019-08-03	*	1204 × 849
P5	Canada	Elephant Hill	2017-07-06	2017-09-20	191,865	2018-05-14	2017-10-03	*	3839 × 4933
P6	Sweden	Enskogen	2018-07-14	2018-07-18	8980	2018-08-07	2018-10-07	*	816 × 861
T1	Greece	Corinthia	2020-07-22	2020-07-26	3282	2020-07-28	2020-07-29	2020-08-23	476 × 544
T2	Sweden	Fågelsjö-Lillåsen	2018-07-13	2018-07-27	3906	2018-07-27	2018-09-02	2018-10-16	409 × 409
T3	Sweden	Trångslet	2018-07-12	2018-07-27	3136	2018-07-27	2018-10-05	2018-10-07	421 × 385

In addition, one large Elephant Hill fire (P5 in Figure 1b) in BC, Canada, is added as a training site to the temperate conifer forests biome. Started on 6 July 2017, the Elephant Hill fire was the largest wildfire in BC during the record-breaking wildfire season in 2017, causing nearly 191,865 ha of land to be burned (see Table 1 for details). The other boreal conifer region of interest is central Sweden in Figure 1c. Due to the extreme and long-lasting drought in forests and windy weather in summer 2018, Sweden suffered many large wildfires, with more than 25,000 ha burned and almost 3 million cubic meters of wood destroyed. One large fire in Enskogen is added for training (P6 in Figure 1c). Three

fire events are selected for independent testing, including two small boreal forests fires in Fågelsjö-Lillåsen and Trängslet (T2 and T3 in Figure 1c) and one wildfire near Corinthia on the Peloponnesian peninsula in Greece (T1 in Figure 1d).



**Figure 1.** Political map (a–d) showing the biome types [63] in regions or countries such as Portugal, Spain, British Columbia in Canada, Sweden, and Greece. Fire events used for training of the burned area mapping methods are marked with points from P1 to P6, while events for independent testing are marked with points of T1–T3. These 3 testing study areas are displayed in false color composites of Sentinel-2 B12, B8A, and B11 bands at the bottom, respectively; burned areas appear in dark red. (T1) Study area in Corinthia, Greece; (T2) study area in Fågelsjö-Lillåsen, Sweden; and (T3) study area near Trängslet, Sweden.

### 3.2. Data Characteristics

#### 3.2.1. Sentinel-2 and Landsat-8 Data Collection

The European Space Agency (ESA) launched twin satellites Sentinel-2A and Sentinel-2B in June 2015 and in March 2017, respectively. They carry the Multispectral Imager (MSI) that gives continuity to high-resolution optical observations over global terrestrial surfaces. These sensors sample 13 spectral bands of pixel size ranging from 10 to 60 meters: Blue (B2), Green (B3), Red (B4), and NIR (B8) at 10 m; red edge bands (B5–B7), narrow NIR (B8A), and SWIRs (B11 and B12) at 20 m; and coastal aerosol (B1), water vapor (B9), and cirrus band (B10) at 60 m spatial resolution. Launched on 11 February 2013, Landsat-8 provides a two-sensor payload: the Operational Land Imager (OLI) and the Thermal Infrared Sensor with a 16-day revisit time [64]. Spectral channels consist of coastal aerosol (B1), Blue (B2), Green (B3), Red (B4), NIR (B5), SWIR1 (B6), SWIR2 (B7), and cirrus (B9) with a 30 m spatial resolution, two thermal infrared wavelengths at 100 m resolution, and a panchromatic band (B8) in 15 m resolution.

Although Sentinel-2 MSI has several similar spectral wavelengths with Landsat 8 OLI, two Sentinel-2 satellites can provide higher temporal resolution together (5 days vs. 16 days) and higher spatial resolution (10/20 vs. 30 m) than Landsat-8 OLI. The multispectral scenes of Sentinel-2 MSI L1C TOA and Landsat-8 OLI TOA in this study were downloaded from GEE datasets (i.e., Image Collection “COPERNICUS/S2” and “LANDSAT/LC08/C01/T1\_TOA”, respectively). The data collection process included date filtering, bounds filtering to the regions of interest, and subsets clipping. All images were resampled into 20 m using the nearest neighbor resampling method. The acquisition dates and image size of post-event Sentinel-2 and Landsat-8 images used in this study are listed in Table 1.

#### 3.2.2. Reference Data

Copernicus Emergency Management Service (EMS) provides us with delineation products and grading products (<https://emergency.copernicus.eu/mapping/list-of-activations-rapid>, accessed on 9 April 2021) as precise annotation masks for corresponding training and testing images. These products have been used as reference data in burned area detection or burn severity estimation in previous studies [16,18,23,33,57,65,66]. Most EMS delineation or grading maps are derived from VHR post-fire images using WorldView-2 and/or SPOT6/7 with 1.5–2.0 m resolution under approximately 0% cloud coverage. For instance, the data source of post-fire images used for test sites is SPOT6/7 with 1.5 m ground sampling distance as presented in Table 2. Based on these VHR images, a semi-automatic strategy helps deliver the EMS thematic layer through visual interpretation (<https://emergency.copernicus.eu/mapping/ems/detection-methods>, accessed on 9 April 2021). This approach is common for the analysis of forest fires based on optical satellite data to identify and classify the burned areas in product delivery. Although we aim to map the extent of burned areas in this study, the burn severity provided by EMS grading maps also helps us analyze the characteristics of spatial heterogeneity of the fire scar.

The burned area masks of fire events in Spain and Portugal are directly derived from EMS products whose Activation ID numbers are EMSR207 (Leiria District), EMSR209 (Donana), EMSR227 (Encinedo), and EMSR372 (Castelo Branco), respectively. Moreover, the EMSR447, EMSR298\_05, and EMSR298\_03 products provide the reference data for the test sites in Corinthia, Fågelsjö-Lillåsen, and Trångslet, respectively. Differently, for the Elephant Hill and Enskogen fires, their dNBR images, calculated from cloud-free pre-fire and post-fire Sentinel-2 images, were empirically thresholded to elaborate the precise ground truth mask within the official perimeters from the Copernicus EMS (EMSR298\_01) and BC Wildfire Service (K20637) [67] as de Bem et al. [56] did. Furthermore, we manually refined all the burned area annotations based on visual analysis of VHR post-event optical images (i.e., Google Earth Map).

Some reference dates in Table 1 are later than those of the post-fire image acquisition dates (e.g., one month for the Spanish fire and 7 months for the Canadian fire). For Sweden,



the EMS reference dates are earlier than post-fire image acquisition dates (e.g., 2 months for Enskogen, 1 month for Fågelsjö-Lillåsen, and 2 months for Trängslet). One reason for this is the different data sources of the reference images and post-fire images. The acquisition of Sentinel-2 or Landsat-8 testing images depends on the cloud coverage, while the delivery of reference products mainly relies on the available VHR images and interpretation time. On the other hand, the reference data acquisition would be earlier or later depending on the official emergency applications (rapid activation or grading severity mapping). The Elephant Hill Fire has the official perimeters after the fire event but it would not affect the quality of burned area perimeters due to the slow regrowth of boreal forests. The time gap between multispectral images and reference data has little influence on the detection of burned area extent in this study, especially for boreal forest fires in Sweden and Canada. However, it would be important to use the images with as little time gap as possible to ensure the reliability of the burn severity accuracy assessment.

**Table 2.** Test sites characteristics. PRE IMG.: Pre-fire Image. POST IMG.: Post-fire Image. GSD: Ground Sampling Distance. n.d.: NOT Detected.

Event	EMSR ID	Tran. (km)	Pop. (No.)	Ele. (m)	PRE IMG. Source (GSD)	POST IMG. Source (GSD)
Corinthia	447	54.4	1501	36.9 to 718.3	SPOT6 (1.5 m)	SPOT7 (1.5 m)
Fågelsjö-Lillåsen	298_05	44.3	n.d.	435.0 to 597.9	Sentinel 2A/B (10 m)	SPOT6/7 (1.5 m)
Trängslet	298_03	10.1	n.d.	526.9 to 698.7	Sentinel 2A/B (10 m)	SPOT6/7 (1.5 m)
Event	Res./Ind. (ha)	Forests (ha)	Het. Agric. (ha)	Perm. Crops (ha)	Shrubs/Herb. (ha)	In. Wetlands (ha)
Corinthia	14.3	1373.6	329.7	647.5	920.4	n.d.
Fågelsjö-Lillåsen	n.d.	2661.0	n.d.	n.d.	985.8	240.1
Trängslet	n.d.	1301.2	n.d.	n.d.	1085.3	749.8

### 3.2.3. Test Sites Characteristics

In EMS reports, land use is specified in the grading maps, which refers to the official Copernicus EMS definition (<https://emergency.copernicus.eu/mapping/ems/domains>, accessed on 9 April 2021) according to Corine Land Cover (CLC) [68]. The first subtable in Table 2 reports the transportation affected by fire in length (km) including primary road, secondary road, local road and cart track, the number of inhabitants (population) affected by fire, and the digital elevation range (meter). The second subtable reports the land use affected by the fire in ha, for residential/industrial areas (residential buildings, non-residential farm buildings, and other buildings not elsewhere classified), forests, heterogeneous agricultural areas, permanent croplands, shrubs or herbaceous vegetation areas, and inland wetlands areas.

The Corinthia fire consists of the various heterogeneous landscapes with crops, agricultural areas, forests, and shrubs, while the two Swedish fire sites mainly consist of forests, shrubs, and wetlands. Except for the land use, other artificial impacts and locations of residence are different in these test sites. For example, the Corinthia fire affected more than 1500 people nearby and damaged 14.3 ha of residential or industrial areas, while the Swedish fire did no damage to the residence. Interestingly, the fire in Corinthia has various elevations and hill shade which might result in fluctuating mountain slope. In contrast, Swedish fires are located in relatively flat regions with rolling hills. From the EMSR447 grading map of the Corinthia fire, over 98.9% burned areas are destroyed or damaged, while the Fågelsjö-Lillåsen fire has over 10% burned areas (395.5 ha) that are possibly damaged (i.e., with low burn severity).

### 3.2.4. Spectral Feature Selection

Spectral bands play different roles in burned area discrimination. A previous study concluded that Sentinel-2 NIR (B8 and B8A), red-edge (B5–B7), and SWIR bands (B11 and B12) are most sensitive to the change in spectral reflectance caused by fire [17]. Fire



scar would cause a decrease in NIR and might result in an increase in SWIRs depending on the ecosystem. Therefore, SWIR and NIR bands are widely used for index like NBR and NBR2 in wildfire science [69]. For Sentinel-2 MSI channels, B8A at 20 m resolution is a more suitable NIR band for vegetation monitoring applications than B8 due to a narrower spectral width [70]. Furthermore, Sentinel-2 MSI B8A rather than B8 has similar characteristics to Landsat-8 OLI NIR band (B5) [71]. To facilitate the potential transferring to Landsat-8 data, the combination of bands B8A, B11, and B12 becomes the most suitable input channels for model training.

To further comprehensively support our assumption, the feature selection approaches including AdaBoost [72] and Light Gradient Boosting Machine (LightGBM) [73] were performed on the Castelo Branco Fire dataset for ranking the spectral feature importance (cf. Appendix A). Three out of 10 bands of Sentinel-2 stood out, as expected, namely, B12, B11, and B8A, as the three most important features on the burned area target. Knopp et al. [57] assessed the accuracy for different band combinations of Sentinel-2 data as input for U-Net to support the channel selection. It was demonstrated that only blue, green, and red channels as input data result in worse outcomes. If B8 and two SWIRs bands are additionally involved, the kappa coefficient of the burned area mapping would increase from 0.75 to 0.90 in comparison with three visible channels as input.

Although Sentinel-2 imagery contains 10 bands, supporting the design of distinct architectures, our DL models take as input a 3-channel (in a color composite: B12, B11, and B8A) image patch, where we replace the B8 in [57] with B8A as one of input channels. One benefit of choosing the most representative three bands is that more channel input will increase the computing operations dramatically and limit the future application in some DL models where only three channels are restricted. A three-channel input can be visualized easily and saved as standard image format (e.g., PNG or JPEG) for the web-based application of graphical user interface. The other advantage is that these three bands match the corresponding B7, B6, and B5 of Landsat 8 imagery in the spectrum with high transferability, allowing further cross-sensor application. Abundant Sentinel-2 data can be used to train the DL-based model which can be transferred directly to Landsat-8 data so that we do not need to train a separate model for the Landsat-8 image.

#### 4. Methods

Figure 2 provides an overview of the experimental design of this research. The feature maps extracted from representative layers are visualized in a proper way for each DL model. To evaluate the performance of DL-based models in burned area mapping, several typical supervised ML methods and traditional NBR-based thresholding approaches are carried out as comparisons. These comparison experiments contribute to the selection of the method suitable for detection of burned areas for specific landscapes.

As shown in Figure 3, a fully automatic workflow for burned area mapping using uni-temporal Sentinel-2 and Landsat-8 multispectral data is demonstrated in comparison. Training data from Sentinel-2 are used to train the DL model (U-Net, HRNet, Fast-SCNN, and DeepLabv3+) after data augmentation. Trained DL models are then used to inference the testing data in an end-to-end processing scheme. On the other hand, feature vectors of spectral characteristics are sorted out to train the ML estimators (i.e., LightGBM, RF, and k-Nearest Neighbors (KNN)) separately to get the predictive models for future estimation.

