

Article

Deep Learning Approaches for Wildfire Severity Prediction: A Comparative Study of Image Segmentation Networks and Visual Transformers on the EO4WildFires Dataset

Dimitris Sykas ^{*,†}, Dimitrios Zografakis [†] and Konstantinos Demestichas [†]

Informatics Laboratory, Department of Agricultural Economics & Rural Development, Agricultural University of Athens, 11855 Athens, Greece; dimzog@aua.gr (D.Z.); cdemest@aua.gr (K.D.)

* Correspondence: dsykas@aua.gr

† These authors contributed equally to this work.

Abstract: This paper investigates the applicability of deep learning models for predicting the severity of forest wildfires, utilizing an innovative benchmark dataset called EO4WildFires. EO4WildFires integrates multispectral imagery from Sentinel-2, SAR data from Sentinel-1, and meteorological data from NASA Power annotated with EFFIS data for forest fire detection and size estimation. These data cover 45 countries with a total of 31,730 wildfire events from 2018 to 2022. All of these various sources of data are archived into data cubes, with the intention of assessing wildfire severity by considering both current and historical forest conditions, utilizing a broad range of data including temperature, precipitation, and soil moisture. The experimental setup has been arranged to test the effectiveness of different deep learning architectures in predicting the size and shape of wildfire-burned areas. This study incorporates both image segmentation networks and visual transformers, employing a consistent experimental design across various models to ensure the comparability of the results. Adjustments were made to the training data, such as the exclusion of empty labels and very small events, to refine the focus on more significant wildfire events and potentially improve prediction accuracy. The models' performance was evaluated using metrics like F1 score, IoU score, and Average Percentage Difference (aPD). These metrics offer a multi-faceted view of model performance, assessing aspects such as precision, sensitivity, and the accuracy of the burned area estimation. Through extensive testing the final model utilizing LinkNet and ResNet-34 as backbones, we obtained the following metric results on the test set: 0.86 F1 score, 0.75 IoU, and 70% aPD. These results were obtained when all of the available samples were used. When the empty labels were absent during the training and testing, the model increased its performance significantly: 0.87 F1 score, 0.77 IoU, and 44.8% aPD. This indicates that the number of samples, as well as their respectively size (area), tend to have an impact on the model's robustness. This restriction is well known in the remote sensing domain, as accessible, accurately labeled data may be limited. Visual transformers like TeleViT showed potential but underperformed compared to segmentation networks in terms of F1 and IoU scores.

Keywords: forest wildfires; earth observation; remote sensing; machine learning; deep learning; training dataset; multi-sensor data



Citation: Sykas, D.; Zografakis, D.; Demestichas, K. Deep Learning Approaches for Wildfire Severity Prediction: A Comparative Study of Image Segmentation Networks and Visual Transformers on the EO4WildFires Dataset. *Fire* **2024**, *7*, 374. <https://doi.org/10.3390/fire7110374>

Academic Editors: Futao Guo and Yufei Zou

Received: 3 September 2024

Revised: 17 October 2024

Accepted: 18 October 2024

Published: 23 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Since the beginning of the 21st century, annual burned area has been increased [1,2]. Forest wildfires present a significant global challenge, resulting in extensive harm to the environment, economy, and communities. The growing occurrence and severity of these wildfires has been linked [3,4] to climate change, deforestation, and anthropogenic factors. Climate change has been also related to large heatwaves, droughts, and wildfires [5,6]. Furthermore, for the region of Southern Europe, wildfires are estimated to have a direct impact on a country's annual GDP, resulting in a 0.11–0.18% annual decrease, which can

translate from 13 up to 21 billion EUR annually [7]. Precisely predicting the extent of areas than can be affected by wildfires is vital for efficient disaster response, allocation of resources, and preventative strategies. For these reasons, satellite systems are currently being employed, as they can provide on-time monitoring of large areas while maintaining data integrity and consistency [8]. Our research seeks to predict the severity of forest wildfires before they occur, specifically focusing on the potential extent of fire damage in a given area, considering the current and historical forest conditions.

We present and utilize the EO4WildFires dataset [9], a unique benchmark collection combining multispectral data (Sentinel-2), Synthetic Aperture Radar (SAR) data (Sentinel-1), and meteorological information from 45 countries. This extensive, multi-sensor time series dataset, annotated with the European Forest Fire Information System (EFFIS) for wildfire detection and area estimation, encompasses 31,730 wildfire incidents from 2018 to 2022. By integrating Sentinel-2, Sentinel-1, and meteorological data into a unified data cube for each wildfire event, EO4WildFires facilitates a detailed examination of the variables influencing wildfire severity. Our objective is to contribute to the development of more precise and reliable models for forecasting wildfire severity, ultimately aiding policymakers and enhancing global wildfire management strategies.

The EO4WildFires dataset addresses the scientific challenge of creating a model or a suite of models to predict the severity of imminent wildfire events in specific locations. This prediction is based on a combination of current and historical (30-day) meteorological data, along with multispectral and SAR imagery, to accurately represent the forest conditions before a wildfire. The primary objective is not to predict the occurrence of wildfires (i.e., whether a forest wildfire might ignite) but to forecast their severity (size and shape), especially the extent of the area likely to be affected. Developing models that can effectively assimilate these diverse data types is crucial for generating precise severity forecasts. Successfully addressing this problem could significantly aid forest protection agencies and other relevant stakeholders in preparing for and mitigating the impacts of wildfires on the environment and communities.

2. Related Work

In this section, we explore the current research related to the EO4WildFires dataset and its applications, with a particular interest in feature extraction, machine learning, and image segmentation using satellite imagery. Our work focuses specifically on predicting the potential size and shape of a wildfire if it occurs. In contrast, the main focuses of the current state-of-the-art are (i) the intensity, (ii) the ignition, (iii) the spread, and/or (iv) all of these, which encompass different but still-related works.

2.1. Wildfires and Earth Observation

Wildfires, as natural disasters, are becoming more intense and more devastating [10,11], resulting in the destruction of property, the loss of lives, and significant harm to natural resources, including biodiversity, soil, and wildlife [12] while also increasing the risk of developing mental illness [13,14]. They have also been linked to poor air quality [15–17], as well as leading to other natural disasters such as extreme floods, storms, and hail [18]. Satellites play a crucial role in detecting, monitoring, and characterizing wildfires [19–23], but converting satellite imagery into maps is a strenuous task, especially in the course of a disaster [24], as they typically require manual identification of the phenomenon [25]. The availability of these data has also increased, as many high-resolution products are collected daily [26,27]. This subsection overviews significant recent research efforts that generate or leverage Earth Observation data for wildfire detection purposes (including intensity, spread, size, ignition, etc.). To our knowledge, only a handful of research similar to ours exists.

A data-oriented approach is purposed by Víctor Fernández-García et al. [19], which leverages EFFIS, Sentinel data, and Worldclim climatic data in order to assess the intensity of a wildfire (burn severity). Their methodology evaluates burn severity utilizing the

Normalized Burn Ratio (NBR) [28] both for the pre- and post-fire data from Sentinel2's Level 2 A bands 8a and 12. They observe that burn severity is highly related to indices that depict fuel load signature in time (i.e., NDWI). Likewise, topographic and climatic features are also important for more robust classification results. Similarly, the NBR index pre- and post-fire was also assessed for wildfire burn severity in the Piedmont region in Italy (2017) [29]. It was noticed that temporal constraints when selecting paired images is crucial. Furthermore, a compositing algorithm is provided that is not dependent on a specific optical sensor and/or multisensor data, as it can easily be transferred to other sensors.

MODIS fire product effectiveness has been compared to ground wildfire records in Yunnan Province, Southwest China, from December 2002 to November 2015 [30], in order to understand the disparities in the spatial and temporal patterns of wildfires identified using these two methods, to evaluate the omission error in MODIS fire products, and to examine the impact of local environmental factors on the probability of MODIS detecting wildfires. Findings indicate that MODIS detects more than double the number of wildfires compared to ground records, yet the patterns vary significantly. Only 11.10% of 5145 verified ground records were identified by multiple MODIS fire products. The research identifies wildfire size as a key limitation in MODIS's detection capability, with a 50% likelihood of detecting fires of at least 18 hectares. Other influencing factors include daily relative humidity, wind speed, and the altitude at which wildfires occur. This study underscores the necessity of integrating local conditions and ground inspections into wildfire monitoring and simulation efforts globally.

Big data, remote sensing, and data mining algorithms have also been used to predict the occurrence (i.e., ignition) of wildfires using satellite imagery [31]. They developed a dataset of remote sensing data, encompassing crop conditions (NDVI), meteorological conditions (LST), and the fire indicator "Thermal Anomalies" from MODIS instruments on Terra and Aqua satellites. This dataset, available on GitHub, was tested on the Databricks big data platform, achieving a high prediction accuracy of 98.32%. The results underwent various validation processes and were benchmarked against existing wildfire early warning systems. "Next Day Wildfire Spread" [32] is a comprehensive, multivariate, historical wildfire dataset from the United States. This dataset, which covers nearly a decade of remote sensing data, is unique, as it merges 2D fire data with multiple explanatory variables like topography, vegetation, weather conditions, drought index, and population density. The authors used this dataset to develop a neural network that predicts wildfire spread with a one-day lead time, and they compared its performance with logistic regression and random forest models.

The VIIRS dataset alongside meteorological data (temperature, soil moisture, land cover, etc.) has been also utilized for assessing wildfires in Australia [33]. CNN and biLSTM-based architectures are merged into a single model, which is later trained to predict the spread of potential wildfires. Incorporating meteorological and weather index variables has also highlighted the significance of integrating diverse data types, including Earth observation data, for effective forest fire severity prediction [34].

