

Article

Not peer-reviewed version

---

# Domain Adaptation of Deep Learning Segmentation Model of Small Agricultural Burn Area Detection Using Hi-Resolution Sentinel-2 Observations: A Case Study of Punjab, India

---

[Anamika Anand](#)\*, [Ryoichi Imasu](#)\*, [Surendra Kumar Dhaka](#), [Prabir Patra](#)

Posted Date: 4 September 2024

doi: 10.20944/preprints202409.0333.v1

Keywords: burn areas; crop residue burning; deep learning; neural networks; transfer learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

# Domain Adaptation of Deep Learning Segmentation Model of Small Agricultural Burn Area Detection Using Hi-Resolution Sentinel-2 Observations: A Case Study of Punjab, India

Anamika Anand <sup>1,2,\*</sup>, Ryoichi Imasu <sup>1,\*</sup>, Surendra K Dhaka <sup>3</sup> and Prabir K Patra <sup>2,4</sup>

<sup>1</sup> Atmosphere and Ocean Research Institute, The University of Tokyo, Japan

<sup>2</sup> Research Institute for Humanity and Nature, Kyoto, Japan

<sup>3</sup> Department of Physics, Rajdhani College, University of Delhi, India

<sup>4</sup> Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Yokohama, Japan

\* Correspondence: anandanamika@aori.u-tokyo.ac.jp; imasu@aori.u-tokyo.ac.jp;

† Current address: General Research Building 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba 277-8568, Japan.

**Abstract:** This study investigates the capabilities of high-resolution Sentinel-2 observations and deep learning (DL) segmentation models, specifically convolutional neural networks (CNNs), for accurate mapping of small and fragmented agricultural burn areas in Punjab, India. Initially, the model was trained using ICNF burn area data from Portugal to capture large fire and burn area delineation, achieving moderate accuracy. Subsequent fine-tuning using annotated data from Punjab improved the model's ability to detect small burn patches, demonstrating higher accuracy and precision than the baseline Normalised Burn Ratio (NBR) Index method. On-ground validation using buffer zone analysis and crop field images confirms the DL approach effectiveness. Cloud interference and temporal gaps in satellite data posed some challenges. Despite these limitations, the study highlights the methodological advancements and potential of using DL models for small burn area detection in agricultural settings. The model achieved overall testing accuracy of 0.64 in Dice score assessment. This report presents a promising approach for improved mapping of the burn area, which might engender better estimation of emissions and the management practices in regions with little available ground truth data.

**Keywords:** burn areas; crop residue burning; deep learning; neural networks; transfer learning;

## 1. Background and Introduction

Crop residue burning or prescribed burning is a common agricultural biomass removal practice that is applied in many parts of the world. Globally, farmers adopt this residue management practice for several reasons as a quick and inexpensive means of managing post-harvest crop residue, eliminating pests and weeds, and restoring soil fertility. Open burning of the residue is conducted intensively in many economically developed and developing countries such as the United States, China, Russia, India, and Mexico c. [1,2].

In the context of India, since the success of the green revolution, most farmers have shifted to a double crop system (also known as the rice–wheat system). Access to modern technologies has tripled food production. However, in the double crop system, rice must be harvested earlier to accommodate wheat. To speed up the process, most farmers have switched to mechanised harvesters, which leave smaller residue fragments scattered on the ground, which are very difficult to clear or collect. Therefore, farmers embrace open burning as an easy, cheap, and quick method to dispose of the scattered residue quickly [3–5].

Although this practice is a very popular and effective management strategy, it has raised some severe environmental concerns and controversial debate among scientists and policymakers over the years. Primarily, it is a prominent source of particulate matter and gaseous pollutants (PM 2.5, black carbon BC) in the atmosphere [6]. These emissions degrade air quality, leading to adverse effects on climate and human health [7,8]. The effects are not merely limited to managed ecosystems. In fact, reports have described the presence of BC deposition in the Arctic from agriculture and forest

fires. These deposits have contributed to Arctic warming [9]. Despite the known adverse effects, a remarkable information gap persists regarding the amounts of emissions produced from crop residue burning, especially from small fires, and the rapidly growing degradation of air quality and adverse impact on human health [3,10].

One of the greatest challenges to understanding the emission contribution is the lack of accurate and reliable burn area assessment [11]. Burn area estimates are critical input in climate models and in the development of fire emission inventories[12]. Current emission inventories such as GFED, FINN, and EDGAR present great degrees of uncertainty for the estimation of emissions calculations and sources. A major reason is undetected small fires[13]. The applications of burn area assessments go beyond scientific exploration and expand considerably. Today, burn area assessments are crucially important not only for climate scientists but also for land management, environmental protection agencies, and public health organisations. These assessments provide fundamentally important insights into the immediate and long-term effects of fires on ecosystems. For land managers, understanding the extent of burned areas facilitates the formulation of effective recovery and reforestation strategies. Environmental agencies utilise these data to monitor fire effects on air quality and biodiversity, which are necessary for maintaining ecological balance and for informed policy decisions [14]. Additionally, the global importance of burn area information is necessary to support international efforts such as the Kyoto Protocol and the Paris Agreement (COP21) to manage fire emissions. Accurate burn area assessment helps monitor progress toward the achievement of Sustainable Development Goals (SDGs) as outlined by the United Nations.

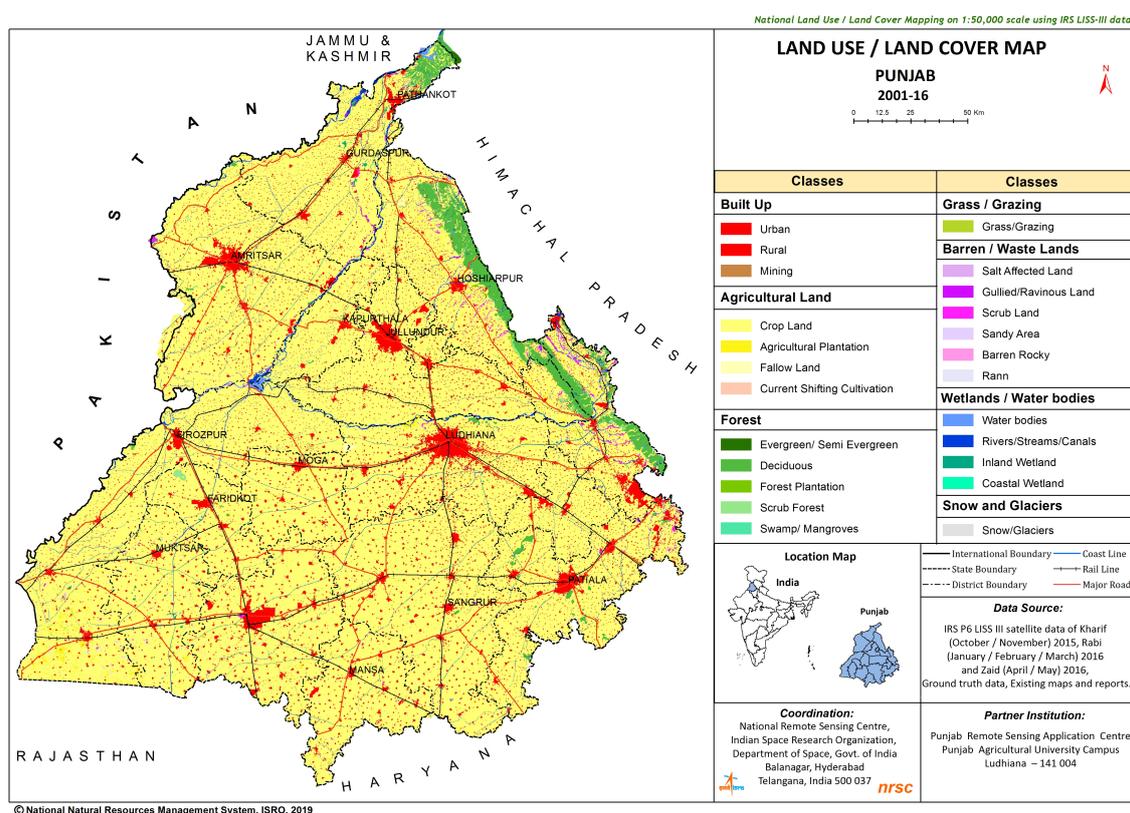
Several factors lead to these inaccuracies, partly because of the dependence on coarse-resolution satellites such as MODIS and VIIRS for global and regional burn area products. For example, the widely used global burn area products MCD64A1 and FIRECC151 respectively have resolutions of 500 m and 250 m. Both are derived from MODIS observation based on thermal anomalies and reflectance data, providing monthly composite burn area datasets. However, these sensors fail to capture small agricultural burn area because of their coarse resolution[11]. A recent study in Africa using Sentinel-2 observations revealed 1.8 times more burn area than that detected by MODIS [15]. Unlike forest fires or wildfires, agricultural burning often occurs in small patches and sometimes without active flames. Consequently, coarse-resolution satellites overlook these small burn area, leading to conservative estimates of burn area. This underestimation emphasises the need for high-resolution satellite data and integrated approaches to improve the accuracy of burn area estimation.

