# Supervised Burned Areas Delineation by Means of Sentinel-2 Imagery and Convolutional Neural Networks

**Alessandro Farasin**
Politecnico di Torino - DAUIN dept.*
and LINKS Foundation - DSISA dept.†
alessandro.farasin@polito.it

**Luca Colomba**
Politecnico di Torino - DAUIN dept.
luca.colomba@polito.it

**Giulio Palomba**
LINKS Foundation - DSISA dept.
giulio.palomba@linksfoundation.com

**Giovanni Nini**
LINKS Foundation - DSISA dept.
giovanni.nini@linksfoundation.com

**Claudio Rossi**
LINKS Foundation - DSISA dept.
claudio.rossi@linksfoundation.com

## ABSTRACT

Wildfire events are increasingly threatening our lands, cities, and lives. To contrast this phenomenon and to limit its damages, governments around the globe are trying to find proper counter-measures, identifying prevention and monitoring as two key factors to reduce wildfires impact worldwide. In this work, we propose two deep convolutional neural networks to automatically detect and delineate burned areas from satellite acquisitions, assessing their performances at scale using validated maps of burned areas of historical wildfires. We demonstrate that the proposed networks substantially improve the burned area delineation accuracy over conventional methods.

### Keywords

Burned Area Delineation, Sentinel-2, U-Net, CuMedVision1, Convolutional Neural Network, Deep Learning, Supervised Learning, Pixel-wise Segmentation.

## INTRODUCTION

In recent years, large wildfires have repeatedly affected Europe with an increasing trend. According to Copernicus' European Forest Fire Information System (EFFIS) and as reported by EuroNews, about 1,600 wildfires have been recorded in the European Union in 2019: more than three times more than the average over the past decade. (*European Forest Fire Information System (EFFIS)* 2019, *Euronews* 2019). The wildfires that broke out in Sokndal (Norway) in 2019 and affected 7500 hectares of lands and forests, or in Catalonia (Spain), which destroyed 6,500 hectares of forests and threatened the town La Torre de l'Espanyol in June 2019, or in Greece, which in August 2019 struck the island of Thassos and blanketed Athens, are unluckily some of the most recent events that caused the European Union and the its countries huge loss of precious resources, lives, and money (*Europe wildfires:*

*Dipartimento di Automatica e Informatica (DAUIN), Corso Duca degli Abruzzi, 24, 10129 Torino, Italy
†Department of Data Science for Industrial and Societal Applications (DSISA), Via Pier Carlo Boggio, 61, 10138 Torino, Italy

*Norway police evacuate hundreds in Sokndal* 2019, *European heatwave: Spain battles major Catalonia wildfire* 2019, *Greece wildfires leave blackened forests in their wake* 2019). The situation is certainly not encouraging and needs to be properly addressed to counteract the evolution of these phenomena. In this regard, as stated in the latest Technical Report made by the European Joint Research Centre (JRC), the European Commission is working to improve information and prevention measures to reduce the damage risk and increase the effectiveness of public authorities and first responders (*JRC Technical Report - Forest Fires in Europe, Middle East and North Africa* 2018).

Nowadays there are powerful tools that can interoperate to support the prevention of emergencies in a better way: data exploitation is certainly one of these. In the last years, satellites have been collecting huge amounts of data for different purposes, starting from the American missions (i.e. Landsat) to the European programme Copernicus (*Copernicus* 2019). The latter belongs to the European Union's Earth Observation Programme, gathering data useful for the preservation of the environment and European citizens' safety: it provides information services based on satellite Earth Observation and in-situ (non-space) data. Through its satellites, it provides daily or hourly, fresh information about lands, seas, air quality, climate analysis, etc. Moreover, the Copernicus Emergency Management System (EMS) provides accurate geo-spatial information and mappings derived from satellite remote sensing and completed by available in-situ or open data sources (*Copernicus Emergency Management System (EMS)* 2019). The creation of these mappings is usually semi-supervised, manually corrected and certified by domain experts. In the wildfires domain, the EMS provides mappings of the Regions of Interest hit by fires, in form of *Delineation Maps*, which are exploited by public authorities for several applications, such as a monetary estimate of damages.

This paper proposes two supervised deep learning approaches, called U-Net (Ronneberger et al. 2015) and CuMedVision (Chen et al. 2017) for the automatic detection and delineation of regions affected by fire, only leveraging on the information provided by Sentinel-2 satellites (*Sentinel-2* 2019). First, we adapted the two techniques to be effective on satellite images: thanks to the acquisition instruments mounted on satellites, we can collect images at different spectral wavelengths - including visible, near-infrared, and short wave infrared part of the spectrum - of specific areas of the Earth. However, in this work, we want to assess the performances of two approaches using either all satellite spectral bands or a specific subset, concerning the visible part of the spectrum. The latter evaluation is useful to provide a preliminary assessment of the theoretical performances that could have been reached by adopting different acquisition systems, like commercial cameras mounted on aircraft or drones. Conversely to satellites, which can inspect a specific region of the Earth only a few times a week, the adoption of aerial vehicles for recognition can sensibly improve the security of lands, accelerating and empowering the early warning phase of a hazardous event.