2.2. Deep Learning Architectures

Deep learning technologies are rapidly evolving, particularly in their application to large-scale visual data, such as remote sensing data. Notably, AI models and satellite imagery can be utilized in modeling forest fire severity across different strata [35]. The utilization of machine learning is also heavily examined and could be adapted for severity assessment using Earth observation data [36]. This subsection briefly overviews some recent significant deep learning architectures in this field.

Machine learning methods that combines both CNN and LSTM networks can be leveraged for predicting the chance of a forest fire happening in Indonesia [37]. This approach is particularly significant for developing countries that may lack the resources for expensive ground-based prediction systems. The model achieves a 0.81 (ROC) curve, significantly surpassing the baseline method's maximum of 0.7. Importantly, the model maintains its effectiveness even with limited data, showcasing the potential of ML methods in establishing efficient and cost-effective forest fire prediction systems. Double-Step U-Net [38] is a "double" CNN network that combines regression and semantic segmentation to enhance accurate detection of the affected areas. It is trained using Copernicus EMS manually annotated data. The model is trained to estimate the final severity level of the wildfire. There are 4 levels spanning from 0—no damage to 4—completely destroyed. Their findings suggest that BCE-MSE loss can, in most cases, outperform the state-of-the-art based on RMSE scores.

Advances in feature extraction from satellite imagery have been propelled by deep learning models. A significant development in this area is the introduction of deep residual learning by He et al. [39]. Their ResNet model is a foundational tool in feature extraction, known for its ability to effectively learn from extensive data without encountering the vanishing gradient problem. The Microsoft COCO [40] dataset offers valuable insights into segmenting complex images; a method that can be applied to satellite imagery to improve the identification and analysis of geographical features. The application of EfficientNet [41] has proven beneficial in satellite imagery analysis. This scalable network is designed to optimize accuracy and efficiency, making it well-suited for handling the varying scales and complexities of satellite data. It achieves this by balancing the network's depth, width, and resolution.

ALLConvNet [42] is a convolutional model that is able to provide a 7-day wildfire burn forecast with the corresponding probability maps. This model is trained using features extracted from various datasets for Victoria and Australia between the years 2006 and 2017. According to the authors, lightning flash density, total precipitation, and land surface temperature tend to be the most significantly contributing features across all their tests, while wind, distance from the power grid, and terrain affected the models' performance the least. Zhang et al. [43], by evaluating different models, likewise suggest that features such as temperature, soil moisture indices (NDVI), and accumulative precipitation have a major impact on model performance throughout the year.

Contrastive Captioner (CoCa) [44], an image–text encoder–decoder foundation model, blends contrastive loss with captioning loss, incorporating elements from CLIP [45] and SimVLM [46]. It focuses initially on unimodal text representations before tackling multi-modal image–text representations. The model, pre-trained on extensive web-scale alt-text data and annotated images, excels in various tasks like visual recognition and image captioning, achieving a remarkable 91.0% top-1 accuracy on ImageNet with a fine-tuned encoder. The Swin Transformer V2 model [47]; a residual post-norm method with cosine attention; a log-spaced continuous position bias method; and SimMIM, a self-supervised pretraining method set new performance records on various benchmarks while being 40 times more data and time-efficient than models of a similar scale developed by Google. Vision Transformers (ViTs) [48] apply the transformer directly to image patches. This model, when pre-trained on large datasets, outperforms traditional convolutional networks in various image recognition benchmarks, with significantly lower training resource requirements.

Chen et al. [49] present a novel approach for discovering algorithms aimed at optimizing deep neural network training. Their technique, termed evolved sign momentum and code-named lion, is a memory-efficient optimizer surpassing popular optimizers like Adam [50] and Adafactor [51] across multiple tasks. Lion [49] enhances ViT's accuracy on ImageNet by up to 2%, reduces pre-training computational requirements on JFT datasets by up to 5×, and outperforms Adam in diffusion models and other learning tasks. It demonstrates better efficiency and accuracy, particularly with larger training batch sizes and reduced learning rates.

3. Dataset

The purpose of creating the EO4Wildfires dataset is to firstly to provide a feature structure in an AI-ready format (satellite images and weather data) that correlates with the wildfire results. Secondly, having this AI-ready format, the purpose is to model how well the size and shape of wildfires can be predicted. This can serve as a risk-planning tool that identifies areas of high risk during wildfire seasons. The EO4WildFires dataset [9], which incorporates data from the European Forest Fire Information System (EFFIS), Copernicus Sentinel-1 and 2, and NASA Power, is utilized. The spatial resolution for all pixels is 10 m, with an exception for the meteorological data, which cover the whole event region. Each event corresponds to a single data cube (netcdf file) that follows a common data loading routine and is then funneled into the various deep learning models. The following sections describe the data components and structure of the EO4WildFires dataset, along with an exploratory analysis section.

3.1. European Forest Fire Information System

EFFIS [52,53] is a key platform providing current and historical data on forest fires in Europe, operated by the European Commission Joint Research Centre (JRC). It offers a wealth of information, including maps, data on fire locations, sizes, intensities, affected vegetation types, and land use. Additionally, EFFIS delivers daily wildfire danger forecasts and a fire danger rating system, all based on meteorological data. This platform is an indispensable tool for forest protection services, researchers, and policymakers who rely on precise and timely wildfire information for monitoring and management purposes in Europe.

3.2. Copernicus Sentinel-1 and 2

Sentinel-1, part of the Copernicus Programme developed by the European Space Agency (ESA), is a mission involving two polar-orbiting satellites that consistently gather Synthetic Aperture Radar (SAR) imagery in high resolution over terrestrial and coastal regions. Its C-band frequency SAR sensor provides all-weather, day-and-night imaging capabilities, useful for various applications like land and ocean monitoring, disaster response, and maritime surveillance. Sentinel-1's data, accessible for free through the Sentinel Data Hub or third-party providers, is vital for long-term, global, environmental, and natural resource management.

Similarly, **Sentinel-2**, also from the Copernicus Programme and developed by ESA, comprises two satellites that capture optical imagery in high resolution. The onboard sensor records data across 13 spectral bands, facilitating the observation and monitoring of various phenomena such as vegetation, land use, natural disasters, and urban growth. Like Sentinel-1, Sentinel-2's data are freely accessible and are pivotal for environmental monitoring, land use mapping, and disaster management due to their global coverage and high revisit frequency.

Sentinel-2 Level 1C data provide top-of-atmosphere (TOA) reflectance values, which are closer to the raw sensor data, instead of **Sentinel-2 Level 2A**, which provide atmospherically corrected bottom-of-atmosphere (BOA) reflectance values. This allows researchers to apply their own correction and preprocessing techniques, potentially leading to more control over the quality and characteristics of the final dataset. In some cases, using S2L1C data can ensure a standard baseline for comparison across different studies and algorithms. Since S2L2A data already include atmospheric correction, the preprocessing steps can vary depending on the algorithms and parameters used by different data providers. By starting with L1C data, researchers can standardize the preprocessing steps, ensuring that comparisons are fair and reproducible.

3.3. NASA Power

NASA Power [54] is a scientific initiative offering solar and meteorological datasets for applications in renewable energy, building energy efficiency, and agriculture. It aims to

make these datasets more accessible and applicable for a diverse range of users, including researchers, policymakers, and practitioners. The project focuses on developing reliable and precise methods for measuring and forecasting solar and meteorological parameters like solar radiation, temperature, precipitation, and wind speed. NASA Power equips users with a variety of datasets, tools, and services to aid research and decision-making in energy, agriculture, and other sectors, contributing to sustainable and resilient future planning based on up-to-date and accurate solar and meteorological information.

3.4. Structure

EO4WildFires [9] comprises a blend of meteorological data and multispectral and SAR satellite imagery, along with wildfire event information sourced from the EFFIS system. EFFIS is tasked with delivering current and dependable data on wildfires across Europe, aiding forest protection efforts.

This dataset was developed utilizing the Sentinel-hub API and the NASA Power API. The Sentinel-hub API facilitates Web Map Service (WMS) and Web Coverage Service (WCS) requests, enabling the downloading and processing of satellite imagery from various sources. Meanwhile, the NASA Power API offers access to solar and meteorological data compiled from NASA's research, which is instrumental in sectors like renewable energy, building energy efficiency, and agriculture.

In the EO4WildFires dataset, each of the 31,730 wildfire events is encapsulated as a data cube in NetCDF format. The process of data collection for each event involves several steps:

- Bounding box coordinates and the event date serve as the initial inputs.
- Meteorological parameters are derived using the central point of the area, collected from the day before the event to 30 days prior.
- Sentinel-2 imagery is cropped according to the bounding box coordinates. To address cloud cover issues, a mosaicing process https://custom-scripts.sentinel-hub.com/sentinel-2/monthly_composite/# (accessed on 15 September 2024) is employed, selecting the optimal pixels from the last 30 days before the event.
- Sentinel-1 images are similarly cropped using the bounding box. Due to SAR images not being affected by cloud cover, only the most recent image before the event is used. Both ascending and descending images are included.
- A burned area mask is provided, representing the burned area as a Boolean mask based on EFFIS vector data rasterized onto the Sentinel-2 grid.

The EO4WildFires dataset, as presented in Table 1, offers a comprehensive range of features suitable for wildfire severity prediction. This includes a fusion of meteorological data and satellite imagery, which sheds light on the environmental conditions conducive to wildfires. The EFFIS-provided wildfire event labels are instrumental for training models to predict and respond to such events.