### *1.1. Study Area*

Although the practice of agricultural burning is not limited to India, the situation there is unique and severe, making it an interesting study area. Several factors contribute to the distinctiveness and intensity of open burning in India. 1) Green Revolution and Double Cropping System: The Green Revolution led to the widespread adoption of the wheat–rice double cropping system in India. This intensive agricultural practice produces vast quantities of crop residue that must be cleared quickly to prepare for the next crop. The narrow window between the harvest of one crop and the planting of the next leaves farmers with limited time to manage the residue effectively[16]. 2) Mechanization: The shift towards mechanised harvesters has exacerbated the difficulties. Mechanised harvesters leave behind huge amounts of crop residue scattered across the fields. This root-bound and scattered residue presents a challenge for farmers, who find open burning to be a quick and cost-effective solution to clear their fields [3,4].

Crop residue burning degrades air quality in greater regions of northern India, especially during the post monsoon period. Meteorological conditions and changing weather patterns also contribute adversely to the air quality. The entire Indo-Gangetic plain, which is home to a huge population, sustains heavy smog during October–November. Light winds fail to disperse the pollutants, and weak north-westerly winds push them toward densely populated urban areas, including cities such as New Delhi, exacerbating air pollution and health issue [17,18].

Every year during March–April and October–November, the entire region is enveloped in smoke and haze as farmers burn the residue from the harvested rice crop. The intensive open burning over such a vast area produces tremendous amounts of various pollutants including  $\text{CO}_2$ ,  $\text{CO}$ ,  $\text{CH}_4$ ,  $\text{N}_2\text{O}$ ,  $\text{NO}_x$  and NMHCs. These emissions not only alter regional atmospheric chemistry but also pose important public health concerns for the nearby densely populated areas [19,20]. The study area, Punjab state (Figure 1), is located in the northern part of India. Punjab has a geographical area of  $50,362 \text{ km}^2$ , of which around  $40,230 \text{ km}^2$  is under cultivation (Dept. of Agriculture and Farmers Welfare, Govt. of Punjab). Therefore, it is often designated as the "breadbasket" of the country, producing two-thirds of the nation's food requirements. In Punjab alone, this seasonal burning practice engenders the burning of approximately 7 million tons of crop residue during post monsoon[21].



**Figure 1.** Land Use Land Cover map over Punjab, 2018. The yellow region represents the amount of area under intense agricultural cultivation. *Image credit: National Natural Resource Management, ISRO [22].*

### 1.2. Evolution of Burn Area Detection Methods

Over the years, methodologies for identifying burn areas have evolved considerably, transitioning from traditional indices based on remote sensing to sophisticated deep-learning (DL) approaches. These methods can be categorised into rule-based and machine-learning approaches [23].

Rule-based approaches depend on the spectral response of different bands, especially middle infrared (MIR) and thermal domains. This reliance is mainly attributable to the fact that biomass burning has evident effects that can be observed easily in these domains. For instance, biomass burning leads to decreased Near Infrared (NIR) reflectance and dryness because of fire, leading to a slight increase in Shortwave Infrared (SWIR) [24]. The spectral response is the backbone of most commonly used indices such as the Normalized Burn Ratio (NBR) Index and its variations, such as the delta NBR (dNBR) and relative NBR (rNBR).

Traditional indices using a rule-based approach, such as the NBR, present several limitations affecting their capability to capture burn areas precisely. Sensitivity to environmental conditions,

including soil moisture and atmospheric variations, can produce false positive or false negative results of burn area detection. Spectral confusion further complicates accuracy because burned areas might resemble water bodies, shadows, or harvested fields[25,26]

Temporal resolution constraints of satellite imagery, reliant on pre-fire and post-fire comparisons, hinder timely detection, especially in the absence of cloud-free images. Additionally, the moderate spatial resolution of commonly used sensors such as MODIS might overlook small-scale burns and fail to capture the full spatial detail of burn scars in heterogeneous landscapes [27]. Finally, the lack of generalisation across ecosystems and fire regimes necessitates region-specific calibration, limiting the broad applicability of traditional indices [28]. These challenges underscore the need for more sophisticated methodologies such as machine learning and deep learning, to improve burn area detection. Unlike traditional indices, machine learning approaches are based on characteristics of burned and unburned pixels from reference or labelled data. Most commonly used machine learning methods include Support Vector Machines (SVM) [29] and Random Forests (RF) [30].

At present, mapping algorithms are limited to active fires and global burn area products. Major challenges are 1) heterogeneous agricultural landscapes, 2) fire durations and patterns, and 3) human-driven changes (ploughing and seeding)[11]. The present mapping algorithms are more efficient at capturing the hotspots, and large fire events. Existing global burn area products derived from MODIS, such as MCD45A1 [27] and MCD64A1 [31], mainly emphasise large wildfire events. They do not take into account small fires. However, regional burn area products MCD14ML [32], which can capture the full extent of cropland burn area, have been found to underestimate the total burn area [33].

These BA products harness the capabilities of MODIS (Moderate Resolution Imaging Spectroradiometer), leveraging its thermal anomalies and reflectance data. However, a major challenge while working with MODIS-based fire products is the coarse resolution, small fires or burned ground often go undetected which compromises the accuracy. Popular BA products like MCD64A1 and MCD45A1, which utilize MODIS data to identify burned areas, frequently miss detecting burns smaller than 100 hectares. This oversight is particularly significant in agricultural regions, where failing to account for smaller burned areas can lead to an underestimation by up to tenfold [4]. While MODIS offers significant advantages in terms of coverage and frequency, its relatively coarse resolution can be a big limitation. Such inaccuracies in detection underscore the need for complementary methods and use of higher-resolution sensors to improve burn area estimations.

In recent years, the use of neural networks has shown enormous success in remote sensing applications. One of the first studies concerning Burn Area mapping using neural networks was in 2001 [34], where a supervised ART neural network was used to detect forest fires in Spain using NOAA-AVHRR imagery. ART neural networks are based on the adaptive resonance concept and classify the burn area image pixel by pixel. Following that, in a later year, several other studies emerge using similar neural networks for burn area mapping. For instance, Brivio [35] used supervised image classification based on Multilayer Perceptron (MLP) and SPOT-VEGETATION images to map burn areas in sub Saharan Africa. This neural network model exploited the spectral signals of burnt areas and provide spatial and temporal information on the burnt areas. Another type of neural network that has shown promising success in remote sensing application is Convolutional Neural Networks (CNNs). CNNs are neural network inspired by human brains, consist of neurons that are not fully connected rather only connected to those with receptive neighbourhood. Their deep architecture enables them to represent non linear function and generalize the learned features, hence they are also referred as deep learning (DL) model architectures [36]. These deep learning (DL) model architectures have been frequently used and show enormous potential in several image segmentation tasks such as clouds [37], water [38], urban areas [39] and vegetation types [40].

Despite its immense potential, this DL has limited applicability in burn area segmentation, mainly because of a lack of reliable qualitative and quantitative data for training. However, few studies in recent years have used the DL approach for burn area mapping. Some notable studies that used DL approaches are: [41] developed the BA-Net model based on CNN with Long Short Term Memory

(LSTM) based on U-Net framework and VIIRS reflectance data. They mapped the wildfire affected areas successfully. Later, BA-Net was developed further for rapid and accurate mapping of forest fires [42]. Although the results demonstrated remarkable improvement, their application was limited mostly to wildfires. [23] used a U-Net architecture DL model to develop a fully automatic processing chain to segment vegetation fires. The trained model with Sentinel-2 imagery as an input showed better results against the Random Forest classification. Another recent work by [43], trained a U-Net based DL model over Portugal and tested Transfer learning applicability of the model to analyse local fire burn dynamics in West Nile, Northern Uganda. The model showed promising application of U-Net and transfer ability of learned features in capturing small burn area and the details of large burn area extent over a new domain. These studies have been a great source of information and motivation for us to conduct this experiment over the Punjab region in India.

One major challenge to conducting DL-based experiments is the lack of sufficient ground truth data. Fire monitoring in economically developing countries is often more challenging for this very reason. Therefore, we propose estimation of small agricultural burn areas in Punjab using DL segmentation and a transfer learning approach. The transfer learning approach involves training a model on one task and then reusing the pre-trained model to perform a closely related task. This technique can be very useful in cases of data scarcity for model training and use of limited computational resources. Our target domain of Punjab lacks any reliable ground truth data for training the model. We aim to address this issue by first training the DL model over the source domain (Portugal) using the ICNF (The Institute for Nature Conservation and Forests) dataset and applying transfer learning technique and later fine-tuning the model using hand annotated data over the target domain of Punjab. The objectives of the present work are three-fold: the first objective is to train the model over source domain (Portugal), second objective is to apply transfer learning and fine tuning of the model over target domain (Punjab), with subsequent improvement of burn area predictions, third objective is to assess the changes and trends of small agricultural burn area throughout the post-harvesting season in Punjab.