The next sections are organized as follows. *Related works* introduces the state of the art on approaches for burned areas detection and delineation from satellite acquisitions. *Collecting burned areas information* introduces the sources of information we used in this work. *Dataset* describes the information we collected and used for the evaluation of the two approaches. *Methodology* details the overall approaches, describing the data pre-processing and the adjustments we applied to the deep networks for both the case studies (using all the available spectral bands and the only bands related to the additive pure colors of the visible light). *Experiments* details the validation of the techniques in the two case studies, making a comparison between the standard indices used for enhancing burned regions. Lastly, *Conclusions* briefly summarize the scientific contribution, highlighting the pros and cons of the approaches and the possible future works to improve this work.

## RELATED WORKS

Burned area delineation, as anticipated in the previous section, is an activity currently conducted by domain experts, using both semi-automated and manual approaches. Moreover, the official European service is activated after sending a formal request to the European response Coordination Centre (ERCC) and the methodology of the approach for the generation of the delineation maps is not public (RMSRF 2019).

In literature, the use of indices computed from the combination of satellite spectral bands to highlight burned regions is a common approach. Among the wide range of studies we mention the Burned Area Index (BAI) (Chuvieco et al. 2002), the Normalized Burned Ratio (NBR) (Escuin et al. 2008) and the Normalized Burned Ratio 2 (NBR2) (M. L. Garcia and Caselles 1991), which were first used within the Landsat acquisitions and BAIS2 (Filipponi 2018), specifically defined for Sentinel-2. From them, human inspections try to create accurate delineations of burned areas, together with several automatic or semi-automatic approaches. Most of the literature in this field is related to Landsat data and the approaches combine rule-based thresholds specifically designed by domain experts to combinations of classical Machine-Learning algorithms (such as Decision Trees, Random

Forest, Support Vector Machines). Among them, it is worth mentioning: (i) approaches for specific regions, such as forests, or deserts (Hardtke et al. 2015, De Araujo and Ferreira 2015, Hughes et al. 2017), (ii) generally applicable approaches (Boschetti et al. 2015, Ramo and Chuvieco 2017, Ramo, M. Garcia, et al. 2018, Shan et al. 2017). Those approaches are extremely valid for the resolution of the acquisitions, ~500m per pixel, but in our case studies, in which the resolution is brought up to 10m per pixel, the approaches need to be reassessed. Regarding Sentinel-2 data, the literature generally presents techniques based on thresholds on the aforementioned indices (Bin et al. 2019, Verhegghen et al. 2016, Roy et al. 2019, Roteta et al. 2019, Stavrakoudis et al. 2019, Filipponi 2019) or through neural networks (Farasin et al. 2019). Generally, they are validated on a few samples, obtained from Landsat sensors (e.g. MODIS) or a limited set of certified delineation maps obtained from Copernicus EMS.

In this work, we exploit a particular subset of Deep neural networks: the Convolutional Neural Networks (CNNs), which in the last years outperformed most of the challenges in image processing. We intend to bring their potential in finding burned regions. Moreover, we provide a complete performance evaluation, over a dataset of 147 wildfires delineated and certified by domain experts, contributing to Copernicus EMS. In literature, approaches such as AlexNet (Krizhevsky et al. 2012), Inception (Szegedy, Vanhoucke, et al. 2016), VGG (Simonyan and Zisserman 2014), GoogleNet (Szegedy, Liu, et al. 2015), ResNet (He et al. 2016) and Xception (Chollet 2017) are robust and largely adopted in a variety of applications related to classification tasks related to images. For instance, given a picture, these techniques are very precise in determining information for the whole image, i.e. whether there is a cat, a dog or - in our context - a wildfire. However, their application is limited to a single classification for the whole picture. Certainly, this is useful for detection, but not for the delineation of an area, where every pixel needs to be labeled as a burned or unburned region. Therefore, we decided to experiment with two techniques, U-Net and CuMedVision, which were invented - and currently applied - for bio-medical topics, with the same purpose: given a picture, state the belonging class of each pixel. Variations of the U-Net are currently applied in the geospatial context of land cover mapping for urban environments (McGlinchy et al. 2019), for sea and land segmentation (Chu et al. 2019) or time-series of satellite acquisitions (Stoian et al. 2019).

## COLLECTING BURNED AREAS INFORMATION

This section covers the main methods and data sources that have been exploited to retrieve the information related to specific geographic areas affected by wildfires.

### ESA Sentinel-2 satellites

The Copernicus Programme, an initiative led by the European Commission (EC) in partnership with the European Space Agency (ESA), is an earth observation program created to improve the surveillance and management of the environment, to understand and mitigate the effects of climate change and ensure civil security through the information gathered by different families of satellites, called Sentinels (*United space in Europe - Overview* 2019).

The information exploited in this work is acquired from the family of Sentinel-2 satellites. They are equipped with high-resolution, multi-spectral imaging sensors and a high revisit time (~2-3 days at European latitudes), aimed at monitoring variability in land surface conditions. Each satellite carries an optical instrument payload that samples 13 spectral bands, at different spatial resolutions: four bands at 10 m, six bands at 20 m and three bands at 60 m. The 13 bands are described in Table 1 (*Sentinel 2 Spatial Resolution* 2019). The orbital swath width is 290 km (*MultiSpectral Instrument (MSI) overview* 2019).