The mosaicing process for Sentinel-2 images creates a monthly composite, selecting the best pixel for each day in the preceding 31 days. This selection, based on band ratios, aims to minimize cloud cover. Depending on the level of blue in the image, different criteria are applied for pixel selection, with adjustments for the presence of water or snow. The resulting composite offers a clear representation of the last 31 days in the selected area.

The dataset encompasses 31,730 wildfire events from 2018 to 2022 across 45 countries, affecting 8707 level-4 administrative areas. It integrates the GDAM database to align detected EFFIS events with administrative boundaries. Analysis of the data reveals that the median wildfire size is 31 hectares, with an average of 128.77 hectares. The largest recorded wildfire was 54,769 hectares in Antalya, Turkey (2021), followed by a 51,245-hectare fire in Evoia, Greece (2021).

Table 1. Parameters included in each wildfire event for each data source.

Channel	Meteorological Data	Sentinel-1	Sentinel-2
1	Ratio of actual partial pressure of water vapor to the partial pressure at saturation (RH2M)	VV	Band 02
2	Average temperature (T2M)	VH	Band 03
3	Bias-corrected average total precipitation (PRECTOTCORR)	$\frac{VV - VH}{VV + VH}$	Band 04
4	Average wind speed (WS2M)	-	Band 05
5	Fraction of land covered by snowfall (PRECSNOLAND)	-	Band 08
6	Percent of root zone soil wetness (GWETROOT)	-	Band 11
7	Snow depth (SNODP)	-	-
8	Snow precipitation (FRSNO)	-	-
9	Soil moisture (GWETTOP)	-	-

Figure 1 maps the level-4 administrative boundaries of the wildfires recorded in the dataset from 2018–2022, highlighting areas like the Mediterranean with high wildfire concentrations. Figure 2 aggregates the five countries that experience the most wildfires per year (for the years 2018–2022) in the dataset. Ukraine (UA) and Turkey (TR) account for approximately 25% of the total fires in the dataset. Adding Algeria (DZ), Spain (SP), Romania (RO), and Syria (SY) increases the total fire percentage to 45%.

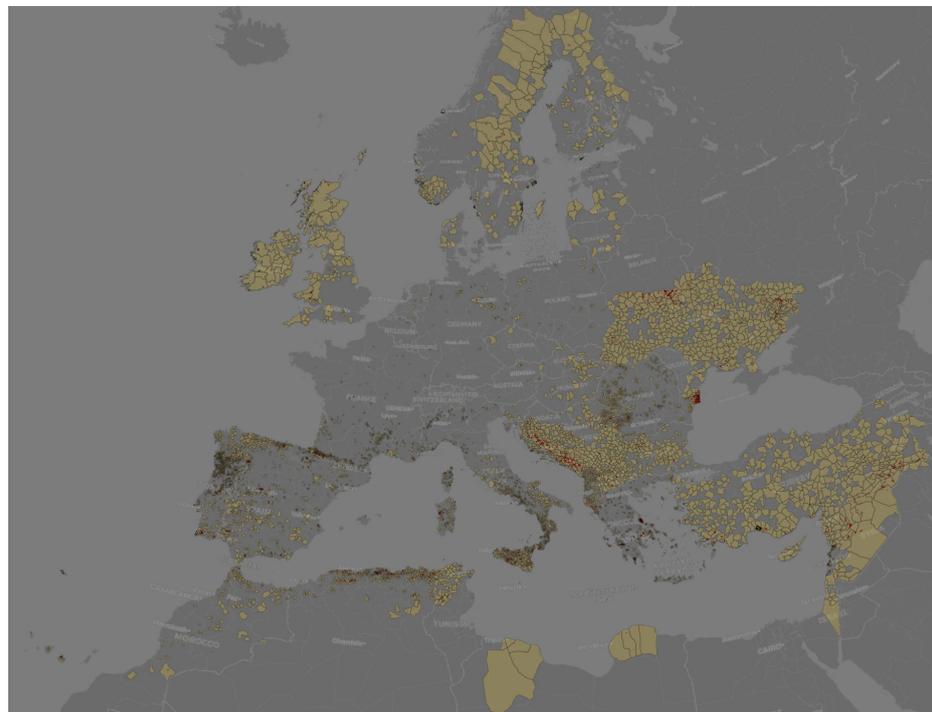


Figure 1. Wildfire events (2018–2022) with the corresponding affected level-4 administration boundaries—yellow polygons: administrative boundaries (level 4) of affected areas, red points: locations of wildfires events.

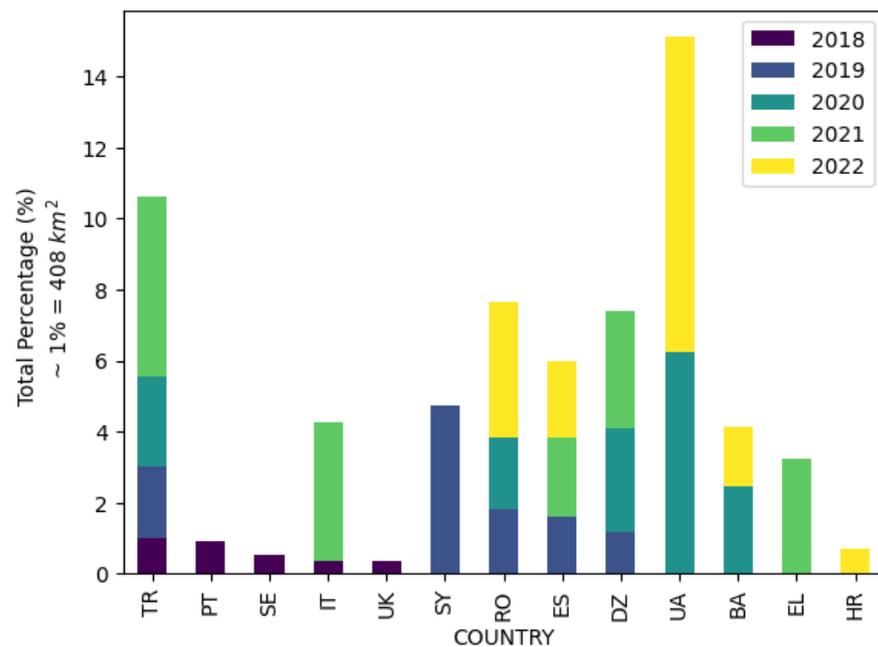


Figure 2. Top contributing countries to wildfires annually. TR: Turkey, PT: Portugal, SE: Sweden, IT: Italy, UK: United Kingdom, SY: Syria, RO: Romania, ES: Spain, DZ: Algeria, UA: Ukraine, BA: Bosnia Herzegovina, EL: Greece, HR: Croatia.

3.5. Explanatory Analysis

This section presents a comprehensive analysis of the EO4WildFires dataset, presenting key metrics and statistical figures to better understand the characteristics and quality of the data. The full dataset, including training, validation, and test sets, comprises 31,740 samples.

Each samples is resized to match a 224×224 pixel size. Samples that contain less than 10 burned pixels, which is equivalent to less than 1 km^2 of burned area, have been excluded from certain experiments for a more focused analysis.

- The median value of burned pixels metric is calculated to provide insights into the typical extent of fire damage per event.
- Average fire area: An average measure of the area affected by fire in the dataset.
- Percentage of unaffected pixels: Indicates the proportion of pixels in each sample that were not impacted by fire, offering a perspective on the spatial extent of wildfires.

The dataset records a total burned area of $38.275.473.851 \text{ m}^2$. The median percentage of burned pixels in all events is 8.8%. The percentage of total burned pixels represents 7.4% of the total pixels across all events, offering a perspective on the overall impact of wildfires in the dataset.

Furthermore, the dataset includes a considerable number of NaN (Not a Number) values across various variables, particularly in the burned_mask, which has a total of 478.047.563 (56%) NaN values. These represent pixels unaffected by fire, which are replaced with zeros to maintain consistency. Moreover, Sentinel-1 products for both ascending and descending orbit directions have a 1.8 percentage of NaN values of the total pixels. These values are likewise replaced with zeros.

$$\bar{m} = \sum_{i=1}^n \frac{s_i m_i}{s} \quad (1)$$

where s_i is the group size, s is the total size, n is the number of groups, and m_i is the group median value.

An analysis using a weighted mean across different channels (Table 2) enables a more detailed understanding of central values in the dataset. Note that the median is used instead of the mean due to its significantly biased mean value, possibly resulting from anomalies in data collection or processing.

Table 2. Weighted mean of median values (See Equation (1)).

Channel	S1GRDA	S1GRDD	S2L2A	
1	0.09	0.08	0.05	
2	0.02	0.02	0.07	
3	0.61	0.63	0.09	
4	-	-	0.13	
5	-	-	0.23	
6	-	-	0.23	
RH2M	T2M	PRECTOTCORR	WS2M	FRSNO
70.82	11.97	0.24	2.2	0.01
GWETROOT	SNODP	FRECSNOLAND	GWETTOP	
0.43	2.16	0	0.4	

In addition to the aforementioned metrics, the average standard deviation using Equation (2) is calculated. This measure provides insights into the variability within the dataset, considering different sample sizes and standard deviations for each channel:

$$\bar{s}d = \sqrt{\frac{\sum_{i=1}^n (s_i - 1)sd_i^2}{s - n}} \tag{2}$$

where s_i is the group size, s is the total size, n is the number of groups, and sd_i is the group std value.

