# Wildfire Detection From Multisensor Satellite Imagery Using Deep Semantic Segmentation

Dmitry Rashkovetsky, Florian Mauracher, Martin Langer, and Michael Schmitt, *Senior Member, IEEE*

***Abstract*—Deriving the extent of areas affected by wildfires is critical to fire management, protection of the population, damage assessment, and better understanding of the consequences of fires. In the past two decades, several algorithms utilizing data from Earth observation satellites have been developed to detect fire-affected areas. However, most of these methods require the establishment of complex functional relationships between numerous remote sensing data parameters. In contrast, more recently, deep learning has found its way into the application, having the advantage of being able to detect patterns in complex data by learning from examples automatically. In this article, a workflow for the detection of fire-affected areas from satellite imagery acquired in the visible, infrared, and microwave domains is described. Using this workflow, the fire detection potentials of four sources of freely available satellite imagery were investigated: the C-SAR instrument on board Sentinel-1, the multispectral instrument on board Sentinel-2, the sea and land surface temperature instrument on board Sentinel-3, and the MODIS instrument on board Terra and Aqua. For each of them, a single-input convolutional neural network based on the well-known U-Net architecture was trained on a newly created dataset. The performance of the resulting four single-instrument models was evaluated in presence of clouds and in clear conditions. In addition, the potential of combining predictions from pairs of single-instrument models was investigated. The results show that fusion of Sentinel-2 and Sentinel-3 data provides the best detection rate in clear conditions, whereas the fusion of Sentinel-1 and Sentinel-2 data shows a significant benefit in cloudy weather.**

***Index Terms*—Data fusion, deep learning, remote sensing, wildfire detection.**

## I. Introduction

WILDFIRES are a natural component of the Earth system. They are important for vegetation growth, release of nutrients on the forest floor, and help in maintaining a balanced forest ecosystem. However, wildfires are also one of the most devastating natural hazards in the world. They contribute to global warming [1], destroy property, and lead to tremendous economical losses and ultimately to loss of human and animal lives and destruction of communities. In 2016, the annualized economic burden of wildfires in the USA alone was estimated to between $71.1 billion and $347.8 billion [2].

In many cases, wildfires occur in remote locations, which makes their early detection by means of *in situ* observations difficult. Thus, remote sensing has become a go-to solution for wildfire detection with satellite data [3]–[5]. However, wildfires are complex physical processes that involve different interrelated factors, including weather, topography, soil moisture, type of fuel, and the location and type of the source. In addition, wildfires are dynamic and can occur in different spatial and temporal scales. Thus, accurate detection and segmentation of wildfires using traditional remote sensing algorithms is often difficult as it requires modeling of complex functional relationships between numerous data. Cloud coverage and smoke, generated during the burning process, furthermore restrict the ability to detect fires from space [6]. All this makes wildfires a challenging phenomenon to map, which is a crucial element in an effective response of the responsible authorities and understanding the consequences of fires [7]. Luckily, the increasing abundance of remote-sensing data together with the dramatic improvement in computational capabilities that occurred in recent years have led to an improved possibility to exploit data-centric approaches for wildfire detection and mapping [8], which make use of machine learning (ML) and multisensor data fusion. The ability of ML algorithms to automatically uncover complex, spatiotemporal patterns in the data, with less need for hand-crafted expert descriptors makes them a favorable candidate for the task [9].

The main challenges in segmentation of wildfires can be summarized as follows.

1) Wildfires are a complex phenomenon varying in time and space that involve numerous interrelated factors difficult to model by classical non-ML models.
2) Because of the dynamic nature of wildfires, following acquisitions over the same area observe the wildfire at different states, which makes fusion of multiple acquisitions, labeling, and validation of results difficult [10].
3) Often, thick smoke and clouds obscure the fire signal observed by optical instruments.
4) In most cases, fire segmentation is an imbalanced classification problem, where the number of pixels affected by fire is significantly smaller than the number of pixels, which are not [11].

With this work, we investigate the use of modern deep learning techniques to detect wildfires in both cloudless and cloudy

Manuscript received March 4, 2021; revised May 19, 2021 and June 24, 2021; accepted June 27, 2021. Date of publication June 30, 2021; date of current version July 22, 2021. This work was supported by the Munich University of Applied Sciences and the German Research Foundation through the "Open Access Publishing" program. *(Corresponding author: Michael Schmitt.)*

Dmitry Rashkovetsky, Florian Mauracher, and Martin Langer are with the Orbital Oracle Technologies GmbH, 80992 Munich, Germany (e-mail: dmitry.rashkovetsky@ororatech.com; florian.mauracher@ororatech.com; martin.langer@ororatech.com).

Michael Schmitt is with the Department of Geoinformatics, Munich University of Applied Sciences, 80335 Munich, Germany (e-mail: michael.schmitt@hm.edu).

Digital Object Identifier 10.1109/JSTARS.2021.3093625

conditions using four sources of satellite imagery with different spatial and spectral resolutions in a combined manner: the C-SAR instrument on board Sentinel-1 A/B, the multispectral instrument (MSI) on board Sentinel-2 A/B, the sea and land surface temperature instrument (SLSTR) on board Sentinel-3 A/B, and the MODIS instrument on board Terra and Aqua. Specifically, we aim to exploit multisensor data to improve results in cloudy conditions and deep segmentation to overcome the necessity of physically modeling wildfires. In summary, the main contributions of our work are the following.

1) We demonstrate the generation of an annotated wildfire dataset combining openly available satellite imagery and information from a public wildfire database.
2) In this context, we provide information about possibilities and limitations to combine data from multiple sensors for a single wildfire given different resolutions and repeat cycles.
3) We investigate the predictive potential of the individual satellite data sources using a fairly standard deep semantic segmentation architecture.
4) We investigate the use of decision fusion to combine the individual single-sensor predictions and quantify its benefit, in particular having the problem of cloud cover in mind.

## II. Related Work

Active fire (AF) detection and burned area (BA) segmentation using remote sensing techniques have been the focus of many research works since as early as the 1970s with the launch of Landsat-1 in 1972. To better understand fire detection from space, it is important first to distinguish between two types of products that are the goal of most research works in the field. The definitions used here are based on [4].

1) AF products describe geographic locations, which are actively burning as a result of a fire. AF products are usually detected by thermal sensors as thermal anomalies; however, the detection is only possible if the satellite overpasses the area that is actively burning.
2) BA products describe geographic locations where fire led to the burning of biomass, which resulted in deposit of char and ash on the ground. The resulting patterns are sometimes also referred to as "burn scars," and they are typically more persistent in time than the thermal anomalies caused by ongoing fires.

Throughout this work, the term fire-affected pixel is used to describe pixels that are either actively burning or are burnt as a result of a fire. Additionally, the terms instrument and sensor are used interchangeably.

In recent years, several algorithms that attempt to combine sensors from multiple spectral domains were suggested. These methods have several potential advantages: First, they allow us to exploit the cloud-penetrating capabilities of microwave sensors in combination with the ability of sensors in the visible light and infrared domains to detect thermal anomalies. Second, combining multiple sensors improves the satellite overpass frequency over a burning area. Finally, an improved temporal and spatial resolution BA and AF products can potentially be achieved by combining images from high-spatial-resolution sensors such as Sentinel-2 with high-temporal-resolution sensors such as Sentinel-3 or sensors on board geostationary satellites. Verhegghen *et al.* [12] were one of the first to suggest combining time series of SAR $C$-band images from Sentinel-1 with multispectral imagery from Sentinel-2 to improve BA estimation in cloudy conditions. BAs were detected in Sentinel-2 images using thresholding of spectral indices, and significant changes in the backscattering coefficient of Sentinel-1 images were used as a gap-filling detector. Crowley *et al.* [13] applied Bayes' theorem approach [14] to combine potential BA detections from Landsat-8 OLI, Sentinel-2 MSI, and MODIS instruments to delineate the BA and monitor the progress of the 2017 Elephant Hill fire in California. One of the most recent studies by Lasko [15] combined Sentinel-1 SAR imagery with a monthly averaged MODIS BA product to reduce the burn date uncertainty of the MODIS BA product caused by observations obscured by clouds. Significant decrease in backscatter between two SAR images was used to detect BA-affected pixels in SAR images, which allowed to decrease the burning date uncertainty.

One can also approach the problem of wildfire detection as an anomaly or target detection problem. In essence, wildfires can be considered as spatial and thermal anomalies that can be detected. Classic anomaly detection algorithms that are applied to wildfire detection can usually be described as a combination of one or more of the following approaches: 1) global thresholding algorithms that identify global anomalies by applying absolute global thresholds to spectral indices that are derived from the spectral data recorded by the satellite [16], [17]; 2) contextual algorithms that identify local anomalies by applying dynamic thresholds that are calculated based on the spectral data in the pixel's neighborhood [18], [19]; and 3) time-series algorithms that detect significant changes in one or more of the measured properties between two acquisitions taken at different times [20], [21]. In recent years, several deep learning methods were used to identify anomalies with spatiotemporal networks (STNs) attempting to combine learning spatial features using convolutional neural networks and learning temporal features using long short-term memory networks [22]. These methods, however, are usually aimed at detecting outliers with a limited spatial signature, which is not the case for large wildfires. Additionally, STNs usually require time-series data, which are not available for most short-burning wildfires. Finally, performance of CNN-based methods for anomaly detection, which lie in the core of most STNs, is still an active and relatively young area of research [23].

This application-oriented work aims to explore the potential of applying simple deep-learning-based image segmentation for the detection of wildfires using single-temporal data from instruments measuring in the visible light, infrared, and microwave spectral regions. In this context, we aim to investigate if combining data from more than one instrument can improve the detection results. While there are several state-of-the-art approaches for multimodal learning methodology that allow us

to fuse multi-instrument images on the data level [24]–[27], our goal is not a proposal for a new multimodal learning approach but rather to show that fusion of openly available multisensor satellite data can improve the detection of wildfires, and how it can be achieved by using standard deep convolutional semantic segmentation networks.

## III. Satellite Instruments and Reference Data

In the presented study, we have employed the visible, infrared, and microwave instruments onboard the Sentinel-1, Sentinel-2, Sentinel-3, Terra, and Aqua satellites. Key properties of the instruments used are described in the next subsections.

### A. Sentinel-1 Mission and C-SAR Instrument

The Sentinel-1 mission is based on a constellation of two twin satellites. Both satellites operate in a near-polar Sun-synchronous orbit with 12-day repeat cycle and are equipped with a $C$-band SAR instrument [28]. For the study area used in this work (California, USA), the coverage frequency of the constellation, without considering repetitive or relative orbits, is two to four days. The revisit frequency with an ensured same repetitive orbit is 12 days. The C-SAR instrument operates in the 5.405-GHz frequency $C$-band, which corresponds to a 5.54-cm wavelength. The instrument operates in one of the four exclusive imaging modes, of which the so-called *interferometric wide swath* (IW) mode is employed in this work. The IW mode is the default operation mode of the mission. In this mode, wide swaths composed of three subswaths are acquired by electronically steering of the antenna. A single-look full-resolution IW image is 5-m ground range resolution and 20-m azimuth resolution. In the IW mode, the C-SAR instrument is able to acquire data in either single copolarization (HH or VV) or dual polarization (HH+HV or VV+VH). For the purpose of this work, medium-resolution IW imagery with $40 \times 40$ $\mathrm{m}^2$ ground resolution and VV+VH polarization [29] was chosen for establishing the Sentinel-1 dataset.