In particular, among the two final available products, we downloaded the Sentinel-2 Level-1C product, composed of 100x100 $km^2$ tiles (ortho-images in UTM/WGS84 projection), resulting from the use of a Digital Elevation Model (DEM) to project the image in cartographic geometry. Per-pixel radiometric measurements are provided in Top Of Atmosphere (TOA) reflectances along with the parameters to transform them into radiances (*Level-1C products* 2019). Currently, those products are very large (600 MB for each tile) and thus hard to manage: therefore, we used Sinergise Sentinel-Hub Service (*Sentinel-Hub* 2019), an engine for processing petabytes of satellite data that handles the complexity of management of raw data internally, making Earth observation imagery easily accessible for browsing, visualization, and analysis through a standard Web Service (*EO Browser* 2019) or API.

### Copernicus Emergency Management System

The Copernicus Programme provides several services for a range of different applications (i.e. analysis, monitoring and forecasting of natural hazards or air-quality) to manage and protect the environment and its natural resources

**Table 1. Bands description**

| Band | Description | Central Wavelength ($\mu m$) | Spatial resolution ($m$) |
|---|---|---|---|
| 1 | Coastal aerosol | 0.443 | 60 |
| 2 | Blue | 0.490 | 10 |
| 3 | Green | 0.560 | 10 |
| 4 | Red | 0.665 | 10 |
| 5 | Vegetation red edge | 0.705 | 20 |
| 6 | Vegetation red edge | 0.740 | 20 |
| 7 | Vegetation red edge | 0.783 | 20 |
| 8 | Near Infrared (NIR) | 0.842 | 10 |
| 8A | Narrow NIR | 0.865 | 20 |
| 9 | Water vapour | 0.945 | 60 |
| 10 | Short wavelength infrared (SWIR) | 1.375 | 60 |
| 11 | SWIR | 1.610 | 20 |
| 12 | SWIR | 2.190 | 20 |

and ensure civil security. Among these, the Copernicus Emergency Management System (EMS) (*Copernicus Emergency Management System (EMS)* 2019) is the service intended to map, gather and provide information for emergency management covering natural hazards such as flood, earthquake, and fire. Copernicus EMS allows Authorized Users (National Focal Points in the EU Member States and countries participating in the Copernicus Programme, European Commission services and the European External Action Service) to issue the request of information and mappings over a specific AOI, triggering a so-called EMS Activation, concerning an emergency. The EMS offers two types of services: Rapid Mapping supplies maps of the AOI within hours or days from the request, to support emergency management activities immediately following a disaster, while Risk & Recovery Mapping is related to prevention, preparedness, disaster risk reduction, and recovery phases activities. Moreover, Copernicus EMS collects and stores all these maps produced in response to EMS activations, making them available for free.

For each *activation*, a variable number of maps is produced, depending on the number of AOIs and the time interval specified in the initial request. The maps used in this work are Delineation Maps: they provide an assessment of the event's geographic extent, based on satellite images acquired immediately after the disaster event (*EMS Rapid Mapping Product Portfolio* 2019). They are created through a semi-automatic approach, that requires human intervention for fine-tuning and validation.

## DATASET

This section introduces the sources of information used in this work. Starting from a past wildfire event, reported through a Delineation Map, we accessed to the information about: (i) the area/s of interest, determined by a couple of coordinates (*latitude*, *longitude*), which delineates the region/s affected by fire and (ii) the date on which Copernicus EMS based for creating the delineation map. Following this information, we collected: (i) Sentinel-2 data for the acquisition of raw data about the soil at the time the burned region was present and (ii) the delineation map itself, to create a binary image (a picture containing only black or white pixels), used for the training and evaluation phases of the approach we propose in this work.

### Sentinel-2 L1C data

A total of 223 areas of interest were downloaded from Copernicus EMS. Starting from these, we downloaded the corresponding satellite acquisitions on all the 13 spectral bands from SentinelHub. We performed all the requests by setting a constraint on the maximum cloud coverage value at 10% to limit the presence of clouds in the pictures. Furthermore, we manually discarded all the images (i) where clouds partially or hid the burned area and (ii) with missing information. The latter option is caused by how satellite imagery is acquired: based on the specified AOI, a satellite image may be obtained by combining different stripes according to the orbit of the satellite and sometimes, for a portion of the specified AOIs, there is no updated information available. After the data cleaning operations, the number of AOIs was reduced to 147.

From the downloaded data, we generated two different datasets: (i) the first one, containing all the 13 spectral bands, in which, for each band, the values were in the range [0, 255] and (ii) the second one, containing only the information in the visible spectrum related to the three additive primary colors (Red, Green, Blue) and corresponding to the bands 4, 3, and 2, combined to generate a single RGB picture for each acquisition.

### Ground Truth Extraction

The ground truth chosen for the evaluation of the proposed approach was taken from the official Delineation Maps provided by Copernicus EMS. From each delineation map, we generated a binary picture fully superimposable on the Sentinel-2 acquisitions, in which each pixel states whether it belongs to a burned or unburned region. In the first case, the pixels assume value 255 (white color); otherwise, they assume value 0 (black color).

## METHODOLOGY

### Data pre-processing

The Data pre-processing phase is needed to suitably prepare the data to increase the effectiveness of the deep learning approach used in this work. Firstly, for each satellite acquisition, we transformed the domain range of each spectral band, from $[0, 255]$ to $[-1, 1]$, through the Min-Max normalization. Furthermore, to improve the model generalization and to enhance its potential accuracy, we applied data augmentation techniques to the training set (Wang and Perez 2017). Considering the context and the irregular shapes of burned areas, different kinds of transformations can be applied. Each transformation is applied to both the input image (either RGB or multispectral acquisition) and the binary mask. We implemented four transformations: random rotation, random horizontal flip, random vertical flip, and random shear. The aforementioned operations are well-known in literature as *traditional transformations* (Mikołajczyk and Grochowski 2018). They are commonly used to increase the variability of the dataset to improve the model generalization during the training phase. The parameters we used in this phase are reported in Table 2.

