CaBuAr: California burned areas dataset for delineation [Software and Data Sets]

This article has been accepted for publication in IEEE Geoscience and Remote Sensing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MGRS.2023.3292467

CaBuAr: California Burned Areas dataset for delineation

Daniele Rege Cambrin, Luca Colomba, Paolo Garza

***Abstract*—Forest wildfires represent one of the catastrophic events that, over the last decades, caused huge environmental and humanitarian damage. In addition to a significant amount of carbon dioxide emission, they are a source of risk to society in both short-term (e.g., temporary city evacuation due to fire) and long-term (e.g., higher risks of landslides) cases. Consequently, the availability of tools to support local authorities in automatically identifying burned areas plays an important role in the continuous monitoring requirement to alleviate the aftereffects of such catastrophic events. The great availability of satellite acquisitions coupled with computer vision techniques represents an important step in developing such tools. This paper introduces a novel open dataset that tackles the burned area delineation problem, a binary segmentation problem applied to satellite imagery. The presented resource consists of pre- and post-fire Sentinel-2 L2A acquisitions of California forest fires that took place from 2015 to 2022. Raster annotations were generated from the data released by California's Department of Forestry and Fire Protection. Moreover, in conjunction with the dataset, we release three different baselines based on spectral indexes analyses, SegFormer, and U-Net models.**

***Index Terms*—Earth Observation, Deep Learning, Burned Area Delineation, Semantic segmentation**

I. Introduction

The Earth Observation (EO) field has greatly increased the number of applications in the last decades thanks to the greater data availability, storage capacity, and computational power of modern systems. In fact, leveraging data acquired by Sentinel [1], Landsat [2], and MODIS [3] missions as an example, it is possible to retrieve information at a continental scale in a short amount of time. This, in conjunction with the development of modern methodologies in the field of machine learning and deep learning, represents an extremely interesting area of research for scientists and authorities from different fields, such as governments and first responders involved in disaster response and disaster recovery missions. Phenomena such as climate change and extreme climate events have a tremendous societal, economic, and environmental impact, also leading to humanitarian and environmental losses (e.g., higher risk of landslide due to a forest fire). Indeed, leveraging EO and modern deep learning methodologies can provide useful tools in the area of disaster management and disaster recovery. Within the research community, numerous previous works proved the effectiveness of computer vision architectures in the field of disaster response, such as flood delineation [4], change detection [5], [6] and burned area delineation [7]–[9]. This paper fits in the last mentioned context. Specifically, we release a dataset to tackle the burned area delineation problem, i.e., a binary image segmentation problem that aims to identify areas damaged by a forest wildfire. Tackling such a problem with modern methodologies requires great data availability. However, the time and cost needed to produce high-quality annotations severely limit the ability to investigate ad-hoc solutions in the EO field. For these reasons, we propose a new dataset related to forest fires in California, collecting data from the Sentinel-2 mission [10]. The dataset is publicly available to the research community at https://huggingface.co/datasets/DarthReca/california_burned_areas. Compared to the few other datasets about wildfires [11], [12], our dataset covers a larger area and spans more years.

Ground truth masks for the task of binary image segmentation were generated starting from the public vector data provided by California's Department of Forestry and Fire Protection [13] and rasterized. Satellite acquisitions, i.e., the raw input data, were instead collected from the Sentinel-2 L2A mission through Copernicus Open Access Hub. More precisely, we collected and released both pre-fire and post-fire information associated with the same area of interest.

The contributions of this paper can be summarized as follows:

- A novel image segmentation dataset tailored to burned area delineation consisting of Sentinel-2 pre- and post-fire acquisitions. We provided more samples than existing datasets to facilitate the training of (large) deep learning models.
- Three different baselines were evaluated on the proposed dataset: one consisting of the evaluation of several burned area indexes and the Otsu's automatic thresholding method [14], one based on the SegFormer model [15], and one based on the U-Net model [16].

The paper is structured as follows. Section II introduces the related works; Section III introduces the collected dataset and the preprocessing steps performed, whereas Section IV and Section V formally introduce the task and the experimental settings and results. Finally, Section VI concludes the paper.