### B. Sentinel-2 Mission and MSI

The Sentinel-2 mission is based on a constellation of two twin satellites, flying in a Sun-synchronous orbit with an average altitude of 786 km, phased at 180°, and a descending node at 10:30 local time. The revisit frequency of the constellation is five days at the equator [30]. Both satellites are equipped with an MSI. The MSI measures the reflected radiance from Earth in 13 spectral bands from VIS to SWIR, with spatial resolution varying from 10 to 60 m, depending on the spectral band. The instrument acquires images with a swath of 290 km. Sentinel-2 products are a compilation of elementary granules of fixed size ($100 \times 100\,\mathrm{km}^2$) orthorectified in UTM/WGS84 projection. The products are available in two levels of processing: Level-1C and level-2A. Level-1C product contains measurements of the top-of-atmosphere (TOA) reflectance along with the parameters to transform them into radiances, with subpixel multispectral registration. Level-2A provides orthorectified bottom-of-atmosphere reflectance, with subpixel multispectral registration. In addition to the reflectance measurements, both products contain cloud masks [31]. For the purpose of this work, Level-1C TOA data were used.

### C. Sentinel-3 Mission and SLSTR

The Sentinel-3 mission is based on a constellation of two multi-instrument satellites aimed at measuring sea-surface topography, sea- and land-surface temperature, ocean color, and land color. Both satellites operate in a Sun-synchronous orbit at an average altitude of 814.5 km and a descending node of 10:00 local time. The orbit of Sentinel-3B is identical to the orbit of Sentinel-3A but with a phase difference of 140°. From the four main instruments onboard Sentinel-3, we have used the SLSTR, capable of acquiring images in 11 spectral bands ranging from 0.55 to 10.95 $\mu$m for the purpose of this work. The SLSTR instrument is a conical scanning imaging radiometer with a dual (oblique and nadir) view technique, employed in the along-track direction. The orbit of the constellation allows a revisit time of less than a day. The Level-1B observation product of the instrument contains TOA radiances in the VIS and SWIR spectral regions, and TOA brightness temperatures in the TIR spectral regions (bands 7–9 and F1, F2). As data basis for the presented study, geolocated, calibrated radiances, acquired in a nadir swath and recorded in Level-1B observation product, were used.

### D. Terra and Aqua Satellites and MODIS Instrument

The two MODIS instruments on board the Terra and Aqua satellites provide data in 36 spectral bands in wavelengths ranging from 0.4 to 14.4 $\mu$m, with ground resolution ranging from 250 to 1000 m. Both satellites orbit the Earth in a Sun-synchronous, near-polar, circular orbit at an average altitude of 705 km, with descending nodes of 10:30 and 13:30 for Terra and Aqua, respectively [32]. The Terra satellite passes from north to south over the equator at approximately 10:30 local time and Aqua passes from south to north over the equator at 13:30. The swath of both satellites is 2330 km cross track by 10 km along track at nadir, and in constellation, they provide a revisit time of between one and two days. This work makes use of the Level-1B MODIS 1-km products from both Terra (MOD021KM) and Aqua (MYD021KM) satellites. The products contain calibrated and geolocated radiances resampled to 1-km ground resolution at nadir for all 36 bands and reflectances for the reflective solar bands (bands 1–19 and 26). It is important to mention here that due to its observation geometry, the nominal 1-km ground resolution pixel size expands to about 4 km because of the change in the observation angle moving from nadir toward the edge of the swath [33]. Table I summarizes the ground resolution and swath of the different types of satellite data used in this work.

### E. California Fire Perimeter Database

To provide a ground reference, this work makes use of the California Fire Perimeter Database (CAL FIRE[1]), provided by

[1]CAL FIRE https://frap.fire.ca.gov/frap-projects/fire-perimeters/ (Accessed on July 23, 2020)

TABLE I
NADIR GROUND RESOLUTION AND SWATH OF THE SATELLITE DATA USED IN THIS WORK

| Instrument | Satellites | Spectral regions | Ground Resolution [m] | Swath [km] |
|---|---|---|---|---|
| MODIS | Terra, Aqua | VIS, SWIR, MWIR, LWIR | 1000 | 2230 |
| SLSTR | Sentinel-3B, Sentinel-3A | SWIR, MWIR, LWIR | 1000 | 1400 |
| MSI | Sentinel-2B, Sentinel-2A | VIS, SWIR | 20 | 290 |
| C-SAR | Sentinel-1B, Sentinel-1A | Microwave C-Band | 40 | 250 |

the Fire and Resource Assessment Program. The database includes records of perimeters of wildfires that occurred in the state of California between the years 1950 and 2019 (inclusive). An important note on the timestamps of the data records is that the fire perimeters were collected after the fire had been contained. Consequently, geographic coordinates of fire perimeter are not available for any date between the alarm and containment dates.

## IV. WORKFLOW

The general workflow of this work is as follows: In the preprocessing and data preparation stage, imagery and metadata from each of the four instruments are downloaded based on alarm and containment dates extracted from the CAL FIRE wildfire perimeter database. Both imagery and reference perimeter data are then cleaned to reduce data uncertainty and potential errors. Next, training, validation, and test datasets are generated from the imagery and reference data, which are later used to train four instrument-specific U-Net segmentation networks. Pseudo-probability rasters are then predicted using the instrument-specific U-Nets. Finally, fusion of pairs of pseudo-probability rasters is performed using simple weighted averaging.

### A. Preprocessing and Data Preparation

The steps for the preparation of a dataset capable of training and evaluating deep learning models for wildfire detection from the observations provided by the satellite instruments described in the previous section are illustrated in Fig. 1. The actual process starts from fire perimeter data extracted from the CAL FIRE database. Specifically, fire perimeters from January 1, 2017 to December 31, 2019 were used. A total of 1324 records of fire perimeters were recorded within the specified time period, with a median fire perimeter area of 0.12 $\mathrm{km}^2$ (cf. Fig. 2).

Based on alarm and containment dates for those fires, imagery and metadata from each of the four instruments are utilized after records with missing alarm or containment date were excluded. Additionally, records with a fire perimeter area smaller than 0.01 $\mathrm{km}^2$ were excluded as well. To further reduce the uncertainty in the reference data, only data that were collected using GPS measurements, hand drawn, or manually interpreted were used. Fire perimeters obtained through analysis of infrared data were excluded to avoid the possibility of the reference data being created using the same imagery as the input data. Additionally, fires represented by more than one overlapping perimeters (usually as a result of collection of the same fire by two different agencies) were treated by giving preference to the perimeter belonging to the record with more complete metadata. Finally, fire perimeters that were not observed by any of the four instruments were also excluded. After applying the aforementioned filtering methods, 961 fire perimeters remained.

For each of the 961 fire perimeters, data were downloaded starting from one day prior to the fire alarm date and ending one day after the date the fire was contained. The two-day buffer was selected to allow generation of negative examples, where no fire was present. Next, data from each instrument were preprocessed and saved into a database. For Sentinel-3 and MODIS instruments, a cloud mask was generated from the imagery. At this stage, the preprocessed database contains all necessary imagery, metadata, and cloud masks georeferenced and projected to a WGS84 geographic coordinate system in the original ground resolution. Each Sentinel-2 image contains TOA reflectance values for all bands together with a provided cloud mask, Sentinel-3 and MODIS images contain TOA reflectance and TOA brightness temperature values for all bands together with the calculated cloud masks. Sentinel-1 imagery contains backscatter coefficient values in two polarization modes (VV and VH). As Fig. 3 illustrates, for each fire, all images are interpolated to the same grid with a pixel size on the ground of 110 $\times$ 130 m, and specific spectral bands are subsampled from each image. These data, together with data from the CAL FIRE database, are then used to generate the base datasets for each instrument. Fig. 4 presents the number of time windows with different combinations of instruments observing the same fire.

### B. Selection of Spectral Bands

For the three optical instruments, selection of the relevant spectral bands, which were used to construct the training examples, was decided based on empirical test of the performance of models trained with different combinations of bands. To determine the final combination of the spectral bands for each instrument dataset, a vanilla U-Net network [34] was trained and evaluated on a subset of 2000 examples from the validation dataset. The combination of bands that led to the best segmentation result were selected. Evaluation of the results was done using Cohen's kappa coefficient, precision, recall, and overall accuracy metrics. Two additional considerations in the process of deciding on the combination of bands were taken into account.

1) Memory and inference time limitations. The maximum allowed number of spectral bands in a training example was limited to six, to allow completion of the training processes of each instrument within reasonable time and with the available hardware resources.
2) For SLSTR and MODIS instruments, preference was given to combinations with bands that were available in both daytime and nighttime illumination conditions.

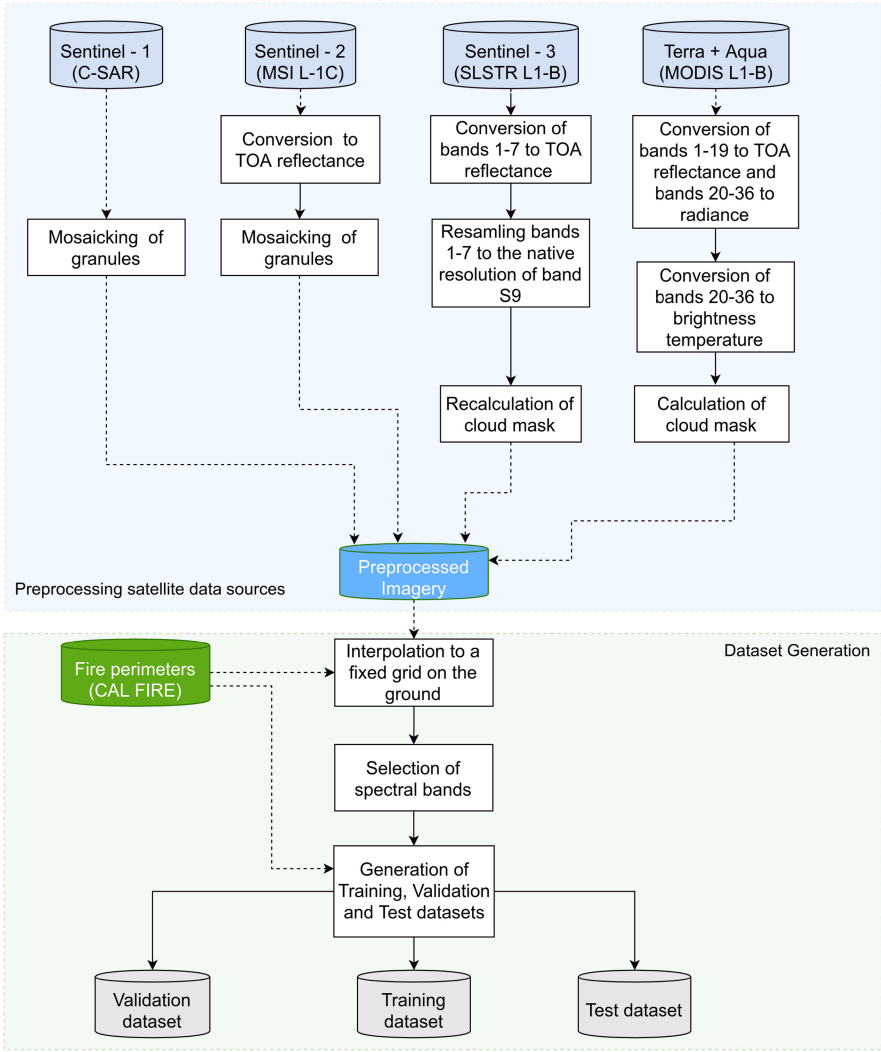


Fig. 1. Workflow of the dataset generation process. The upper part describes the necessary preprocessing steps for each of the satellite data sources. The lower part describes the necessary steps for creating the final dataset completed with reference fire annotations.