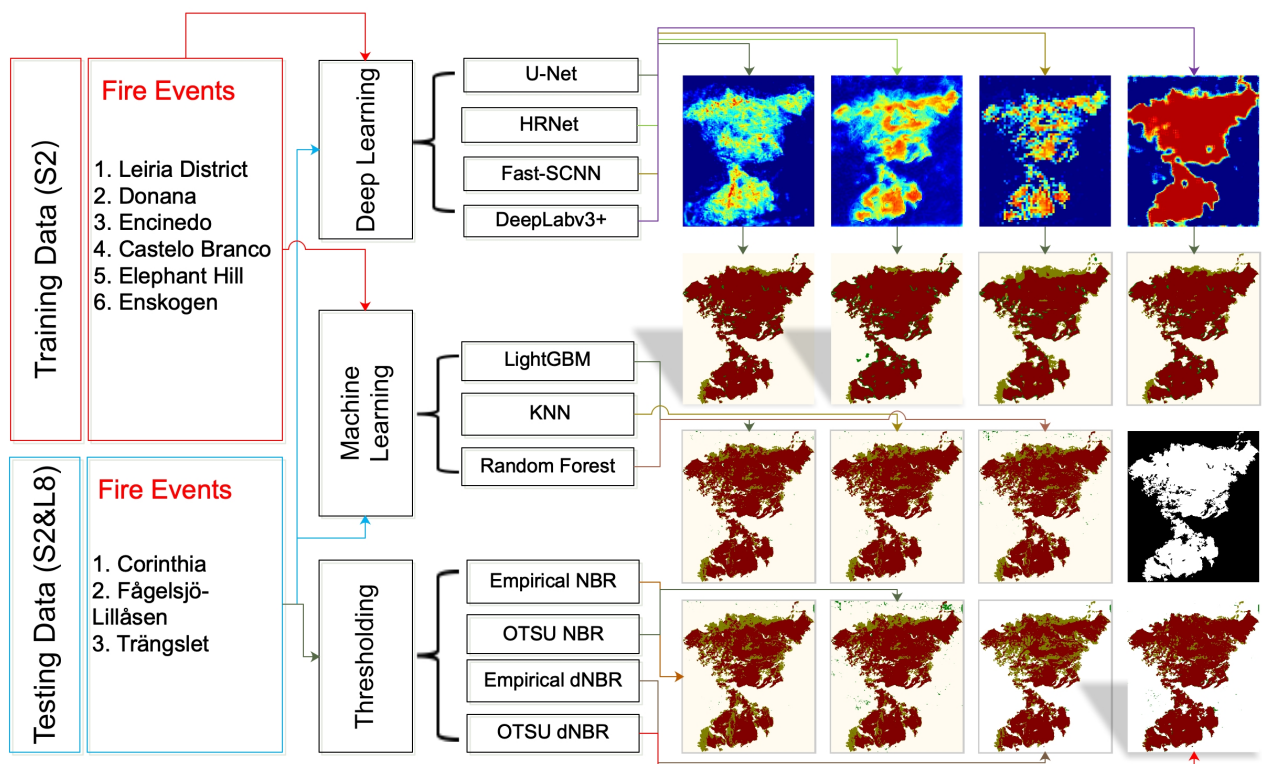


Figure 2. Overview of the experimental design.

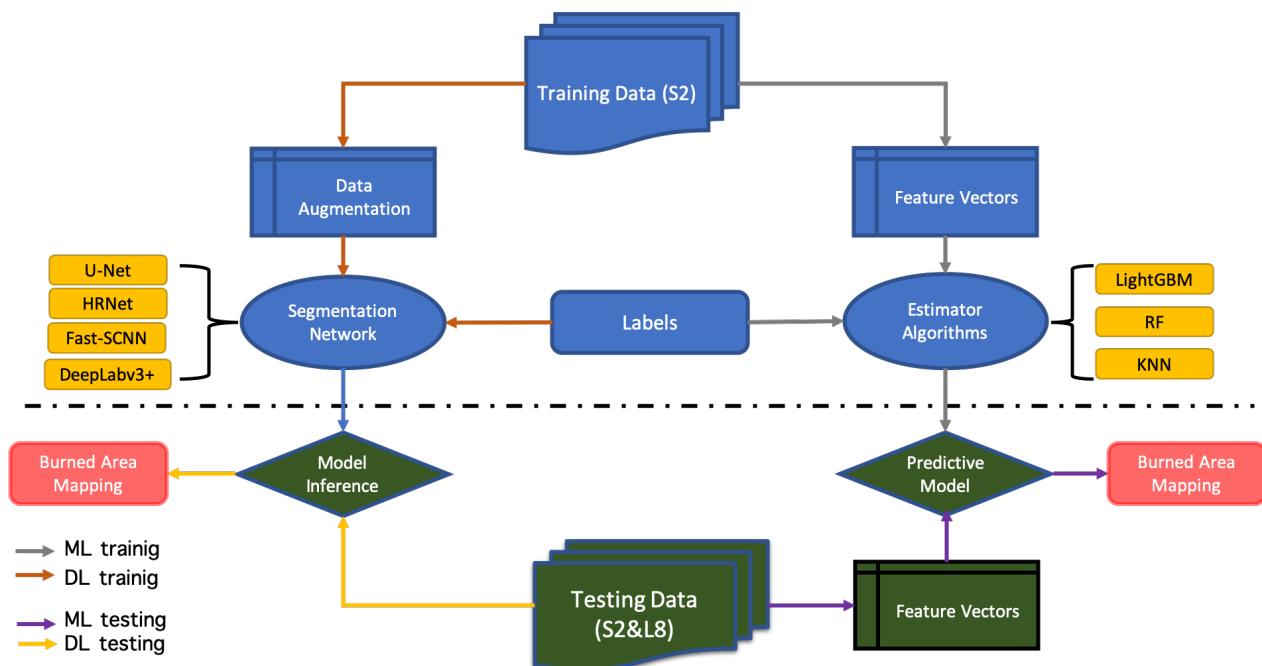


Figure 3. Basic workflow to automatically map burned areas with available multispectral data i.e., Sentinel-2 (S2) and Landsat-8 (L8) using deep learning (DL) and machine learning (ML) methods, respectively.

#### 4.1. Threshold-Based Approaches

NBR has been widely used in fire-related research in a uni-temporal way as in Equation (1):

$$NBR = \frac{NIR - SWIR}{NIR + SWIR} \quad (1)$$

where *NIR* and *SWIR* are the reflectance of B8A and B12 for Sentinel-2 data (in accordance with B5 and B7 for Landsat-8 data), respectively. One drawback using NBR-based methods is that the common false positives like water body have to be removed manually. The further bitemporal difference (i.e., dNBR) can be computed to highlight the burned areas as Equation (2).

$$dNBR = NBR_{pre} - NBR_{post} \quad (2)$$

where *pre* and *post* denote the pre-fire and post-fire images, respectively. A dNBR index between  $-0.1$  and  $0.1$  indicates that few changes over time interval are accounted for [74]. If there exists fire disturbance between the *pre* and *post* images, dNBR might have a bimodal distribution which usually needs a further thresholding process (in an empirical or automatic way) to filter out burned areas pixels. The empirical thresholding approach is used to generate reference data for training images of Elephant Hill and Enskogen as described in Section 3.2.2.

Nobuyuki Otsu [22] proposed a classical threshold selection method for the gray image in an automatic way to define the threshold in a bimodal distribution of NBR, which has been used for burned area mapping [75,76]. In this study, the performance of the OTSU method is compared with ML-based and DL-based methods. Specifically, the pixel values less than the NBR thresholds (empirical or OTSU-based) are clustered as burned area pixels.

#### 4.2. ML-Based Approaches

To select suitable ML methods, the commonly used ML algorithms were compared for the classification of burned areas based on the PyCaret implementation [77] with standard parameters; 100,000 pixel data sampled from training imageries with feature vectors of B12, B11, and B8A were fed as estimator inputs. The ratio between the pixel amounts of the burned areas and unburned areas is around two-thirds in the whole training data. Seventy percent of samples were used for training based on a 10-fold cross-validation approach [78], while 30% were used for validation to assess the performance of the various ML methods. The comprehensive results can be found in Table A1. LightGBM, KNN, and RF performed better than others, thus were selected (Note: the top two methods are both boosting-based approaches so the first one is selected.).

The LightGBM is a new gradient boosting framework that employs tree-based learning algorithms. Different from traditional boosting tools applied to fire-related work (e.g., XGBoost [79], AdaBoost [32], and LogitBoost [32]) that use pre-sort-based algorithms, LightGBM is a histogram-based algorithm, which buckets continuous feature (attribute) values into discrete bins. This speeds up training and reduces memory usage. We consider the Gradient Boosting Decision Tree boosting method with 170 iterations, 60 leaves, and 0.1 as a learning rate.

KNN is a data classification algorithm based on the premise that similar data exist in close proximity to each other according to some metric [80]. It has been used for burned area mapping in France [81], fire occurrence prediction [82], and wildfire damage assessment [83]. The parameters used in the experiments are 49 neighbors, Minkowski metric, and uniform weight.

RF is one of the most popular classifiers within the remote sensing community, vigorously handling dimensionality and multicollinearity in high data [84,85]. Ramo and Chuvieco [86] developed a global burned area mapping algorithm with MODIS data based on the RF classifier. Ramo et al. [87] evaluated the ability of four algorithms, including RF, SVM, Neural Networks, and a decision tree algorithm, for classifying burned areas at a global scale using MODIS data. It was shown that RF offered the best performance. The RF method consists of an ensemble of individual decision trees and is more robust than single classifiers. Each tree employs a set of decision rules to spit out a class prediction. The ensemble votes are used to determine the final model's prediction. A larger number of trees can help the generalization error converge [88]. In this study, we consider the

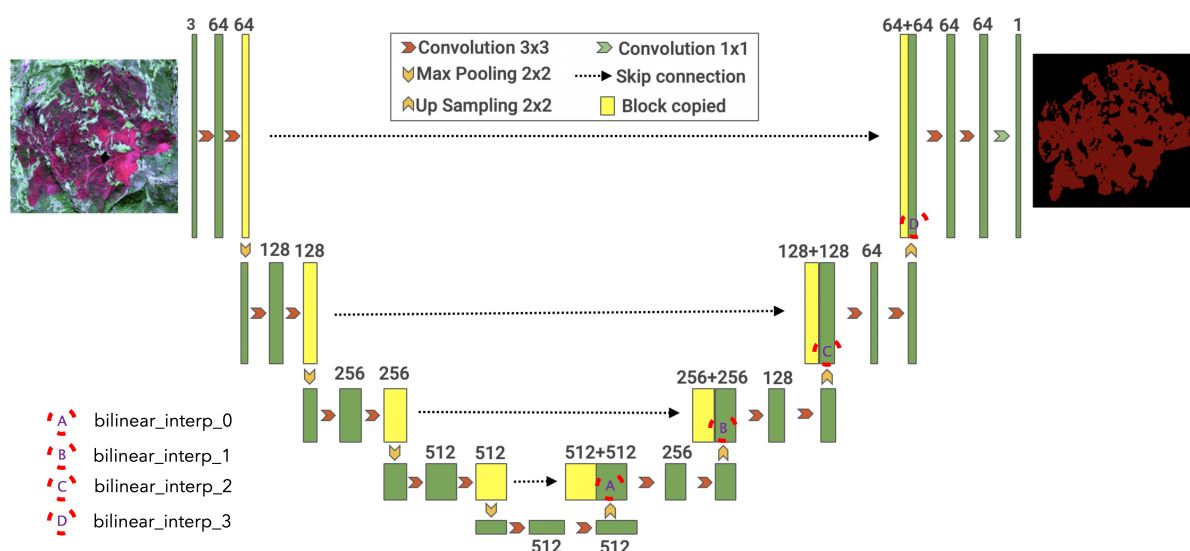
maximum depth as 5, 150 trees in the forest, and 5 as the minimum number of instances in leaves.

### 4.3. DL-Based Approaches

Semantic segmentation can be categorized into several classes according to the common concept that underlays their architectures [37]. The architecture would usually determine the network's performance. In general, combining the low-resolution feature with larger receptive fields and the high-resolution feature with smaller receptive fields can help extract contextual information. To compare the ability of different DL architectures for burned area mapping, we evaluate four typical models among popular categories: HRNet, U-Net, Fast-SCNN, and DeepLabv3+.

#### 4.3.1. U-Net

Upsampling/deconvolution-based methods are dominant approaches to extract feature maps based on stacked convolutional layers, ReLu layers, and pooling layers [37]. The U-Net connects low-level details and high-level information, achieving better performance than classical ML classification methods [49,59]. Recent papers [56,57] mostly employed U-Net series architecture for burned area mapping with multispectral imagery. The U-Net structure applied here is shown in Figure 4, which is adapted from [59]. The encoder part could extract the features at different scales using five convolutional blocks from input data. In each convolutional block, there exist two  $3 \times 3$  convolutions with ReLu activation layers. The output is followed by batch normalization and a max-pooling operation to downsample the feature maps. Therefore, the size of feature maps is divided by 4 after the convolutional block, while the number of feature channels will be doubled. Through the other five blocks in the decoder part, these feature maps are upsampled to the input size. Each block has  $2 \times 2$  transpose convolution, a concatenation of feature maps from the corresponding encoder part, and two  $3 \times 3$  convolutions with ReLu activation and batch normalization, with the feature maps in A–D after bilinear interpolation (*bilinear\_interp*), respectively. The final layer additionally has a  $1 \times 1$  convolution to calculate the probability for burned area prediction.