II. Related Works

Before the development of deep learning-based methodologies, domain experts based their analyses on satellite imagery leveraging spectral index computation and evaluation. Considering the SAR context, thresholding-based techniques have been adopted to distinguish between flooded and unflooded areas [17]. Different analyses have been performed on various tasks concerning several spectral indexes such as in cloud

Daniele Rege Cambrin, Luca Colomba and Paolo Garza are with Politecnico di Torino, Turin, Italy

This article has been accepted for publication in IEEE Geoscience and Remote Sensing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MGRS.2023.3292467

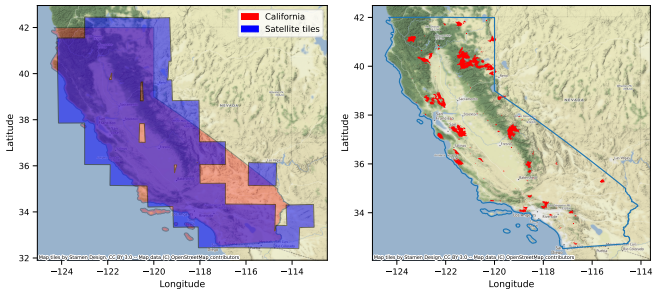
detection (cloud mask) [18], water presence (WP, NDWI) [19], [20] and vegetation analysis (NDVI) [21].

Considering the burned area delineation problem, domain experts have developed several indexes: NBR, NBR2, BAI, and BAIS2 [19]. They are computed using different spectral bands to generate an index highlighting the affected areas of interest. Such techniques are often coupled with thresholding methodologies: either fixed or manually calibrated threshold values are chosen [22], or automatic thresholding algorithms are used [23]. Additional studies evaluate index-based techniques with additional in-situ information, namely the Composite Burned Area Index (CBI), which indeed provides insightful information but does not represent a scalable solution because in-situ data are incredibly costly to collect. Furthermore, studies confirmed that finding a unique threshold that is region- and vegetation-independent is difficult [24]. These methods assume that burned and unburned areas are linearly separable, which is usually untrue.

More recently, researchers started adopting supervised learning techniques to solve several tasks in computer vision and EO. More precisely, CNN-based models proved their effectiveness in image classification and segmentation tasks, achieving state-of-the-art performances compared to index-based methodologies [25], [26]. Deep models proved their effectiveness in similar tasks covering wildfire detection [27] and spreading [28], too. The main drawback is the need for a significant amount of labeled data, possibly covering heterogeneous regions with different morphological characteristics, to learn better representations. Over the years, many of the proposed frameworks limit their analyses to a few samples collected from a limited number of countries or locations [29]. In a few cases, larger datasets were adopted to tackle the semantic segmentation problem, without disclosing the dataset [30].

In the EO domain, different public datasets are available to the research community tackling different problems, such as flood delineation [17], [31], deforestation [32], wild areas monitoring [33], sustainable development goal monitoring [34], crop classification and segmentation [35] but, to the best of our knowledge, only two public datasets are available for the burned area delineation problem covering some countries in Europe [11] and Indonesia [12]. Our dataset collects more data than the formers, considering more wildfires and a larger area. It comprises pre- and post-fire Sentinel-2 L2A data about California forest fires.

Table I shows a comparison between the three datasets. The proposed dataset consists of the highest number of considered wildfires (340), globally covering the largest amount of burned areas (28 million pixels covering 11.000 km^2) and a higher total covered surface (450000 km^2). Figure 1a shows the covered areas. Even though the proposed dataset has the greatest amount of burned surface, it achieves the lowest percentage of burned area compared to the others. However, the CaBuAr dataset provides the highest number of training samples in supervised (supervised learning in binary segmentation, highest amount of burned areas) and unsupervised cases (self-supervised learning, highest area covered). It is notorious that more curated data provide better machine learning models,



(a) Tiles map (b) Wilfires map

Fig. 1: In (a) is shown the satellite tiles coverage: California administrative boundaries (red) vs satellite tiles of the proposed dataset (blue). In (b) is shown the location of the wildfires (red) inside the California boundaries (blue).

TABLE I: Comparison between datasets. TD is the time difference between pre-fire and post-fire acquisitions. MP expresses the number of burned pixels in millions. The highest value in each line is highlighted in bold, except for the resolution case in which the lowest numerical value is highlighted.