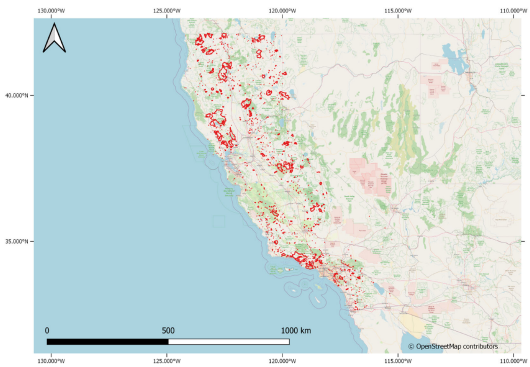


Fig. 2. Perimeters of fires in CAL FIRE between January 1, 2017 and December 31, 2019. Fire perimeters are delineated in red.

Table II summarizes the performance achieved by applying a vanilla U-Net trained with different band combinations. We observe from Table II that the assessed performance of combination MSI_1 is noticeably better than combinations MSI_2 and MSI_3. Since Sentinel-2 does not operate during nighttime, MSI_1 can be used without excluding any images in the training process. The assessed performance of SLSTR and MODIS varies only slightly between the combinations; therefore, combinations with minimum number of bands that allow inclusion of nighttime imagery were selected. For the SLSTR instrument, "fire" bands F1 and F2 were preferred to bands S7 and S8, respectively, because of their increased dynamic range, which prevents them from saturating over strong fires. An additional benefit of using the F1 and F2 bands is their far more limited growth in pixel area in off-nadir scan angles [35]. For the C-SAR instrument on board Sentinel-1, VV and VH polarization bands were used.

### C. Data Splits

For the training of single-instrument models, only images where the fire perimeter was not covered by clouds and with percentage of corrupted pixels lower than 20% were used. Corrupted pixels were identified using metadata provided for each image by each instrument.

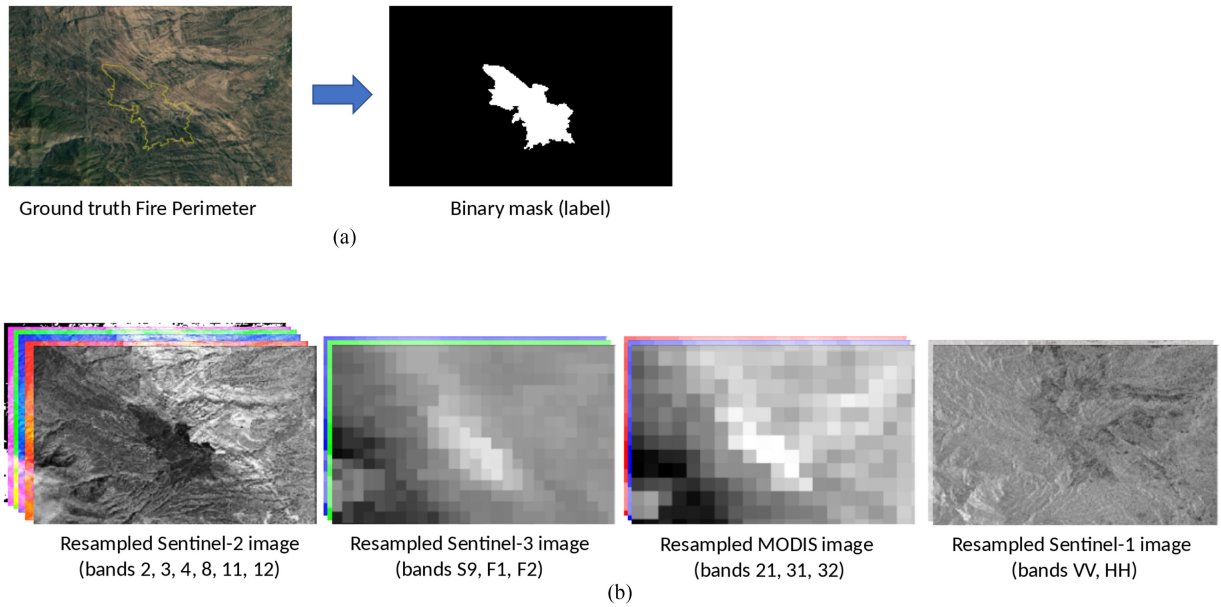


Fig. 3. Generation of a training examples. (a) Label generation—on the left in yellow, the perimeter of a fire from the CAL FIRE database that is used to generate the binary reference mask on the right. (b) Selected bands from the four instruments are resampled to a grid based on the perimeter of the fire.

TABLE II
ASSESSMENT OF PERFORMANCE OF A VANILLA U-NET MODEL TRAINED ON DIFFERENT COMBINATIONS OF SPECTRAL BANDS

| Instrument | Combination Name | Bands | Day/Night | Kappa | Precision | Recall | Overall Acc. |
|---|---|---|---|---|---|---|---|
| MSI | **MSI_1** | **2, 3, 4, 8, 11, 12** | **Day** | **0.61** | **0.74** | **0.63** | **0.89** |
| | MSI_2 | 2, 3, 4 | | 0.54 | 0.66 | 0.70 | 0.84 |
| | MSI_3 | 3, 8, 12 | | 0.58 | 0.69 | 0.74 | 0.88 |
| SLSTR | SLSTR_1 | S7, S8, S9, F1, F2 | Day+Night | 0.54 | 0.60 | 0.62 | 0.87 |
| | SLSTR_2 | S1, S2, S3, S5, S6, F1 | Day | 0.55 | 0.61 | 0.64 | 0.89 |
| | **SLSTR_3** | **S9, F1, F2** | **Day+Night** | **0.54** | **0.59** | **0.62** | **0.87** |
| MODIS | MOD_1 | 1, 2, 3, 21, 31, 32 | Day | 0.52 | 0.54 | 0.63 | 0.85 |
| | MOD_2 | 1, 2, 3, 4, 6, 7 | Day | 0.50 | 0.56 | 0.60 | 0.84 |
| | **MOD_3** | **21, 31, 32** | **Day+Night** | **0.52** | **0.52** | **0.59** | **0.84** |

For each optical instrument, performance of three combinations was evaluated. The final combinations chosen for training are in bold.

Since different fires are observed by different instruments at different times, and due to the relatively limited amount of data, establishment of the datasets used in this work was not trivial. Usually, when comparing between models, the model is evaluated using the same test dataset. Here, however, this was impossible. Moreover, different images of the same fire cannot be shared between the training, validation, and test datasets because of the ability of deep segmentation networks to incorporate spatial information in their predictions, which may lead to the network "remembering" spatial correlations in training examples from the same fire. Finally, assessment of results of combining two individual models into a multisensor output must be done using images that were not involved in the training process of both models and were acquired in the same scene within a time window of less than 12 h. To deal with the complexities described above, the datasets were established based on the following guiding principles:

1) designation of at least 80% of the available data to be used for training;
2) images of the same fire must be present in either the training, validation, or test dataset;
3) the same fires and the same number of images from each fire was used to establish the test dataset for each instrument. However, not all images of the same fire were acquired at the same time;
4) the validation and test datasets for each model must contain short-lasting (less than five days), average-lasting (between five days and 30 days), and long-lasting (longer than 30 days) fires.
5) the validation and test datasets for each model must contain small (less than 5 $\mathrm{km}^2$), average (between 5 and 40 $\mathrm{km}^2$), and large (over 40 $\mathrm{km}^2$) fires.
6) images of the same fire in the test set that were acquired by different instruments should be taken within a time period of 12 hours.

For the test dataset, 60 images that were acquired by all four instruments, each within the same time window of 12 h, were used. Table III summarizes the established datasets. Fig. 4

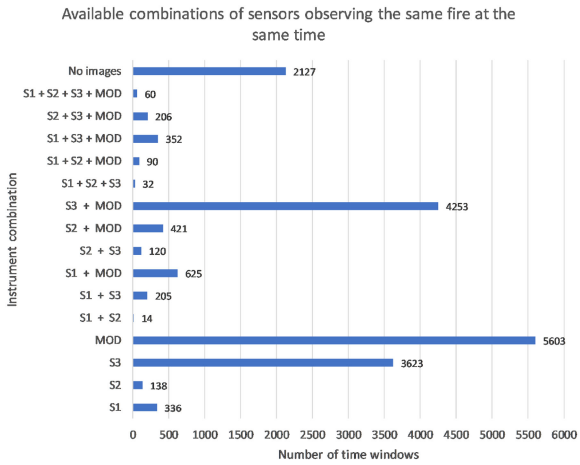


Fig. 4. Number of time windows with different combinations of instruments observing the same fire. The duration of every fire is split into 12-h-long time windows. The chart shows the number of time windows that were observed by different combinations of instruments. The horizontal axis represents the number of time windows and the vertical axis represents the possible combinations of instruments.

TABLE III
NUMBER OF IMAGES IN THE TRAINING, VALIDATION, AND TEST DATASETS FOR SINGLE-INSTRUMENT MODELS

| | Training Dataset | | Validation Dataset | | Test Dataset | |
|---|---|---|---|---|---|---|
| **Instrument** | # Fires | # Images | # Fires | # Images | # Fires | # Images |
| C-SAR | 686 | 2231 | 34 | 328 | 32 | 60 |
| MSI | 354 | 1639 | 34 | 193 | 32 | 60 |
| SLSTR | 730 | 13852 | 34 | 1602 | 32 | 60 |
| MODIS | 853 | 16851 | 34 | 1961 | 32 | 60 |

presents the different combinations of instruments that observed the same fire.

## V. DEEP LEARNING FOR MULTISENSOR WILDFIRE DETECTION

Four instrument-specific convolutional neural networks are created for the detection of wildfires by semantic segmentation. In a second step, the individual predictions are combined to evaluate whether multisensor fusion supports the detection process.