## 2. Data and Method

For our study, we used Sentinel-2 Multispectral Instrument (MSI) Level 1C data, which provide Top of Atmosphere (TOA) reflectance at 20 m resolution, as our training dataset. We use reference data of two types. For initial training over Portugal, we used the ICNF dataset for 2016 and 2017 to train our model during the fire season. The deep learning model used is based on the Resnet18 architecture, a robust framework known for its effectiveness in image segmentation tasks, especially when distinguishing between burned and unburned areas [44]. Initially, we trained our model on the ICNF dataset to capture the general features of burned areas. Later, we fine tune the model using hand annotated data to enhance the model's performance and to adapt it to local conditions of Punjab. The hand-annotated data is created over a 200 km<sup>2</sup> area in Punjab, roughly equivalent to two Sentinel-2 tiles. Fine-tuning step involved adjusting hyperparameters of the model based on the Punjab data and re-evaluating its accuracy to ensure optimal performance in detecting burn areas under different geographical and environmental conditions. The model's accuracy was assessed using two key metrics: the Intersection over Union (IoU) and the Dice coefficient. These metrics helped us gauge the model's ability to identify and delineate burned areas accurately. The next section presents detailed discussion of each step involved in the data pipeline and model experiment.

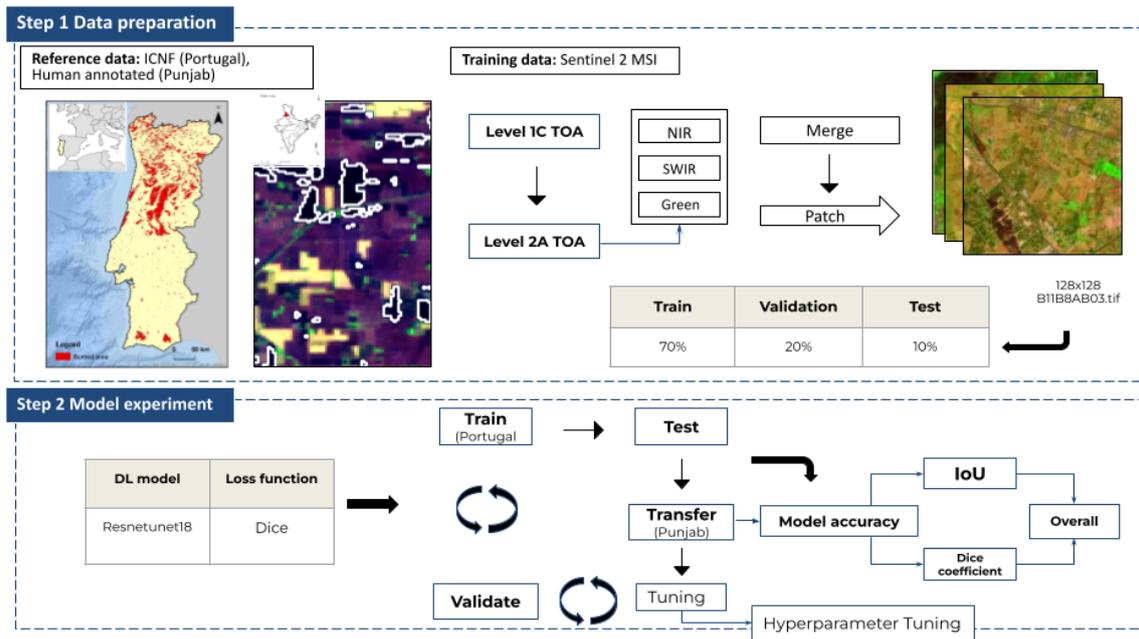


Figure 2. Workflow.

### 2.1. Data Acquisition and Atmospheric Correction

To analyse the burn areas over source domain (Portugal) and target domain (Punjab), we used Sentinel-2 Level 1C Top of Atmosphere (TOA) reflectance data. The Copernicus Sentinel-2 program consists of two sun-synchronous polar-orbiting satellites, Sentinel-2A and Sentinel-2B, which have a swath width of 290 km and a revisit time of 5 days. The Multispectral Instrument (MSI), a key payload onboard the Sentinel satellites, provides high spatial resolution imagery. MSI covers 13 spectral bands with spatial resolutions of 10–60 m. Sentinel-2 MSI data are available in two product levels: Level 1C (L1C) and Level 2A (L2A). L1C data represents TOA reflectance and L2A provides atmospherically corrected bottom-of-atmosphere (terrain-corrected and cirrus-corrected reflectance images) as per `sen2cor` algorithm [45]. The global availability of L1C data starts from 2015, whereas L2A data are available from 2017.

To conserve computation resources for our analysis, we opted to use L2A products whenever available. Alternatively, we acquired L1C data and converted it into L2A using the `sen2cor` algorithm [45]. In the case of Portugal, we aimed at reproducing similar model accuracy to that of an earlier study [43]. Portugal experienced numerous fire events in August; especially in 2017, Portugal experienced a record-breaking number of fire events, accounting for approximately 500,000 ha of total burned area [47,48]. This dataset with unusually numerous burned areas can be useful to ensure models can learn burn area features rigorously. In Portugal, we specifically examined periods of peak fire activity during 2016–2017 to ensure a maximum burn area. Specifically, we targeted August for Portugal (2016 and 2017). In Punjab, we particularly examined the post-monsoon burning season October–November for Punjab (2020–2023). The data were downloaded using the Python API from the Copernicus Data Space. The API used to download the data can be found in an earlier report of the literature [49].

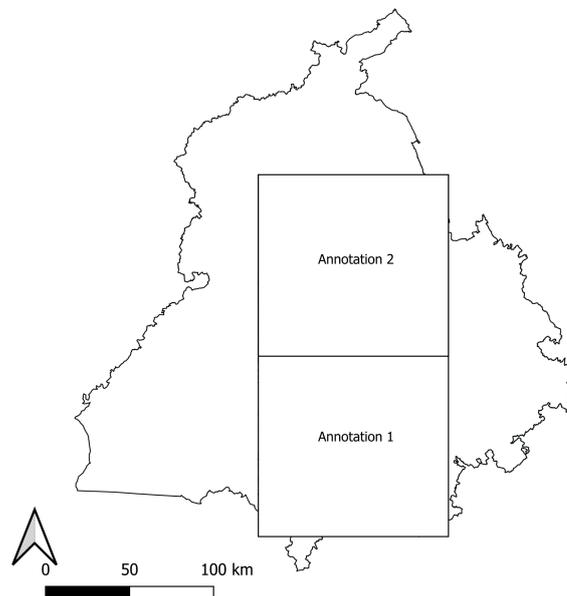
### 2.2. Preparation of Reference Data

ICNF provides a regional database in vector files. The ICNF burn area dataset is based on a cartographic method by which burn area polygons are created using a semi-automatic classification process using Sentinel, LANDSAT, and MODIS satellite images. Additional details can be found in the literature [50]. The hand-labelled dataset has high quality, giving a good idea of burn extents.

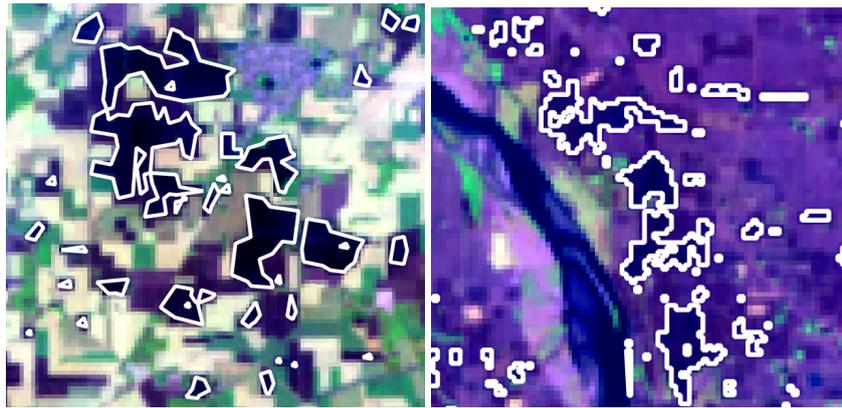
Moreover, it is publicly accessible. For our analyses, the ICNF data were acquired for August 2016 and August 2017.