**Figure 4.** The U-Net structure used in this study, adapted from the original architecture in [59].

#### 4.3.2. HRNet

As a typical enhancement of feature-based methods, state-of-the-art HRNet maintains stronger high-resolution representations through high-to-low resolution parallel convolu-

tions [62]. According to the semantic segmentation results on PASCAL-context [89] and LIP [90], HRNet outperforms DeepLabv3+ and U-Net++ with lighter computation cost and fewer parameters [62].

There are four stages in HRNet to connect the high-to-low resolution convolutions in parallel, whose final Stage 4 is mainly illustrated in Figure 5. It comprises a horizontal multi-resolution parallel convolution and a crossing multi-resolution fusion. The multi-resolution parallel convolution is adapted from the group convolution that conducts a regular convolution over a subset clipped from input channels in various spatial resolutions separately. The multi-resolution fusion resembles the fully connected multi-branch convolutions. The input and output channels are both divided into several subsets along with the stages deepening. The four output representations are mixed from four resolutions using a  $1 \times 1$  convolution followed by a classifier with softmax loss to obtain the segmentation maps. Furthermore, these maps are upsampled into the size same as the original input image using bilinear interpolation. Four outputs representations from low to high resolutions are named *bilinear\_interp\_33*, *bilinear\_interp\_32*, *bilinear\_interp\_31*, and *relu\_152*, respectively.

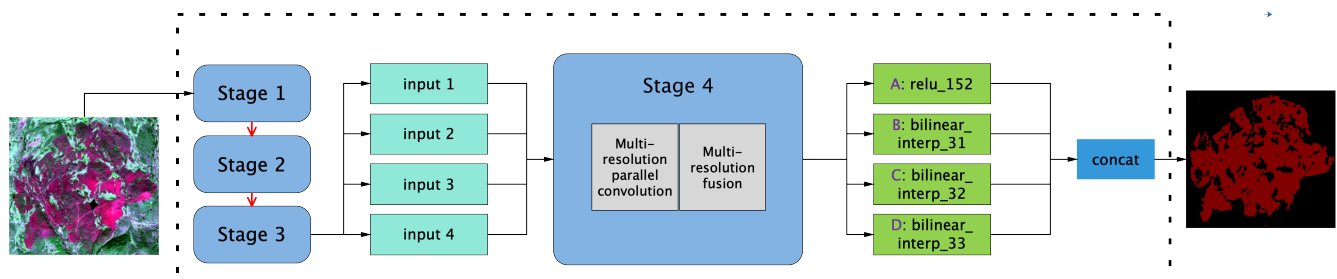


Figure 5. The HRNet framework, adapted from the original architecture in [62].

#### 4.3.3. Fast-SCNN

In addition, one of the feature encoder-based methods is Fast-SCNN [60], a real-time semantic segmentation model, which extracts multiple low-level resolution features simultaneously, making it suitable on embedded devices offline. As Figure 6 demonstrated, Fast-SCNN has four parts (details can be found in [60]):

- (1) A learning to downsample module with a standard convolutional layer (Conv2D) and two depthwise separable convolutional layers (DSConv). The output of feature maps is named *relu\_4* in A.
- (2) A coarse global feature extractor to capture the contextual information for segmentation using a bottleneck block that employs the depthwise separable convolution. It can reduce the number of parameters to train floating-point operations. The end of the extractor is a pyramid pooling module that aims to aggregate the different-region-based context. The context information from the extractor part is given in *relu\_6* before the fusion operation.
- (3) A feature fusion module with simple addition of the features with high-level and low-level representations (*relu\_7* in C).
- (4) A standard classifier consists of two DSConv (*relu\_11* in D), one Conv2D to boost the accuracy, and a softmax layer to get the segmentation results.

#### 4.3.4. DeepLabv3+

The last one is the DeepLabv3+ model, an increased resolution feature-based method [61]. A fully connected Conditional Random Field on the final layer of DeepLabv3+ improved localization performance both quantitatively and qualitatively. We mainly introduce the DeepLabv3+ framework in Figure 7 (more details can be referred to in [61]). Multi-scale contextual information is extracted by atrous convolution at an arbitrary resolution in the encoder module. In the decoder module, first, the features from the encoder



atrous convolution ( $relu\_8$  in A) are bilinearly upsampled by 4 and then concatenated with low-level features ( $relu\_9$  in B) from the backbone with the same resolution after the  $1 \times 1$  convolution to obtain the multi-level feature fusion ( $relu\_10$  in C). Further, a few  $3 \times 3$  convolutions are applied to refine the features ( $relu\_12$  in D). They are upsampled by a factor of 4 in a simple bilinear way.

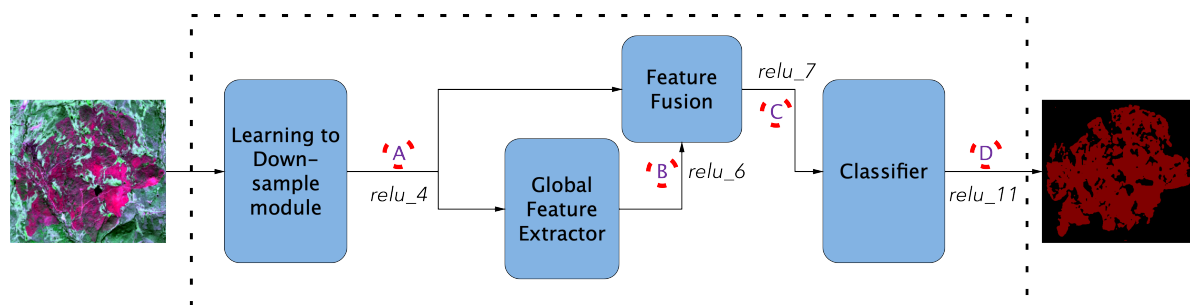


Figure 6. The Fast-SCNN framework, adapted from the original architecture in [60].

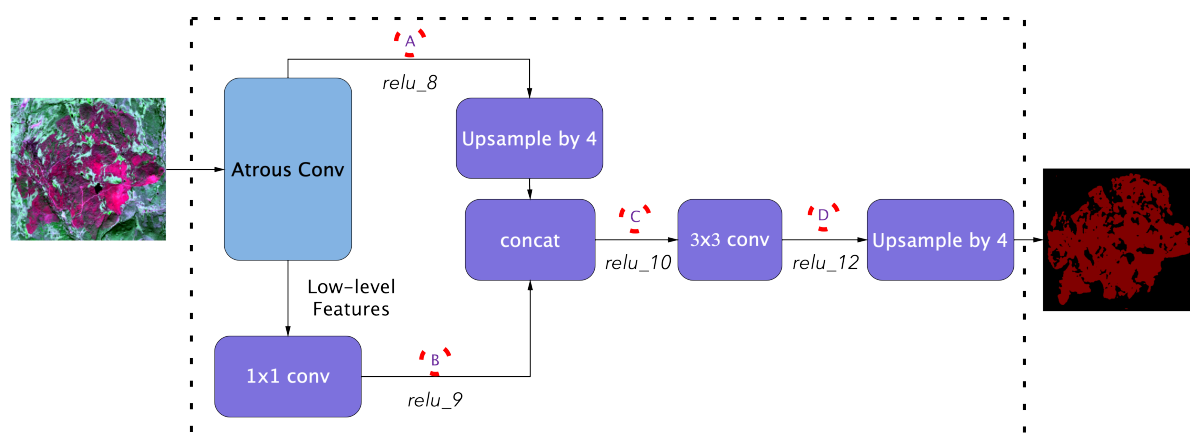


Figure 7. The DeepLabv3+ framework, adapted from the original architecture in [61].

#### 4.3.5. Data Augmentation

Training images were cropped into patch tiles of  $256 \times 256$  pixels randomly to raise the volume of this dataset and to reduce classification problems around edges. There is a total of 1837 training patches and 197 validation patches. Data augmentation was employed to enhance the data set and avoid overfitting as listed in Table 3. We adopted step-scaling resize between 0.7 and 1.2 with step 0.1; flip operation with a probability of 0.5; mirror operation with a probability of 0.5; and additional operations including rotation, area crop, aspect, and color jitter (brightness, saturation, and contrast with a probability of 0.2) on the input images. These augmented images were standardly normalized between 0 and 1 by removing the mean and scaling to unit variance.

**Table 3.** Data augmentation operations.

Methods	Parameters	Define
SCALING	min scale factor = 0.7; max scale factor = 1.2; step = 0.1	Resize image with a random scale within maximum and minimum in a fixed step.
FLIP	ratio = 0.5	Flip (horizontally and vertically) the input data randomly with a given probability (i.e., ratio).
ROTATION	0°–75°	Rotate the image by angle between 0° and 75°
AREA	min = 0.2	A crop of random size of the original size.
ASPECT	min = 0.2	A random aspect ratio of the original aspect ratio.
COLOR JITTER	brightness jitter ratio = 0.5; saturation jitter ratio = 0.5; contrast jitter ratio = 0.5	Randomly change the brightness, contrast, saturation of an image with a given probability.

#### 4.4. Accuracy Assessment

For model validation and accuracy assessment, overall accuracy (OA) and the mean intersection over union (mIoU) of two classes (unburned pixels and burned pixels) are used to measure the comprehensive performance of various methods, see Equations (3) and (4) ( $fp$  = false positive,  $tp$  = true positive,  $tn$  = true negative,  $fn$  = false negative in confusion matrix). The Cohens' kappa coefficient [91], commission errors (Ce), and omission errors (Oe) are used to assess the performance of testing process, see Equations (5) and (6).

$$OA = \frac{tp + tn}{tp + tn + fp + fn} \quad (3)$$

$$IoU = \frac{tp}{tp + fn + fp} \quad (4)$$

$$Ce = \frac{fp}{tp + fp} \quad (5)$$

$$Oe = \frac{fn}{tp + fn} \quad (6)$$

To compensate for the shortcomings of kappa in map comparison, allocation disagreement (AD) and quantity disagreement (QD) [92] are also evaluated in this study. AD denotes the proportion of the disagreement (i.e., error) associated with pixels in the wrong spatial location, as Equation (5) in [92]. QD means the proportion of the disagreement associated with the amount classified, presented by Equation (3) in [92]. Overall, the sum of OA, AD, and QD should be equal to 1. The way to interpret these two indexes is that if AD is high but QD is low, then this reveals that the area classified is correct but the locations classified are incorrect; conversely, if AD is low but QD is high, then this represents that the locations classified are correct but the amounts of pixels classified are incorrect.

## 5. Results

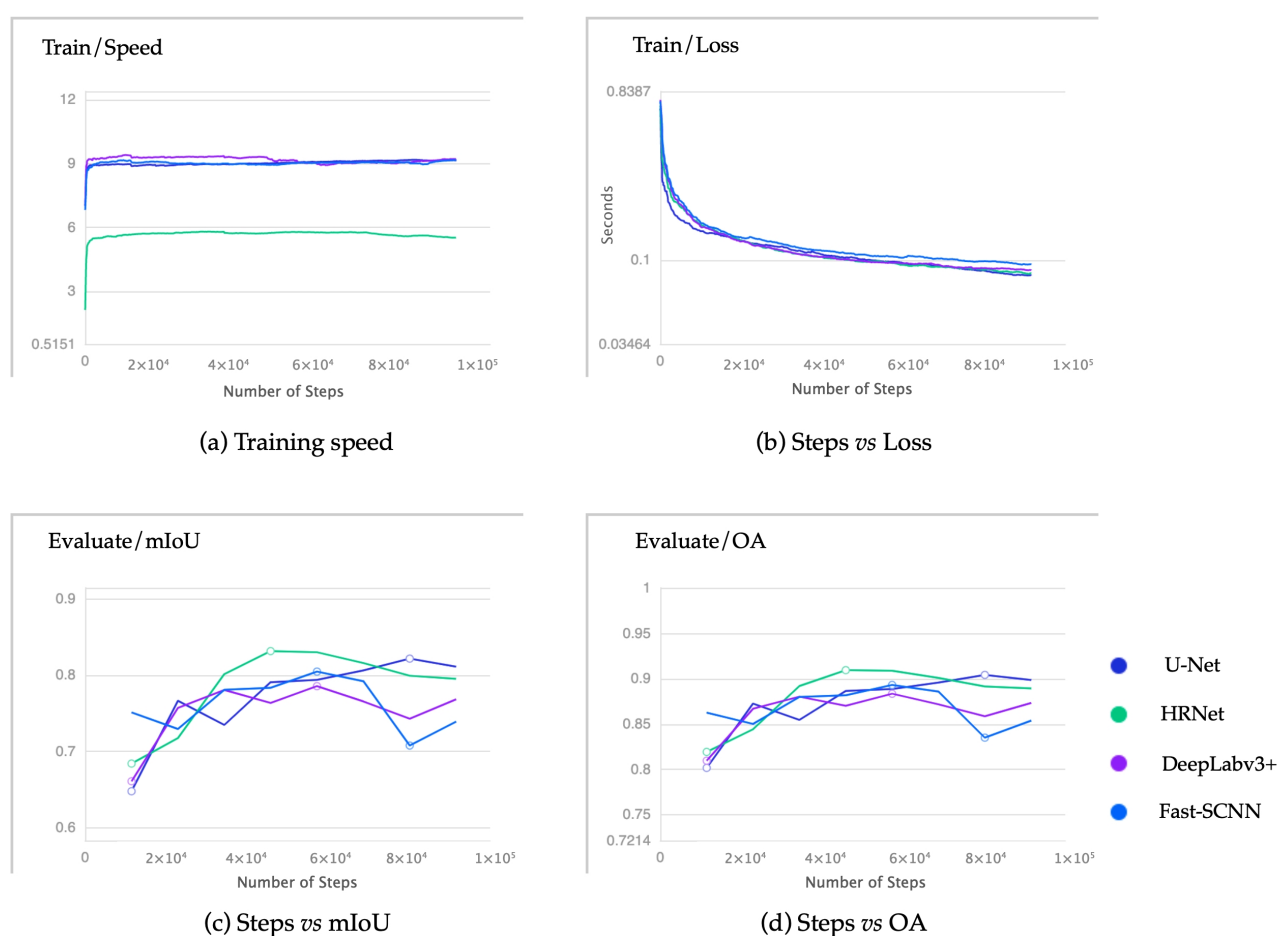
The capabilities of DL-based models for burned area mapping with Sentinel-2 and Landsat-8 data are presented and analyzed in the following sections, subject to the DL

network evaluation and quantitative assessment in comparison with ML-based models and threshold-based approaches.