### A. Wildfire Detection by Deep Semantic Segmentation

We use a standard U-Net architecture [34] as the deep semantic segmentation network in this study. However, several modifications to the original U-Net architecture were made to adjust it to the goals of this work.

1) The dimensions of the input layer were changed from $572 \times 572$ to $256 \times 256$ to fit the input data.
2) Batch normalization was introduced after each convolution operation to reduce the problem of updating weights and biases across many layers.
3) A dropout of 10% was introduced in the bottleneck stage for regularization purposes.
4) The increase in number of channels in the downsampling path was made more gradual by first convolving the input data with 32 filters.
5) To reduce the number of total parameters, the bottleneck was introduced at a depth of 512 channels.
6) The outputs of the network were fed to a sigmoid function to achieve a mask of probability values between 0.0 and 1.0.

The structure of the modified U-Net architecture used in this work can be seen in Fig. 5.

Since in most training examples, the number of fire-affected pixels comprises only a small portion of all the pixels in the scene, there is a risk that the learning process converges to a local minimum of the loss function, leading to predictions, which are favorable toward the background, i.e., to pixels not affected by fire. To overcome this problem, this work exploits a combination of two loss functions: binary cross entropy (BCE) loss and Dice loss. The Dice loss originates from the Dice coefficient (also known as F1 score) and is especially suitable for segmentation problems with uneven class distributions [36]. This renders the final loss function as follows:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{m}\sum_{i=1}^{m}\frac{1}{n}\sum_{j=1}^{n}\left(y_{ij} \times \log \hat{y_{ij}}\right) + (1 - y_{ij}) \times (1 - \log \hat{y_{ij}}) \tag{1}$$

$$\mathcal{L}_{\text{dice}} = 1 - \frac{1}{m}\sum_{i=1}^{m}\frac{2 \times \sum_{j=1}^{n}(y_{ij} \times \hat{y_{ij}}) + \epsilon}{\sum_{j=1}^{n} y_{ij} + \sum_{j=1}^{n} \hat{y_{ij}} + \epsilon} \tag{2}$$

$$\mathcal{L}_{\text{final}} = W_{\text{BCE}} \times BCE\ Loss + W_{\text{DL}} \times Dice\ Loss \tag{3}$$

where $y$ is the reference pixels, $\hat{y}$ are the pixels network predictions, indices $i$ and $j$ denote the current mini-batch and example, respectively, $m$ is the number of mini-batches, $n$ is the number of examples per mini-batch, and $\epsilon$ is a small value added to the dice loss to avoid division by zero. The final loss is a linear combination of the BCE and Dice losses, with $w_{\text{BCE}}$ and $w_{\text{DL}}$ being the weights adjusting the influence of each component of the loss function. They are treated as hyperparameters and are adjusted accordingly. Final classification of pixels to the classes *fire-affected* or *background* is decided by applying a threshold $\tau$ to the output of the sigmoid layer. If the value of the sigmoid function at the pixel exceeds $\tau$, the pixel is classified as fire-affected. The value of $\tau$ is tuned as an additional hyperparameter using the validation dataset.

### B. Fusion of Single-Instrument Predictions

The output of each model trained on a single instrument is a 2-D raster, where the values of each pixel represent the probability that a pixel is affected by fire. Given outputs from two models that were trained individually on data from a single instrument, the combined probability raster is calculated based on the following rationale.

1) In case a pixel is covered by clouds in one of the outputs, the probability of the pixel being affected by fire is determined by the model, where the pixel is not covered by clouds.
2) For Sentinel-1 individual output, all pixels are considered to be not covered by clouds because of the cloud penetrating capabilities of the C-SAR instrument.

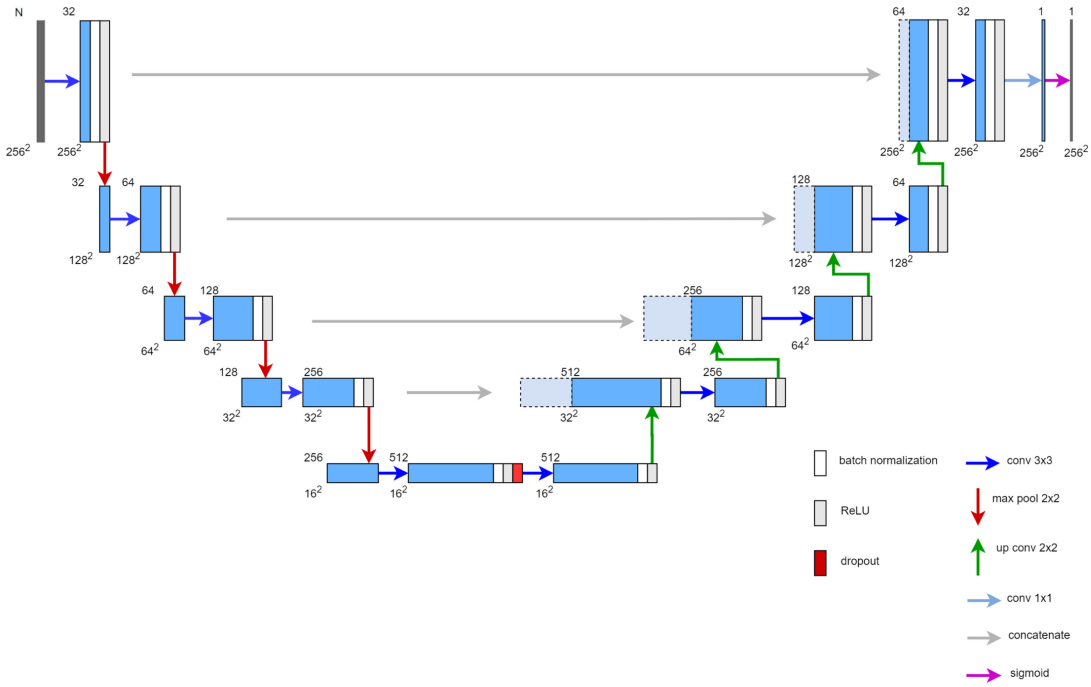


Fig. 5. Modified U-Net architecture used in this work. Each blue box corresponds to a multichannel feature layer. Dark gray box represents the input image with $N$ channels (2 for Sentinel-1 and 3 for all other instruments). The number of channels is denoted on top of the box. The spatial dimensions of the feature are provided at the lower left edge of the box. Pale blue boxes represent copied feature information. The arrows denote the convolution, pooling, and concatenating operations. The final output is a probability raster with values between 0.0 and 1.0 (modified from [34]).

3) In case the pixel is either covered or not covered by clouds in both single-instrument prediction rasters, the value of the pixel in the combined raster is calculated as a weighted average of the individual outputs.
4) The weight of each individual output in the weighted average is determined by the average F1 score achieved by the model when evaluated on the validation dataset, with the sum of the weights is normalized to 1.0.
5) A "fusion" threshold $\tau_f$ is applied to the result of the linear combination of two prediction rasters. Pixels with a value above the threshold are considered to be affected by fire.
6) The value of the threshold is adjusted using a grid search, with the optimum value selected as the one leading to the best kappa coefficient.

Fig. 6 describes the methodology for combining outputs of two models.

### C. Model Training

To adjust the parameters of each of the four single-instrument models, training on augmented datasets was performed. Due to the relatively limited total number of training examples and observed fires, each example was augmented by applying four translations (north, south, east, and west) and four rotations around a randomly selected axis point. Translation in each direction was applied in the range $[0.1, 0.5]$ of the image dimensions, and rotation varied between $[-180°, +180°]$. In an attempt to increase the variability of fire perimeter area in the training database, scaling of fires smaller than 5 km$^2$ by a random factor between 1.5 and 3.0 and scaling of fires larger than 40 km$^2$ by a random factor between 0.33 and 0.66 was applied. The weights $w_{\text{BCE}}$ and $w_{\text{DL}}$ of the final loss function defined in (3) were tuned during the hyperparameter tuning phase to 0.6 and 0.4, respectively.

The adaptive moment estimation (Adam) optimization algorithm [37] was used with default parameters $\beta_1 = 0.9$ and $\beta_1 = 0.999$. The initial learning rate was set to 0.001, and a learning rate decay by a factor of 0.5 was implemented after every five epochs. Training examples were fed to the network in mini-batches of size 8. To calculate the validation loss, average validation F1 score and Cohen's kappa coefficient, examples from the validation dataset were fed to the network during the training process. Training was stopped if the validation loss did not improve after ten epochs. At the end of the training process, the model that achieved the highest F1 score on the validation set was selected as the final model. The process was implemented using Keras[2] with TensorFlow backend as the deep learning framework, on a PC desktop with Intel Core i-5-8600 K CPU @ 3.60 GHz, six cores, 16-GB RAM, and NVIDIA GeForce GTX 1060 6-GB graphic card. Table IV summarizes the average training duration of each single-instrument model. Fusion of two single-instrument prediction rasters with size $256 \times 266 \times 1$ takes less than 5.5% of the average prediction time of a single-instrument model.

[2]Keras https://keras.io/ (Accessed: November 14, 2020)

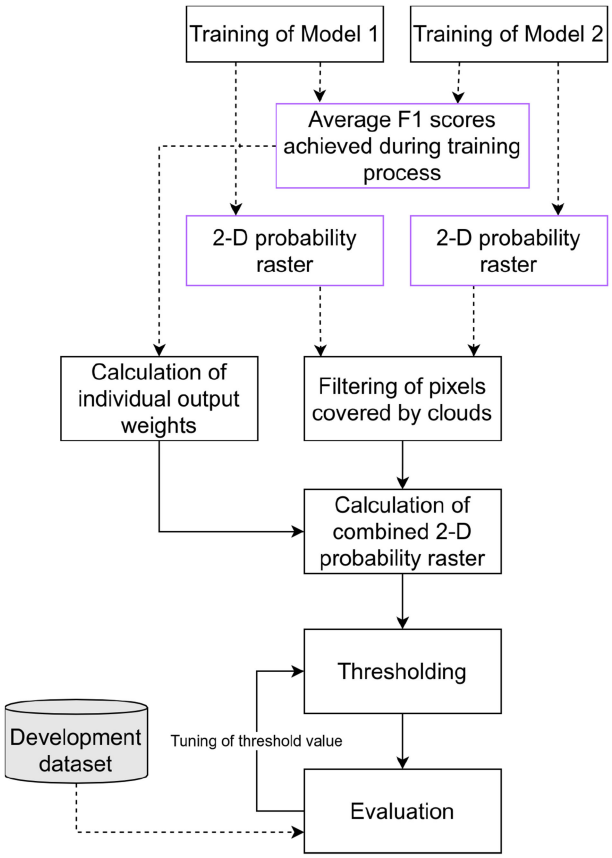


Fig. 6. Methodology for combining predictions of two single-instrument models. 2-D probability rasters are predicted by two single-instrument models, each trained on data from a different instrument. Pixels that are covered by clouds are identified and removed from each output. A combined 2-D probability raster is calculated as weighted average of the two single-instrument probability rasters. The weights used in the combination are derived from the average F1 score achieved during training of the single-instrument models. A threshold $\tau_f$ is applied to the combined raster and is iteratively tuned on the validation dataset. Finally, pixels with values exceeding the tuned $\tau_f$ are predicted as being affected by fire. In magenta: products of single-instrument training. Dashed lines represent usage of products of single-instrument training, and solid lines represent the processing order.