An important domain difference exists between Punjab and Portugal in terms of climate, topography, and fire event types. To compensate for that gap, hand annotation of burn areas over 200 km<sup>2</sup> (equivalent to two sentinel tiles) was performed. Our objective was to encompass various burn areas. For that reason, we selected regions with a high density of MODIS and VIIRS fire hotspots. MODIS fire hotspot data give a good approximate idea about where most fire events are occurring. Therefore researchers have often relied on fire hotspot information to assess burn area [51,52].

To ensure good quality and better acclimatisation of the model, we considered several factors for hand annotation of the data: a) use of RGB Sentinel images for visualisation and manual checking of the presence and absence of burn scars, b) human capability based on past experience and observation to distinguish burned and unburned areas, c) presence of active flame and smoke, d) cross-verification of hand annotations with the closest available Google Earth (Alphabet Inc.) images. To consider annotation over agricultural land explicitly, we used land use land cover data to exclude all other land types, especially for urban areas and aquatic bodies. The annotation was performed in two sets (Figure 4). Annotation 1 (Ano1) and Annotation 2 (Ano2), over two adjacent areas covering 100 km<sup>2</sup> each. Ano 1 had more common features such as large uninterrupted patches of agricultural land, absence of big water bodies and more observable burned agricultural fields, whereas Ano 2 had more diverse features such as river streams and big urban areas and more observable non-burned agricultural fields. The two sets of annotations were performed on close dates and during peak fire seasons. The motive behind this step is to study the model's learning and transferable capacity of learned features while fine-tuning with different types of annotation data. The ultimate goal was to enhance model robustness not only by making accurate predictions over burned areas but also to make minimum false predictions over non-burned or non-agricultural areas.



**Figure 3.** Location and area for hand annotated data over Punjab.



**Figure 4.** Example of Ano 1 data (left) showing the presence of agricultural burn areas (common features) and Ano 2 (right) showing the presence of diverse features such as rivers, unburned areas and non-agriculture areas.



**Figure 5.** Cross-verification of annotated burn areas using the closest available Google Earth (Alphabet Inc.) images.

### 2.3. Data Pipeline

The data pipeline is divided into four stages, adapted from an earlier study [43]).

For these analyses, we are interested in exploiting the spectral properties of the following three bands: near-infrared (8a), green (3), and shortwave-infrared (11). The bands are selected because of their spectral sensitivity toward change in unburned and burned vegetation. Selected bands were stacked together to create false-colour images for both Portugal and Punjab. Merged false-colour images were extracted into patches of  $128 \times 128$  pixels. Patch size of  $128 \times 128$  was selected as a good optimal middle ground between detailed retention of spatial context and the need for intensive computational resources. Extracted patches were stacked along with the corresponding reference dataset. Incomplete or broken patches were discarded.

The dataset was split into a 70% for training, 20% for validation and 10% for testing. Details on the four stages are discussed hereinafter. Merge bands (Stage 1): Combine the atmospherically corrected Green, NIR, and SWIR bands of Sentinel-2 data to create false-colour images that emphasize burned areas. Normalize the values to the 0–255 colour scale. Extract patches (Stage 2): Match ground truth data to false-colour images based on area and date. Extract small  $128 \times 128$  patches from both datasets to reduce memory usage during training. For Portugal, we set a time frame of 30 days to consider a maximum observation by Sentinel-2. Filter patches (Stage 3): Remove empty or damaged patches from the datasets to ensure data quality. Split patches (Stage 4): Randomly split training patches into a 70% training dataset and a 20% validation dataset and 10% test dataset.

### 2.3.1. Baseline Model

To put the model performance into context, a baseline model was tested on the same validation dataset as the DL model. The baseline model is based on calculation of the NBR index method [53]. The NBR is a popular index to calculate burn area and burn severity along with its derivatives dNBR and rNBR [54]. NBR is based on the spectral response difference between Near Infrared (NIR) and Shortwave Infrared (SWIR). In this experiment, we applied NBR as baseline model to the same validation dataset used for the DL model. The NBR value for each pixel was calculated using the following formula:

$$\text{NBR} = \text{NIR} - \text{SWIR} / \text{NIR} + \text{SWIR}. \quad (1)$$

The NBR values range from -1 to +1. Higher NBR values indicate healthy vegetation, while lower values suggest recently burned vegetation or bare ground. To generate the baseline prediction, we initialized an array and iterate over each validation image. For pixels meeting the burn threshold (NBR < 0.3) [55], the baseline prediction array was updated to classify the pixel as burned. The baseline predictions for the entire dataset were then evaluated using the IoU and Dice coefficient metrics, similarly to the DL model evaluation.

### 2.4. Model Framework

We used a deep learning segmentation model based on U-Net architecture from the PyTorch lightning framework. The training of the model was conducted using specific parameters and strategies to ensure the optimal performance and generalisation. The model has an encoder-decoder structure, with an encoder based on Resnet [56] and a decoder based PSPnet [57], a popular choice for image-related tasks because of its balance between performance and computational efficiency. The decoder parameters included a scaling factor (multiple: 2) and the number of output classes (n\_classes: 1), indicating a binary segmentation task (e.g., burn area vs. non-burn area).

**Training Environment:** The model training was conducted using computational resources of Fujitsu PRIMEHPC FX1000 and FUJITSU PRIMERGY GX2570 (Wisteria/BDEC-01). Details are presented in the literature [58].

This powerful computing environment, equipped with (A100-SXM4-40GBs; NVIDIA Corp.), provided the necessary computational resources to accommodate the intensive demands of deep learning. Regarding the model hyperparameters, training was conducted initially with batch sizes of 16 and 32. This smaller batch size might present implications for the stability of batch normalisation, but it allows for higher-resolution images or deeper models within the same memory constraints. The epoch size was set to 50, but it was later increased to 60. We set the learning rate at 0.0003. The training utilised the Adam optimizer, known for its efficiency in handling sparse gradients and adaptive learning rate capabilities [59]. Weight decay of 0.0001 was set, helping in regularising the model and preventing overfitting. A step scheduler was used, which adjusts the learning rate based on the rate of change of validation loss.

### 2.5. Evaluation Metrics

To evaluate the model performance, the spatial accuracy of the semantic segmentation task was evaluated using the IoU and Dice score method. These two metrics are used frequently in semantic segmentation tasks [60,61].

#### 2.5.1. Intersection over Union

The Jaccard Index, also known as IoU, is a commonly used metric for evaluating segmentation tasks [62]. IoU is calculated as the ratio of the intersection to the union of the ground truth data and the model's predictions. Mathematically, it is defined as

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}, \quad (2)$$

where  $A$  represents the set of ground truth pixels and  $B$  represents the set of predicted pixels. The importance of IoU lies in its ability to provide a clear measure of how well the predicted segmentation matches the ground truth. A higher IoU value (close to 1) represents a greater degree of overlap, signifying that the model's predictions are closely aligned with the actual data. IoU handles cases with class imbalance, which is common in segmentation tasks, by particularly addressing the regions of interest (i.e., the intersection).

### 2.5.2. Dice Similarity Coefficient

The Dice Similarity Coefficient (DSC), commonly designated as the Dice Score, is another standard metric used to evaluate the accuracy of the image segmentation mode [63]. It represents the spatial overlap between two segmentations: the ground truth data and the model predictions. Mathematically, the Dice Score is defined as

$$\text{DSC}(A, B) = \frac{2|A \cap B|}{|A| + |B|}. \quad (3)$$

The Dice Score can quantify the overlap between two binary sets, making it particularly useful for assessing segmentation performance [61]. A higher Dice Score (close to 1) shows greater similarity between the predicted segmentation and the ground truth, reflecting more accurate and precise model predictions. The IoU and Dice score metrics often penalise the zero overlap, which can misrepresent the segmentation accuracy [64]. Therefore, to ensure a balanced evaluation system and to avoid possibly penalising the model for correct non-detection. We assign a perfect score of 1 to cases of zero overlap based on the fact that not detecting non-burned areas as burned is also a good model performance indicator.

## 2.6. Training and Fine-Tuning

As discussed in section 2.3, we used the Adam optimizer for training with an initial learning rate of 0.003. Other hyperparameters were kept at their default values. Validation and testing were performed after each training epoch, with the corresponding loss calculated and monitored through Wandb [65]. To avoid overfitting, training was stopped if the validation loss did not improve after 30 epochs or earlier. The model with the lowest validation and training loss was saved. The DL experiment was conducted in two main phases: 1) model initial training over source domain (Portugal) and 2) transfer learning and fine tuning over target domain (Punjab). During the first phase, the model was trained for two consecutive fire seasons in Portugal: 2016 and 2017. The model associated with the lowest training loss was saved. In the second phase, we used transfer learning of the features learned during initial training and fine-tuned the pretrained model for the Punjab region under varying hyperparameters and weight update conditions.