### 5.1. DL Network Evaluation

#### 5.1.1. Test Results and Analysis

DL models were trained on an Nvidia Tesla V100 GPU using the Adam optimizer [93] and a batch size of eight patches. The training speed is illustrated in Figure 8a. The HRNet has an average lower speed (5.8 step/s) with a total of 4 h and 48 min in the training of 400 epochs, but other DL models take approximately 2 h and 30 min. All models use a polynomial decay learning rate policy with a gradual warmup strategy [94] based on the PaddlePaddle backend DL framework (<https://github.com/PaddlePaddle/Paddle>, accessed on 9 April 2021). The initial learning rate (0.000005) increases to the final learning rate (0.001) after 2000 warmup steps (one of the super parameters) and then decreases to the initial learning rate gradually when the training process ends with a total of 91,600 steps.



**Figure 8.** The training and evaluation process of DL networks used in this study. OA: overall accuracy. mIoU: mean intersection over union.

The loss is a summation of the errors produced by each batch in training or validation sets, which indicates how properly or badly a trained model performs after each iteration of optimization. A weighted sparse softmax cross-entropy loss function [95] is employed to track the performance in the training stage. It particularly picks the weights according to the current labels and applies them as batch weights, and then combines the calculation of the softmax operation and the cross-entropy loss function [96] to provide a more numerically stable gradient. The respective curves with motion average of the training loss are presented in Figure 8b. It can be observed that the training loss stably decreases with an increase in training steps. All networks show convergence towards zero with some minimal jitter

between 0.01 and 0.05. U-Net reaches a low loss value much faster than the other models but seems to cause quick overfitting, whereas the HRNet seems to converge reasonably.

We evaluated the segmentation accuracy of DL models on the test dataset in every 50 epochs. The accuracy curves are shown in Figure 8c,d, where HRNet in green obtains the highest mIoU after the training of 200 epochs rapidly, and the accuracy of U-Net increases gradually and peaks in 350th epochs. The other two models (DeepLabv3+ and Fast-SCNN) represent a fluctuation in curves during training, and they achieve a high value in the 250th epoch. With an increase in the number of training epochs, the accuracy of the DL network decreases after some epochs and even the training loss continues decreasing in Figure 8b, and therefore we selected the trained models (parameters and weights) with the best accuracy rather than the final model to avoid overfitting.

### 5.1.2. Feature Analysis

The test images are input into the trained DL models with an arbitrary size, and then features extracted from a few representative layers can be visualized. The visualization results with Sentinel-2 and Landsat-8 images as input for each test fire event are listed in Figure 9. Four outputs of feature maps are marked with A–D in each DL architecture in Section 4.3. The features selected in each DL model can reflect different levels of semantic information from low to high. The deep convolution layers during the encoder branch gradually extract the low-resolution representations (or high-level features) such as contour features and hotspot distribution caused by burned areas. The decoder branch with upsampling subnetwork aims to recover high-resolution representations (i.e., precise segmentation) such as burned confidence, accurate burned delineation, and unburned areas around burned areas. Note that not all the features can be visualized well or can be understood intuitively, so here we select the most important feature in some typical layers (upsampling layer or ReLU activation layer) in the feature visualization results, which are normal phenomena in the DL field.

Regarding the U-Net in Figure 9, the first *bilinear\_interp\_0* is the name of the feature outputs from the bilinear upsampling layer after the convolution layer with 512 filters at the end of the encoder branch. It perceives a high-level representation that contains general characteristics regarding the semantic information of the input image. Therefore, we observed the heatmap from one typical feature matrix that implies the general distribution of burned areas. Similarly, the *bilinear\_interp\_1* of the decoder branch fuses the low-level spatial features from the encoder branch, and one of its feature maps showed the contour features (i.e., fire delineation) of burned areas for each fire event. One feature channel of the further *bilinear\_interp\_2* indicates the inner burned areas independently. Finally, the last one of the upsampling layer (i.e., *bilinear\_interp\_3*) provides us with reliable visualization results close to the burned area segmentation, even reflecting the burned severity to some extent that needs to be investigated in the future study.

HRNet contains multi-resolution group convolution with four outputs as presented in Figure 5. We visualized these four representations from low resolution to high resolution in Figure 9. The *bilinear\_interp\_33* denotes the output features of low-resolution subsets. It learns the low-resolution representations successfully with the raw spatial distribution of possible burned areas in Figure 9. Furthermore, one of the feature maps in the *bilinear\_interp\_32* output shows an obvious highlight in the unburned regions, which conversely facilitate the unburned area segmentation. This group convolution could capture mid-level representations. In parallel, high-resolution representations are maintained through the whole process (i.e., *bilinear\_interp\_31* and *relu\_152*). The repeatedly conducting multi-scale fusions across parallel convolutions connects high-to-low resolution convolutions. Therefore, it can be observed that one feature of *relu\_152* can capture burned areas with more high-resolution details, compared to one of the feature maps from the *bilinear\_interp\_31* outputs.



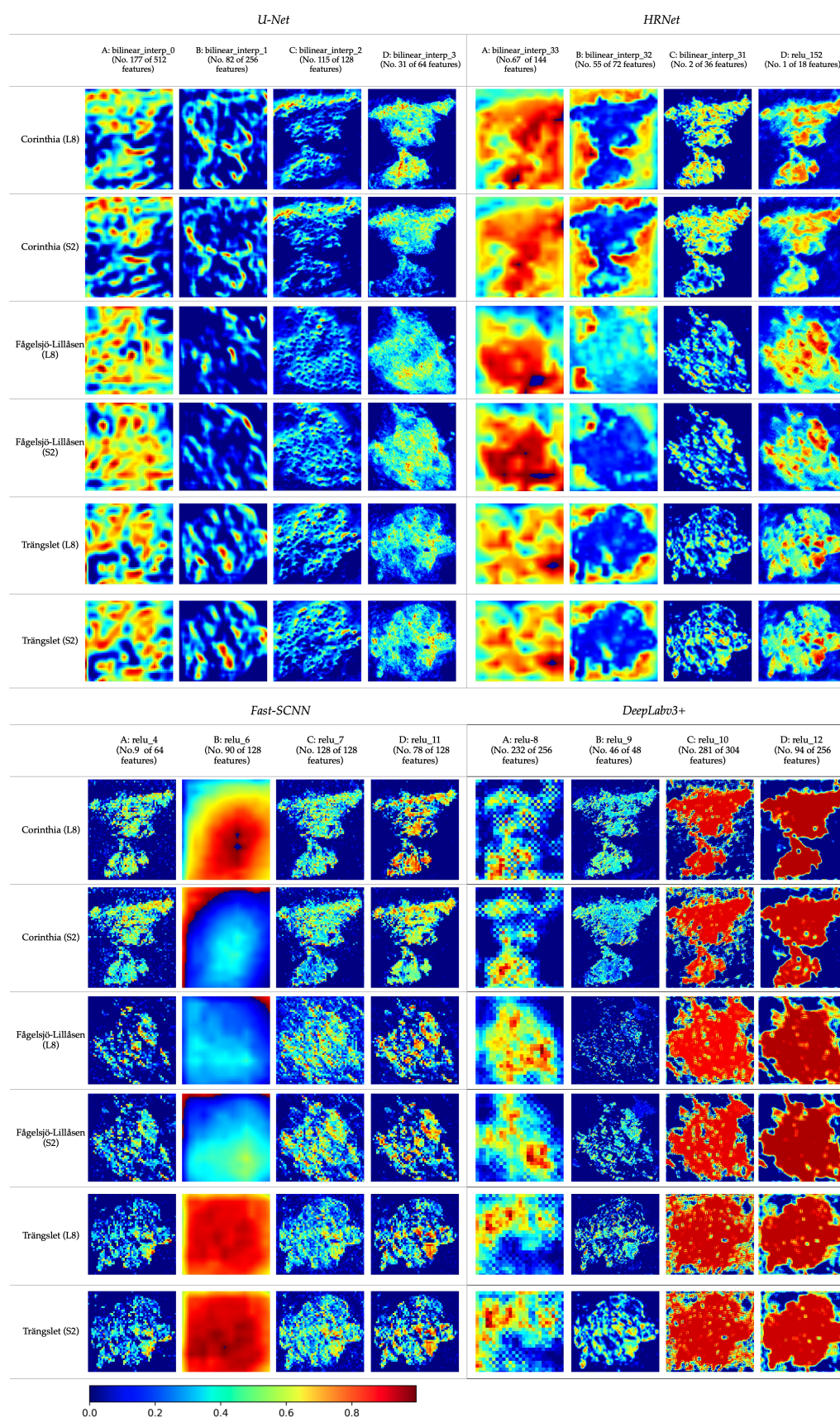


Figure 9. The feature maps of DL models.



Fast-SCNN learns the features through downsampling, and its output in *relu\_4* preserves some semantic information but lacks the contextual relationship, due to the missing of high-level features with large receptive fields. Then, the global feature extractors produce the feature maps in the ReLU activation layer (i.e., *relu\_6*). However, it fails to learn the high-level features, as Figure 9 showed. Therefore, the feature maps in *relu\_7* have similar representations to *relu\_4*. As a real-time semantic segmentation model, Fast-SCNN is more suitable for embedded devices with low memory and power. DeepLabv3+ shows better visualization results in low-level feature extraction from the *relu\_9* activation layer than that of *relu\_4* in the Fast-SCNN model. In addition, its high-level features in the *relu\_8* passing ASPP module also look more reasonable than the *relu\_6* of Fast-SCNN. Therefore, the combination of low-level and high-level features produces contextual information. On the other hand, one of the features in the final ReLU activation layer (*relu\_12*) represents obvious burned areas but with some overestimation.

Overall, HRNet shows more promising visualization results, while U-Net looks at overfitting in the training process taking the feature maps of from *bilinear\_interp\_2* output as an example. Fast-SCNN performs the worst among the four DL models due to poor representations from low resolution to high resolution. DeepLabv3+ has reasonable results, but they are not so precise to some extent.

## 5.2. Burned Area Mapping with Sentinel-2 Data

### 5.2.1. Corinthia Fire

Table 4 summarizes a quantitative comparison of different methods for burned area detection in Corinthia area (Mediterranean forests). DL algorithms (except for Fast-SCNN) generally outperform ML-based models. U-Net results in significantly better metrics than LightGBM: mIoU is 0.04 higher and the kappa coefficient is 0.06 higher.