3.6. Data Loading

The EO4WildFires dataset consists of distinct NetCDF (.nc) files, each representing a unique wildfire event. The dataset structure is designed to facilitate comprehensive analysis and to serve as a standalone structure:

- **Dataset composition:** Every file in the dataset encapsulates a comprehensive data cube representing a single wildfire event. This data cube is comprised of multiple channels:
 - Sentinel-2 Level 2A product: Six channels from Sentinel-2, providing detailed multispectral imagery.
 - Sentinel-1 product: Six channels, including VV, VH, and the ratio $(VV - VH) / (VV + VH)$, split evenly between ascending and descending products. This accounts for a total of twelve unique channels from both Sentinel-1 and Sentinel-2.
 - Weather metrics: Nine meteorological variables covering the 31 days leading up to the event, offering a holistic view of the environmental conditions prior to each wildfire.
- **Binary classification mask:** An integral part of each file is the binary mask that delineates the wildfire’s footprint. This mask is crucial for the classification and severity analysis of the event.
- **Geospatial encoding:** Drawing inspiration from the work of Prapas et al. [55], we employ longitude and latitude information for pixel positional encoding. A sine and cosine transformation is applied, resulting in four-channel encoding. Consequently, each loaded sample in the dataset is a 16-channel data cube with varying dimensions due to the differing widths and heights of individual samples.

To accommodate state-of-the-art architectural requirements, each data cube Figure 3 ($C \times H \times W$, where C is channels, H is height, and W is width) undergoes a process of padding or segmentation to conform to a standard resolution of 224×224 pixels. This ensures uniformity across all input samples, with each one being a $16 \times 224 \times 224$ data cube, accompanied by a 31×9 weather data matrix. During the data loading process, any missing values are substituted with zeros. This approach is also applied to the padding process, ensuring data integrity and consistency. The dataset does not employ data augmentation techniques, maintaining the authenticity of the original satellite and meteorological data. The dataset, when adjusted to a 224×224 resolution, includes 24,794 samples with fewer than 10 burned pixels (equivalent to less than 1 km^2 of the burned area). To refine the dataset for more effective experiments, these samples are excluded from the analysis. The reason for excluding the labels that contain less than 1 km^2 is due to the fact that the resolution of the EFFIS is 1 km^2 ; thus, it is the limit of the labels details.

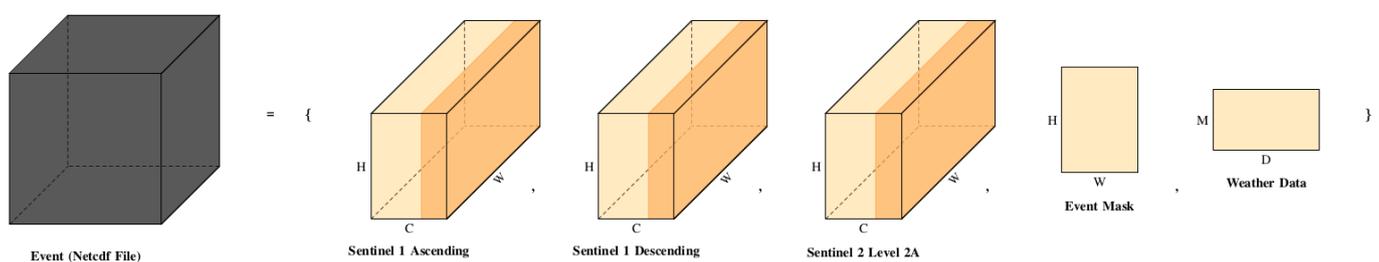


Figure 3. EO4Wildfires data cube structure.

4. Experiments

In this section, we present the experiments performed using different deep learning architectures. In every experiment, the target is the same to maximize the accuracy of the estimation of the burned area. Two main types of network architectures (with varying parameters) are examined:

- Image segmentation networks
- Visual transformers

4.1. Data Processing and Machine Learning Pipeline

Figure 4 presents the dataflow for training and evaluating the models in each experiment. The input data streams are initialized: Sentinel-1 ascending and descending, Sentinel-2 Level 2A, event metadata (latitude and longitude), weather data. The event metadata are encoded to transform the latitude and longitude into a format that is usable by machine learning models. The encoded metadata and the Sentinel-2 imagery are concatenated (joined together) with the Sentinel-1 data to create an aggregated input. This combines all the different types of data into a single, cohesive dataset. Weather data are also prepared to be fed into a Fully Connected (FC) network, a type of neural network used to process structured data. The aggregated input is then passed through various deep learning architectures. The FC network processes the weather data in parallel to the image data processing. The deep learning models output a “Prediction Mask”, which is a binary mask predicting the burned area. This predicted output is then compared against the “Event Mask”, which is the ground truth/actual observed data of the burned area. The results of the comparison (how close the prediction is to the event mask) are monitored, and various metrics are calculated to evaluate the performance of the models. These metrics include F1 score, IOU score, and aPD.

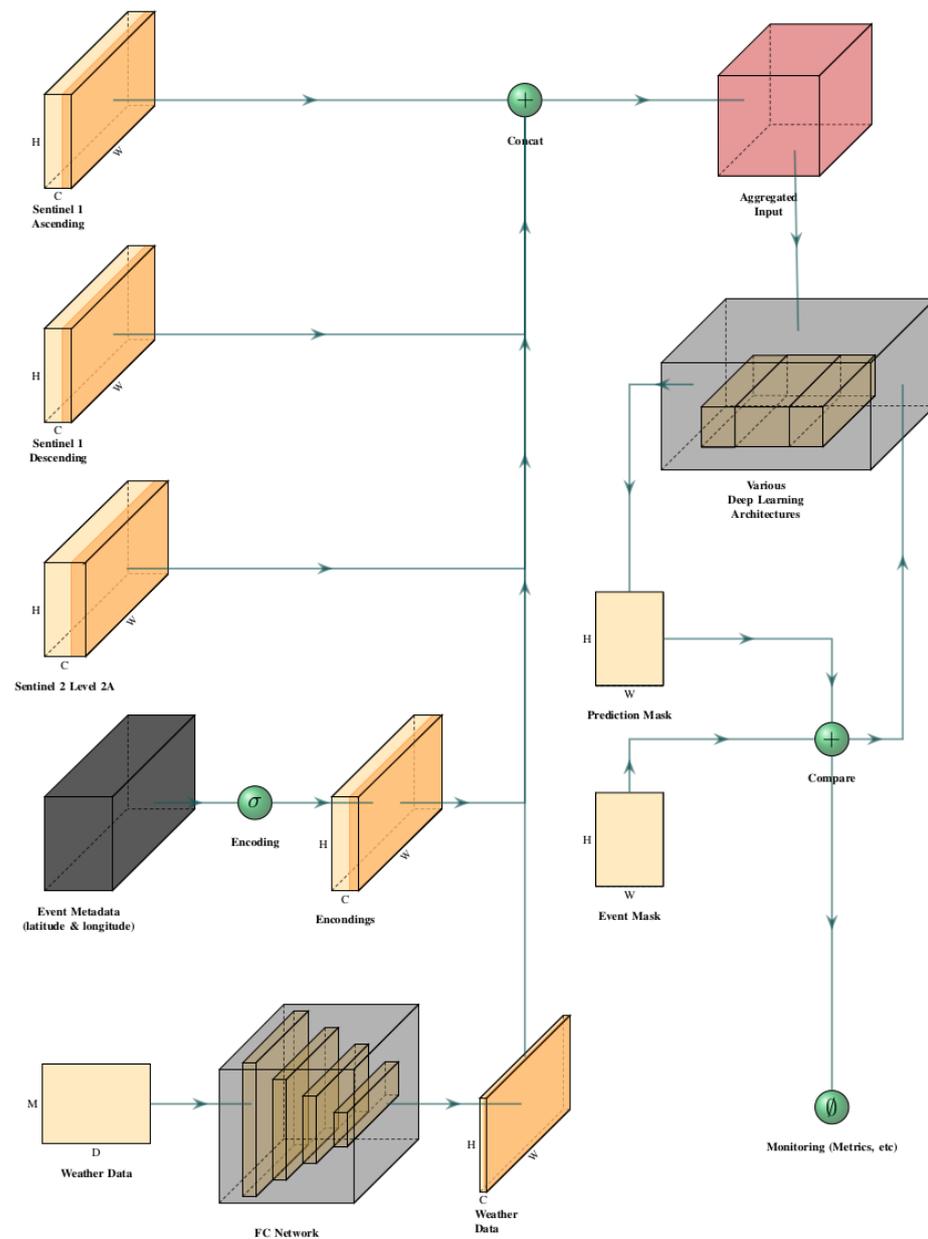


Figure 4. Schematic representation of the data processing and machine learning pipeline for wildfire event analysis using the EO4WildFires dataset.

In a nutshell, Figure 4 describes the complex machine learning workflow that takes in satellite imagery, event metadata, and weather data to predict the extent of wildfire events. We emphasize data integration and transformation before applying machine learning models for prediction and subsequent performance evaluation.

The original EO4WildFires dataset is split into training, validation, and testing subsets, with 20,307 events in the training set, 5077 events in the validation set, and 6346 events in the test set. The division of the dataset is the same for all three experiments to ensure comparability of the results. The goal of the experiments is to determine which input features provide the best prediction performance for the size of a wildfire event. Heavily inspired by popular datasets like coco [40], three index files (train/val/test) are created that operate as literal file catalogs. Each row in the index files refers to a specific file in the disk.

4.2. Image Segmentation

Central to the architecture of every image segmentation model is the encoder, a component typically built upon well-known deep convolutional networks such as ResNet and EfficientNet. The encoder's primary function is to extract features from the input data, a critical step in the segmentation process. Although transfer learning is frequently employed to leverage pre-trained weights on three-dimensional images—thereby retaining knowledge from extensive prior training—this approach often falters when applied to satellite imagery. To counteract this issue, we have retrained the encoders' weights from scratch for each experiment within our study.