The first phase involved Experiments A and B. In Experiment A, a raw CNN network was trained over Portugal for 2016. In Experiment B, the network was trained further with more data from 2017. Our goal in this was to reproduce similar accuracy by the experiment conducted by [43]. This step was crucially important to instil the model with a general capability to segment burned and unburned areas. The second phase involved numerous experiments with the aim of monitoring the sensitivity and precision over detection of small burn areas. Experiment C involved direct application of transfer learning to the pretrained model to analyse how well the pretrained model can capture burn areas over Punjab. Experiment D involved fine tuning of the pretrained model over Portugal using the first set of hand-annotated data over Punjab, where the learning rate was reduced to 0.0003, while other parameters were left at their default values. For Experiment E, the batch size was reduced to 16 to

analyse its effects on model accuracy, particularly addressing small delineations of burned areas. For Experiment F, the batch size was reduced further to 8 while other parameters were kept the same.

Experiment G, we used model weights from experiment E and a second set of annotation data (Ano 2). For experiment H, we used model weights from experiment B and both sets of annotation data together (Ano1 and Ano2). From Experiments I–K, all hyperparameters were kept at default values. For Experiment I, we used model weights from Experiment B and the first set of annotation data (Ano 1), froze the layers of encoder for the 25 epochs and unfroze the last layer of the encoder and ran fine tuning for another 5 epochs. For Experiment J, we also used the model weights from Experiment B and Ano1 data. The encoder layers were kept frozen for the whole 30 epochs.

Experiment K was identical to Experiment J, except that Ano2 was used as the input data. For the final Experiment L, we used the model weights from Experiment B and used Ano1 as an input and trained model with 45 epochs where the encoder layers were kept frozen for 15 epochs followed by unfreezing of the last layer and retraining of the whole network for another 25 epochs. For the final Experiment M, we used the model weights from Experiment L and used Ano2 as an input. The model was trained for a total of 40 epochs, where the encoder layers were kept frozen for the first 15 epochs followed by unfreezing of the last layer and training of the whole network for the remaining 25 epochs.

The main goal of running the experiment in different phases and conditions was to monitor the transferability of learned features, effectiveness in capturing the boundaries of small fires, the ability to distinguish between aquatic and burned areas, and sensitivity towards other diverse features in the classification process. This multistage training process allowed us to explore different training configurations and their effects on model performance, leading to an optimised approach for burned area detection. Details of all the experiments discussed above are presented in Table 1.

**Table 1.** A comprehensive overview of all the experiments conducted, highlighting the different stages, conditions, and key parameters modified during the training process.

Expt.	Descr.	Domain	GT-Data	lr	batch size	w	Encoder layers	Epochs
A	raw model	Portugal (2016)	ICNF	0.0003	32	0.0001	unfrozen	50
B	Pretrained	<b>Portugal (2017)</b>	ICNF	0.0003	32	0.0001	unfrozen	50
C	Pretrained	Punjab (2020)	ICNF	0.0003	32	0.0001	unfrozen	60
D	TL-Pretrained	Punjab (2020)	Ano 1	<b>3E-5</b>	32	0.0001	unfrozen	60
E	TL-Pretrained	Punjab (2020)	Ano 1	0.0003	<b>16</b>	0.0001	unfrozen	50
F	TL-Pretrained	Punjab (2020)	Ano 1	0.0003	<b>8</b>	0.0001	unfrozen	50
G	TL-Pretrained	Punjab (2020)	<b>Ano 2</b>	0.0003	16	0.0001	unfrozen	45
H	TL-Pretrained	Punjab (2020)	<b>Ano 1</b> +	0.0003	16	0.0001	unfrozen	60
I	TL-Pretrained	Punjab (2020)	<b>Ano 2</b>	Ano 1	0.0003	<b>0.001</b>	<b>frozen</b>	<b>25 + 5*</b>
J	TL-Pretrained	Punjab (2020)	Ano 1	0.0003	16	0.001	<b>frozen</b>	30
K	TL-Pretrained	Punjab (2020)	Ano 2	0.0003	16	0.001	<b>frozen</b>	30
L	TL-Pretrained	Punjab (2020)	Ano 1	0.0003	16	0.001	<b>frozen</b>	<b>15 + 30*</b>
M	TL-Pretrained	Punjab (2020)	Ano 2	0.0003	16	0.001	<b>frozen</b>	<b>15 + 25*</b>

**Note:** \*The last layer of the encoder was unfrozen and the whole network was retrained. For example, 25+5\* means the encoder layers were frozen for 25 epochs and the last layer was unfrozen and the whole network was retrained for 5 more epochs.

### 3. Result and Discussion

During the first phase, when the model was trained over Portugal for 2016 with no pre-trained weights, the average IoU was 0.51 and the average Dice was 0.60. Following that, we trained the model again for 2017 with pre-trained weights from the previous training. This time we achieved accuracy of IoU 0.52 and Dice of 0.61. Because we aim to capture small fires over Punjab and to avoid too much generalization over Portugal, we decided not to train the model further over Portugal. The second phase of the model experiment involves transfer learning, but before that, just to put the performance of the present model into context, we tested the pre-trained model directly over Punjab before fine-tuning. The testing accuracy was IoU 0.15 and Dice 0.13. This marked dip in the testing

accuracy indicates that the model is struggling with the domain gap and the new domain has new features which cannot be comprehended by the model.

As a result, we observed several over-predicted burned areas over Punjab (Figure 7(a) and (b)). In addition, several false predictions were made, e.g., in Figure 7(c) an urban center is misunderstood as burned by the model, most likely because of the similarity in the reflectance value. In another case, an urban center (top right) and a large area of green vegetation (left) are falsely detected as a burned area (Figure 7(d)). Finally, the model is tuned with annotated data and pre-trained weights, the accuracy of the final best-performing model validation accuracy, after adjusting the hyperparameters (batch size, weight decay, and learning rate) is IoU 0.48, with Dice of 0.64 over Punjab. The model's overall capacity to detect small and fragmented fire patches also improved.

### 3.1. Model Performance



(a)



(b)

**Figure 6.** Training history of the final model: (a) model best performance over Portugal and (b) model best performance over Punjab.

Table 2. Model accuracy assessment for all experiments.

Phase	Expt.	Domain	IoU (Val)	Dice (Val)	Model Accuracy IoU (Test)	Dice (Test)	Val_loss	Train_loss
I	A	Portugal	0.62	0.72			0.34	0.33
	B	<b>Portugal</b>	<b>0.66</b>	<b>0.75</b>			<b>0.33</b>	<b>0.32</b>
II	C	Punjab			0.10	0.13		
	D	Punjab	0.44	0.60	0.45	0.60	0.42	0.42
	E	Punjab	0.48	0.64	0.45	0.60	0.38	0.38
	F	Punjab	0.49	0.64	0.46	0.61	0.38	0.37
	G	Punjab	0.49	0.60	0.52	0.64	0.51	0.43
	H	Punjab	0.40	0.53	0.42	0.54	0.45	0.40
	I	Punjab	0.40	0.55	0.39	0.54	0.44	0.51
	J	Punjab	0.44	0.59	0.43	0.59	0.48	0.49
	K	Punjab	0.47	0.57	0.48	0.58	0.58	0.56
	L	Punjab	0.43	0.59	0.43	0.58	0.44	0.45
	M	<b>Punjab</b>	<b>0.49</b>	<b>0.60</b>	<b>0.54</b>	<b>0.64</b>	<b>0.56</b>	<b>0.51</b>