	CaBuAr (ours)	[11]	[12]
Region	California	Europe	Indonesia
Mission	Sentinel-2	Sentinel-1/2	Landsat-8
Resolution	20m	**10m** (S2)	30m
Image size	**5490 × 5490**	up to 5000 × 5000	512 × 512
Raw data	✓	✗	✓
Channels	**12**	**12**	8
Forest Fires	**340**	73	227
Start-End date	**Jan, 2015 - Dec, 2022**	July, 2017 - July, 2019	Jan, 2019 - Dec, 2021
Total surface	∼**450000 km^2**	∼19000 km^2	∼46000 km^2
Burned surface	∼**28 MP/∼11000 km^2**	∼20 MP/∼2000 km^2	∼8 MP/∼7000 km^2
Post-fire	✓	✓	✓
Pre-fire	✓	✓	✗
TD	∼1 year	$\leq$ 2 months	/

and the dataset provides many ready-to-use samples without leveraging other sources. Images are larger in terms of pixels (5490) and disclosed as raw data in the original and unaltered state, as directly collected from satellite instrumentation. On the other hand, the European dataset provides data collected from a third-party service for which preprocessing operations are performed. The availability of raw data enables researchers to apply the preferred preprocessing steps without any loss of information. Furthermore, the monitored range of dates of the new dataset spans from 2015 to 2022, whereas the other two datasets span a smaller time period.

III. Dataset

The newly created dataset comprises L2A products of Sentinel-2, a European Space Agency (ESA) mission. The area of interest is California, with the geographical distributions of the events shown in Figure 1b. We collected images of the same area before and after the wildfire. It is essential to note that the L2A product contains RGB channels and other spectral bands in the infrared region and ultra blue for a total of 12 channels. Depending on the band, they have a resolution of $10m$, $20m$, or $60m$ per pixel.

This article has been accepted for publication in IEEE Geoscience and Remote Sensing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MGRS.2023.3292467

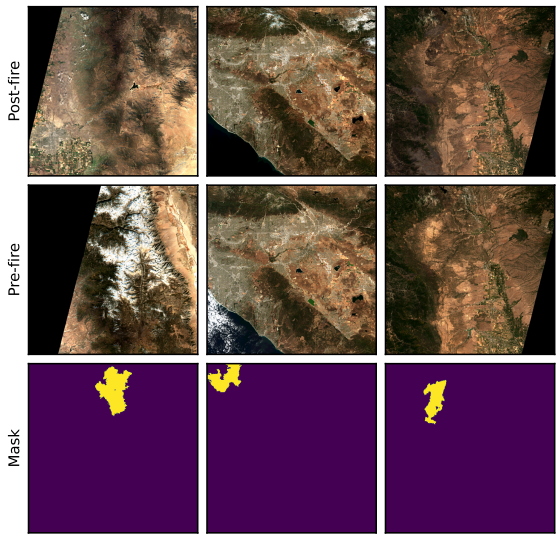


Fig. 2: Example of pre-fire and post-fire RGBs and relative masks.

A. Preprocessing

The California Department of Forestry and Fire Protection publicly provides the ground truth vector information, which we converted into raster images. Each pixel contains a binary value: 1 in the burned area and 0 in the case of undamaged areas. Although the registered wildfires span from 1898 to 2022, we collected data only for wildfires from 2015 to 2022 because there were no Sentinel-2 images before 2015. We gathered the Sentinel-2 images directly from Copernicus Open Access Hub.

To minimize the effects of vegetation regrowth and post-wildfire modifications, images are collected within one month after the wildfires are fully contained and extinguished. A total number of 340 acquisitions associated with 340 wildfires were downloaded, each being of size 5490×5490 pixels with a resolution of $20m$ per pixel. The few Sentinel-2 bands with different resolutions were either upsampled or downsampled with bicubic interpolation to reach the target resolution.

Pre-fire images have the same size and resolution as the post-fire acquisitions. To enforce coherence and similar seasonal and phenological conditions, we downloaded pre-fire data considering a temporal window of 4 weeks, centered one year before the date post-fire data were collected. For example, given a post-fire acquisition collected in 2018/04/01, we downloaded the products available between 2017/03/18 and 2017/04/15, with center 2017/04/01. This ensures similar climatic and seasonal conditions, thus limiting environmental changes as much as possible. In some cases, retrieving these products was impossible due to data unavailability, i.e., not all wildfires have a pre-fire acquisition satisfying such a constraint. Given the 340 wildfires considered in this study, 208 have pre-fire availability satisfying the abovementioned constraint.