TABLE IV
AVERAGE TRAINING AND PREDICTION TIME OF THE FOUR SINGLE-INSTRUMENT U-NET MODELS

| Input instrument | Average training time per 1000 examples [m] | Average prediction time [s] |
|---|---|---|
| C-SAR | 10.9 | $0.014 \pm 0.002$ |
| MSI | 12.2 | $0.018 \pm 0.006$ |
| SLSTR | 11.1 | $0.018 \pm 0.006$ |
| MODIS | 11.2 | $0.017 \pm 0.005$ |

Prediction time refers to applying a trained single-instrument model to a single $256 \times 256 \times n$ input, with $n$ being the number of channels (two for C-SAR, six for MSI, and three for MODIS and SLSTR).

## VI. EXPERIMENTAL RESULTS

### A. Single-Instrument Detection Results

To assess the performance of each model, the model was evaluated using the examples from the test dataset. Pixels affected by fire were selected by applying a threshold $\tau$ to the output of the final sigmoid layer. Results were obtained by varying $\tau$ from 0.05 to 0.95 in increments of 0.05. The resulting binary prediction was then compared to the reference mask and evaluated using standard evaluation metrics. Fig. 7 exemplifies how changing the threshold affects the final segmentation mask in the form of precision–recall curves (PRC), where the precision achieved by the model is plotted against the achieved recall, as the threshold $\tau$ is varied. This form of presentation was chosen because it is more suitable for imbalanced problems, due to the fact that pixels that were classified as true negatives, and occupy the majority of the scene, are not taken into account. Thus, every point on the PRC represents the precision and recall scores obtained by applying a single-instrument model to the test dataset. Generally, it can be said that the closer a PRC to the upper right corner, the better the performance of a model is. A summary of the quantitative assessment of the results of each model is summarized in Table V.

Examining Fig. 7 and Table V, one can observe a clear separation in the ability of the developed models to correctly identify fire-affected pixels. In clear conditions, the worst results were obtained using the S1 model with an achieved F1 score of 0.46% and 26% of detected fires, whereas the S2 model led to the best results with F1 score of 0.83 and 92% of the fire perimeters detected. Model S3 with F1 score of 0.71 and 68% of perimeters detected exhibited slightly better performance than the MOD model with F1 score of 0.67 and detection of 58% of the perimeters. Examination of the PRC shapes suggests that the S2 model is less sensitive to small changes in $\tau$, whereas results obtained using models S1 and MOD are highly dependent on the chosen value of $\tau$, with S1 displaying almost a linear association between precision and recall. Another clear distinction can be made between the performance of S2, S3, and MOD models in cloudy and clear conditions. For each model, the influence of clouds on the segmentation performance can be approximately estimated by examining the distance between two points: the upper right-most point in the PRC of the clear conditions and its equivalent in the PRC of the cloudy conditions. The larger the distance between the points, the larger the difference between the performance in cloudy and clear conditions is. All models performed worse in cloudy conditions with model S2 exhibiting the largest deterioration of performance with $\tau$ influencing the results by approximately 5%. Performance of both S3 and MOD models seems to be less affected in cloudy conditions and depends more on the chosen value of $\tau$.

### B. Multi-Instrument Detection Results

To combine the results of a pair of two single-instrument models, a dataset consisting of pairs of images was created for every combination of instruments. Images of the same fire that were acquired within the same, 12-h-long, time window were identified and used to create an image pair. As with the assessment of single-instrument models, a distinction was made between images acquired in clear and cloudy conditions. All image pairs were constructed from images that did not participate in the training and hyperparameters tuning

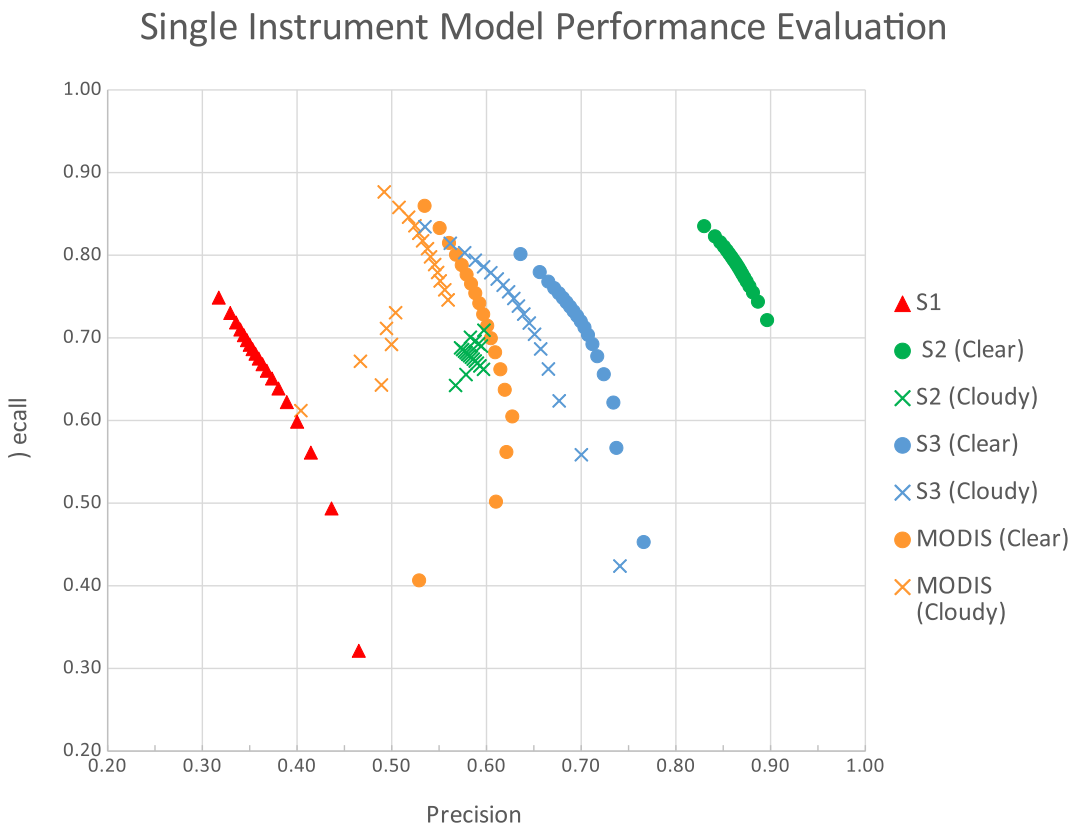


Fig. 7. Precision–recall curves of trained single-instrument models. Each point on the curve represents a single pair of recall–precision values obtained by applying a threshold $\tau$ to the output of the sigmoid layer in a single-instrument model and calculating the recall and precision scores on the test dataset in both clear and cloudy conditions. For each single-instrument model, $\tau$ was varied between 0.05 and 0.95 with increments of 0.05. Generally, it can be said that the closer a PRC to the upper right corner, the better the performance of a model is. The following abbreviations representing a model trained on specific satellite/instrument data are used: S1—Sentinel-1/C-SAR, S2—Sentinel-2/MSI, S3—Sentinel-3/SLSTR, MODIS—Terra and Aqua/MODIS.

TABLE V
ASSESSED PERFORMANCE OF SINGLE-INSTRUMENT MODELS

| Conditions | Instrument (Model) | # Images | $\tau$ | Precision | Recall | F1 Score | K | $\overline{IOU}$ | Detection [%] |
|---|---|---|---|---|---|---|---|---|---|
| Clear | C-SAR (S1) | 60 | 0.20 | 0.39 | 0.57 | 0.46 | 0.38 | 0.39 | 0.26 |
| | MSI (S2) | 39 | 0.05 | 0.83 | 0.84 | 0.83 | 0.67 | 0.71 | 0.92 |
| | SLSTR (S3) | 44 | 0.10 | 0.67 | 0.76 | 0.71 | 0.57 | 0.57 | 0.68 |
| | MODIS (MOD) | 41 | 0.10 | 0.57 | 0.80 | 0.67 | 0.48 | 0.50 | 0.58 |
| Cloudy | MSI (S2) | 21 | 0.05 | 0.60 | 0.71 | 0.65 | 0.51 | 0.45 | 0.50 |
| | SLSTR (S3) | 16 | 0.10 | 0.60 | 0.76 | 0.67 | 0.54 | 0.47 | 0.48 |
| | MODIS (MOD) | 19 | 0.10 | 0.54 | 0.81 | 0.65 | 0.48 | 0.44 | 0.45 |

Reported values are obtained with $\tau$ that led to the highest F1 score. S1, S2, and S3 refer to the models trained on Sentinel-1 A/B, Sentinel-2 A/B, and Sentinel-3 A/B data accordingly. MOD refers to the model trained on MODIS data.

stages. In total, four combinations of bi-instrument models were generated:

1) C-SAR and MSI;
2) C-SAR and SLSTR;
3) MSI and SLSTR;
4) SLSTR and MODIS.

Combinations 1 and 2 were generated to analyze how combining VIS, SWIR, and microwave spectral bands affects the segmentation results. Combination 3 aims at investigating how combining VIS, SWIR, MIR, and TIR spectral bands with varying ground resolutions affects the results. Finally, combination 4 was created in an attempt to explore how combining MIR and TIR spectral bands with similar ground resolution affects the results. Similarly to the single-instrument models, for every combination, a threshold value $\tau$ was applied to the combined probability mask. The value of $\tau$ was tuned using a grid search method. Segmentation results were generated with $\tau$ ranging from 0.1 to 0.9. Again, results are presented in form of a PRC with distinction between cloudy and clear conditions. To compare between results of bi-instrument and single-instrument models, PRCs of results obtained using bi-instrument models were plotted together with the best precision and recall values obtained by the relevant single-instrument models. The combined plots are presented in Fig. 8 , and the accompanying