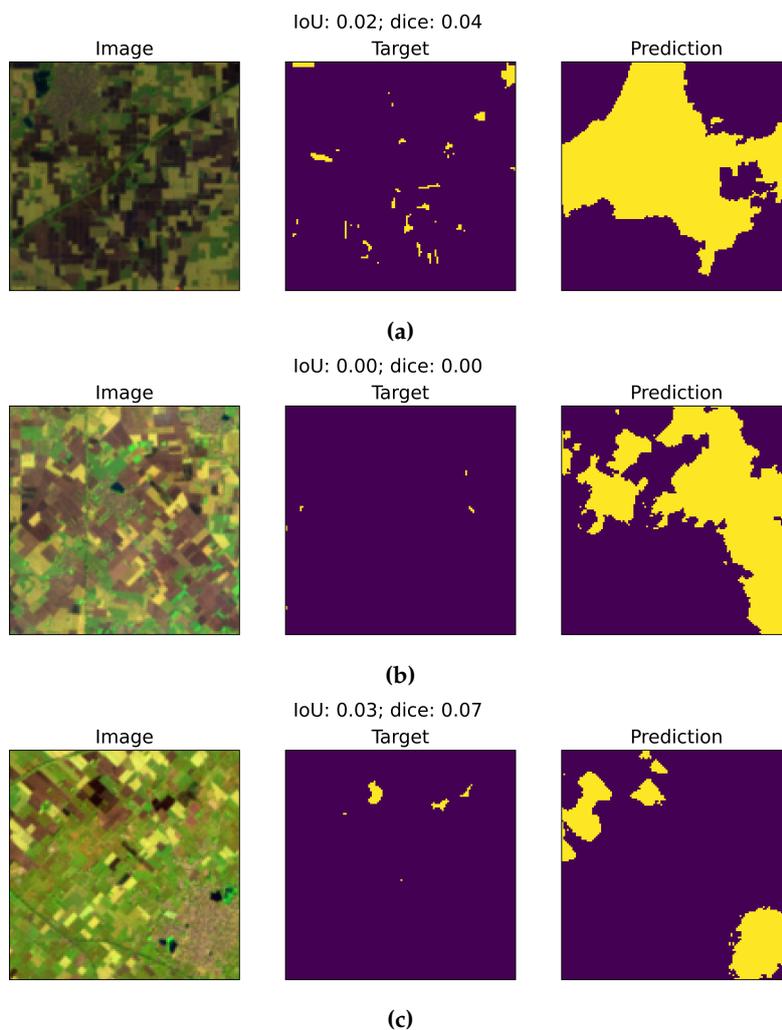
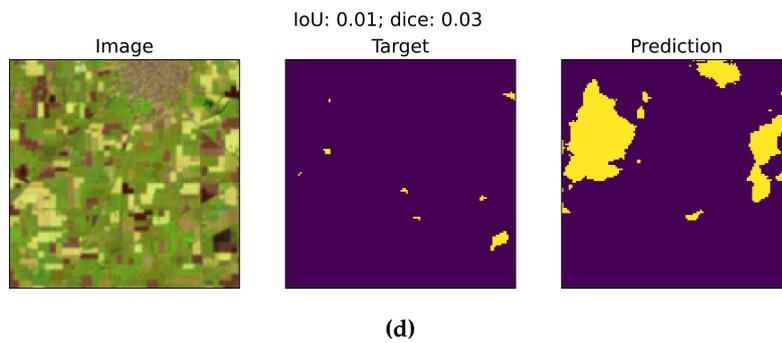
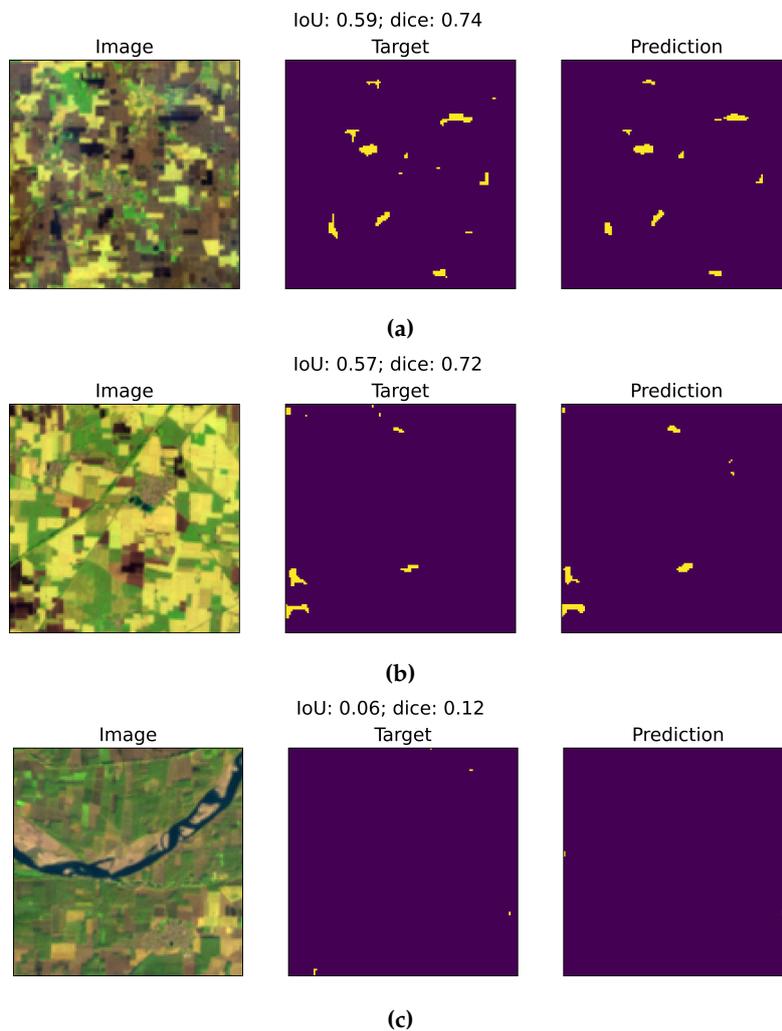


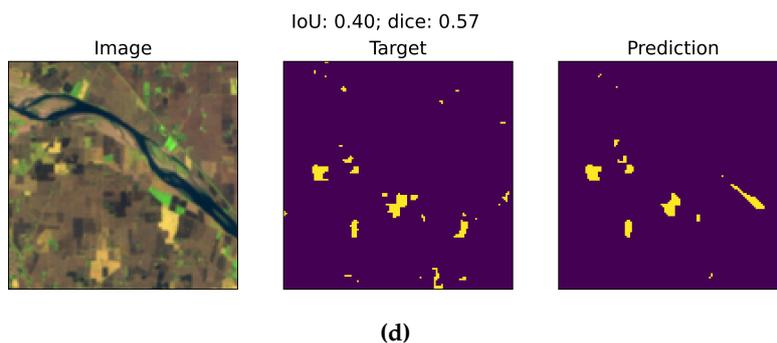
Figure 7. Cont.



**Figure 7.** Burn Area detection over Punjab (Before fine-tuning). Left panel shows Sentinel-2 false-color images (Green, NIR, and SWIR) over Punjab. Middle panel display hand-annotated reference data. Right panel present model predictions. IoU and Dice scores quantify similarities between ground truth and the model output. (a) and (b): display the model's overestimation of burn areas; (c) and (d): show examples of false burn area predictions over urban areas (bottom right).



**Figure 8.** *Cont.*



**Figure 8.** Burn Area detection over Punjab (After fine-tuning). Left panel shows Sentinel-2 false-color images (Green, NIR, and SWIR) over Punjab. Middle panel display hand-annotated reference data. Right panel present model predictions. IoU and Dice scores quantify similarities between ground truth and the model output. (a) shows the model's improvement, with less false detection over urban area. (b) shows very precise capture of small fragmented burned crop fields, with less false detection over bare/ ploughed ground. (c) and (d) show considerable improvement in fewer false predictions over aquatic bodies.

### 3.2. Model Prediction over Punjab (before Tuning vs. after Tuning)

As described in this section, we particularly examine the importance of fine tuning the DL model and improved segmentation of burned areas by examining Punjab before and after fine-tuning. Comparing these predictions highlights marked improvements in accuracy and precision when detecting active fires and smoke in satellite imagery after tuning with hand-annotated data. Figure 9(a), (c) shows segmentation results obtained before fine-tuning, which suggests that the pre-tuning model struggled with smaller fires, indicating a lack of sensitivity to finer details. Figure 9(b) and (d) portray segmentation results for the same area after fine-tuning the DL - model, where the burn area delineations have been captured better, resulting in high IoU and Dice scores.

In the pre-tuning predictions, the model overestimates fire presence, as indicated by the large yellow patches. The poor overlap between predicted and target areas is reflected in the IoU (0.10) and the Dice score (0.19). Although the model detects areas of fire and smoke, it fails to delineate boundaries accurately, leading to numerous false positive results. Another case in Figure 9 (e) segmentation results before fine tuning shows the detection of green and possibly bare ground as a burned area (right side of the figure). Figure 9(f) is the segmentation result of the same area after fine-tuning. The model was able to capture the small fragmented burned patch (lower right) with no severe false overlap with more than green vegetation or urban areas, resulting in a high IoU (0.73) and Dice score (0.84). These improvements indicate that fine-tuning of the model through various combinations of annotated data and training environment indeed helped the model to reflect the burned and unburned features better.

A core objective of this study is to enhance the model's ability to capture small fires and burn patches. Figure 9(c) shows that the pre-tuning model struggled with smaller fires, indicating a lack of sensitivity to finer details. The post-tuning prediction (bottom image) demonstrates significant improvement in detecting small fires and burn patches, with an IoU score of 0.40 and a Dice coefficient of 0.57.

The prediction map reflects the distribution and size of small fires more accurately, as indicated in the target. The improved accuracy and precision of the post-tuning model are important because they enhance the model's ability to provide reliable data for monitoring and managing agricultural fires. This improvement is crucially important for addressing environmental and public health concerns associated with burn areas, as discussed in earlier sections of this paper.