The dataset was randomly split into five non-overlapped folds to perform cross-validation.

TABLE II: Association between code and comment.

comment	meaning
0	Affected area is in the incomplete region
1	Image is incomplete
2	Small burned area
3	Mask has a small offset
4	Mask is totally wrong
5	Extensive burned area
6	Clouds over the burned area
7	Too many clouds over the image
8	Wildfire ongoing
9	Snow on the burned area
10	Mask seems smaller than the burned area
11	Mask seems bigger than the burned area
12	Mask is in the missing data area
13	Part of the mask is outside the area

B. Manual inspection

After collecting data, we manually evaluated each post-fire image using RGB channels. This was done to (i) discard invalid samples and (ii) enrich the dataset with metadata and comments based on our subjective evaluation. We remark that such comments are not helpful for the final prediction task but can be used to better characterize the data. Our evaluation is associated with each satellite acquisition.

Each image has a metadata field with a list of numeric codes generated from the manual inspection. Figure II reports the code-to-comment association. As can be seen, different climatic conditions can be found in the dataset. Figure 3 reports some examples of post-fire images. For each post-fire acquisition, Figure 3 reports its RGB version (first line), its binary mask (second line), and the comment(s) assigned to it (on top of the RGB image). For instance, the second acquisition has two comments: 2 and 11. We noted that some masks seem to overestimate the burned area. However, our perception refers to the RGB version of the images, i.e., to a subset of the available information. Moreover, our subjective perception can be biased also because the regions at the borders of burned areas are usually less damaged than the central ones. These notes must be extended to other mask-related comments, but they are rarer. Comments are almost equally distributed among the folds. All the 340 acquisitions do not include any comments that can negatively affect results and the dataset's quality (i.e., 4, 8, and 12).

Finally, each pre-fire image was manually inspected to verify its validity, but no new comment types were added. All invalid pre-fire acquisitions were discarded.

IV. Tasks

The proposed dataset can be used as a benchmark for different tasks in supervised and unsupervised scenarios. Having at our disposal two sets of images, called PS and PR, containing post-fire samples and pre-fire ones, respectively, the tasks we considered in this paper can be formulated as follows:

(a) Binary segmentation through machine learning methods based on post-fire acquisitions only. It involves a supervised learning algorithm (A_S) to perform pixel-level prediction based on samples from PS. A_S labels pixels of images as burned or undamaged, creating a binary mask for each new image.

This article has been accepted for publication in IEEE Geoscience and Remote Sensing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MGRS.2023.3292467

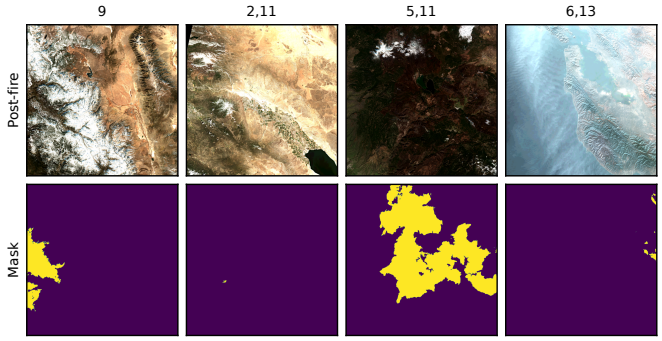


Fig. 3: Sample of post-fire RGBs and masks with the associated comments.

(b) Binary segmentation through machine learning models based on pre-fire and post-fire acquisitions. It involves a supervised learning algorithm (A_S) to perform pixel-level prediction considering samples from PS and PR.

(c) Binary segmentation through spectral indexes. It involves a spectral index (S_I) designed for burned area identification. Taking samples from PS, S_I outputs a value for each pixel, creating a matrix PS'. Then a binary mask burned-unburned can be made thresholding PS'.