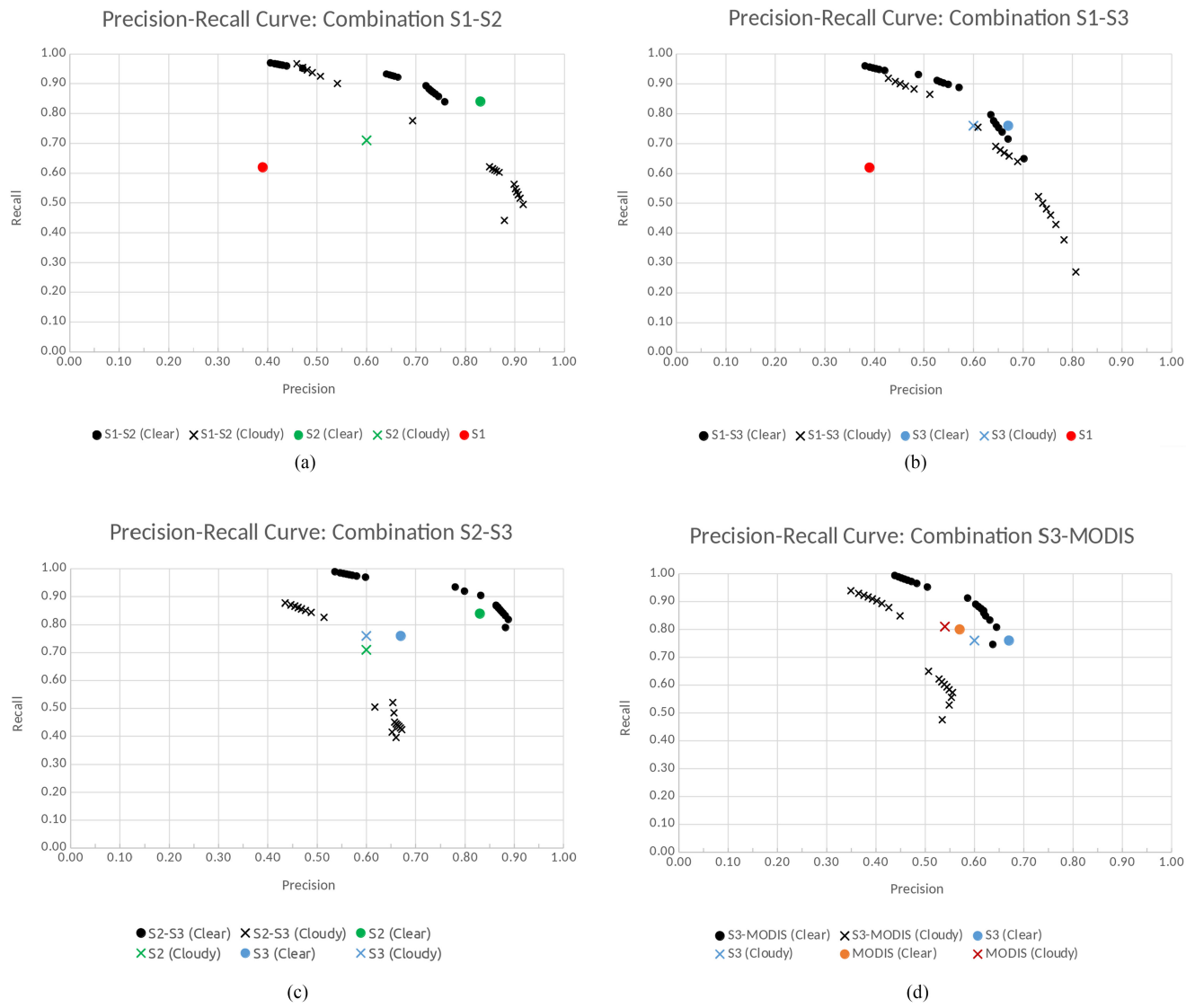


Fig. 8. Precision–recall curves of bi-instrument segmentation results. Best precision–recall results obtained with single-instrument models are plotted in comparison. (a) S1–S2. (b) S1–S3. (c) S2–S3. (d) S3–MOD.

values of the achieved evaluation metrics are summarized in Table VI.

Fig. 8(a) shows that combining S1 and S2 models generally leads to improved segmentation results in cloudy conditions compared to results obtained using the S2 model only. Combining models S1 and S2 improves precision by 25% while causing a reduction of 9% in recall, leading to an increase of 9% in the intersection over union (IOU) and 6% in detection percentage. In clear conditions, however, combination S1–S2 does not seem to improve upon the results obtained with S2. The combined result causes a deterioration of 17% in precision, while increasing the recall by 8% leads to a 2% decrease in the number of detected fires. It can be seen in Fig. 8(b) that the combined S1–S3 model does not significantly affect the overall segmentation results when compared to results obtained using the S3 model only. Compared to the S3 model, in clear conditions, the bi-instrument model achieves 11% higher recall and 10% lower precision with a resulting increase of 1% in the average IOU. In cloudy conditions, S1–S3 achieves the same average IOU, number of detected perimeters, and F1 score as S3 with an increase of 5% in precision and decrease of 8% in recall. Fig. 8(c) shows that the combined S2–S3 model performs better than both S2 and S3 models in clear conditions. Comparing Table VI with Table V one can see an increase of 4% and 3% in precision and recall, respectively, and 4% increase in IOU and detection percentage . In cloudy conditions, it appears that combining models S2 and S3 does not improve the results. The PRC plotted in Fig. 8(d) shows that combining the outputs of S3 and MOD models does not lead to a significant improvement in clear conditions. In cloudy conditions, the S3–MOD combination performed worse than single-instrument S3 and MOD models. Fig. 9 exemplifies segmentation results of the Detwiler fire (2017) obtained using single-instrument and bi-instrument models.

Fig. 9(a) presents an example where most of the fire perimeter is covered by clouds, which hinders the results of S2. Combining the results with prediction of the S1 model improves the overall

TABLE VI
ASSESSED PERFORMANCE OF BI-INSTRUMENT MODELS

| Conditions | Combination | # Images | $W_1$ | $W_2$ | $\tau_f$ | Precision | Recall | F1 score | K | $\overline{IOU}$ | Detection [%] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Clear | S1-S2 | 39 | 0.34 | 0.66 | 0.65 | 0.66 | 0.92 | 0.77 | 0.65 | 0.70 | 0.92 |
| | S1-S3 | 44 | 0.39 | 0.61 | 0.60 | 0.57 | 0.87 | 0.71 | 0.58 | 0.58 | 0.68 |
| | S2-S3 | 39 | 0.55 | 0.45 | 0.55 | 0.87 | 0.87 | 0.87 | 0.76 | 0.78 | 0.96 |
| | S3-MOD | 41 | 0.50 | 0.50 | 0.70 | 0.62 | 0.87 | 0.72 | 0.59 | 0.58 | 0.70 |
| Cloudy | S1-S2 | 21 | 0.34 | 0.66 | 0.45 | 0.62 | 0.85 | 0.72 | 0.59 | 0.54 | 0.56 |
| | S1-S3 | 16 | 0.39 | 0.61 | 0.45 | 0.65 | 0.68 | 0.67 | 0.54 | 0.47 | 0.48 |
| | S2-S3 | 21 | 0.55 | 0.45 | 0.40 | 0.51 | 0.83 | 0.63 | 0.53 | 0.46 | 0.48 |
| | S3-MOD | 19 | 0.50 | 0.50 | 0.45 | 0.45 | 0.85 | 0.59 | 0.51 | 0.44 | 0.42 |

Reported values are obtained with $\tau_f$ that led to the highest F1 score. $W_1$ and $W_2$ refer to the weights used in the linear combination of single-instrument models. Specified average IOU value represents the average intersection over union ratio across all images in the tested dataset.

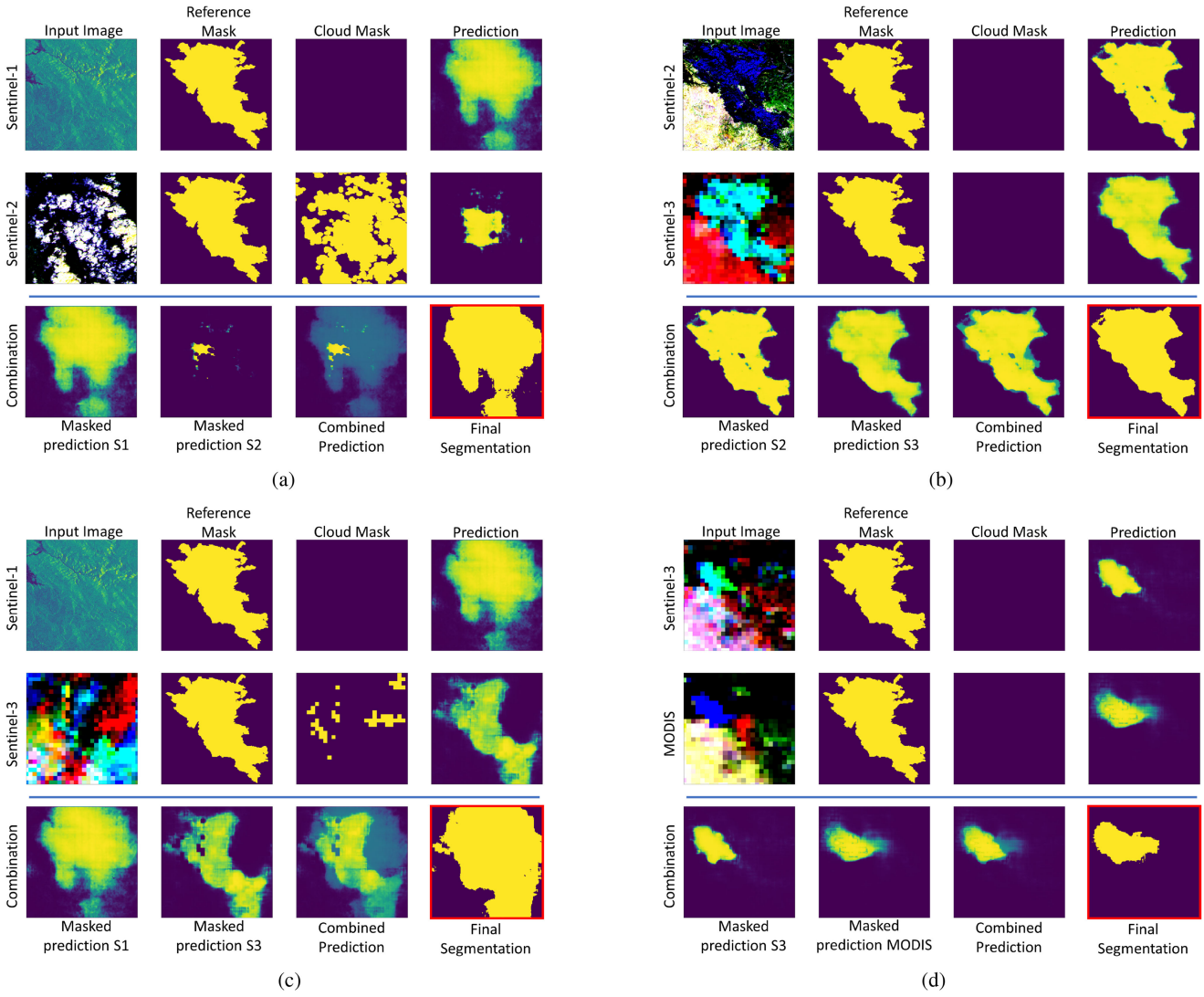


Fig. 9. Examples of segmentation results of the Detwiler fire obtained using single-instrument and bi-instrument models. *Order of displayed results from top to bottom*: The first two rows display results of single-instrument segmentation and the last row displays the combined segmentation process. The final result in the lower right most corner is highlighted in red. *Order of displayed results from left to right*: Single-instrument results: input image, reference mask from CAL FIRE, cloud mask, and prediction of the last sigmoid layer of the trained model. Combined results: prediction from the first single-instrument model after filtering of cloud-covered pixels, prediction from the second single-instrument model after filtering of cloud-covered pixels, combined prediction prior to applying threshold $\tau_f$, and final combined segmentation result after thresholding with $\tau_f$. (a) S1–S2. (b) S2–S3. (c) S1–S3. (d) S3–MOD.