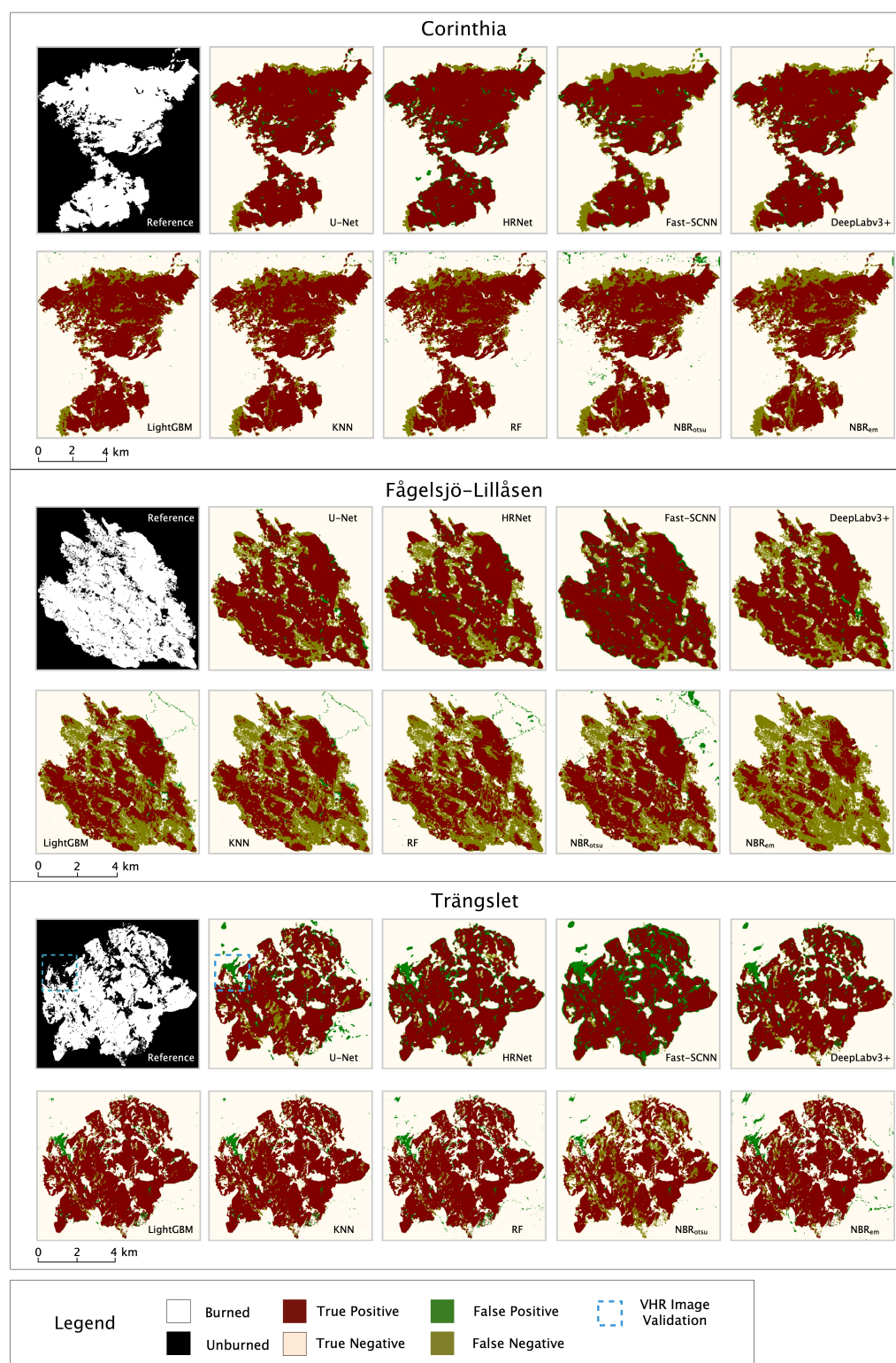
**Table 4.** Testing accuracies of burned area detection for the Corinthia fire with Sentinel-2 data. The lowest omission errors (Oe), commission errors (Ce), allocation disagreement (AD), and quantity disagreement (QD) and the highest mIoU and Kappa in each model group are highlighted in bold.

Corinthia						
Model	Oe (%)	Ce (%)	AD (%)	QD (%)	mIoU	Kappa
U-Net	7.71	<b>4.35</b>	<b>3.38</b>	<b>1.41</b>	<b>0.90</b>	<b>0.90</b>
HRNet	<b>4.43</b>	7.98	3.57	1.55	<b>0.90</b>	0.89
Fast-SCNN	17.05	5.53	3.90	4.90	0.83	0.81
DeepLabv3+	9.56	5.00	3.83	1.93	0.89	0.88
LightGBM	<b>17.00</b>	1.38	0.93	<b>6.37</b>	<b>0.86</b>	<b>0.84</b>
KNN	18.93	<b>1.33</b>	<b>0.88</b>	7.17	0.84	0.83
RF	19.58	1.88	1.24	7.26	0.83	0.82
NBR <sub>otsu</sub>	<b>20.12</b>	3.57	2.38	<b>6.90</b>	<b>0.82</b>	<b>0.80</b>
NBR <sub>em</sub>	29.87	<b>0.79</b>	<b>0.45</b>	11.79	0.76	0.73

HRNet and U-Net display similar results, with higher kappa and mIoU values than Fast-SCNN and DeepLabv3+. On the other hand, the U-Net model provides a good balance between AD and QD, indicating that it classifies large amounts of pixels and locations correctly. The overall performance of U-Net shows high agreement with recent literature in [57], whose U-Net could reach kappa of 0.86 for mapping Mediterranean forest fires in Pantelleria, Italy, with uni-temporal Sentinel-2 data, but resulted in a poor balance between commission errors (23%) and omission errors (3%).

U-Net reconstructs more accurate delineations around the mountain valleys as it learned the features about hillshade in convolution feature maps as Figure 9 showed. The HRNet model produces lower omission errors (false negative) but higher commission errors (false positive) than U-Net, as presented in Figure 10. The highest omission errors

observed by using the Fast-SCNN are mostly related to the underestimation of burned areas in complex terrain surfaces in the northern valleys.



**Figure 10.** Burned area detection results using various methods with Sentinel-2 data.

The pixel-based ML classification methods seem to systematically underestimate the burned area within the fire perimeter as Figure 10 presented, resulting in higher omission error (see Table 4). These phenomena are in agreement with the classification results with uni-temporal Landsat TM imagery on three Mediterranean test sites [28]. Within the three

ML methods in our study, LigthGBM has higher mIoU than RF and KNN. The low kappa is mainly related to high omission errors in the northern regions. Moreover, several false positives outside of the perimeters are mainly related to the croplands, as they show very similar spectral behavior to burned areas. Automated threshold-based methods ( $NBR_{otsu}$ ) yield lower accuracy than all DL and ML methods with huge omission errors due to relatively strict thresholds being applied. In general, NBR-based methods result in false negatives and false positives that occur in the burned class and unburned class, respectively (see Figure 10).

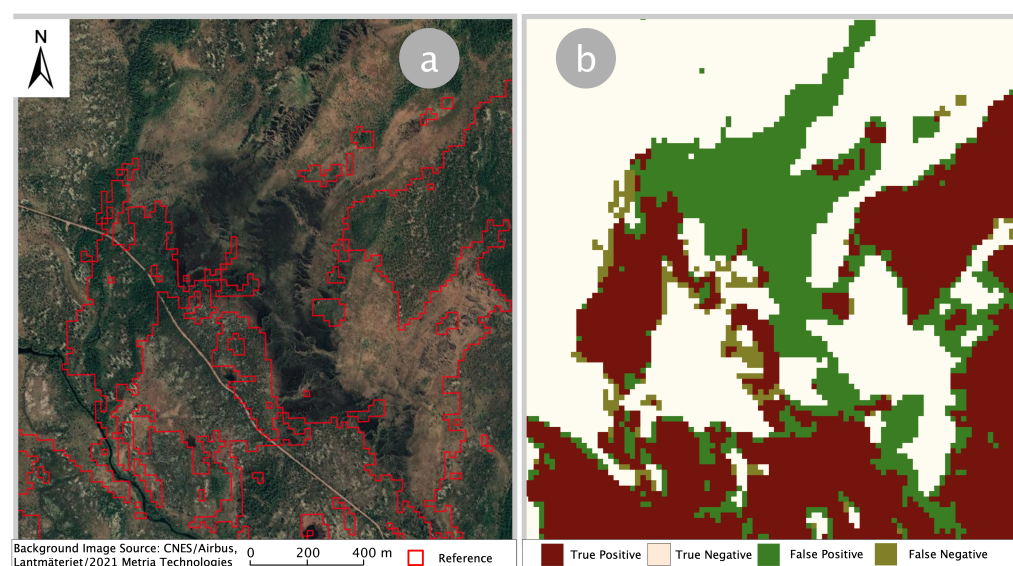
### 5.2.2. Fågelsjö-Lillåsen Fire and Trängslet Fire

Table 5 demonstrates a quantitative algorithm comparison for burned area detection in Fågelsjö-Lillåsen and Trängslet with Boreal conifer trees. Each DL algorithm significantly outperforms the ML methods and NBR-based approaches for the Fågelsjö-Lillåsen fire. Interestingly, Fast-SCNN results in the best metrics: mIoU is 0.81 and the kappa value is 0.79 as it can highlight most low burned severity areas (covered over 10% burned areas from EMSR 298\_05 grading map), which are greatly underestimated by other DL models and ML methods as shown in Figure 10. Fast-SCNN also keeps the lowest omission errors but the largest commission errors among all methods due to the compact consistency of the burned area delineation in mapping burned areas. ML methods would misclassify the shoreside soils as burned areas due to their similar spectral behaviors to burned areas.

**Table 5.** Testing accuracies of burned area detection for Fågelsjö-Lillåsen and Trängslet fires with Sentinel-2 data. The lowest errors (Oe and Ce) and the highest accuracy (mIoU and Kappa) in each model group are highlighted in bold.

Sites	Fågelsjö-Lillåsen				Trängslet			
Model	Oe (%)	Ce (%)	mIoU	Kappa	Oe (%)	Ce (%)	mIoU	Kappa
U-Net	20.91	4.26	0.77	0.75	10.36	<b>12.61</b>	0.82	0.80
HRNet	22.10	4.78	0.76	0.73	5.33	13.91	<b>0.84</b>	<b>0.82</b>
Fast-SCNN	<b>11.82</b>	8.47	<b>0.81</b>	<b>0.79</b>	<b>2.99</b>	20.24	0.79	0.77
DeepLabv3+	26.32	<b>4.16</b>	0.73	0.69	8.63	12.66	0.83	0.81
LightGBM	44.41	<b>3.56</b>	0.60	0.52	6.95	6.04	<b>0.89</b>	<b>0.89</b>
KNN	46.58	3.73	0.58	0.50	7.54	<b>5.74</b>	<b>0.89</b>	0.88
RF	<b>40.71</b>	3.99	<b>0.63</b>	<b>0.56</b>	<b>6.37</b>	7.35	<b>0.89</b>	0.88
$NBR_{otsu}$	<b>34.92</b>	5.17	<b>0.66</b>	<b>0.60</b>	25.94	<b>2.81</b>	0.77	0.74
$NBR_{em}$	65.03	<b>0.73</b>	0.47	0.34	<b>8.59</b>	7.31	<b>0.87</b>	<b>0.86</b>

Different from the Fågelsjö-Lillåsen fire, the Trängslet fire caused the disperse heterogeneous burned areas due to several wetlands within the study areas rather than homogeneous land cover. DL models tend to consider the connectivity between different burned area patches based on the convolutional feature extraction in low-resolution. Regarding the Trängslet fire, the dispersed burned areas mislead the general contextual information extraction of DL models. A similar conclusion also applies to the work in [57]. Most of the commission errors are located over the unburned patches close to the perimeters that are related with the wetland and bare soils in Figure 11. Importantly, all ML methods surpass all DL methods greatly with a kappa value of approximately 0.88. Every ML method represents similar visual results in Figure 10, which can inevitably misclassify the burned areas in the subset with wetland and bare soils. The empirical NBR-based thresholding method can reach a high accuracy in detecting Trängslet fire (kappa over 0.86).



**Figure 11.** Analysis of commission errors in detecting Trängslet fire. (a) the cropped VHR image subset in 2021 with the reference polygons in red; (b) the cropped subset of U-Net burned area detection result from Figure 10.

### 5.3. Transferring Phase with Landsat-8 Data

DL and ML models trained with Sentinel-2 data are then transferred to the corresponding Landsat-8 images on the same test sites so that we could assess the prediction performance and similarity with identical references. The performance evaluation with Landsat-8 data in Table 6 keeps the same tendency as Sentinel-2 data. DL models outperform other methods in Corinthia and Fågelsjö-Lillåsen fires, whereas ML methods (i.e., LightGBM) perform best in the Trängslet fire. The accuracy of HRNet dominates all DL models in the three test sites with Landsat-8 data, rather than U-Net and Fast-SCNN in Tables 4 and 5, respectively.

**Table 6.** Testing results with Landsat-8 data. The lowest errors (Oe and Ce) and the highest Kappa in each model group are highlighted in bold.