(d) Binary segmentation through differential spectral indexes. It involves a spectral index (S_I) designed for burned area identification based on the comparison of pre and post-wildfire images. Taking samples from PS and PR, S_I outputs a value for each pixel creating a matrix PS' and PR'. Then a binary mask burned-unburned can be made thresholding the difference $PS' - PR'$.

V. Experiments

Our experiments test various classical threshold-based and deep learning-based methods considering three different data settings:

1) Usage of all the available post-fire images (Setting 1, tackling Tasks (a) and (c)).
2) Usage of the subset of post-fire images for which the corresponding pre-fire image is available, without using the pre-fire image to train the models (Setting 2, tackling Tasks (a) and (c)).
3) Usage of post-fire and pre-fire images. Thus, two input images are considered for each area. Spectral indexes in this setting were evaluated by computing the difference between pre-fire and post-fire indexes (Setting 3, tackling Tasks (b) and (d)).

The code for the experiments can be found at https://github.com/DarthReca/CaBuAr.

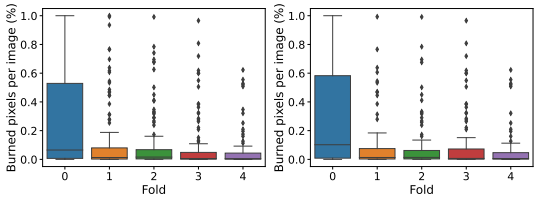
A. Experimental settings

The encoder of SegFormer is initialized with the original weights for Image-Net duplicated four times to handle the 12 available channels for Sentinel-2 L2A acquisitions. U-Net is instead randomly initialized. The batch size was set to 8. We used the AdamW optimizer with an initial learning rate of 0.001, decreased by a factor of 10 every 15 epochs, and a weight decay of 0.01 for every considered model. We used the well-known Dice loss [36] as the loss function. All models were trained on one Tesla V100 32GB GPU. The testing was made using the weights associated with the best validation loss.

Due to the size of the original input images (5490×5490), we split them into patches of size 512×512. Furthermore, due to class imbalance, we kept only those patches containing at least one pixel associated with the positive class and no clouds over the area of interest (Comment 6 in Table II). A total of 534 patches for setting (1) and 356 for settings (2) and (3) were obtained.

The statistics and performances reported in the remainder of the paper refer to the data obtained after the split-and-filter process mentioned earlier. All training and evaluation procedures were performed with a cross-validation approach. The same criterion was applied for spectral indexes methodologies to obtain comparable results, despite the absence of a trainable model. The reported values are expressed as mean and standard deviation computed over the five folds.

In Figure 4, we highlight the percentage of burned pixels per image in each fold. Even if data were split randomly, Fold 0 is characterized by a larger variability in terms of the number of burned pixels per image.



(a) Setting (1). (b) Settings (2) and (3).

Fig. 4: Burned pixels percentage per image per fold.

TABLE III: Spectral bands used by indexes. VNIR (Visible and Near InfraRed), SWIR (ShortWave InfraRed) and Visible are exploited.

Index	Ultra-Blue	Visible	VNIR	SWIR
NBR	✗	✗	B8	B12
NBR2	✗	✗	✗	B11,B12
BAI	✗	B4	B8	✗
BAIS2	✗	B4	B8A,B7,B6	B12

B. Spectral Indexes

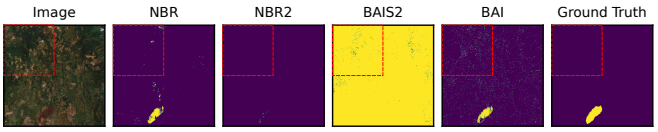
We evaluated several spectral indexes (NBR, NBR2, BAI, BAIS2) for the burned area delineation task. In Table III, we summarized the spectral bands exploited by the various indexes, and it is possible to note many bands are common to many of them. They take as input some bands of a Sentinel product and output a value for each pixel. This value is generally thresholded to create a binary mask providing the burned/unburned information for each pixel.

In particular, to assess the performances on the dataset, we computed the Separability Index [19] (see Table IV),

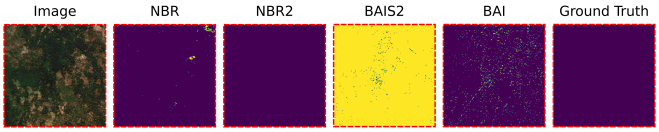
This article has been accepted for publication in IEEE Geoscience and Remote Sensing Magazine. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MGRS.2023.3292467

TABLE IV: Separability Indexes (SI) and metrics computed for each setting and each evaluated index.