**Table 2. Data augmentation parameters**

| Transformation | Probability | Parameters |
|---|---|---|
| Random rotation | 50% | Angle: [-50°, +50°] |
| Random horizontal flip | 50% | - |
| Random vertical flip | 50% | - |
| Random shear | 50% | Angle: [-20°, +20°] |

All the transformations have a 50% probability of being applied. In the positive case for random rotation and random shear, a random angle is generated for each tuple image and mask within the specified range.

### Models

In this work, the Deep Learning models considered to solve the burned area identification problem are namely the U-Net (Ronneberger et al. 2015) and the CuMed Vision (Chen et al. 2017) architectures, based on Convolutional Neural Networks (CNNs) and created in the context of biomedical image segmentation to identify biological cells, thus aimed at pixel-wise classification.

#### *U-Net*

The U-Net architecture is composed of two subsequential paths of convolutional and pooling layers, composing a U-shaped form: in the contraction path, the number of feature channels increases while input dimensions decrease; in the expansion path, instead, dimensionality increases and the number of feature channels decreases, returning in this way a spatial localization of the features. In this work we followed, for the greatest part, the original version: it is composed of 10 Convolutional Layers in the Contracting path and 8 in the Expanding path. It uses Rectified Linear Units (ReLUs) as activation functions, with max-pooling layers operating on a 2x2 Pool Size with Stride=2 with Batch Normalization. Variations from the original version are related to (i) the net input and output dimensions and (ii) the loss function. The net input is adapted to the dimension of the acquired data, namely 512x512x13 in the first dataset (13 bands) and 512x512x3 in the second one (bands 4, 3 and 2). Consequently, the net output has

the same number of pixels of the input, but with a single channel. Therefore, the net output dimension will be 512x512x1, through the last Convolutional Layer and using the Sigmoid function as activation.

Concerning the loss function, the original version used a Cross-Entropy loss combined with a pixel-wise weight map, which penalized more errors made in contour pixels. In the biological context, cells used to be of similar dimensions and shapes, with regular "borders" and very close to each other. Instead, in our context there might be only one burned region per acquisition; otherwise, the burned regions might be of very different dimensions and shapes, in both cases with very irregular borders. Computing weight maps in our context would lead most of the values near zero, thus making training neural networks more difficult and more error-prone. Therefore, we chose the Binary Cross-Entropy as the loss function.

#### *CuMed Vision1*

CUMed Vision1 architecture is made of two main components: a single downsampling path, consisting of 5 max-pooling layers and 5 convolutional layers extracts the high-level abstraction information reducing the resolution of the input images; then, 3 upsampling path that aim to produce an end-to-end output, which means outputting a labeled binary map for each input image, with its same resolution, through a convolutional layer using the Sigmoid Function as Activation. This is made to improve performances during the training process: through multi-task learning, better intermediate features are learned such that the performances of the overall model are boosted. The labeled maps obtained by the decoders are lastly summed together, and a final Sigmoid function is applied, resulting in the final Segmentation map. The loss function used, given the aforementioned reasons, is the Binary CrossEntropy. The input is adapted to the dimension of the acquired data (as explained for the U-Net model), respectively 512x512x3 for the RGB input or 512x512x13 for the Multi-Spectral input, while the output consists of a binary picture with the dimension of 512x512x1 pixels.

## EXPERIMENTS

### Baseline

As said, Sentinel-2 acquires information at 13 different wavelengths, including not visible spectral bands. Combining them, it is possible to obtain indices useful for the identification of some specific areas, like burned area or water presence. As introduced in *Related works*, indices like BAI, NBR, NBR2, and BAIS2 are widely adopted in this domain. We adapted the indices computed specifically for Landsat (BAI, NBR, NBR2) to use the same spectral wavelengths acquired by Sentinel-2. Therefore, we report the indices formulas as follows:

$$\mathrm{BAI} = \frac{1}{(0.1 - \mathrm{B04})^2 + (0.06 - \mathrm{B08})^2}$$

$$\mathrm{NBR} = \frac{\mathrm{B08} - \mathrm{B12}}{\mathrm{B08} + \mathrm{B12}}$$

$$\mathrm{NBR2} = \frac{\mathrm{B11} - \mathrm{B12}}{\mathrm{B11} + \mathrm{B12}}$$

$$\mathrm{BAIS2} = \left(1 - \sqrt{\frac{\mathrm{B06} \cdot \mathrm{B07} \cdot \mathrm{B8A}}{(B04)}}\right) \cdot \left(\frac{\mathrm{B12} - \mathrm{B8A}}{\sqrt{\mathrm{B12} + \mathrm{B8A}}} + 1\right)$$

Given these burned area indices, we computed the Separability Index to quantify the burned-unburned regions separability. Its values are in the range $[0, \infty)$, generally values greater than 1 express an easier separability and, therefore, potentially better results for classification (Lasaponara 2006).