Sites	Corinthia			Fågelsjö-Lillåsen			Trängslet		
Model	Oe (%)	Ce (%)	Kappa	Oe (%)	Ce (%)	Kappa	Oe (%)	Ce (%)	Kappa
U-Net	16.24	2.30	<b>0.85</b>	21.85	4.74	0.73	8.45	12.12	0.82
HRNet	<b>14.17</b>	4.08	<b>0.85</b>	12.62	9.19	<b>0.78</b>	5.63	13.23	<b>0.83</b>
Fast-SCNN	20.39	3.68	0.80	<b>10.92</b>	10.92	0.77	<b>2.64</b>	20.05	0.77
DeepLabv3+	20.50	<b>2.13</b>	0.81	40.31	<b>2.87</b>	0.57	12.51	<b>10.68</b>	0.80
LightGBM	<b>21.57</b>	1.09	<b>0.75</b>	54.66	<b>3.20</b>	0.43	10.05	6.53	<b>0.86</b>
KNN	29.07	<b>1.01</b>	0.74	56.46	3.56	0.41	10.84	<b>6.15</b>	0.85
RF	28.86	1.14	0.74	<b>49.07</b>	3.32	<b>0.48</b>	<b>9.21</b>	7.52	0.85
NBR <sub>otsu</sub>	<b>22.53</b>	4.24	<b>0.78</b>	<b>49.05</b>	6.42	<b>0.46</b>	25.94	<b>3.47</b>	0.74
NBR <sub>em</sub>	38.92	<b>0.46</b>	0.65	72.16	<b>1.62</b>	0.26	<b>8.62</b>	9.47	<b>0.84</b>

DL models indicate good generalization ability in cross-sensor satellite images with acceptable accuracy as listed in Table 6. The performance of ML models seems to be unstable. Interestingly, the segmentation accuracies of DL models (except Fast-SCNN) are increased at the Trängslet site. On the other hand, the classification accuracies of ML methods decrease greatly in Corinthia and Fågelsjö-Lillåsen, even though they still perform well in the Trängslet fire with kappa over 0.85. In detail, kappa values decrease by 0.04 (HRNet) and 0.09 (LightGBM) in comparison to Table 4 for the Corinthia fire. Regarding the Fågelsjö-Lillåsen fire, it only sees a slight decrease of 0.02 in kappa values using the Fast-



SCNN model while RF drops from 0.56 to 0.48. Finally, for the Trängslet fire, LightGBM has a small decrease of 0.03 in kappa, while HRNet conversely shows a little improvement in kappa from 0.82 to 0.83. The reason might be that the commission errors caused by the Landsat-8 data are lower than the Sentinel-2 data.

Compared to other methods for mapping burned areas from Landsat-8, our results for the Corinthia fire reached a comparative accuracy with the results in [97], which used a two-phase algorithm (i.e., spectral–temporal rule and region-growing) to balance omission and commission errors (11.1% and 16.5%, respectively) with a kappa value of 0.85 for a wildfire in Portugal. HRNet in our study also reaches a kappa of approximately 0.85 but much lower commission errors (4.08%).

From the graphical results in Figure A2, DL models show more robust segmentation on burned areas with accurate delineation for the large burned patch in the Corinthia and Fågelsjö-Lillåsen fires. Due to the abundant lightly burned areas, ML methods fail to highlight the burned areas in Fågelsjö-Lillåsen, while DL models like HRNet and Fast-SCNN could still depict the perimeters of burned areas. On the other hand, disperse burned patches in Trängslet also affect the performance of DL with high commission errors due to their similar contextual connectivity in the spatial domain.

## 6. Discussion

Test results have shown that DL models achieve acceptable performance in mapping burned areas. They make full use of the background–foreground context in single-date imagery and capture the object features in multiple scales, making discrimination of the burned area possible. DL models can successfully keep a strong feature representation within neighboring burned pixels. Overall, DL models show good results on the compact Mediterranean mountain forests and Swedish boreal forests, while ML methods can obtain more accurate results in the dispersed Swedish site.

Although U-Net and Fast-SCNN perform best with Sentinel-2 data in the testing sites of the Corinthia fire and Fågelsjö-Lillåsen fire, respectively, HRNet stands out with higher mIoU and kappa than other methods when the testing data are from Landsat-8. In other words, all DL models need to be treated with some caution as their selection may depend on the data source, what landscapes and terrain the study areas are located in, and what kinds of biomass was burned.

ML methods employ pixel-wise classification that might lack spatial contextual information. They would produce some commission errors outside perimeters and huge omission errors within delineations, leading to the lack of balance between the commission and omission errors. Using a uni-temporal image, NBR can reach good performance in some specific sites (e.g., Trängslet fire) but is not recommended to map burned areas on a large scale with various landscapes. ML methods show similar results, and their performance is deeply affected by variance imposed by local study sites rather than the selection of methods as reported in [33]. Moreover, the parameters of ML methods in this study can be further optimized to improve the accuracy to some extent.

On the other hand, the bitemporal index, namely, dNBR, requires the pre-fire cloud-free image; thus, it was expected to perform better than NBR but was somewhat unexpected to be worse if using OTSU-based automated method, taking the fire in Corinthia as an example ( $NBR_{otsu}$  in Table 4 vs.  $dNBR_{otsu}$  in Table 7). That might be caused by the sub-optimal threshold computed using the OTSU approach (see the Figure 12c). Not surprisingly, empirically thresholding dNBR in Table 7 can obtain a similar kappa value (0.90) to the U-Net model in Table 4. These comparative experiments might imply that some DL models based on one uni-temporal post-event image could reach the same high accuracy as bitemporal index. However, we aim to avoid the usage of bitemporal images and manual assistance, as the advantages of automated algorithms are easy to deploy in practice, widely used, and without any human intervention.

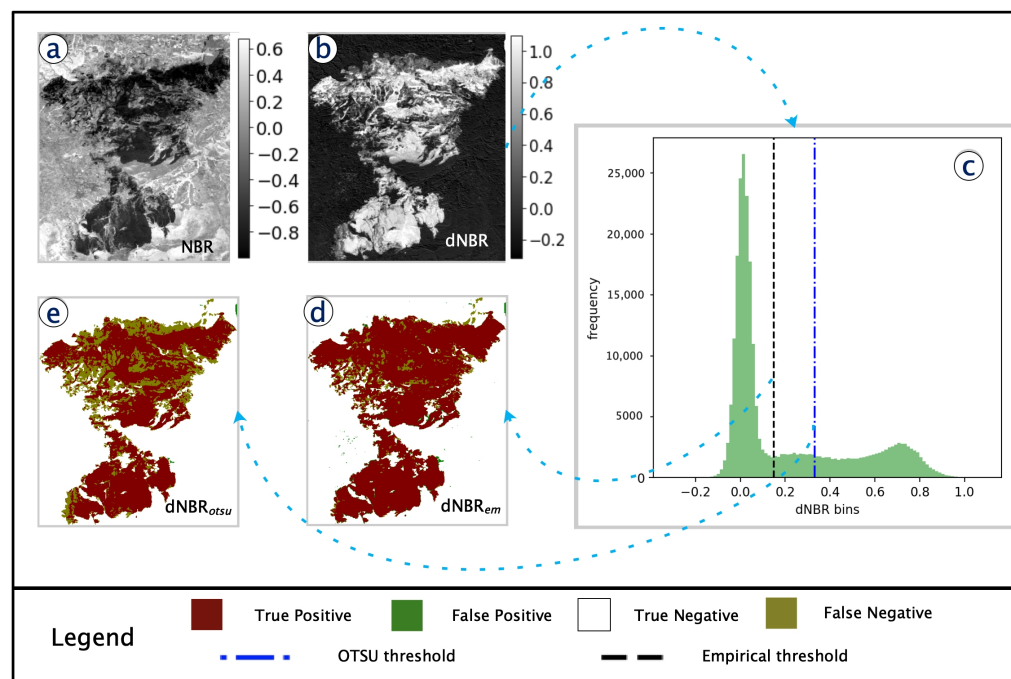
The pretrained DL models on Sentinel-2 were successfully transferred to map the burned area on Landsat-8 data with acceptable accuracy. We relate this observation to the



fact that DL network (e.g., HRNet) has a high level of generalization on multi-scale representative features to keep consistency between Sentinel-2 and Landsat-8 data. By fusing available multispectral data, DL models show promise for use in future NRT applications.

**Table 7.** dNBR-based results with Sentinel-2 data in Corinthia.

	AD (%)	QD (%)	OA	mIoU	Kappa
dNBR <sub>otsu</sub>	0.33	11.99	0.88	0.76	0.73
dNBR <sub>em</sub>	2.18	2.51	0.95	0.91	0.90



**Figure 12.** Burned area mapping results of Corinthia using the threshold-based methods on dNBR with Sentinel-2 data. Two thresholds are determined by human intervention and OTSU automated approaches, respectively, as shown in the (c) histogram map. (a) The NBR map to threshold the burned area of the Corinthia fire in Figure 10. (b) dNBR map that is binarized to get the burned areas of dNBR<sub>em</sub> (d) and dNBR<sub>otsu</sub> (e) by empirical and automatic approaches, respectively.

## 7. Conclusions

In this study, a series of experiments is conducted to evaluate the capabilities of several DL models for mapping burned areas with uni-temporal Sentinel-2 and Landsat-8 images. These DL approaches achieve acceptable performance in comparison to ML algorithms and NBR-based methods, especially when the burned areas are compact. It can be observed that the DL network model tends to increase the mapping accuracy and thematic consistency of the final burned area delineation due to the fusion of multi-scale features rather than a pixel-based classification. This research highlights that DL models have several advantages: (i) automated mapping with high overall accuracy without the need for a pre-fire cloud-free image or the use of fixed or empirical thresholds and (ii) cross-sensor ability to detect the burned area in the various biomes.

From an operational viewpoint, although the present results show a very promising potential using DL models for burned area mapping, surely further study is needed towards the extension of creating larger data sets in more diverse fire-disturbed regions around the globe. An important direction of future work would be a specific investigation of the hybrid approaches to fuse DL and ML methods to improve the accuracy. Furthermore, fusion of optical and SAR imagery could be explored to improve burned area mapping.

**Author Contributions:** Conceptualization, X.H. and Y.B.; data acquisition, X.H.; experimental design and investigation, X.H., Y.B. and A.N.; methodology, X.H.; writing—original draft preparation, X.H. and Y.B.; writing—review and editing, X.H., Y.B. and A.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by KTH Digital Futures, grant number VF-2020-0260, entitled “EOAI4GlobalChange: Earth Observation Big Data and Deep Learning for Global Environmental Change Monitoring”.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

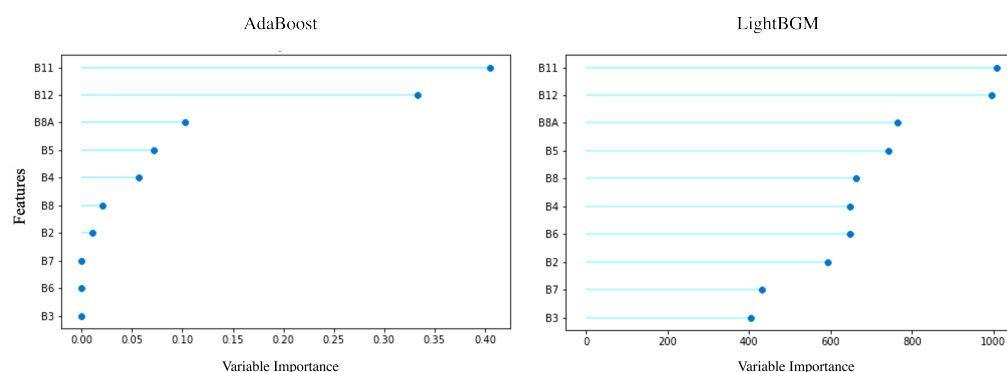
**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** This research is part of the project “EO-AI4GlobalChange” funded by KTH Digital Futures. Xikun Hu was funded by the China Scholarship Council (CSC). Xikun Hu acknowledged the financial and technical support from FOSS4G and ECMWF for the 2019 EO Data Challenge Copernicus Award. We acknowledge the use of data from Sentinel-2 operated by the Copernicus Programme and Landsat-8 data by NASA. Xikun Hu also thank Maryam Rahnemoonfar for her comments and suggestions that help to improve the manuscript.

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

## Appendix A. Feature Selection on Sentinel-2 Spectral Bands

Feature selection aims at reducing the dataset dimensionality by removing irrelevant and redundant attributes while keeping important ones. It reduces the risk of overfitting caused by a lack of model generalization [87]. AdaBoost and LightGBM are used to rank the score of importance for each spectral feature. B12, B11, and B8A shared the top three places in Figure A1.