Setting	Index	SI	F1 Score	IoU
(1)	NBR	**0.294**	0.150±0.231	0.103±0.180
	NBR2	0.224	**0.226±0.269**	**0.159±0.209**
	BAI	0.044	0.040±0.121	0.026±0.086
	BAIS2	0.027	0.194±0.292	0.148±0.252
(2)	NBR	0.320	0.106±0.196	0.071±0.150
	NBR2	**0.349**	**0.243±0.278**	**0.172±0.218**
	BAI	0.052	0.037±0.115	0.024±0.079
	BAIS2	0.002	0.086±0.174	0.057±0.138
(3)	dNBR	**0.247**	0.114±0.212	0.079±0.168
	dNBR2	0.189	**0.218±0.281**	**0.157±0.225**
	dBAI	0.040	0.066±0.161	0.045±0.127
	dBAIS2	0.027	0.047±0.126	0.030±0.099



(a) Predictions



(b) Zoomed predictions

Fig. 5: Example of segmentation using Otsu's method for different indexes. In (b), the zoom was applied to the top-left corner of images in (a), boxed in *red*. This is done to show more clearly an area with false positive predictions.

which quantifies how well the index under analysis discerns between burned and unburned regions, i.e., a higher value of SI implies that classes are more separable from each other. We apply Otsu's thresholding method to quantify the indexes' segmentation performances. Results are shown in Table IV, which confirm the poor performances in terms of F1-Score and IoU. Additionally, the availability of pre-fire images (Setting (3)) does not significantly improve the evaluation metrics.

Figure 5 shows an example of predictions for the cited indexes applying Otsu's method. In this example case, BAI and NBR achieve the best scores, but many "disturbances" affect the final result in the unburned regions. Figure 5b, which is a zoom on an actual unburned area, shows that many false positive points are inside the considered unburned area. We refer to this situation as "disturbances".

C. Deep learning models

We tested two deep learning architectures for semantic segmentation: a CNN (U-Net [16]) and a Vision Transformer (SegFormer [15]). We decided to take into account two different versions of SegFormer (*B0*, the smallest version, and *B3*, a mid-range version) that differ only in size and so in the number of parameters. U-Net, SegFormer-B0, and SegFormer-B3 consist of 31M, 3.8M, and 47M parameters, respectively.

TABLE V: Metrics calculated for each deep learning model evaluated.

Setting	Metric	SegFormerB3	SegFormerB0	U-Net
(1)	F1 Score	0.620±0.009	0.686±0.004	**0.707±0.004**
	IoU	0.497±0.008	0.563±0.004	**0.583±0.004**
(2)	F1 Score	0.583±0.014	0.654±0.003	**0.705±0.002**
	IoU	0.447±0.012	0.535±0.003	**0.577±0.002**
(3)	F1 Score	0.533±0.003	0.499±0.009	**0.625±0.002**
	IoU	0.401±0.003	0.370±0.007	**0.494±0.002**

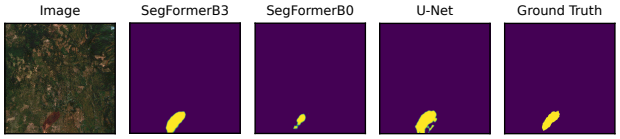


Fig. 6: Examples of prediction with deep learning models.

To deal with Setting (3), the two input images (pre- and post-fire) are concatenated along the channel axis, creating patches of size $24 \times 512 \times 512$ ($C \times H \times W$). We reported the results for the different settings and models in Table V.

Without any specific pre-training, U-Net provides in every setting the best performance. SegFormer-B0, which is also lighter than U-Net, provides comparable performance, having some difficulties only with Setting (3). SegFormer-B3 does not justify the greater complexity considering its results. In this case, pre-fire images do not provide any improvements too. This can be justified because of the curse of dimensionality that affects almost all machine learning models. In fact, the concatenation approach we applied increases the number of features without increasing the number of input samples. An open research direction is the design of more sophisticated models to exploit both images effectively.