$$\mathrm{SI} = \frac{|\mu_a - \mu_b|}{\sigma_a + \sigma_b}$$

As shown in Table 3, the analysis over the whole dataset showed that the NBR2 index is the one giving the best SI among all the others (SI=0,928). Therefore, we implemented an approach from literature (Bin et al. 2019), which leverages on the Otsu Method (Otsu 1979), to determine a threshold value (for each acquisition) to generate delineation maps which distinguish between burned and unburned areas. The F1-Scores resulting from these with respect to the ground truth were then used as a baseline for our models' performance evaluation.

| | BAI | NBR | NBR2 | BAIS2 |
|---|---|---|---|---|
| SI | 0.773 | 0.553 | **0.928** | 0.862 |

**Table 3. Separability Index computed on the images produced by the indices NBR, NBR2, BAI and BAIS2.**

### Training phase

The training of both models was performed through the use of a standard Adam Optimizer, using the Binary CrossEntropy as Loss Function. For each of the cross-validation folds, the training is performed over the entire dataset with Batch Size=8, the maximum value possible due to computation limitations, over a variable number of epochs established through the Early Stopping technique, that prevent overfitting of the models ending the training when the performance has stopped improving on a held-out validation set. The probability maps produced by the models are converted in labeled binary mask through threshold operations on the probability value of 0.5, is the most conservative among the possible choices.

### Testing phase

In order to estimate the results achieved by each model on different morphologies and geographic areas, the performances of the networks are computed using a cross-validation approach (Stone 1974), specifically a k-Fold technique, setting k to the number of groups considered, i.e. $k = 7$. For each fold, we train the models on $k - 1$ groups and we compute the F1-Score on the $k^{th}$ one. The same approach is used for the evaluation of each model, to compare their performances.

**Table 4. Folds description**

| Fold | Activation codes | Number of images |
|---|---|---|
| 1 | EMSR370, EMSR299, EMSR252, EMSR288, EMSR316, EMSR365, EMSR344, EMSR363, EMSR303, EMSR362, EMSR278, EMSR373, EMSR371, EMSR247, EMSR221, EMSR173, EMSR360, EMSR331, EMSR307, EMSR367, EMSR230 | 21 |
| 2 | EMSR207, EMSR133 | 21 |
| 3 | EMSR169, EMSR209, EMSR290 | 21 |
| 4 | EMSR213, EMSR217, EMSR302 | 20 |
| 5 | EMSR239, EMSR237, EMSR214, EMSR219, EMSR368, EMSR305, EMSR353, EMSR175, EMSR295 | 21 |
| 6 | EMSR369, EMSR171, EMSR250, EMSR248, EMSR210, EMSR254, EMSR211 | 21 |
| 7 | EMSR132, EMSR216, EMSR227, EMSR298, EMSR291, EMSR300 | 22 |

As described in Table 4, reporting also the cardinality of each group, the 7 groups are divided according to EMS Activation codes: each geographical region belongs to one single group to avoid overlapping issues and the presence of morphologically similar regions in different groups.

### Results evaluation

In this subsection, we first introduce the evaluation measures and the overall results on both the case studies: (i) only on Visible Wavelengths (RGB) and (ii) Multispectral case, in which all the 13 bands are used. Then, we provide deeper details about the approach which scored best in our evaluation.

#### *Evaluation criteria*

This task deal with imbalanced data: the majority of pixels do not belong to burned areas (85,23% compared to just 14,77% of burned area pixels). Since a model could reach 85% of Accuracy just labeling all the pixels as 'not

(a) (b) (c)

**Figure 1. Binary UNet RGB predictions from best fold: (a) RGB input; (b) Ground Truth; (c) Model prediction**

burned', Accuracy is not a reliable metric to evaluate results. Therefore, the principal metric used to make this evaluation is the F1-Score, defined as the harmonic mean of Precision and Recall metrics.

### *Results*

The overall results are shown in Table 5. The baseline, which leverages on the NBR2 index, shows an acceptable SI and the highest Recall measure among all the examined approaches. However, it resulted to be very imprecise, bringing to the lowest F1-Score (= 49,1%). The CuMedVision1 resulted to have good separability indices and showed similar performances in both the case studies, increasing more than 6% of F-Score when using all the spectral bands. However, the U-Net was the technique that performed best under almost all the considered measures, showing reliable performances (~70%) using information from visible light and increasing more than 12% of F1-Score when all the spectral bands are available. Therefore, in the remaining part of this section, we provide further details on this last technique.

**Table 5. Evaluation metrics for the two case studies. The Separability Index is in range $[0, \infty)$, while Precision, Recall and F1-score are in range $[0, 100]$.**

| Approach | Visible wavelengths (Bands 4, 3, 2) | | | | All spectral bands | | | |
|---|---|---|---|---|---|---|---|---|
| | SI | Precision | Recall | F1-Score | SI | Precision | Recall | F1-Score |
| Baseline | - | - | - | - | 0.93 | 34,0% | **88,4%** | 49,1% |
| U-Net | **1.11** | **70,6%** | **69,8%** | **70,2%** | **1.58** | 88,5% | 77,1% | **82,4%** |
| CuMedVision1 | 1.09 | 67,8% | 68,8% | 68,3% | 1.36 | **93,0%** | 60,9% | 73,6% |

Table 6 reports the performance reached by the U-Net, with a mean F1-score of 70,2%, with the best score of 91,6%.