TABLE VII
COMPARISON BETWEEN RESULTS ACHIEVED BY SINGLE-INSTRUMENT MODELS TRAINED ON MSI, C-SAR, AND MODIS DATA AND RESULTS OF NOTICEABLE WORKS IN THE FIELD

| Instrument | Author | Pixel Spacing [m] | Temporal Resolution | Precision | Recall | Test dataset |
|---|---|---|---|---|---|---|
| C-SAR | This Work | ∼118 | Mono-temporal | 0.39 | 0.57 | 32 Fire perimeters from CAL-FIRE |
| | * Belenguer-Plomer et al., 2019 | 20-50 | 12 days | 0.24-0.84 | 0.19-0.86 | Burned area perimeters in 18 globally spread MGRS tiles obtained from Landsat-8 images using a RF classifier trained on manually annotated data |
| | Ban et al., 2020 | 20-50 | 12 days | 0.72-0.90 | 0.93-0.99 | Burned areas perimeters manually derived using Worldview-3 imagery and field data in two large fires in California |
| MSI | This Work | ∼118 | Mono-temporal | 0.83 | 0.84 | 32 Fire perimeters from CAL-FIRE |
| | Farasin et al., 2020 | 10-60 | Mono-temporal | 0.45 - 0.91 | 0.61 - 0.95 | Copernicus Emergency Management Service damage severity maps of 21 fires in five European regions |
| | Knopp et al., 2020 | 10-20 | Mono-temporal | 0.77- 0.96 | 0.97- 0.99 | Three manually refined Copernicus Emergency Management Service in Europe |
| MODIS | This Work | 1000 | Mono-temporal | 0.57 | 0.80 | 32 Fire perimeters from CAL-FIRE |
| | Mithal et al., 2018 | 500 | Time series | 0.29-0.78 | 0.26-0.74 | Automatically generated Global Landsat reference maps in 19 Landsat scenes |
| | * Ramo and Chuvieco, 2017 | 500 | Mono-temporal | 0.75 | 0.75 | MODIS BA product in 3 test sites (Canada, Australia, California) |

Conversion to precision and recall values was applied to results of papers marked with *. Ranges of precision and recall are used to report results obtained in different geographical regions.

segmentation result. The result is still inaccurate with a significant number of commission errors introduced by S1; however, the overall IOU is improved. Fig. 9(b) shows an example in which combining Sentinel-2 and Sentinel-3 predictions leads to a more continuous segmentation with an improved IOU. In Fig. 9(c), combining the predictions from Sentinel-1 and Sentinel-3 led to worsening of the final segmentation due to introduction of new commission errors. In Fig. 9(d), combining predictions of Sentinel-3 and MODIS improved the segmentation; here, however, it is not clear whether the reference mask represents accurately the extent of fire-affected pixels at the time of the acquisition.

## VII. DISCUSSION

### A. Performance of the Individual Image Sources

The results obtained in Section VI have shown that supervised deep learning methods, applied to satellite imagery in the visible light and infrared domain, can be used to detect fire-affected areas and perform segmentation to classify fire-affected pixels. The degree of the achieved success varied between the models, depending on the instrument. The model trained on Sentinel-2 data achieved the best results, followed by Sentinel-3 and MODIS. An attempt to perform similar tasks using a model trained on Sentinel-1 C-SAR data did not lead to satisfying results.

Several possible explanations for the difference in performance between Sentinel-2 MSI, Sentinel-3 SLSTR, and MODIS models are possible. The MSI produces pixels with significantly smaller spatial resolution on the ground compared to SLSTR and MODIS instruments. For the same spatial extent and a single band, Sentinel-2 produces up to 2500 times more information than both SLSTR and MODIS. Thus, the number of observations contained in each training example is significantly larger, which allows the network to train on more data. Another influence of the ground resolution is the ability of the network to correctly identify the border between fire-affected pixels and background. With smaller ground resolution, the border between fire-affected pixels and background can be identified with higher level of detail. Second, six spectral bands of the MSI were used to predict the final segmentation map, while only three spectral bands of SLSTR and MODIS were used for the same task. The decision to use SLSTR and MODIS bands available both in

TABLE VIII
COMPARISON BETWEEN RESULTS ACHIEVED BY BI-INSTRUMENT MODELS AND RESULTS OF NOTICEABLE WORKS IN THE FIELD

| Instrument | Author | Cloud Coverage | Pixel Spacing [m] | Temporal Resolution | Precision | Recall | Reference data used |
|---|---|---|---|---|---|---|---|
| C-SAR+ MSI | This work | >20%<br><20% | ∼118 | Mono-temporal | 0.62<br>0.66 | 0.85<br>0.92 | CAL-FIRE |
| C-SAR + MSI + MODIS | Verhegghen et al., 2016 | NR | 20 | Monthly | NR | NR | NR |
| MSI+ SLSTR | This work | >20%<br><20% | ∼118 | Mono-temporal | 0.51<br>0.87 | 0.83<br>0.87 | CAL-FIRE |
| MSI+ MODIS | * Roteta et al., 2019 | NR | 20 | 10 days | 0.82 | 0.75 | Manually annotated burned area perimeters delineated from Landsat-7 and Landsat-8 imagery in 45 tiles in Sub-Saharan Africa |

Conversion to precision and recall values was applied to results of papers marked with *. NR abbreviation stands for not reported information.

daytime and nighttime was made to increase the total number of available imagery, and three bands only were used based on the results summarized in Table II, which showed that contrary to the authors' expectations, adding additional bands to SLSTR and MODIS does not improve the results significantly. This finding might be explained by the fact that the spectral signal of AFs and BAs occurs mostly in the three bands selected or in reflective bands in the NIR and SWIR regions, which are available during daytime only [38].

The marginally better performance of the SLSTR model compared to the MODIS model could possibly be explained by expansion of the ground resolution of MODIS pixels toward the edge of the swath, which can reach 4 km, and further reduce the amount of information measured in each scene [33]. The second possible explanation is that SLSTR has a lower minimum detection limit for actively burning pixels with lower FPR. In their recent work, Xu *et al.* [35] have found SLSTR to detect 44% more AF pixels than MODIS.

With respect to the comparably bad results achieved with Sentinel-1 SAR imagery, there are several possible explanations for this outcome. The first is that unlike many research works in the field [6], [12], [39], this work attempted to detect fire-affected pixels in $C$-band SAR images without comparing them to previous imagery of the scene that was acquired prior to the fire. The primary challenge of this method is to establish a model based on direct measurements of the backscatter and not on the change in backscatter. The second possible explanation for the poor performance of the C-SAR model is insufficient training data. The model was trained on examples from 686 fires; however, only 2231 images, consisting of two bands only, participated in training of the model.

### B. Usefulness of Multisource Data Fusion

The experiments furthermore showed that combining segmentation maps predicted by S1 and S2 models in cloudy conditions increases the percentage of detected fires and achieves higher average IOU, kappa, and F1 score (cf. Table VI). This happens because the S1 model, despite its large commission errors, is generally able to detect areas affected by fire. As long as the predictions made by the model are only used in pixels covered by clouds, the overall segmentation result improves. Unexpectedly, combining outputs of the S1 and S3 models did not achieve similar results. Looking at the results, it is difficult to pinpoint the reason for this behavior. One possible explanation can be insufficient tuning of the weights $w_1$ and $w_2$ and the threshold $\tau_f$ that were used in computing the mixed probability mask and the resulting segmentation map.

### C. Comparison of Results to Other Works in the Field

It is important to emphasize once more that this work aims at exploring whether combining results of single-instrument models can improve semantic segmentation of wildfires. Therefore, comparison of the results of this work obtained with single-instrument models with the results of other works in the field is presented to merely put the results of this work in context. Table VII presents a comparison between the results achieved by single-instrument models in this work and results achieved in [38]–[43]. Only three out of the four single-instrument models are compared because, to the best of the authors' knowledge, only one comprehensive evaluation of performance of Sentinel-3 SLSTR data for detection of fire-affected areas exists to date [35], which, unfortunately, does not report any accuracy or precision metrics.

Belenguer-Plomer *et al.* [39] detected anomalies of the backscattering coefficient in Sentinel-1 dual-polarized backscatter image time series, which were combined with thermal anomalies derived from MODIS to detect BAs. The performance of the algorithm was assessed using reference perimeters derived from optical Sentinel-2 and Landsat imagery and reached a mean F1 score of 0.59 with 0.62 and 0.72 in terms of recall and precision, respectively. Ban *et al.* [40] used an 18-layer CNN applied to dual-polarization time series of Sentinel-1 images and achieved an F1 score of 0.81 and a kappa of 0.67 when evaluated over three large fires in California. Compared to the aforementioned works, the results of this work underachieve with 0.39 and 0.57 precision and recall, respectively. It has to be recalled, however, that we have only used single-temporal Sentinel-1 imagery as input to the fire detection workflow. Furthermore, our

work demonstrated that combining a model trained on Sentinel-2 data even with the underachieving monotemporal Sentinel-1 model improves segmentation results in the presence of clouds. For the single-instrument models trained on MSI and MODIS data, respectively, the results of this work compare well with the works in the field—especially when considering the difference in the test datasets that were used.

Providing context for the results that were obtained using bi-instrument models is challenging due to the limited number of works that implement multisensor data fusion to segment wildfires and even fewer works reporting evaluation metrics. Table VIII compares the data fusion results obtained in this work with two noticeable works from recent years and yet again confirms comparable performance.

## VIII. Conclusion

With this article, we have confirmed that deep learning shows great potential for the automatic and robust detection of wildfires from openly available multisensorial remote sensing imagery. While in clear, cloud-free weather condition detection rates up to 92% are achieved using just the most suitable sensor (Sentinel-2), and up to 96% when employing the best multisensor fusion scenario (Sentinel-2 and Sentinel-3), we have been able to show that data fusion can generally be confirmed as beneficial. This holds, in particular, if different optical sensors are combined during clear noncloudy conditions or if Sentinel-1 SAR observations are utilized to support the optical observations during cloudy weather.

## References


[1] S. C. Coogan, F. N. Robinne, P. Jain, and M. D. Flannigan, "Scientists' warning on wildfire—A Canadian perspective," *Can. J. Forest Res.*, vol. 49, no. 9, pp. 1015–1023, 2019.

[2] D. Thomas, D. Butry, S. Gilbert, D. Webb, and J. Fung, "The costs and losses of wildfires: A literature survey (NIST Special Publication 1215)," 2017. [Online]. Available: https://doi.org/10.6028/NIST.SP.1215

[3] E. Chuvieco *et al.*, "Historical background and current developments for mapping burned area from satellite Earth observation," *Remote Sens. Environ.*, vol. 225, pp. 45–64, 2019.