Figure 6 reports the predictions of these models on the same input sample shown in Figure 5. The most evident difference is deep models tend to be more precise and less affected by false positives in the unburned areas (i.e., there are fewer "disturbances"). Looking attentively at Figure 5b, many false positive points can be seen. The greatest problem of thresholding techniques is that they try to find a linear separation between classes, which is frequently unrealistic. This is why deep learning models, which support non-linearities, perform better [37], [38]. The substantial variability of the threshold technique, shown by the higher variance (see Table IV), could be caused by the fact some areas are more separable than others.

VI. Conclusion

This paper introduced a new dataset for burned area delineation containing samples with different morphological features of the terrain and different sizes of burned regions. The dataset includes both pre-fire and post-fire data.

We provided baselines based on different approaches to assess the quality of basic methods and encourage further research activities. This publicly available dataset can benefit researchers and public authorities for further tasks, such as recovery planning, constant monitoring of affected areas, and

developing deep learning models for burned area delineation. We plan to extend the dataset to new regions and satellite acquisitions continuously. The collection of satellite acquisitions is made publicly available to encourage future use and research activities.

REFERENCES

[1] Sentinel missions. https://sentinel.esa.int/web/sentinel/missions.

[2] Landsat mission. https://landsat.gsfc.nasa.gov/.

[3] MODIS. https://www.earthdata.nasa.gov/sensors/modis.

[4] Z. Dong, G. Wang, S. O. Y. Amankwah, X. Wei, Y. Hu, and A. Feng, "Monitoring the summer flooding in the Poyang Lake area of China in 2020 based on Sentinel-1 data and multiple convolutional neural networks," *International Journal of Applied Earth Observation and Geoinformation*, vol. 102, p. 102400, 2021.

[5] A. Asokan and J. Anitha, "Change detection techniques for remote sensing applications: a survey," *Earth Science Informatics*, vol. 12, no. 2, pp. 143–160, 2019.

[6] J. Sublime and E. Kalinicheva, "Automatic post-disaster damage mapping using deep-learning techniques for change detection: Case study of the Tohoku tsunami," *Remote Sensing*, vol. 11, no. 9, p. 1123, 2019.

[7] R. Lasaponara and B. Tucci, "Identification of burned areas and severity using SAR Sentinel-1," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 917–921, 2019.

[8] D. R. Cambrin, L. Colomba, and P. Garza, "Vision transformers for burned area delineation," in *Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 2022.

[9] M. A. Tanase, M. A. Belenguer-Plomer, E. Roteta, A. Bastarrika, J. Wheeler, Á. Fernández-Carrillo, K. Tansey, W. Wiedemann, P. Navratil, S. Lohberger *et al.*, "Burned area detection and mapping: Intercomparison of Sentinel-1 and Sentinel-2 based algorithms over tropical Africa," *Remote Sensing*, vol. 12, no. 2, p. 334, 2020.

[10] Sentinel-2 mission guide. https://sentinel.esa.int/web/sentinel/missions/sentinel-2. Accessed on: 2023/04/14.

[11] L. Colomba, A. Farasin, S. Monaco, S. Greco, P. Garza, D. Apiletti, E. Baralis, and T. Cerquitelli, "A dataset for burned area delineation and severity estimation from satellite imagery," in *CIKM2022*, ser. CIKM '22. ACM, 2022, p. 3893–3897.

[12] Y. Prabowo, A. D. Sakti, K. A. Pradono, Q. Amriyah, F. H. Rasyidy, I. Bengkulah, K. Ulfa, D. S. Candra, M. T. Imdad, and S. Ali, "Deep learning dataset for estimating burned areas: Case study, indonesia," *Data*, vol. 7, no. 6, 2022.

[13] California department of forestry and fire protection. https://www.fire.ca.gov/. Accessed on: 2023/04/14.

[14] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[15] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, "SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers," in *Advances in Neural Information Processing Systems*, vol. 34. Curran Associates, Inc., 2021, pp. 12 077–12 090.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Lecture Notes in Computer Science*. Springer International Publishing, 2015, pp. 234–241.

[17] D. Bonafilia, B. Tellman, T. Anderson, and E. Issenberg, "Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020.