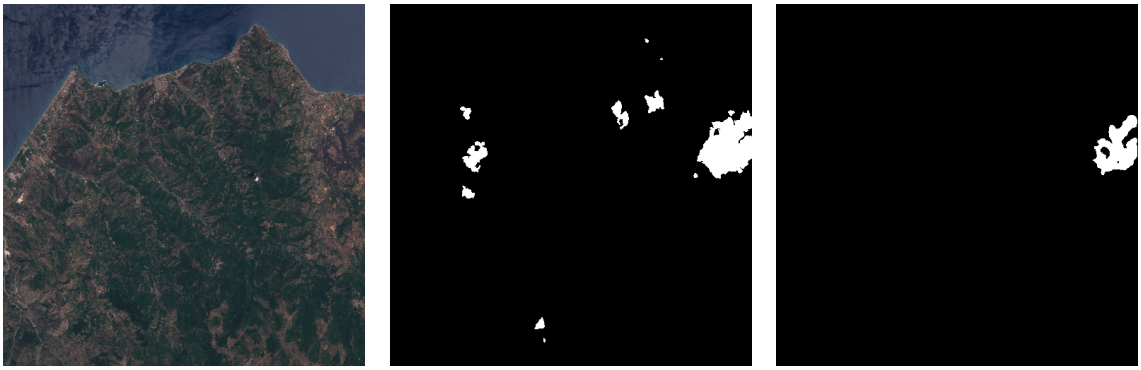
| | Accuracy | Recall | Precision | F1-score |
|---|---|---|---|---|
| Mean | 90,4% | 69,8% | 70,6% | 70,2% |
| Best | 96,3% | 93,4% | 89,9% | 91,6% |
| Worst | 82,4% | 39,0% | 49,2% | 43,5% |

**Table 6. UNet model evaluation metrics, RGB input**

Figure 1 and Figure 2 shows examples of the model outputs obtained from the best and the worst performing folds respectively.

Light conditions affect prediction outcomes and training process: depending on the time of the day at which the satellite acquisition is made and the morphology of the area, the presence of shadows caused by mountainous areas may lead the neural network to label undamaged areas as burned.

Analyzing the performance obtained by the same model over the Multiple bands' input, including all the 13 input spectral bands, the same model achieved an overall mean F1-score over all the folds of 82,4%, a substantial increase compared to 70,2% obtained in the RGB input case. Considering the best performance obtained over the folds, F1-score incremented from 91,6% to 95,8%, as described in Table 7.



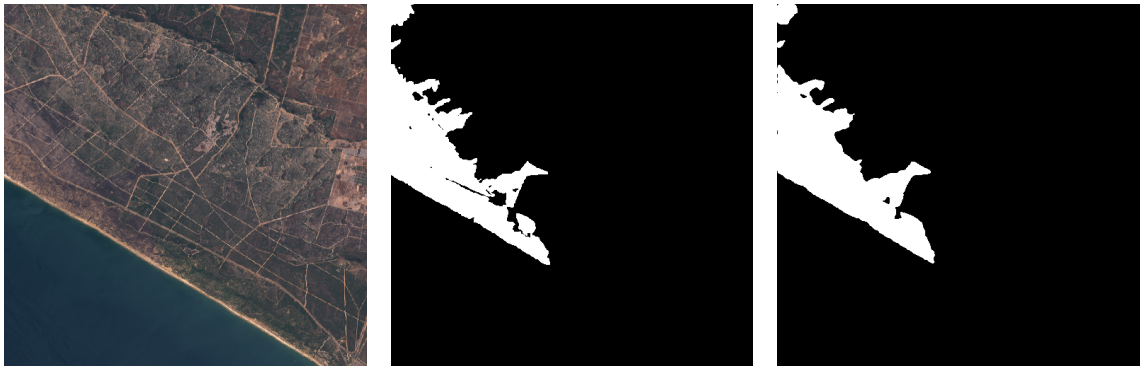
(a) (b) (c)

**Figure 2. Binary UNet RGB predictions from worst fold: (d) RGB input; (e) Ground Truth; (f) Model prediction**

| | Accuracy | Recall | Precision | F1-score |
|---|---|---|---|---|
| Mean | 95,6% | 77,1% | 88,5% | 82,4% |
| Best | 97,6% | 96,4% | 95,3% | 95,8% |
| Worst | 93,4% | 44,4% | 79,6% | 57,0% |

**Table 7. UNet model evaluation metrics, Multi-spectral input**

In particular, using this input data, the model was able to determine burned areas more precisely, clearly identify the presence of water in the image, not misclassifying it as a burned area and, lastly, become more robust for atmospheric noise and light conditions, which were limiting factors when considering performances in the RGB scenario. Figure 3 shows an example of the output generated by the U-Net (trained with multi-spectral input), which can be qualitatively compared with the ground truth.



(a) (b) (c)

**Figure 3. Binary UNet Multi-spectral input predictions from best fold: (a) Multi-spectral input; (b) Ground Truth; (c) Model prediction**

In general, the possibility to use the multi-spectral input bands severely improved performances and burned area delineation proficiency in contour regions, but the UNet model still lacks the ability to precisely determine small burned areas.