Alongside re-initializing the weights, we have standardized the training approach across experiments. A consistent optimizer schedule was applied, and CrossEntropy loss was selected as the training criterion. This decision was made despite initial trials with Dice loss, which was quickly abandoned due to its subpar performance, likely aggravated by class imbalance within the dataset. Dice loss tends to underperform when class imbalances are significant [56]. The models were built using the segmentation models library https://github.com/qubvel/segmentation_models.pytorch, (accessed on 15 September 2024), chosen for its robustness and ease of use. The library provided a solid foundation upon which we could construct and evaluate our models. Our segmentation task is binary, determining whether a pixel will be impacted by fire or remain unaffected. This classification allows us to calculate True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). From these values, we derive several metrics:

- F1 score: A measure combining precision and sensitivity.
- IoU score: The Intersection over Union (IoU), or Jaccard index, computed using the formula $\frac{TP}{TP+FP+FN}$.
- Average Percentage Difference (aPD): Metric indicating the model's deviation from the actual observed data, derived from the percentage difference between the ground truth and predicted values.

For experiments utilizing direct models from the segmentation models library, an additional layer is incorporated into the architecture. This layer transforms the weather data matrix into an extra channel, facilitating the combination of meteorological data with the Sentinel data cube. The resultant input size of $13 \times 224 \times 224$ provides a comprehensive view of each wildfire event, integrating satellite and weather data.

4.3. Visual Transformers

Building on the work of Prapas et al. [55], our experiments explore the adaptation of Visual Transformers (ViT) for the task of image segmentation within the context of satellite imagery. The architecture under consideration, TeleViT, presents a novel approach specifically tailored to satellite data. To utilize their model effectively, we meticulously followed the guidelines set forth in their publication, making necessary adjustments to accommodate our dataset.

The encoding channel was implemented as delineated in their code repository, ensuring that our model retains the intended data format throughout training. Consistent with previous steps that anticipated input sizes of 224×224 , we maintained this dimensionality to guarantee that the data fed into the network would be evenly divisible by 224. Consequently, we opted for a patch size of 56, thereby subdividing each 224×224 sample into smaller segments of 56×56 . This process effectively augmented the dataset size by a factor of 16.

A notable deviation from the TeleViT model is the requirement for weather data. Their model is configured to process 10 metrics spanning 10 months, which differs from our dataset's structure of 9 metrics across the last 31 days. To reconcile this discrepancy, we appended a dedicated fully connected network to the model architecture. This ancillary network's sole purpose is to transform our unique weather data dimensions into a compatible format for the TeleViT model. The transformed vector is then reshaped to align with

the model's channel size and integrated into the model's input, akin to the methodology applied in segmentation model experiments.

For training, we employ CrossEntropy loss in conjunction with the Adam optimizer. A learning rate scheduler on plateau is utilized to modulate the learning rate based on the validation loss, ensuring that the model's learning is both effective and efficient.

4.4. Example Use Cases

In this section, a number of example case studies are presented to demonstrate the usability and potential of forecasting wildfire size in a more qualitative manner. The case studies are based on the Copernicus Emergency Management Service (EMS)—Mapping service, specifically its Rapid Mapping Portfolio and the EFFIS dataset.

EMS offers rapid geospatial information in response to emergencies, utilizing satellite imagery and other data. It aims to support emergency management activities by providing standardized products, such as pre-event and post-event analysis, including fast impact assessment and detailed damage grading. The service operates under two production modes to cater to urgent and less-urgent needs, ensuring timely delivery of critical information for disaster response efforts. Among the geospatial products of the EMS are the burned areas, which are produced using very-high-resolution satellite images (<0.5 m) with high-quality procedures that involve manual digitization and corrections to produce the final maps. For this purpose, we utilized the following:

- EMSN077: Post-disaster mapping of forest fires in De Meinweg National Park, Germany–Netherlands border.
- EMSN090: Wildfires in the Piedmont region, Italy.

Figure 5 shows the location of the case studies on a map. The cases of Northern Italy case and The Netherlands are based on EMSN090 and EMSN077, respectively, while the rest rely on the EFFIS.



Figure 5. Use case map overview.

Table 3 and Figures 6 and 7 depict the number of pixels predicted versus the ground truth for wildfire areas while providing a detailed comparison of the predictive capabilities of the models used in the study against actual observed wildfire impacts. Table 3 and Figures 6 and 7 compare the predicted area affected by wildfires (in terms of pixel count) to the actual (ground truth) area impacted. For each case study, the table shows a percentage difference between the model's predictions and the actual observed data. The percentage differences range from very close estimations of the ground truth to larger underestimations, illustrating the challenges in accurately predicting the extent of wildfire damage

using satellite imagery and machine learning models. The accompanying figure visually represents the data provided in Table 3 and Figures 6 and 7, plotting the predicted pixel counts against the ground truth for each case study.

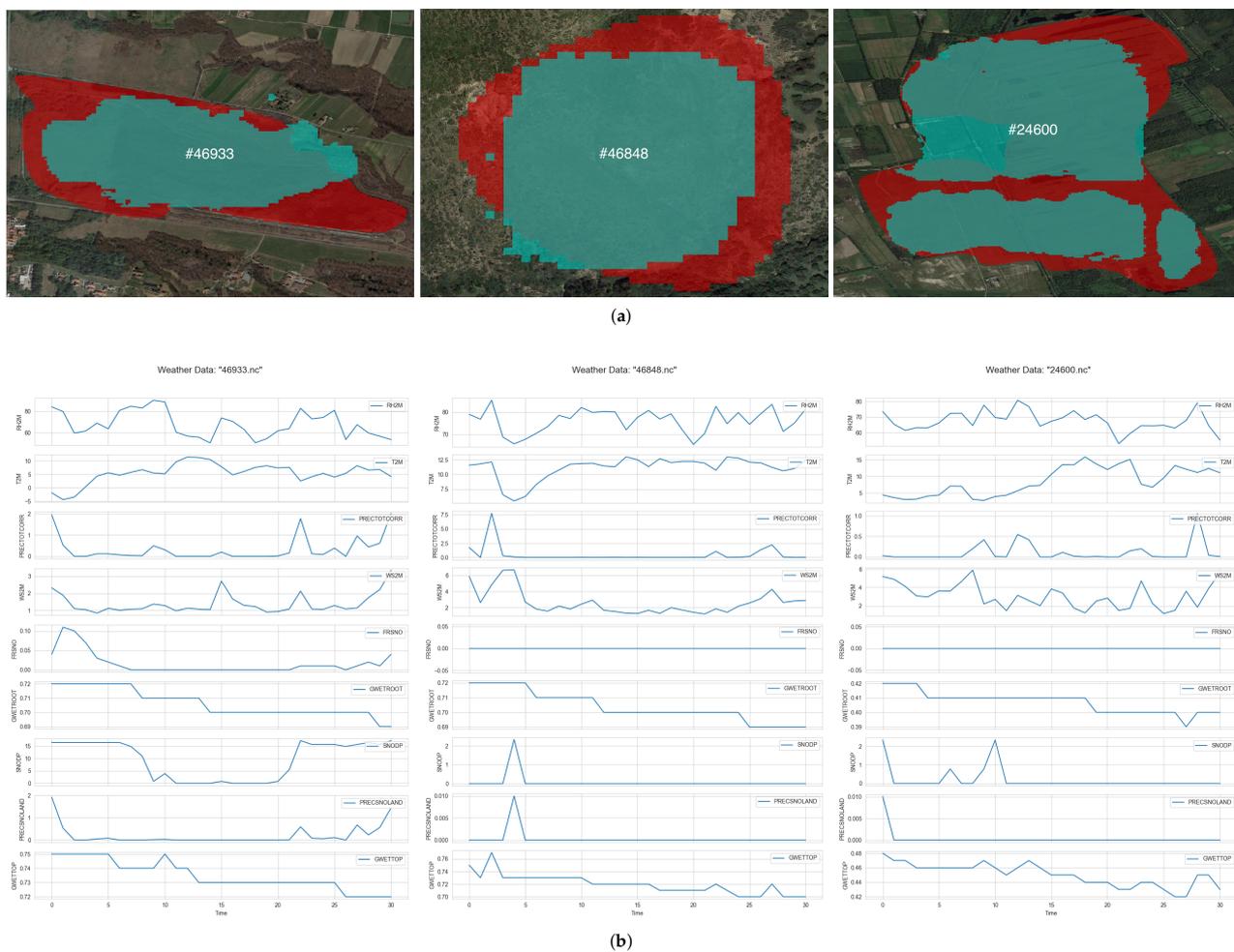


Figure 6. Case #46933, #46848, #24600. (a) Predicted (cyan) vs. ground truth (red); (b) 30-day past meteorological time series variables.

The key takeaway from the case studies’ examination is that the developed methodology can be used to forecast the wildfire size if it actually ignites. Table 3 shows that although the errors will be in the range of 20–25%, this is a constant underestimation of the predictor. Thus, this gives a minimum baseline for evaluating upcoming risks during the fire season. Our proposed methodology is not intended to be used as a precision wildfire spread model but rather as a tool to forecast the potential size and shape of the wildfire if it occurs to act as a utility tool that helps in the planning phase. Although shape is not the explicit optimization parameter, the models learn the relevant pattern to predict the shape as they are trained, since they are image segmentation models.

Table 3. Number of pixels predicted, ground truth, and corresponding % difference.