[18] R. A. Frey, S. A. Ackerman, R. E. Holz, S. Dutcher, and Z. Griffith, "The continuity modis-viirs cloud mask," *Remote Sensing*, vol. 12, no. 20, 2020.

[19] F. Filipponi, "Bais2: Burned area index for sentinel-2," *Proceedings*, vol. 2, no. 7, 2018.

[20] A. Fisher, N. Flood, and T. Danaher, "Comparing landsat water index methods for automated water classification in eastern australia," *Remote Sensing of Environment*, vol. 175, pp. 167–182, 2016.

[21] N. Pettorelli, J. O. Vik, A. Mysterud, J.-M. Gaillard, C. J. Tucker, and N. C. Stenseth, "Using the satellite-derived ndvi to assess ecological responses to environmental change," *Trends in Ecology & Evolution*, vol. 20, no. 9, pp. 503–510, 2005.

[22] L. Saulino, A. Rita, A. Migliozzi, C. Maffei, E. Allevato, A. P. Garonna, and A. Saracino, "Detecting burn severity across mediterranean forest types by coupling medium-spatial resolution satellite imagery and field data," *Remote Sensing*, vol. 12, no. 4, 2020.

[23] W. Bin, L. Ming, J. Dan, L. Suju, C. Qiang, W. Chao, Z. Yang, Y. Huan, and Z. Jun, "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*, 2019, pp. 9953–9955.

[24] C. A. Cansler and D. McKenzie, "How robust are burn severity indices when applied in a new region? Evaluation of alternate field-based and remote-sensing methods," *Remote sensing*, vol. 4, no. 2, pp. 456–483, 2012.

[25] L. Knopp, M. Wieland, M. Rättich, and S. Martinis, "A deep learning approach for burned area segmentation with Sentinel-2 data," *Remote Sensing*, vol. 12, no. 15, p. 2422, 2020.

[26] A. Farasin, L. Colomba, G. Palomba, G. Nini, and C. Rossi, "Supervised Burned Areas delineation by means of Sentinel-2 imagery and Convolutional Neural Networks," in *ISCRAM 2020, Virginia Tech, Blacksburg, VA, USA*, 2020, pp. 24–27.

[27] D. Rashkovetsky, F. Mauracher, M. Langer, and M. Schmitt, "Wildfire detection from multisensor satellite imagery using deep semantic segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7001–7016, 2021.

[28] F. Huot, R. L. Hu, N. Goyal, T. Sankar, M. Ihme, and Y.-F. Chen, "Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[29] Q. Safder, H. Zhang, and Z. Zheng, "Burnt area segmentation with densely layered capsules," in *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 2022, pp. 2199–2202.

[30] D. Rashkovetsky, F. Mauracher, M. Langer, and M. Schmitt, "Wildfire detection from multisensor satellite imagery using deep semantic segmentation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7001–7016, 2021.

[31] F. Montello, E. Arnaudo, and C. Rossi, "Mmflood: A multimodal dataset for flood delineation from satellite imagery," *IEEE Access*, vol. 10, pp. 96 774–96 787, 2022.

[32] D. John and C. Zhang, "An attention-based u-net for detecting deforestation within satellite sensor imagery," *International Journal of Applied Earth Observation and Geoinformation*, vol. 107, p. 102685, 2022.

[33] B. Ekim, T. T. Stomberg, R. Roscher, and M. Schmitt, "Mapinwild: A remote sensing dataset to address the question of what makes nature wild [software and data sets]," *IEEE Geoscience and Remote Sensing Magazine*, vol. 11, no. 1, pp. 103–114, 2023.

[34] C. Yeh, C. Meng, S. Wang, A. Driscoll, E. Rozi, P. Liu, J. Lee, M. Burke, D. B. Lobell, and S. Ermon, "SustainBench: Benchmarks for Monitoring the Sustainable Development Goals with Machine Learning," in *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 12 2021. [Online]. Available: https://openreview.net/forum?id=5HR3vCylqD

[35] D. Sykas, M. Sdraka, D. Zografakis, and I. Papoutsis, "A sentinel-2 multi-year, multi-country benchmark dataset for crop classification and segmentation with deep learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022.

[36] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer International Publishing, 2017, pp. 240–248.

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

[38] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," *Applied Soft Computing*, vol. 70, pp. 41–65, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494618302813