## CONCLUSIONS AND FUTURE WORKS

The analysis of satellite imagery exploiting machine learning and especially deep learning techniques is still at the early stages of development and many improvements are possible. The development of machine learning algorithms for emergency management systems is extremely important, especially nowadays, due to the high number of forest wildfires which are endangering the planet environment. This work shows the feasibility of an automatic burned area detection but also some limits of the considered architectures in this context: the U-Net and CuMed

architectures were created in the context of biomedical image segmentation, where cells and organs are defined by specific membranes or tissues, whereas burned areas characterized by different shapes and extensions. To overcome performances issues, it is possible to work either on input data or on the network architecture, and possible future works will be focused on the following purposes:

- Sentinel2-L1C products are affected by atmospheric noise which can negatively influence the training process and the outcome. To overcome this issue, it could be possible to use Sentinel2-L2A products, which provide extra information about the bottom of atmosphere reflectance, which sensibly reduces the noise from acquisitions; Another possibility is to train a neural network to remove noise from Sentinel2-L1C products: if both L1C and L2A data are available, it is possible to train a custom neural network which learns how to clean the input data. Then, the cleaned data is used to feed and train a UNet. Eventually, the input feature space could be enriched with data from the Sentinel-1 mission.
- To reduce misclassification errors due to light conditions and to enhance the identification of small burned areas, another possibility is to modify the current neural network architecture, either by changing the encoder-decoder structure or by developing ad-hoc solutions for the satellite imagery. Moreover, many misclassification errors were made in mountainous and cultivated areas due to the presence of shadows and great changes during summer and winter season respectively: the development of a custom loss function may also provide benefits both in terms of training time required and in model performances.

This work proves the possibility to exploit satellite data to detect regions hit by wildfires in an automated way, through the use of Deep Learning techniques. They achieve considerable results in comparison with certified mappings of the events, which are more precise, but more expensive in terms of manual intervention and time required to be created. The 13 spectral bands acquired by Sentinel-2 sensors exploited by UNet achieve the best performances in terms of F1-Score (82, 4% in the mean case). However, considering only a limited number of features (3 spectral bands of the visible light) allows achieving acceptable performances of F1-score (70, 2% in the mean case) in burned area delineation. In the latter case, the performances are acceptable to encourage the assessment of the proposed approach with other sources of information, such as cameras mounted on aircraft or drones, to achieve a more frequent and versatile observation of hazard events.

## ACKNOWLEDGEMENT

This work was partially funded by the European Commission through the SHELTER project, grant agreement n.821282.

## REFERENCES

Bin, W., Ming, L., Dan, J., Suju, L., Qiang, C., Chao, W., Yang, Z., Huan, Y., and Jun, Z. (2019). "A Method of Automatically Extracting Forest Fire Burned Areas Using Gf-1 Remote Sensing Images". In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 9953–9955.

Boschetti, L., Roy, D. P., Justice, C. O., and Humber, M. L. (2015). "MODIS–Landsat fusion for large area 30 m burned area mapping". In: *Remote Sensing of Environment* 161, pp. 27–42.

Chen, H., Qi, X., Yu, L., Dou, Q., Qin, J., and Heng, P.-A. (2017). "DCAN: Deep contour-aware networks for object instance segmentation from histology images". In: *Medical image analysis* 36, pp. 135–146.

Chollet, F. (2017). "Xception: Deep learning with depthwise separable convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258.

Chu, Z., Tian, T., Feng, R., and Wang, L. (2019). "Sea-Land Segmentation With Res-UNet And Fully Connected CRF". In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 3840–3843.

Chuvieco, E., Martin, M. P., and Palacios, A. (2002). "Assessment of different spectral indices in the red-near-infrared spectral domain for burned land discrimination". In: *International Journal of Remote Sensing* 23.23, pp. 5103–5110.

*Copernicus* (2019). https://www.copernicus.eu/en.

*Copernicus Emergency Management System (EMS)* (2019). http://emergency.copernicus.eu/.

De Araujo, F. M. and Ferreira, L. G. (2015). "Satellite-based automated burned area detection: A performance assessment of the MODIS MCD45A1 in the Brazilian savanna". In: *International Journal of Applied Earth Observation and Geoinformation* 36, pp. 94–102.

*EMS Rapid Mapping Product Portfolio* (2019). https://emergency.copernicus.eu/mapping/sites/default/files/files/CopernicusEMS-Service_PortfolioRapid$_M apping.pdf$.

*EO Browser* (2019). https://www.sentinel-hub.com/explore/eobrowser.

Escuin, S., Navarro, R., and Fernandez, P. (2008). "Fire severity assessment by using NBR (Normalized Burn Ratio) and NDVI (Normalized Difference Vegetation Index) derived from LANDSAT TM/ETM images". In: *International Journal of Remote Sensing* 29.4, pp. 1053–1073.

*Euronews* (2019). https://www.euronews.com/2019/08/15/there-have-been-three-times-more-wildfires-in-the-eu-so-far-this-year.

*Europe wildfires: Norway police evacuate hundreds in Sokndal* (2019). https://www.bbc.com/news/world-europe-48035682.

*European Forest Fire Information System (EFFIS)* (2019). https://effis.jrc.ec.europa.eu/.

*European heatwave: Spain battles major Catalonia wildfire* (2019). https://www.bbc.com/news/world-europe-48790242.

Farasin, A., Nini, G., Garza, P., and Rossi, C. (2019). "Unsupervised Burned Area Estimation through Satellite Tiles: A multimodal approach by means of image segmentation over remote sensing imagery". In:

Filipponi, F. (2018). "BAIS2: Burned Area Index for Sentinel-2". In: *Multidisciplinary Digital Publishing Institute Proceedings*. Vol. 2. 7, p. 364.