**Figure A1.** Spectral feature importance based on AdaBoost and LightGBM on Sentinel-2 data of Castelo Branco fire.

## Appendix B. Comparison Results of ML Algorithms

The equations of the precision, recall, and F1 score are given as followed:

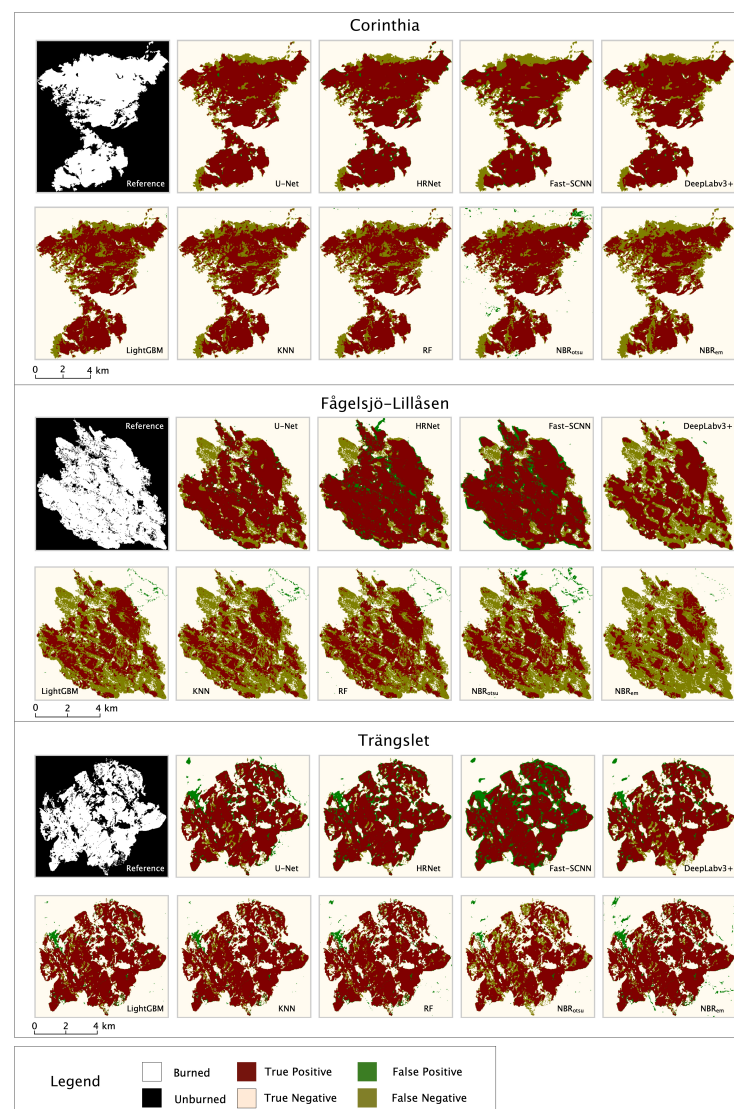
$$\text{Precision} = \frac{tp}{tp + fp} \quad (\text{A1})$$

$$\text{Recall} = \frac{tp}{tp + fn} \quad (\text{A2})$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{A3})$$

**Table A1.** Preliminary evaluation of ML methods on burned area classification.

Model	OA	Recall	Precision	F1	Kappa
Light Gradient Boosting Machine	0.9664	0.9011	0.9201	0.9105	0.8898
Gradient Boosting Classifier	0.9656	0.8955	0.9209	0.908	0.8868
K Neighbors Classifier	0.9638	0.8933	0.9136	0.9033	0.881
Random Forest Classifier	0.9633	0.8911	0.9132	0.902	0.8794
Extra Trees Classifier	0.9619	0.8875	0.9093	0.8982	0.8748
Ada Boost Classifier	0.9603	0.8617	0.9239	0.8917	0.8674
Quadratic Discriminant Analysis	0.9567	0.8481	0.9173	0.8813	0.8549
Logistic Regression	0.9538	0.8213	0.9266	0.8707	0.8427
Decision Tree Classifier	0.9484	0.8621	0.865	0.8636	0.8317
SVM - Linear Kernel	0.9476	0.7791	0.9338	0.849	0.8176
Linear Discriminant Analysis	0.9456	0.7827	0.9183	0.8451	0.8124
Ridge Classifier	0.9415	0.7424	0.9355	0.8278	0.7931
Naive Bayes	0.9379	0.6934	0.9705	0.8088	0.773

**Appendix C. Burned Area Mapping Results with Landsat-8 Data****Figure A2.** Burned area detection results with Landsat-8 data.

## References

- Bowman, D.M.; Williamson, G.J.; Abatzoglou, J.T.; Kolden, C.A.; Cochrane, M.A.; Smith, A.M. Human exposure and sensitivity to globally extreme wildfire events. *Nat. Ecol. Evol.* **2017**, *1*, 58. [\[CrossRef\]](#)
- Mangeon, S.; Field, R.; Fromm, M.; McHugh, C.; Voulgarakis, A. Satellite versus ground-based estimates of burned area: A comparison between MODIS based burned area and fire agency reports over North America in 2007. *Anthr. Rev.* **2016**, *3*, 76–92. [\[CrossRef\]](#)
- Chuvieco, E.; Congalton, R.G. Mapping and inventory of forest fires from digital processing of tm data. *Geocarto Int.* **1988**, *3*, 41–53. [\[CrossRef\]](#)
- Chuvieco, E.; Mouillot, F.; van der Werf, G.R.; San Miguel, J.; Tanasse, M.; Koutsias, N.; García, M.; Yebra, M.; Padilla, M.; Gitas, I.; et al. Historical background and current developments for mapping burned area from satellite Earth observation. *Remote Sens. Environ.* **2019**, *225*, 45–64. [\[CrossRef\]](#)
- Giglio, L.; Justice, C.; Boschetti, L.; Roy, D. MCD64A1 MODIS/Terra+Aqua Burned Area Monthly L3 Global 500m SIN Grid V006; Distributed by NASA EOSDIS Land Processes DAAC. 2015. Available online: <https://doi.org/10.5067/MODIS/MCD64A1.006> (accessed on 11 April 2021).
- Lizundia-Loiola, J.; Otón, G.; Ramo, R.; Chuvieco, E. A spatio-temporal active-fire clustering approach for global burned area mapping at 250 m from MODIS data. *Remote Sens. Environ.* **2020**, *236*, 111493. [\[CrossRef\]](#)
- Roteta, E.; Bastarrika, A.; Padilla, M.; Storm, T.; Chuvieco, E. Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa. *Remote Sens. Environ.* **2019**, *222*, 1–17. [\[CrossRef\]](#)
- Li, J.; Roy, D.P. A global analysis of Sentinel-2a, Sentinel-2b and Landsat-8 data revisit intervals and implications for terrestrial monitoring. *Remote Sens.* **2017**, *9*, 902. [\[CrossRef\]](#)
- Toukiloglou, P.; Gitas, I.Z.; Katagis, T. An automated two-step NDVI-based method for the production of low-cost historical burned area map records over large areas. *Int. J. Remote Sens.* **2014**, *35*, 2713–2730. [\[CrossRef\]](#)
- Roy, D.P.; Huang, H.; Boschetti, L.; Giglio, L.; Yan, L.; Zhang, H.H.; Li, Z. Landsat-8 and Sentinel-2 burned area mapping—A combined sensor multi-temporal change detection approach. *Remote Sens. Environ.* **2019**, *231*. [\[CrossRef\]](#)
- Chen, Y.; Lara, M.J.; Hu, F.S. A robust visible near-infrared index for fire severity mapping in Arctic tundra ecosystems. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 101–113. [\[CrossRef\]](#)
- Kontoes, C.C.; Poilvé, H.; Florsch, G.; Keramitsoglou, I.; Paralikidis, S. A comparative analysis of a fixed thresholding vs. a classification tree approach for operational burn scar detection and mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 299–316. [\[CrossRef\]](#)
- Quintano, C.; Fernández-Manso, A.; Stein, A.; Bijker, W. Estimation of area burned by forest fires in Mediterranean countries: A remote sensing data mining perspective. *For. Ecol. Manag.* **2011**, *262*, 1597–1607. [\[CrossRef\]](#)
- Trigg, S.; Flasse, S. An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah. *Int. J. Remote Sens.* **2001**, *22*, 2641–2647. [\[CrossRef\]](#)
- Chu, T.; Guo, X. Remote sensing techniques in monitoring post-fire effects and patterns of forest recovery in boreal forest regions: A review. *Remote Sens.* **2013**, *6*, 470–520. [\[CrossRef\]](#)
- Fernández-Manso, A.; Fernández-Manso, O.; Quintano, C. SENTINEL-2A red-edge spectral indices suitability for discriminating burn severity. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *50*, 170–175. [\[CrossRef\]](#)
- Huang, H.; Roy, D.P.; Boschetti, L.; Zhang, H.K.; Yan, L.; Kumar, S.S.; Gomez-Dans, J.; Li, J.; Huang, H.; Roy, D.P.; et al. Separability analysis of Sentinel-2A Multi-Spectral Instrument (MSI) data for burned area discrimination. *Remote Sens.* **2016**, *8*, 873. [\[CrossRef\]](#)
- Navarro, G.; Caballero, I.; Silva, G.; Parra, P.C.; Vázquez, Á.; Caldeira, R. Evaluation of forest fire on Madeira Island using Sentinel-2A MSI imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *58*, 97–106. [\[CrossRef\]](#)
- Quintano, C.; Fernández-Manso, A.; Fernández-Manso, O. Combination of Landsat and Sentinel-2 MSI data for initial assessing of burn severity. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 221–225. [\[CrossRef\]](#)
- Filipponi, F. BAIS2: Burned Area Index for Sentinel-2. *Proceedings* **2018**, *2*, 364. [\[CrossRef\]](#)
- Loboda, T.; O'Neal, K.J.; Csiszar, I. Regionally adaptable dNBR-based algorithm for burned area mapping from MODIS data. *Remote Sens. Environ.* **2007**. [\[CrossRef\]](#)
- Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66. [\[CrossRef\]](#)
- Pulvirenti, L.; Squicciarino, G.; Fiori, E.; Fiorucci, P.; Ferraris, L.; Negro, D.; Gollini, A.; Severino, M.; Puca, S. An automatic processing chain for near real-time mapping of burned forest areas using sentinel-2 data. *Remote Sens.* **2020**, *12*, 674. [\[CrossRef\]](#)
- Smith, A.M.; Drake, N.A.; Wooster, M.J.; Hudak, A.T.; Holden, Z.A.; Gibbons, C.J. Production of Landsat ETM+ reference imagery of burned areas within Southern African savannahs: Comparison of methods and application to MODIS. *Int. J. Remote Sens.* **2007**, *28*, 2753–2775. [\[CrossRef\]](#)
- Jain, P.; Coogan, S.C.; Subramanian, S.G.; Crowley, M.; Taylor, S.; Flannigan, M.D. A review of machine learning applications in wildfire science and management. *Environ. Rev.* **2020**, *28*, 478–505. [\[CrossRef\]](#)
- Hawbaker, T.J.; Vanderhoof, M.K.; Beal, Y.J.J.; Takacs, J.D.; Schmidt, G.L.; Falgout, J.T.; Williams, B.; Fairaux, N.M.; Caldwell, M.K.; Picotte, J.J.; et al. Mapping burned areas using dense time-series of Landsat data. *Remote Sens. Environ.* **2017**, *198*, 504–522. [\[CrossRef\]](#)
- Liu, J.; Heiskanen, J.; Maeda, E.E.; Pellikka, P.K.E. Burned area detection based on Landsat time series in savannas of southern Burkina Faso. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *64*, 210–220. [\[CrossRef\]](#)