[4] D. Fornacca, G. Ren, and W. Xiao, "Performance of three MODIS fire products (MCD45A1, MCD64A1, MCD14ML), and ESA Fire_CCI in a mountainous area of Northwest Yunnan, China, characterized by frequent small fires," *Remote Sens.*, vol. 9, no. 11, pp. 1–20, 2017.

[5] J. Wang *et al.*, "Review of satellite remote sensing use in forest health studies," *Open Geography J.*, vol. 3, no. 1, pp. 28–42, 2010.

[6] J. Engelbrecht, A. Theron, L. Vhengani, and J. Kemp, "A simple normalized difference approach to burnt area mapping using multi-polarisation C-band SAR," *Remote Sens.*, vol. 9, no. 8, pp. 9–11, 2017.

[7] M. Caggiano, "Using community base maps to improve the safety and effectiveness of wildfire response," 2018. Accessed: Oct. 7, 2020. [Online]. Available: https://fireadaptednetwork.org/using-community-base-maps-to-improve-the-safety-and-effectiveness-of-wildland-fire-response/

[8] P. Jain, S. C. Coogan, S. G. Subramanian, M. Crowley, S. W. Taylor, and M. D. Flannigan, "A review of machine learning applications in wildfire science and management," *Environ. Rev.*, vol. 28, pp. 1–70, 2020.

[9] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, pp. 195–204, 2019.

[10] S. Hantson *et al.*, "The status and challenge of global fire modelling," *Biogeosciences*, vol. 13, no. 11, pp. 3359–3375, 2016.

[11] Z. Langford, J. Kumar, and F. Hoffman, "Wildfire mapping in interior Alaska using deep neural networks on imbalanced datasets," in *Proc. IEEE Int. Conf. Data Mining Workshops*, 2018, pp. 770–778.

[12] A. Verhegghen *et al.*, "The potential of Sentinel satellites for burnt area mapping and monitoring in the Congo Basin forests," *Remote Sens.*, vol. 8, no. 12, pp. 1–22, 2016.

[13] M. A. Crowley, J. A. Cardille, J. C. White, and M. A. Wulder, "Multi-sensor, multi-scale, Bayesian data synthesis for mapping within-year wildfire progression," *Remote Sens. Lett.*, vol. 10, no. 3, pp. 302–311, 2019.

[14] J. A. Cardille and J. A. Fortin, "Bayesian updating of land-cover estimates in a data-rich environment," *Remote Sens. Environ.*, vol. 186, pp. 234–249, 2016.

[15] K. Lasko, "Incorporating Sentinel-1 SAR imagery with the MODIS MCD64A1 burned area product to improve burn date estimates and reduce burn date uncertainty in wildland fire mapping," *Geocarto Int.*, vol. 36, no. 3, pp. 340–360, 2019.

[16] M. J. García and V. Caselles, "Mapping burns and natural reforestation using thematic mapper data," *Geocarto Int.*, vol. 6, no. 1, pp. 31–37, 1991.

[17] C. O. Justice *et al.*, "The MODIS fire products," *Remote Sens. Environ.*, vol. 83, nos. 1/2, pp. 244–262, 2002.

[18] M. J. Wooster, W. Xu, T. Nightingale, "Sentinel-3 SLSTR active fire detection and FRP product: Pre-launch algorithm development and performance evaluation using MODIS and ASTER datasets, Remote Sensing of Environment," vol. 120 pp. 236–254, 2012.

[19] L. Giglio, J. Descloitres, C. O. Justice, and Y. J. Kaufman, "An enhanced contextual fire detection algorithm for MODIS," *Remote Sens. Environ.*, vol. 87, nos. 2/3, pp. 273–282, 2003.

[20] M. Gimeno, J. San-Miguel-Ayanz, and G. Schmuck, "Identification of burnt areas in Mediterranean forest environments from ERS-2 SAR time series," *Int. J. Remote Sens.*, vol. 25, no. 22, pp. 4873–4888, 2004.

[21] C. Carmona-Moreno *et al.*, "Characterizing interannual variations in global fire calendar using data from earth observing satellites," *Global Change Biol.*, vol. 11, no. 9, pp. 1537–1555, 2005.

[22] R. Chalapathy and S. Chawla, "Deep learning for anomaly detection: A survey," 2019, *arXiv preprint arXiv:1901.03407v2*.

[23] D. Kwon, K. Natarajan, S. Suh, H. Kim, and J. Kim, "An empirical study on network anomaly detection using convolutional neural networks," in Proc.*IEEE 38th Int. Conf. Distrib. Comput. Syst.*, 2018, pp. 1595–1598.

[24] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "CoSpace: Common subspace learning from hyperspectral-multispectral correspondences," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4349–4359, Jul. 2019.

[25] D. Hong, N. Yokoya, N. Ge, J. Chanussot, and X. X. Zhu, "Learnable manifold alignment (LeMA): A semi-supervised cross-modality learning framework for land cover and land use classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 147, pp. 193–205, 2019.

[26] D. Hong, J. Yao, D. Meng, Z. Xu, and J. Chanussot, "Multimodal GANs: Toward crossmodal hyperspectral-multispectral image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5103–5113, Jun. 2021.

[27] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

[28] N. Yague-Martinez *et al.*, "Interferometric processing of Sentinel-1 TOPS data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2220–2234, Apr. 2016.

[29] R. Torres *et al.*, "GMES Sentinel-1 mission," Remote Sens. Environ., vol. 120, pp. 9–24, 2012

[30] M. Drusch *et al.*, "Sentinel-2: ESA's optical high-resolution mission for GMES operational services," *Remote Sens. Environ.*, vol. 120, pp. 25–36, 2012.

[31] F. Gascon *et al.*, "Copernicus Sentinel-2A calibration and products validation status," *Remote Sens.*, vol. 9, no. 6, 2017, Art. no. 584.

[32] X. Xiong, K. Chiang, J. Esposito, B. Guenther, and W. Barnes, "MODIS on-orbit calibration and characterization," *Metrologia*, vol. 40, no. 1, p. S 89, 2003.

[33] L. Giglio, L. Boschetti, D. P. Roy, M. L. Humber, and C. O. Justice, "The collection 6 MODIS burned area mapping algorithm and product," *Remote Sens. Environ.*, vol. 217, pp. 72–85, 2018.

[34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, vol. 9351. New York, NY, USA: Springer, 2015, pp. 234–241.

[35] W. Xu, M. J. Wooster, J. He, and T. Zhang, "First study of Sentinel-3 SLSTR active fire detection and FRP retrieval: Night-time algorithm enhancements and global intercomparison to MODIS and VIIRS AF products," *Remote Sens. Environ.*, vol. 248, 2020, Art. no. 111947.

[36] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis.*, 2016, pp. 565–571.
[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017, *arXiv preprint arXiv:1412.6980v9*.
[38] R. Ramo and E. Chuvieco, "Developing a random forest algorithm for MODIS global burned area classification," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1193.
[39] M. A. Belenguer-Plomer, M. A. Tanase, A. Fernandez-Carrillo, and E. Chuvieco, "Burned area detection and mapping using Sentinel-1 backscatter coefficient and thermal anomalies," *Remote Sens. Environ.*, vol. 233, 2019, Art. no. 111345.
[40] Y. Ban, P. Zhang, A. Nascetti, A. R. Bevington, and M. A. Wulder, "Near real-time wildfire progression monitoring with Sentinel-1 SAR time series and deep learning," *Sci. Rep.*, vol. 10, no. 1, pp. 1–15, 2020.
[41] L. Knopp, M. Wieland, M. Rättich, and S. Martinis, "A deep learning approach for burned area segmentation with Sentinel-2 data," *Remote Sens.*, vol. 12, no. 15, 2020, Art. no. 2422.
[42] A. Farasin, L. Colomba, and P. Garza, "Double-step U-Net: A deep learning-based approach for the estimation of wildfire damage severity through Sentinel-2 satellite data," *Appl. Sci.*, vol. 10, no. 12, pp. 1–22, 2020.
[43] V. Mithal, G. Nayak, A. Khandelwal, V. Kumar, R. Nemani, and N. C. Oza, "Mapping burned areas in tropical forests using a novel machine learning framework," *Remote Sens.*, vol. 10, no. 1, pp. 1–22, 2018.

**Dmitry Rashkovetsky** received the M.Sc. degree in Earth-oriented space and technology from the Technical University of Munich, Munich, Germany, in 2021.

He is currently a Remote Sensing Engineer with Orbital Oracle Technologies GmbH, Munich. He was a Mapping and Geo-Information Engineer with the Israeli Mapping and Geo-Information Centre, where he developed digital elevation data production workflows from electrooptical, synthetic aperture radar, and LiDAR sensors, developed and implemented geographic information system (GIS) database strategies and headed the GIS and Digital Elevation Data Departments.

**Florian Mauracher** received the M.Sc. degree in computer science, in 2019.

He is one of the Co-Founders and Lead Software Engineer with Orbital Oracle Technologies GmbH (OroraTech), Munich, Germany. He has experience in agile product development, embedded systems, and system architecture. He is responsible for the development of OroraTech's core product, a satellite-based wildfire intelligence platform the startup offers to clients worldwide.

**Martin Langer** received the Dipl.-Ing. and Dr.-Ing. degrees in aerospace engineering from the Technical University of Munich (TUM), Munich, Germany in 2011 and 2018 respectively.

He is currently the Chief Technical Officer of Orbital Oracle Technologies GmbH, Munich. At TUM, he was leading the Small Satellite Research Group and was part of three CubeSat Missions, leading the development of TUM's second and third CubeSat. His research interests include the development of robust SmallSat hardware and software for space use and reliability assurance and risk mitigation of small satellites.

**Michael Schmitt** (Senior Member, IEEE) received the Dipl.-Ing. degree in geodesy and geoinformation, the Dr.-Ing. degree in remote sensing, and the Habilitation in data fusion from the Technical University of Munich (TUM), Munich, Germany, in 2011 and 2018, respectively.

Since 2020, he has been a Full Professor of Applied Geodesy and Remote Sensing with the Department of Geoinformatics, Munich University of Applied Sciences, Munich. From 2015 to 2020, he was a Senior Researcher and the Deputy Head with the Professorship for Signal Processing in Earth Observation, TUM. In 2019, he was additionally appointed as an Adjunct Teaching Professor with the Department of Aerospace and Geodesy, TUM. In 2016, he was a Guest Scientist with the University of Massachusetts at Amherst, Amherst, MA, USA. His research interests include image analysis and machine learning applied to the extraction of information from multimodal remote sensing observations. In particular, he is interested in remote sensing data fusion with a focus on synthetic aperture radar (SAR) and optical data.

Dr. Schmitt is the Co-Chair of the Working Group "SAR and Microwave Sensing" of the International Society for Photogrammetry and Remote Sensing, and also of the Working Group "Benchmarking" of the IEEE Geoscience and Remote Sensing Society Image Analysis and Data Fusion Technical Committee. He frequently serves as a Reviewer for a number of renowned international journals and conferences and has received several best reviewer awards. He is an Associate Editor for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.