Filipponi, F. (2019). "Exploitation of Sentinel-2 Time Series to Map Burned Areas at the National Level: A Case Study on the 2017 Italy Wildfires". In: *Remote Sensing* 11.6, p. 622.

Garcia, M. L. and Caselles, V. (1991). "Mapping burns and natural reforestation using Thematic Mapper data". In: *Geocarto International* 6.1, pp. 31–37.

*Greece wildfires leave blackened forests in their wake* (2019). https://www.bbc.com/news/world-europe-49343572.

Hardtke, L. A., Blanco, P. D., Valle, H. F. del, Metternicht, G. I., and Sione, W. F. (2015). "Semi-automated mapping of burned areas in semi-arid ecosystems using MODIS time-series imagery". In: *International Journal of Applied Earth Observation and Geoinformation* 38, pp. 25–35.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Hughes, M., Kaylor, S., and Hayes, D. (2017). "Patch-based forest change detection from Landsat time series". In: *Forests* 8.5, p. 166.

*JRC Technical Report - Forest Fires in Europe, Middle East and North Africa* (2018). https://effis.jrc.ec.europa.eu/media/cms_page_media/40/Annual_Report_2018_final_pdf_06.11.2019_n208KFB.pdf.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*, pp. 1097–1105.

Lasaponara, R. (2006). "Estimating spectral separability of satellite derived parameters for burned areas mapping in the Calabria region by using SPOT-Vegetation data". In: *Ecological Modelling* 196.1-2, pp. 265–270.

*Level-1C products* (2019). https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types/level-1c.

McGlinchy, J., Johnson, B., Muller, B., Joseph, M., and Diaz, J. (2019). "Application of UNet Fully Convolutional Neural Network to Impervious Surface Segmentation in Urban Environment from High Resolution Satellite Imagery". In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, pp. 3915–3918.

Mikołajczyk, A. and Grochowski, M. (2018). "Data augmentation for improving deep learning in image classification problem". In: *2018 international interdisciplinary PhD workshop (IIPhDW)*. IEEE, pp. 117–122.

*MultiSpectral Instrument (MSI) overview* (2019). https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument.

Otsu, N. (1979). "A threshold selection method from gray-level histograms". In: *IEEE transactions on systems, man, and cybernetics* 9.1, pp. 62–66.

Ramo, R. and Chuvieco, E. (2017). "Developing a random forest algorithm for MODIS global burned area classification". In: *Remote Sensing* 9.11, p. 1193.

Ramo, R., Garcia, M., Rodriguez, D., and Chuvieco, E. (2018). "A data mining approach for global burned area mapping". In: *International journal of applied earth observation and geoinformation* 73, pp. 39–51.

RMSRF (2019). *Copernicus EMS Rapid Mapping Service Request Form*. URL: https://emergency.copernicus.eu/mapping/ems/how-use-service.

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer, pp. 234–241.

Roteta, E., Bastarrika, A., Padilla, M., Storm, T., and Chuvieco, E. (2019). "Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa". In: *Remote sensing of environment* 222, pp. 1–17.

Roy, D. P., Huang, H., Boschetti, L., Giglio, L., Yan, L., Zhang, H. H., and Li, Z. (2019). "Landsat-8 and Sentinel-2 burned area mapping-A combined sensor multi-temporal change detection approach". In: *Remote Sensing of Environment* 231, p. 111254.

*Sentinel 2 Spatial Resolution* (2019). https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial.

*Sentinel-2* (2019). https://sentinel.esa.int/web/sentinel/missions/sentinel-2.

*Sentinel-Hub* (2019). https://www.sentinel-hub.com/.

Shan, T., Wang, C., Chen, F., Wu, Q., Li, B., Yu, B., Shirazi, Z., Lin, Z., and Wu, W. (2017). "A Burned Area Mapping Algorithm for Chinese FengYun-3 MERSI Satellite Data". In: *Remote Sensing* 9.7, p. 736.

Simonyan, K. and Zisserman, A. (2014). "Very deep convolutional networks for large-scale image recognition". In: *arXiv preprint arXiv:1409.1556*.

Stavrakoudis, D., Katagis, T., Minakou, C., and Gitas, I. Z. (2019). "Towards a fully automatic processing chain for operationally mapping burned areas countrywide exploiting Sentinel-2 imagery". In: *Seventh International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2019)*. Vol. 11174. International Society for Optics and Photonics, p. 1117405.

Stoian, A., Poulain, V., Inglada, J., Poughon, V., and Derksen, D. (2019). "Land cover maps production with high resolution satellite image time series and convolutional neural networks: Adaptations and limits for operational systems". In: *Remote Sensing* 11.17, p. 1986.

Stone, M. (1974). "Cross-validatory choice and assessment of statistical predictions". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2, pp. 111–133.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2015). "Going deeper with convolutions". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826.

*United space in Europe - Overview* (2019). https://www.esa.int/Applications/Observing$_t he_E arth/Copernicus/Overview$3.

Verhegghen, A., Eva, H., Ceccherini, G., Achard, F., Gond, V., Gourlet-Fleury, S., and Cerutti, P. (2016). "The potential of Sentinel satellites for burnt area mapping and monitoring in the Congo Basin forests". In: *Remote Sensing* 8.12, p. 986.

Wang, J. and Perez, L. (2017). "The effectiveness of data augmentation in image classification using deep learning". In: *Convolutional Neural Networks Vis. Recognit*.