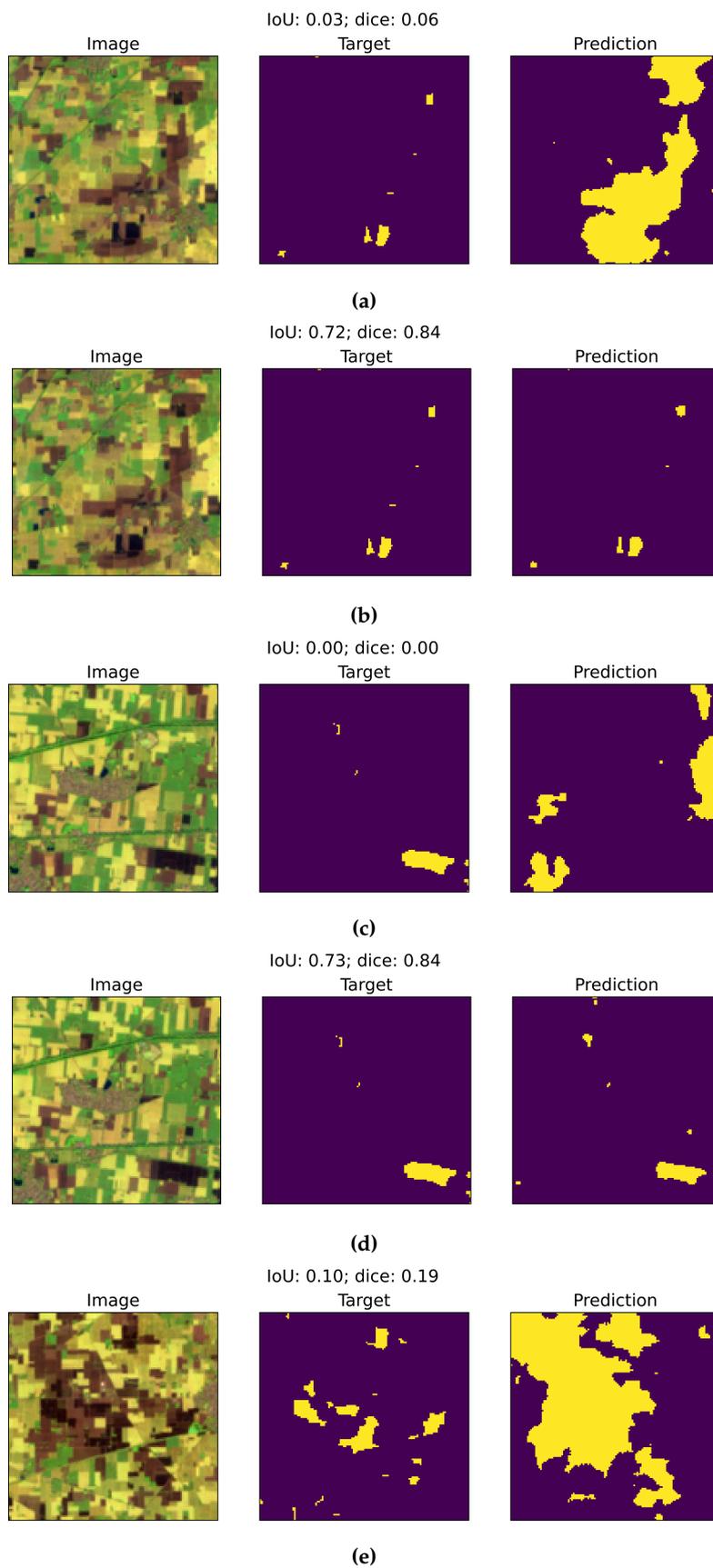
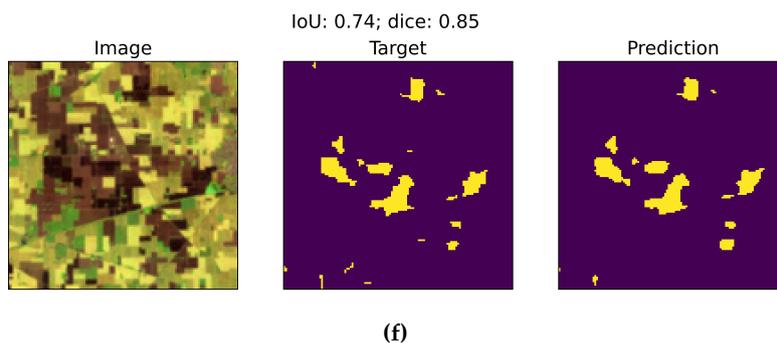


Figure 9. Cont.



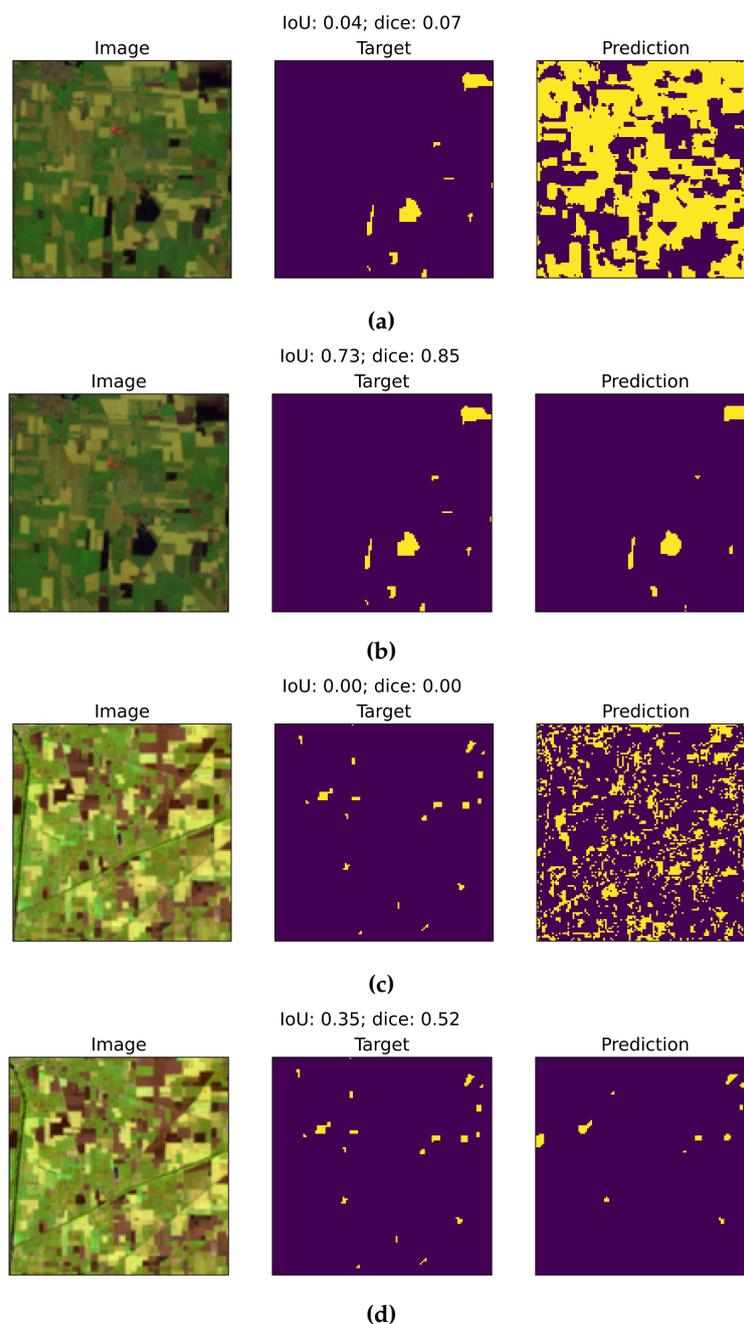
**Figure 9.** Comparison of burn area detection before and after tuning with the annotated data. Left panel shows Sentinel-2 false-color images (Green, NIR, and SWIR) over Punjab. Middle panel display hand-annotated reference data. Right panel present model predictions. (a) and (c) show predictions from the pretuned model; (b) and (d) are predictions from the tuned model. (e) segmentation results before fine tuning shows the detection of green and possibly bare ground as a burned area (right side of the figure). (f) capture the small fragmented burned patch (lower right) with no severe false overlap with more than green vegetation or urban areas

### 3.3. Model Prediction vs. Baseline Model (Normalized Burn Ratio Index Method)

This section presents comparison of the performance of the tuned DL model against the baseline model that uses the NBR index method for detecting burn areas. Our comparative analysis highlights the superiority and realistic approach of using the DL model, demonstrating how it overcomes the limitations inherent in the baseline model. The images illustrate that the baseline model tends to overestimate burn areas, particularly at the boundaries of small fires, whereas the tuned DL model captures these details more precisely. For instance, the baseline model prediction (Figure 10(a), top image) shows marked overestimation of burn areas, with large yellow patches indicating these regions. This overestimation is reflected in the low IoU score of 0.04 and Dice coefficient of 0.07, indicating poor overlap with the ground truth. The baseline model's tendency to exaggerate the burn area extent likely stems from its lower sensitivity to fine details and its inability to distinguish between burned and unburned areas accurately, especially for smaller fires.