28. Stavrakoudis, D.; Katagis, T.; Minakou, C.; Gitas, I.Z.; Stavrakoudis, D.; Katagis, T.; Minakou, C.; Gitas, I.Z. Automated Burned Scar Mapping Using Sentinel-2 Imagery. *J. Geogr. Inf. Syst.* **2020**, *12*, 221–240. [\[CrossRef\]](#)
29. Long, T.; Zhang, Z.; He, G.; Jiao, W.; Tang, C.; Wu, B.; Zhang, X.; Wang, G.; Yin, R. 30m resolution global annual burned area mapping based on landsat images and Google Earth Engine. *Remote Sens.* **2019**, *11*, 489. [\[CrossRef\]](#)
30. Koutsias, N.; Karteris, M. Burned area mapping using logistic regression modeling of a single post-fire Landsat-5 Thematic Mapper image. *Int. J. Remote Sens.* **2000**, *21*, 673–687. [\[CrossRef\]](#)
31. Petropoulos, G.P.; Kontoes, C.; Keramitsoglou, I. Burnt area delineation from a uni-temporal perspective based on landsat TM imagery classification using Support Vector Machines. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 70–80. [\[CrossRef\]](#)
32. Mitrakis, N.E.; Mallinis, G.; Koutsias, N.; Theocharis, J.B. Burned area mapping in Mediterranean environment using medium-resolution multi-spectral data and a neuro-fuzzy classifier. *Int. J. Image Data Fusion* **2012**, *3*, 299–318. [\[CrossRef\]](#)
33. Mallinis, G.; Koutsias, N. Comparing ten classification methods for burned area mapping in a Mediterranean environment using Landsat TM satellite data. *Int. J. Remote Sens.* **2012**, *33*, 4408–4433. [\[CrossRef\]](#)
34. Pu, R.; Gong, P. Determination of burnt scars using logistic regression and neural network techniques from a single post-fire Landsat 7 ETM+ image. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 841–850. [\[CrossRef\]](#)
35. Stroppiana, D.; Bordogna, G.; Carrara, P.; Boschetti, M.; Boschetti, L.; Brivio, P. A method for extracting burned areas from Landsat TM/ETM+ images by soft aggregation of multiple Spectral Indices and a region growing algorithm. *ISPRS J. Photogramm. Remote Sens.* **2012**, *69*, 88–102. [\[CrossRef\]](#)
36. Reichstein, M.; Camps-Valls, G.; Stevens, B.; Jung, M.; Denzler, J.; Carvalhais, N.; Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature* **2019**, *566*, 195–204. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Lateef, F.; Ruichek, Y. Survey on semantic segmentation using deep learning techniques. *Neurocomputing* **2019**, *338*, 321–348. [\[CrossRef\]](#)
38. Milioto, A.; Lottes, P.; Stachniss, C. Real-Time Semantic Segmentation of Crop and Weed for Precision Agriculture Robots Leveraging Background Knowledge in CNNs. In Proceedings of the IEEE International Conference on Robotics and Automation, Brisbane, QLD, Australia, 21–25 May 2018; Volume 338, pp. 2229–2235. [\[CrossRef\]](#)
39. Tseng, Y.H.; Jan, S.S. Combination of computer vision detection and segmentation for autonomous driving. In Proceedings of the 2018 IEEE/ION Position, Location and Navigation Symposium, PLANS, Monterey, CA, USA, 23–26 April 2018; pp. 1047–1052. [\[CrossRef\]](#)
40. Jiang, F.; Grigorev, A.; Rho, S.; Tian, Z.; Fu, Y.S.; Jifara, W.; Adil, K.; Liu, S. Medical image semantic segmentation based on deep learning. *Neural Comput. Appl.* **2018**, *29*, 1257–1265. [\[CrossRef\]](#)
41. Bhuiyan, M.A.E.; Witharana, C.; Liljedahl, A.K. Use of Very High Spatial Resolution Commercial Satellite Imagery and Deep Learning to Automatically Map Ice-Wedge Polygons across Tundra Vegetation Types. *J. Imaging* **2020**, *6*, 137. [\[CrossRef\]](#)
42. Zhang, W.; Liljedahl, A.K.; Kanevskiy, M.; Epstein, H.E.; Jones, B.M.; Jorgenson, M.T.; Kent, K. Transferability of the deep learning mask R-CNN model for automated mapping of ice-wedge polygons in high-resolution satellite and UAV images. *Remote Sens.* **2020**, *12*, 1085. [\[CrossRef\]](#)
43. Bonhage, A.; Eltaher, M.; Raab, T.; Breuß, M.; Raab, A.; Schneider, A. A modified Mask region-based convolutional neural network approach for the automated detection of archaeological sites on high-resolution light detection and ranging-derived digital elevation models in the North German Lowland. *Archaeol. Prospect.* **2021**, *1*–10. [\[CrossRef\]](#)
44. Ma, L.; Liu, Y.; Zhang, X.; Ye, Y.; Yin, G.; Johnson, B.A. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **2019**, *152*, 166–177. [\[CrossRef\]](#)
45. Zhang, R.; Li, G.; Li, M.; Wang, L. Fusion of images and point clouds for the semantic segmentation of large-scale 3D scenes based on deep learning. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 85–96. [\[CrossRef\]](#)
46. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [\[CrossRef\]](#)
47. Wurm, M.; Stark, T.; Zhu, X.X.; Weigand, M.; Taubenböck, H. Semantic segmentation of slums in satellite images using transfer learning on fully convolutional neural networks. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 59–69. [\[CrossRef\]](#)
48. Liu, C.C.; Zhang, Y.C.; Chen, P.Y.; Lai, C.C.; Chen, Y.H.; Cheng, J.H.; Ko, M.H. Clouds classification from Sentinel-2 imagery with deep residual learning and semantic image segmentation. *Remote Sens.* **2019**, *11*, 119. [\[CrossRef\]](#)
49. Li, R.; Liu, W.; Yang, L.; Sun, S.; Hu, W.; Zhang, F.; Li, W. DeepUNet: A Deep Fully Convolutional Network for Pixel-Level Sea-Land Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3954–3962. [\[CrossRef\]](#)
50. Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 749–753. [\[CrossRef\]](#)
51. Pinto, M.M.; Libonati, R.; Trigo, R.M.; Trigo, I.F.; DaCamara, C.C. A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. *ISPRS J. Photogramm. Remote Sens.* **2020**, *160*, 260–274. [\[CrossRef\]](#)
52. Langford, Z.; Kumar, J.; Hoffman, F. Wildfire mapping in interior alaska using deep neural networks on imbalanced datasets. In Proceedings of the IEEE International Conference on Data Mining Workshops, ICDMW, Singapore, 17–20 November 2018; pp. 770–778. [\[CrossRef\]](#)
53. Zhang, P.; Nascetti, A.; Ban, Y.; Gong, M. An implicit radar convolutional burn index for burnt area mapping with Sentinel-1 C-band SAR data. *ISPRS J. Photogramm. Remote Sens.* **2019**, *158*, 50–62. [\[CrossRef\]](#)

54. Ban, Y.; Zhang, P.; Nascetti, A.; Bevington, A.R.; Wulder, M.A. Near Real-Time Wildfire Progression Monitoring with Sentinel-1 SAR Time Series and Deep Learning. *Sci. Rep.* **2020**, *10*. [\[CrossRef\]](#)
55. Bermudez, J.D.; Happ, P.N.; Feitosa, R.Q.; Oliveira, D.A. Synthesis of Multispectral Optical Images from SAR/Optical Multitemporal Data Using Conditional Generative Adversarial Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1220–1224. [\[CrossRef\]](#)
56. de Bem, P.P.; de Carvalho, O.A., Jr.; de Carvalho, O.L.F.; Gomes, R.A.T.; Guimarães, R.F. Performance analysis of deep convolutional autoencoders with different patch sizes for change detection from burnt areas. *Remote Sens.* **2020**, *12*, 2576. [\[CrossRef\]](#)
57. Knopp, L.; Wieland, M.; Rättich, M.; Martinis, S. A deep learning approach for burned area segmentation with Sentinel-2 data. *Remote Sens.* **2020**, *12*, 2422. [\[CrossRef\]](#)
58. Van Der Werff, H.; Van Der Meer, F. Sentinel-2A MSI and Landsat 8 OLI provide data continuity for geological remote sensing. *Remote Sens.* **2016**, *8*, 883. [\[CrossRef\]](#)
59. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; Volume 9351, pp. 234–241. [\[CrossRef\]](#)
60. Poudel, R.P.; Liwicki, S.; Cipolla, R. Fast-SCNN: Fast semantic segmentation network. In Proceedings of the 30th British Machine Vision Conference (BMVC), Cardiff, UK, 9–12 September 2019.
61. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
62. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep High-Resolution Representation Learning for Visual Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [\[CrossRef\]](#)
63. Dinerstein, E.; Olson, D.; Joshi, A.; Vynne, C.; Burgess, N.D.; Wikramanayake, E.; Hahn, N.; Palminteri, S.; Hedao, P.; Noss, R.; et al. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience* **2017**, *67*, 534–545. [\[CrossRef\]](#)
64. Roy, D.P.; Wulder, M.A.; Loveland, T.R.; Woodcock, C.E.; Allen, R.G.; Anderson, M.C.; Helder, D.; Irons, J.R.; Johnson, D.M.; Kennedy, R.; et al. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sens. Environ.* **2014**, *145*, 154–172. [\[CrossRef\]](#)
65. Mallinis, G.; Mitsopoulos, I.; Chrysafi, I. Evaluating and comparing sentinel 2A and landsat-8 operational land imager (OLI) spectral indices for estimating fire severity in a mediterranean pine ecosystem of Greece. *GISci. Remote Sens.* **2018**, *55*, 1–18. [\[CrossRef\]](#)
66. Farasin, A.; Colomba, L.; Garza, P. Double-step U-Net: A deep learning-based approach for the estimation of wildfire damage severity through sentinel-2 satellite data. *Appl. Sci.* **2020**, *10*, 4332. [\[CrossRef\]](#)
67. BC Wildfire Service. Wildfires of Note—Elephant Hill (K20637). Available online: <http://bcfireinfo.for.gov.bc.ca/hprScripts/WildfireNews/OneFire.asp?ID=620> (accessed on 3 February 2021).
68. Matthews, J.A. CORINE land-cover map. In *Encyclopedia of Environmental Change*; SAGE Publications: Thousand Oaks, CA, USA, 2014. [\[CrossRef\]](#)
69. Lutes, D.C.; Keane, R.E.; Caratti, J.F.; Key, C.H.; Benson, N.C.; Gangi, L.J. FIREMON: Fire effects monitoring and inventory system. In *USDA Forest Service, Rocky Mountain Research Station, General Technical Report*; U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station: Fort Collins, CO, USA, 2006. [\[CrossRef\]](#)
70. Gascon, F.; Bouzinac, C.; Thépaut, O.; Jung, M.; Francesconi, B.; Louis, J.; Lonjou, V.; Lafrance, B.; Massera, S.; Gaudel-Vacaresse, A.; et al. Copernicus Sentinel-2A calibration and products validation status. *Remote Sens.* **2017**, *9*, 584. [\[CrossRef\]](#)
71. Chastain, R.; Housman, I.; Goldstein, J.; Finco, M. Empirical cross sensor comparison of Sentinel-2A and 2B MSI, Landsat-8 OLI, and Landsat-7 ETM+ top of atmosphere spectral characteristics over the conterminous United States. *Remote Sens. Environ.* **2019**, *221*, 274–285. [\[CrossRef\]](#)
72. Freund, Y.; Schapire, R.E. Experiments with a New Boosting Algorithm. In Proceedings of the 13th International Conference on Machine Learning, Bari, Italy, 3–6 July 1996; pp. 148–156.
73. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.Y. LightGBM: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *2017*, 3147–3155.
74. Hardtke, L.A.; Blanco, P.D.; Del Valle, F.; Metternicht, G.I.; Sione, W.F. Semi-automated mapping of burned areas in semi-arid ecosystems using MODIS time-series imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 25–35. [\[CrossRef\]](#)
75. Imperatore, P.; Azar, R.; Calo, F.; Stroppiana, D.; Brivio, P.A.; Lanari, R.; Pepe, A. Effect of the Vegetation Fire on Backscattering: An Investigation Based on Sentinel-1 Observations. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4478–4492. [\[CrossRef\]](#)
76. Kato, A.; Thau, D.; Hudak, A.T.; Meigs, G.W.; Moskal, L.M. Quantifying fire trends in boreal forests with Landsat time series and self-organized criticality. *Remote Sens. Environ.* **2020**, *237*, 111525. [\[CrossRef\]](#)
77. Ali, M. PyCaret: An Open Source, Low-Code Machine Learning Library in Python, PyCaret Version 2.3. 2020. Available online: <https://pycaret.org/> (accessed on 11 April 2021).
78. Bengio, Y.; Grandvalet, Y. No unbiased estimator of the variance of K-fold cross-validation. *J. Mach. Learn. Res.* **2004**, *5*, 1089–1105.
79. Xie, Y.; Peng, M. Forest fire forecasting using ensemble learning approaches. *Neural Comput. Appl.* **2019**, *31*, 4541–4550. [\[CrossRef\]](#)
80. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175–185. [\[CrossRef\]](#)

81. Zammit, O.; Descombes, X.; Zerubia, J. Burnt area mapping using Support Vector Machines. *For. Ecol. Manag.* **2006**, *234*, S240. [[CrossRef](#)]
82. Dutta, R.; Das, A.; Aryal, J. Big data integration shows Australian bush-fire frequency is increasing significantly. *R. Soc. Open Sci.* **2016**, *3*. [[CrossRef](#)] [[PubMed](#)]
83. 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 Sens.* **2021**, *13*, 220. [[CrossRef](#)]
84. Belgiu, M.; Drăgu, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
85. Gibson, R.; Danaher, T.; Hehir, W.; Collins, L. A remote sensing approach to mapping fire severity in south-eastern Australia using sentinel 2 and random forest. *Remote Sens. Environ.* **2020**, *240*. [[CrossRef](#)]
86. Ramo, R.; Chuvieco, E. Developing a Random Forest algorithm for MODIS global burned area classification. *Remote Sens.* **2017**, *9*, 1193. [[CrossRef](#)]
87. Ramo, R.; García, M.; Rodríguez, D.; Chuvieco, E. A data mining approach for global burned area mapping. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *73*, 39–51. [[CrossRef](#)]
88. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
89. Mottaghi, R.; Chen, X.; Liu, X.; Cho, N.G.; Lee, S.W.; Fidler, S.; Urtasun, R.; Yuille, A. The role of context for object detection and semantic segmentation in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 891–898. [[CrossRef](#)]
90. Gong, K.; Liang, X.; Zhang, D.; Shen, X.; Lin, L. Look into Person: Self-supervised Structure-sensitive Learning and a new benchmark for human parsing. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [[CrossRef](#)]
91. Cohen, J. A Coefficient of Agreement for Nominal Scales. *Educ. Psychol. Meas.* **1960**, *20*, 37–46. [[CrossRef](#)]
92. Pontius, R.G.; Millones, M. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* **2011**, *32*, 4407–4429. [[CrossRef](#)]
93. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015; pp. 1–15.
94. Goyal, P.; Dollár, P.; Girshick, R.; Noordhuis, P.; Wesolowski, L.; Kyrola, A.; Tulloch, A.; Jia, Y.; He, K. Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour. *arXiv* **2017**, arXiv:1706.02677.
95. Ho, Y.; Wookey, S. The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling. *IEEE Access* **2020**, *8*, 4806–4813. [[CrossRef](#)]
96. Shore, J.E.; Johnson, R.W. Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy. *IEEE Trans. Inf. Theory* **1980**, *26*, 26–37. [[CrossRef](#)]
97. Bastarrika, A.; Chuvieco, E.; Martín, M.P. Mapping burned areas from landsat TM/ETM+ data with a two-phase algorithm: Balancing omission and commission errors. *Remote Sens. Environ.* **2011**, *115*, 1003–1012. [[CrossRef](#)]