Case Study #	Predicted	Ground Truth	% Difference
54,278	4620	4486	2.99
24,600	49,641	63,619	−21.97
46,848	750	1075	−30.23
46,933	5589	8211	−31.93
54,455	844	974	−13.35
55,463	172	217	−20.74

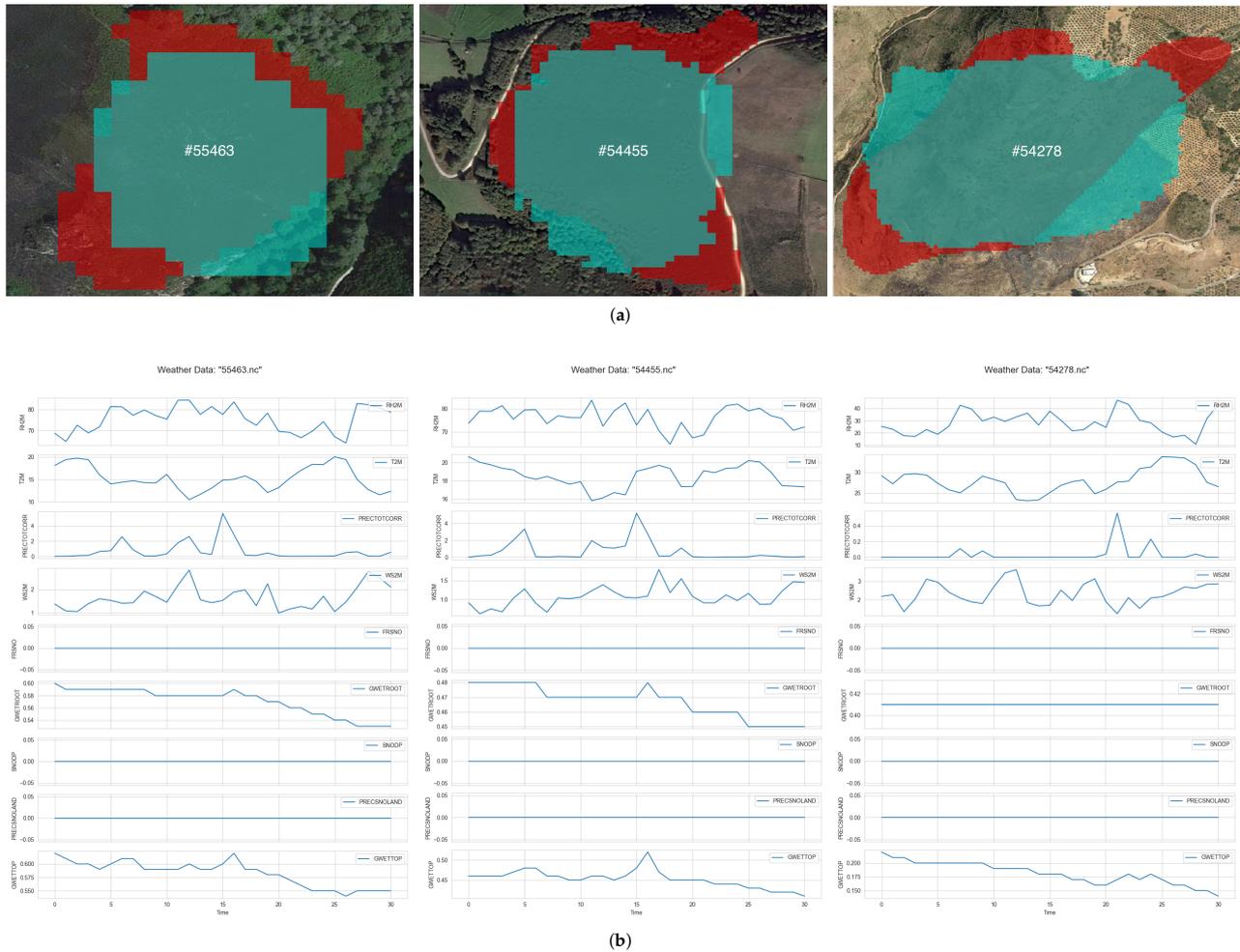


Figure 7. Case #55463, #54455, #54278. (a) Predicted (cyan) vs. ground truth (red); (b) 30-day past meteorological time series variables.

5. Results

Table 4 serves as a benchmark, presenting the initial results of applying various image segmentation networks and a visual transformer model to the EO4WildFires dataset. It lists different configurations of ResNet encoders paired with segmentation models like Unet++ and LinkNet. The table also includes metrics that evaluate the models' performance in predicting the burned area, namely F1 score, Intersection over Union (IoU) score, and average percentage difference (aPD). From this baseline comparison, we can observe that the TeleVIT (global) model outperforms every other model regarding average percentage difference, but it fails in F1 score and IoU. Table 5 further iterates on the experimentation by removing samples with empty labels or below the threshold size (e.g., 1 km²) only from the testing set while keeping the full dataset for training. By keeping empty-labeled data in the training set and removing empty labels in the testing set, we can observe the robustness of each model, as its generalization is tested. Notably, TeleVit (global) error (aPD) is increased while the F1 and IoU metrics are decreased. This indicates that the model might not be good at distinguishing between the two classes. Table 6 refines the experimental approach by removing samples with empty labels (full zeros) or below a certain size threshold (e.g., 1 km²) from both the training and testing phases.

Table 4. Baseline models.

Encoder	Model	F1 Score	IOU Score	aPD
ResNet18	Unet++	0.85	0.74	56.2
ResNet34	Unet++	0.86	0.76	64.2
ResNet50	Unet++	0.85	0.75	77.8
ResNet18	LinkNet	0.87	0.76	66.3
ResNet34	LinkNet	0.86	0.75	70
ResNet50	LinkNet	0.86	0.75	69
-	TeleViT (global)	0.84	0.72	14.9

Table 5. Baseline models trained without empty labels.

Encoder	Model	F1 Score	IOU Score	aPD
ResNet18	Unet++	0.86	0.76	96.5
ResNet34	Unet++	0.86	0.75	97.6
ResNet50	Unet++	0.86	0.76	86.5
ResNet18	LinkNet	0.86	0.76	91.8
ResNet34	LinkNet	0.87	0.76	85.9
ResNet50	LinkNet	0.87	0.75	103
-	TeleViT (global)	0.84	0.72	79.4

Table 6. Baseline models trained and tested without empty labels.

Encoder	Model	F1 Score	IOU Score	aPD
ResNet18	Unet++	0.87	0.77	50.4
ResNet34	Unet++	0.87	0.76	52.1
ResNet50	Unet++	0.87	0.76	44.9
ResNet18	LinkNet	0.87	0.77	48.6
ResNet34	LinkNet	0.87	0.77	44.8
ResNet50	LinkNet	0.86	0.76	55.1
-	TeleViT (global)	0.83	0.71	58.3

The F1 and IoU scores are relatively consistent across the different encoder and model combinations. This suggests that all the models tested have a comparable ability to predict the burned area in wildfire events. The performance seems to slightly improve when the models are trained without empty labels or very small events (Table 6) compared to the baseline models (Table 4). This indicates that filtering out less-informative data can lead to more accurate predictions by focusing on more substantial wildfire events. The ViT model tends to perform better in the APD metric because it tends to predict the non-dominant class, which is non-burned pixels. This is better understood when accessing F1 and IoU scores.

Table 5 shows that when the models are trained without the empty labels but tested on the entire dataset, the performance metrics (F1 and IoU) do not significantly deteriorate, suggesting that the models are capable of generalizing well from more-substantial to less-substantial events. The Visual Transformer (TeleViT) model's performance, as indicated by the F1 and IoU scores, is marginally lower than that of the image segmentation networks, implying that while visual transformers are promising, the segmentation networks are currently more effective for this specific task of wildfire analysis. The aPD values are not consistently better or worse across the experiments, which suggests that the models' ability to estimate the actual size of the burned area varies. This metric could be influenced by factors such as the complexity of the event or the quality of the input data.

The image segmentation networks seem to slightly outperform the visual transformer architecture in this context. Moreover, the practice of filtering training data to remove noise (empty labels or very small events) appears to enhance model performance, indicating the importance of high-quality training data for machine learning tasks in satellite imagery

analysis. The results provide a strong basis for further refining the models and data preprocessing techniques to improve wildfire severity prediction using deep learning.

6. Discussion

This study has presented an approach to use the potential of deep learning architectures in the domain of wildfire severity (size and shape) analysis using the EO4WildFires dataset. The EO4WildFires dataset is a comprehensive collection that includes Sentinel-1 and Sentinel-2 satellite imagery, event metadata (geolocated burned mask), and weather data. These multi-source data have been structured into a unified format, with event metadata encoded to be machine learning compatible. All data streams were aggregated into a cohesive dataset for input into various deep learning models.

Through experimentation with image segmentation networks and visual transformers, we have pursued the goal of maximizing the accuracy of burned area size estimation. Our approach can be used to predict the size and shape of a wildfire if it ignites, but it does not estimate the probability of its ignition or spread. Moreover, the severity is measured in terms of burned area surface and not in terms of potentially incurred costs (socio-economic, environmental, or other, i.e., due to human losses and injuries or damages to properties, infrastructure, and ecosystems).

Traditional image segmentation networks that rely on convolution encoders such as ResNet and EfficientNet have demonstrated a slight edge over the visual transformer architecture in predicting wildfire-affected areas. This suggests that, at least in the context of this dataset, conventional segmentation approaches are currently more effective than their transformer-based counterparts. Although transformers are now being studied for their ability to learn in-context and their ability to solve problems that they were not trained for [57], their enormous number of parameters does not allow them to perform well with the big nature of satellite data. This is because experimenting with bigger and more complex transformer-based architectures tends to require significant computational resources [58], as memory requirements scale exponentially with the attention model's size [59].

Filtering the training data to exclude noise, such as empty labels or events below a certain threshold, has been found to improve model performance across all evaluation metrics. This emphasizes the importance of high-quality training data in machine learning, particularly in the field of satellite imagery analysis, where the vast presence of non-burned areas can pose a significant challenge.

The robustness of the models is reflected in the consistency of F1 and IoU scores across different encoder and model combinations, indicating a reliable ability to predict the burned area across a variety of deep learning approaches. The dataset was split into training, validation, and test, and these splits were common for all experiments. For the first experiment, various models were trained with all the samples in order to establish a baseline for comparison. After that, all empty-label data were removed from the training and validation sets but were kept for the test set. The ResNet34 encoder with LinkNet model was better at generalizing than all the other models, increasing its error rate from 70% aPD to 85.9%, while TeleViT had the greatest increase in errors, from 14.9% to 79.4%. Finally, the last experiment removed empty labels from all training, validation, and test sets. ResNet34 with LinkNet still outperformed the other variants, scoring 44.8% aPD. Consequently, the main advantage of the ResNet34/LinkNet model is its ability to generalize, as it tends to minimize the aPD error to the various concepts tested. On the other side, it seems that choosing the best size is essential for establishing a good prediction model. From Tables 4–6, it is apparent that ResNet34/LinkNet outperforms all the other models, but when the encoder is switched from ResNet34 to either ResNet18 or ResNet50, its performance takes a toll, resulting in the worst models. Therefore, a “model” selection procedure needs to be in place, which requires both computer resources and time and adds complexity.

7. Conclusions and Future Work

Based on the research performed, several key findings are extracted. Excluding empty labels or very small wildfire events improved model prediction accuracy significantly, demonstrating the value of high-quality data. Regarding the model architecture, visual transformers like TeleViT have promising results, but segmentation networks performed better in terms of F1 and IoU scores. The models demonstrated utility for estimating the potential size and shape of wildfires. This can be helpful in aiding resource planning during fire seasons. After conducting numerous experiments, it is apparent that class imbalance can affect the model's final performance. The latter is rather important, as remote sensing-categorized data are limited and not always available to researchers. Other domains eliminate this problem by utilize transfer learning, but in RS, this is not always possible because the model's input is typically adjusted to the area under examination or the available equipment. However, we believe that adjusting traditional image segmentation networks to receive multichannel imagery can still be used to tackle important problems such as measuring the severity of wildfires.

Future directions of this work are essential for establishing a better understanding to the complex mechanisms through which each ecosystem responds to wildfires, therefore enabling better forecasting and management of wildfire events. For instance, taking into consideration the type of forest, the terrain of the area, and various other external factors, i.e., distance from urban fabric, etc., may further enhance the predictor's performance. Another possible direction of this work would be to enrich the dataset used by incorporating global events and consequently training a model using all available samples. Once trained, this model can serve as a base model, which then can be fine-tuned to perform area-specific wildfire severity prediction.

Author Contributions: Conceptualization, D.S. and K.D.; methodology, D.S. and D.Z.; software, D.Z.; validation, D.S., D.Z. and K.D.; formal analysis, D.S.; resources, K.D.; data curation, D.S. and D.Z.; writing—original draft preparation, D.S. and D.Z.; writing—review and editing, D.S. and K.D.; visualization, D.S. and D.Z.; supervision, D.S. and K.D.; project administration, K.D.; funding acquisition, K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This study was conducted in the framework of the SILVANUS project. This project received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 101037247. The contents of this publication are the sole responsibility of the authors and can in no way be taken to reflect the views of the European Commission.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data utilized in this paper can be found at: <https://doi.org/10.5281/zenodo.7762564>, (accessed on 15 September 2024).

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Bousfield, C.; Lindenmayer, D.; Edwards, D. Substantial and increasing global losses of timber-producing forest due to wildfires. *Nat. Geosci.* **2023**, *16*, 1145–1150. [[CrossRef](#)]
2. Shvidenko, A.Z.; Shchepashchenko, D.G.; Vaganov, E.A.; Sukhinin, A.I.; Maksyutov, S.S.; McCallum, I.; Lakyda, I.P. Impact of wildfire in Russia between 1998–2010 on ecosystems and the global carbon budget. *Dokl. Earth Sci.* **2011**, *441*, 1678–1682. [[CrossRef](#)]
3. Flannigan, M.D.; Wagner, C.E.V. Climate change and wildfire in Canada. *Can. J. For. Res.* **1991**, *21*, 66–72. [[CrossRef](#)]
4. Coogan, S.C.; Robinne, F.N.; Jain, P.; Flannigan, M.D. Scientists' warning on wildfire—A Canadian perspective. *Can. J. For. Res.* **2019**, *49*, 1015–1023. [[CrossRef](#)]
5. Krawisz, B. Health Effects of Climate Destabilization: Understanding the Problem. *WMJ Off. Publ. State Med Soc. Wis.* **2020**, *119*, 132–138.

6. Spiller, D.; Carbone, A.; Amici, S.; Thangavel, K.; Sabatini, R.; Laneve, G. Wildfire Detection Using Convolutional Neural Networks and PRISMA Hyperspectral Imagery: A Spatial-Spectral Analysis. *Remote Sens.* **2023**, *15*, 4855. [[CrossRef](#)]
7. Meier, S.; Elliott, R.J.; Strobl, E. The regional economic impact of wildfires: Evidence from Southern Europe. *J. Environ. Econ. Manag.* **2023**, *118*, 102787. [[CrossRef](#)]
8. Giglio, L.; Schroeder, W.; Justice, C.O. The collection 6 MODIS active fire detection algorithm and fire products. *Remote Sens. Environ.* **2016**, *178*, 31–41. [[CrossRef](#)]
9. Sykas, D.; Zografakis, D.; Demestichas, K.; Costopoulou, C.; Kosmidis, P. EO4WildFires: An Earth observation multi-sensor, time-series machine-learning-ready benchmark dataset for wildfire impact prediction. In Proceedings of the Ninth International Conference on Remote Sensing and Geoinformation of the Environment (RSCy2023), Ayia Napa, Cyprus, 3–5 April 2023; Themistocleous, K., Hadjimitsis, D.G., Michaelides, S., Papadavid, G., Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2023; Volume 12786, p. 1278603. [[CrossRef](#)]
10. Toan, N.T.; Thanh Cong, P.; Viet Hung, N.Q.; Jo, J. A deep learning approach for early wildfire detection from hyperspectral satellite images. In Proceedings of the 2019 7th International Conference on Robot Intelligence Technology and Applications (RiTA), Daejeon, Republic of Korea, 1–3 November 2019; pp. 38–45. [[CrossRef](#)]
11. Calkin, D.E.; Thompson, M.P.; Finney, M.A. Negative consequences of positive feedbacks in US wildfire management. *For. Ecosyst.* **2015**, *2*, 9. [[CrossRef](#)]
12. Ghali, R.; Akhloofi, M.A. Deep Learning Approaches for Wildland Fires Using Satellite Remote Sensing Data: Detection, Mapping, and Prediction. *Fire* **2023**, *6*, 192. [[CrossRef](#)]
13. To, P.; Eboreime, E.; Agyapong, V.I.O. The Impact of Wildfires on Mental Health: A Scoping Review. *Behav. Sci.* **2021**, *11*, 126. [[CrossRef](#)] [[PubMed](#)]
14. Belleville, G.; Ouellet, M.C.; Morin, C. Post-Traumatic Stress among Evacuees from the 2016 Fort McMurray Wildfires: Exploration of Psychological and Sleep Symptoms Three Months after the Evacuation. *Int. J. Environ. Res. Public Health* **2019**, *16*, 1604. [[CrossRef](#)] [[PubMed](#)]
15. Li, Y.; Tong, D.; Ma, S.; Zhang, X.; Kondragunta, S.; Li, F.; Saylor, R. Dominance of Wildfires Impact on Air Quality Exceedances During the 2020 Record-Breaking Wildfire Season in the United States. *Geophys. Res. Lett.* **2021**, *48*, e2021GL094908. [[CrossRef](#)]
16. Tao, Z.; He, H.; Sun, C.; Tong, D.; Liang, X.Z. Impact of Fire Emissions on U.S. Air Quality from 1997 to 2016—A Modeling Study in the Satellite Era. *Remote Sens.* **2020**, *12*, 913. [[CrossRef](#)]
17. Bravo, A.; Sosa, E.; Sánchez, A.; Jaimes, P.; Saavedra, R. Impact of wildfires on the air quality of Mexico City, 1992–1999. *Environ. Pollut.* **2002**, *117*, 243–253. [[CrossRef](#)] [[PubMed](#)]
18. Zhang, Y.; Fan, J.; Shrivastava, M.; Homeyer, C.R.; Wang, Y.; Seinfeld, J.H. Notable impact of wildfires in the western United States on weather hazards in the central United States. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2207329119. [[CrossRef](#)]
19. Fernández-García, V.; Beltrán-Marcos, D.; Fernández-Guisuraga, J.M.; Marcos, E.; Calvo, L. Predicting potential wildfire severity across Southern Europe with global data sources. *Sci. Total Environ.* **2022**, *829*, 154729. [[CrossRef](#)]
20. Barmpoutis, P.; Papaioannou, P.; Dimitropoulos, K.; Grammalidis, N. A Review on Early Forest Fire Detection Systems Using Optical Remote Sensing. *Sensors* **2020**, *20*, 6442. [[CrossRef](#)]
21. Rashkovetsky, D.; Mauracher, F.; Langer, M.; Schmitt, M. Wildfire Detection From Multisensor Satellite Imagery Using Deep Semantic Segmentation. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 7001–7016. [[CrossRef](#)]
22. Chuvieco, E.; Mouillot, F.; van der Werf, G.R.; San Miguel, J.; Tanase, M.; Koutsias, N.; García, M.; Yebra, M.; Padilla, M.; Gitas, I.; et al. Historical background and current developments for mapping burned area from satellite Earth observation. *Remote Sens. Environ.* **2019**, *225*, 45–64. [[CrossRef](#)]
23. Wang, J.; Sammis, T.; Gutschick, V.; Gebremichael, M.; Dennis, S.; Harrison, R. Review of Satellite Remote Sensing Use in Forest Health Studies. *Open Geogr. J.* **2010**, *3*, 28–42. [[CrossRef](#)]
24. Mohanty, S.P.; Czakon, J.; Kaczmarek, K.A.; Pyskir, A.; Tarasiewicz, P.; Kunwar, S.; Rohrbach, J.; Luo, D.; Prasad, M.; Fleer, S.; et al. Deep Learning for Understanding Satellite Imagery: An Experimental Survey. *Front. Artif. Intell.* **2020**, *3*, 534696. [[CrossRef](#)] [[PubMed](#)]
25. Pritt, M.; Chern, G. Satellite Image Classification with Deep Learning. In Proceedings of the 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 10–12 October 2017; pp. 1–7. [[CrossRef](#)]
26. Shafaey, M.A.; Salem, M.A.M.; Ebied, H.M.; Al-Berry, M.N.; Tolba, M.F. Deep Learning for Satellite Image Classification. In Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2018, Cairo, Egypt, 1–3 September 2018; Hassanien, A.E., Tolba, M.F., Shaalan, K., Azar, A.T., Eds.; Springer: Cham, Switzerland, 2019; pp. 383–391.
27. Neupane, B.; Horanont, T.; Aryal, J. Deep Learning-Based Semantic Segmentation of Urban Features in Satellite Images: A Review and Meta-Analysis. *Remote Sens.* **2021**, *13*, 808. [[CrossRef](#)]
28. López García, M.J.; Caselles, V. Mapping burns and natural reforestation using Thematic Mapper data. *Geocarto Int.* **1991**, *6*, 31–37. [[CrossRef](#)]
29. Morresi, D.; Marzano, R.; Lingua, E.; Motta, R.; Garbarino, M. Mapping burn severity in the western Italian Alps through phenologically coherent reflectance composites derived from Sentinel-2 imagery. *Remote Sens. Environ.* **2022**, *269*, 112800. [[CrossRef](#)]
30. Ying, L.; Shen, Z.; Yang, M.; Piao, S. Wildfire Detection Probability of MODIS Fire Products under the Constraint of Environmental Factors: A Study Based on Confirmed Ground Wildfire Records. *Remote Sens.* **2019**, *11*, 3031. [[CrossRef](#)]

31. Sayad, Y.O.; Mousannif, H.; Al Moatassime, H. Predictive modeling of wildfires: A new dataset and machine learning approach. *Fire Saf. J.* **2019**, *104*, 130–146. [[CrossRef](#)]
32. Huot, F.; Hu, R.L.; Goyal, N.; Sankar, T.; Ihme, M.; Chen, Y.F. Next Day Wildfire Spread: A Machine Learning Dataset to Predict Wildfire Spreading From Remote-Sensing Data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4412513. [[CrossRef](#)]
33. Marjani, M.; Mahdianpari, M.; Mohammadimanesh, F. CNN-BiLSTM: A Novel Deep Learning Model for Near-Real-Time Daily Wildfire Spread Prediction. *Remote Sens.* **2024**, *16*, 1467. [[CrossRef](#)]
34. Rosadi, D.; Arisanty, D.; Agustina, D. Prediction of forest fire using neural networks with backpropagation learning and extreme learning machine approach using meteorological and weather index variables. *Media Stat.* **2022**, *14*, 118–124. [[CrossRef](#)]
35. Meng, Q.; Meentemeyer, R.K. Modeling of multi-strata forest fire severity using Landsat TM Data. *Int. J. Appl. Earth Obs. Geoinf.* **2011**, *13*, 120–126. [[CrossRef](#)]
36. Ahmad, F.; Waseem, Z.; Ahmad, M.; Ansari, M.Z. Forest Fire Prediction Using Machine Learning Techniques. In Proceedings of the 2023 International Conference on Recent Advances in Electrical, Electronics & Digital Healthcare Technologies (REEDCON), New Delhi, India, 1–3 May 2023; pp. 705–708. [[CrossRef](#)]
37. Yang, S.; Lupascu, M.; Meel, K.S. Predicting Forest Fire Using Remote Sensing Data And Machine Learning. *arXiv* **2021**, arXiv:2101.01975. [[CrossRef](#)]
38. Monaco, S.; Pasini, A.; Apiletti, D.; Colomba, L.; Garza, P.; Baralis, E. Improving Wildfire Severity Classification of Deep Learning U-Nets from Satellite Images. In Proceedings of the 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 10–13 December 2020; pp. 5786–5788. [[CrossRef](#)]
39. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.03385.
40. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2015**, arXiv:1405.0312.
41. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv* **2020**, arXiv:1905.11946.
42. Bergado, J.R.; Persello, C.; Reinke, K.; Stein, A. Predicting wildfire burns from big geodata using deep learning. *Saf. Sci.* **2021**, *140*, 105276. [[CrossRef](#)]
43. Zhang, G.; Wang, M.; Liu, K. Deep neural networks for global wildfire susceptibility modelling. *Ecol. Indic.* **2021**, *127*, 107735. [[CrossRef](#)]
44. Yu, J.; Wang, Z.; Vasudevan, V.; Yeung, L.; Seyedhosseini, M.; Wu, Y. CoCa: Contrastive Captioners are Image-Text Foundation Models. *arXiv* **2022**, arXiv:2205.01917.
45. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision. *arXiv* **2021**, arXiv:2103.00020.
46. Wang, Z.; Yu, J.; Yu, A.W.; Dai, Z.; Tsvetkov, Y.; Cao, Y. SimVLM: Simple Visual Language Model Pretraining with Weak Supervision. *arXiv* **2022**, arXiv:2108.10904.
47. Liu, Z.; Hu, H.; Lin, Y.; Yao, Z.; Xie, Z.; Wei, Y.; Ning, J.; Cao, Y.; Zhang, Z.; Dong, L.; et al. Swin Transformer V2: Scaling Up Capacity and Resolution. *arXiv* **2022**, arXiv:2111.09883.
48. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2021**, arXiv:2010.11929.
49. Chen, X.; Liang, C.; Huang, D.; Real, E.; Wang, K.; Liu, Y.; Pham, H.; Dong, X.; Luong, T.; Hsieh, C.J.; et al. Symbolic Discovery of Optimization Algorithms. *arXiv* **2023**, arXiv:2302.06675.
50. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
51. Shazeer, N.; Stern, M. Adafactor: Adaptive Learning Rates with Sublinear Memory Cost. *arXiv* **2018**, arXiv:1804.04235.
52. Camia, A.; Durrant, T.; San-Miguel-Ayanz, J.; European Commission; Joint Research Centre; Institute for Environment and Sustainability. *The European Fire Database—Technical Specifications and Data Submission—Executive Report*; Publications Office of the European Union: Luxembourg, 2014. [[CrossRef](#)]
53. European Commission; Joint Research Centre; Schulte, E.; Maianti, P.; Boca, R.; De Rigo, D.; Ferrari, D.; Durrant, T.; Löffler, P.; San-Miguel-Ayanz, J.; et al. *Forest fires in Europe, Middle East and North Africa 2016*; Publications Office of the European Union: Luxembourg, 2017. [[CrossRef](#)]
54. The Data Was Obtained from the National Aeronautics and Space Administration (NASA) Langley Research Center (LaRC) Prediction of Worldwide Energy Resource (POWER) Project funded through the NASA Earth Science/Applied Science Program. 2023. Available online: <https://power.larc.nasa.gov/> (accessed on 15 September 2024).
55. Prapas, I.; Bountos, N.I.; Kondylatos, S.; Michail, D.; Camps-Valls, G.; Papoutsis, I. TeleViT: Teleconnection-driven Transformers Improve Subseasonal to Seasonal Wildfire Forecasting. *arXiv* **2023**, arXiv:2306.10940.
56. Sudre, C.H.; Li, W.; Vercauteren, T.K.M.; Ourselin, S.; Cardoso, M.J. 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, Proceedings of the Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, 14 September 2017*; Springer International Publishing: Cham, Switzerland, 2017; pp. 240–248.
57. Singh, A.K.; Chan, S.C.Y.; Moskovitz, T.; Grant, E.; Saxe, A.M.; Hill, F. The Transient Nature of Emergent In-Context Learning in Transformers. *arXiv* **2023**, arXiv:2311.08360.

-
58. Kaddour, J.; Key, O.; Nawrot, P.; Minervini, P.; Kusner, M.J. No Train No Gain: Revisiting Efficient Training Algorithms For Transformer-based Language Models. In *Proceedings of the Advances in Neural Information Processing Systems*; Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S., Eds.; Curran Associates, Inc.: New York, NY, USA, 2023; Volume 36, pp. 25793–25818.
 59. Zhuang, B.; Liu, J.; Pan, Z.; He, H.; Weng, Y.; Shen, C. A Survey on Efficient Training of Transformers. *arXiv* **2023**, arXiv:2302.01107.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.