By contrast, the tuned DL model prediction (Figure 10(b), bottom image) exhibits more precise and accurate detection of burn areas, including the details of small fires. The higher IoU score of 0.73 and Dice coefficient of 0.85 demonstrate markedly better overlap with the ground truth, showcasing the DL model's enhanced performance. Another example (Figure 10(c)) further underscores the limitations of the baseline model. The baseline model prediction (top image) identifies numerous small burn areas scattered throughout the image, even in regions where the ground truth indicates healthy vegetation. These false positives suggest that the baseline model suffers from spectral confusion, particularly between burn areas, bare ground, and healthy vegetation. Conversely, the tuned DL model prediction (bottom image) captures the burn areas with minimal false positives accurately, aligning well with the ground truth. The improved IoU score of 0.35 and Dice coefficient of 0.52 emphasize the DL model's superior ability to identify burn areas correctly.

As discussed in Section 2.3.1, both NIR and SWIR bands are sensitive to changes in vegetation and soil moisture. Burned areas typically exhibit a distinct spectral signature in these bands, characterized by lower NIR reflectance and higher SWIR reflectance compared to healthy vegetation. However, bare ground and burned ground can exhibit similar spectral properties, particularly in the NIR and SWIR bands. This similarity can then engender spectral confusion, where the baseline model misclassifies bare ground as burned areas. Additionally, varying levels of vegetation cover can affect NIR and SWIR reflectance. The NBR index's sensitivity to chlorophyll and moisture content of the ground can influence the spectral response, resembling that of burned ground and leading to false positives [66].

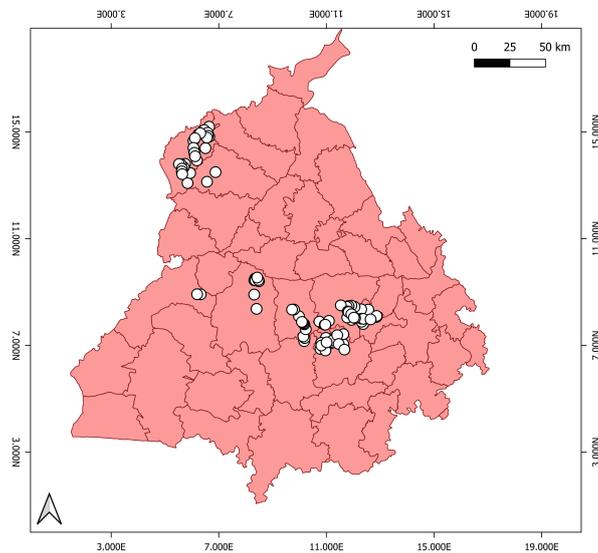


**Figure 10.** Comparison of burn area prediction by fine-tuned model and baseline model (Normalized Burn Area Index). Left panel shows Sentinel-2 false-color images (Green, NIR, and SWIR) over Punjab. Middle panel display hand-annotated reference data. Right panel present baseline model and DL model predictions. **(a)** shows a case of false detection over healthy vegetation by baseline model whereas in **(b)** shows burn area prediction over the same area by DL model, burn area delineations captured more precisely and fewer false predictions over healthy vegetation. In **(c)** burn area prediction by the baseline model shows false detection over bare ground and healthy vegetation, **(d)** shows burn area prediction by the DL model over the same area, small and fragmented burn patches captured with few false predictions.

### 3.4. Site-Validation

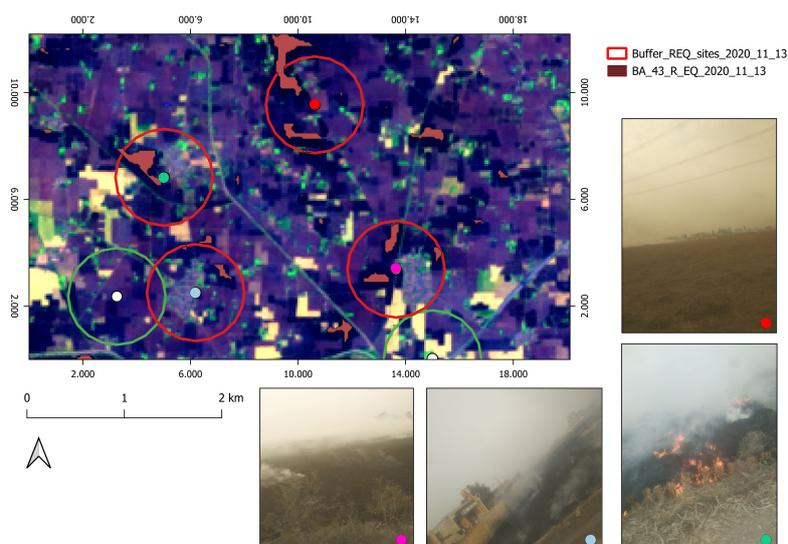
On-ground verification is necessary to elucidate the model predictions in light of actual ground truth data. To be certain that the burn area observed in the images is likely also to be detected by the model, we conducted a buffer zone analysis. To assess the model's effectiveness over the crop-burned

area, we also conducted a site survey across the Punjab region. During this site survey, approximately 110 images were collected on the ground using a geolocated camera. The location and time of capture were recorded meticulously for each image. Considering the quality of images and the presence and absence of a potential burn area site, we selected 94 images suitable for site validation. All the selected 94 sites are shown in Figure 11.

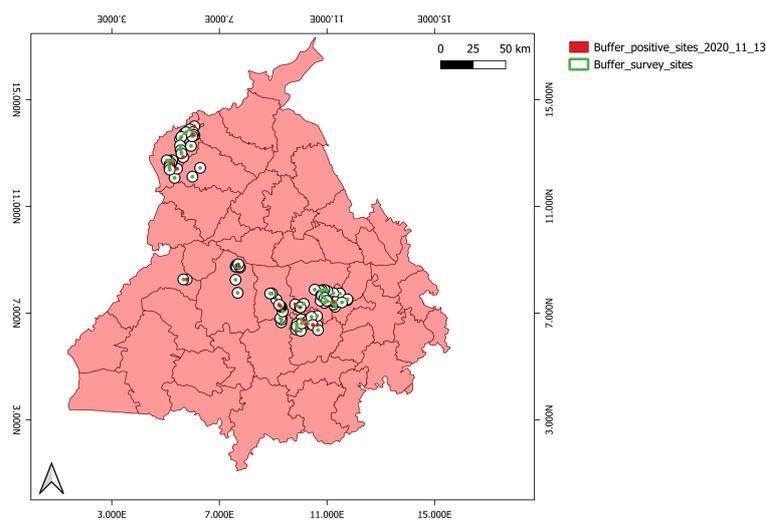


**Figure 11.** On-ground validation sites in Punjab, India. Each point represents a geolocation of an image captured during 9–11 November, 2020.

We conducted buffer zone analysis, a GIS technique by which a specified distance is applied around a point, line, or area to create a zone for spatial analysis [67]. This analysis was used to verify whether the burn area polygons detected by the model correspond to actual burn areas observed on the ground. We created a 500 m radius buffer zone around each location where field images were captured. The model-detected burn area data were then overlaid with these buffer zones, where the centre of each zone represented the point where on-ground images were taken. Because of uncertainty in the direction the images were taken, we assumed that if more than one burn area polygon fell within a zone, it would be considered a positive detection. The assumption underlying this approach is that any burn area detected within a 500 m radius from a given point is likely to fall within the field of view of the camera used to capture the geolocated images.



**Figure 12.** On-ground validation sites in Punjab, India. Red circles show buffer zones with the presence of BA most likely detected by the model.



**Figure 13.** Buffer zone analysis. Red spots signify the presence of a burn area.

Of the 94 sites, 27 showed strong evidence of burned areas detected within the 500 m buffer zone. Several factors contribute to this number. The survey images were captured during November 9–11, 2020, whereas the closest available Sentinel-2 images were taken on November 3 and 13, 2020. This timing discrepancy suggests that burning might have occurred either before or after the available satellite imagery, thereby affecting the validation accuracy. The positive observations from these 27 sites are important because they confirm the model's capability to detect burned areas accurately in real-world scenarios. However, the lower number of positive validations can be attributed to the timing mismatch between ground truth data collection and satellite image acquisition. Despite these challenges, the buffer zone analysis has been an important step in validating and refining our model, ensuring its reliability and effectiveness in real-world applications.

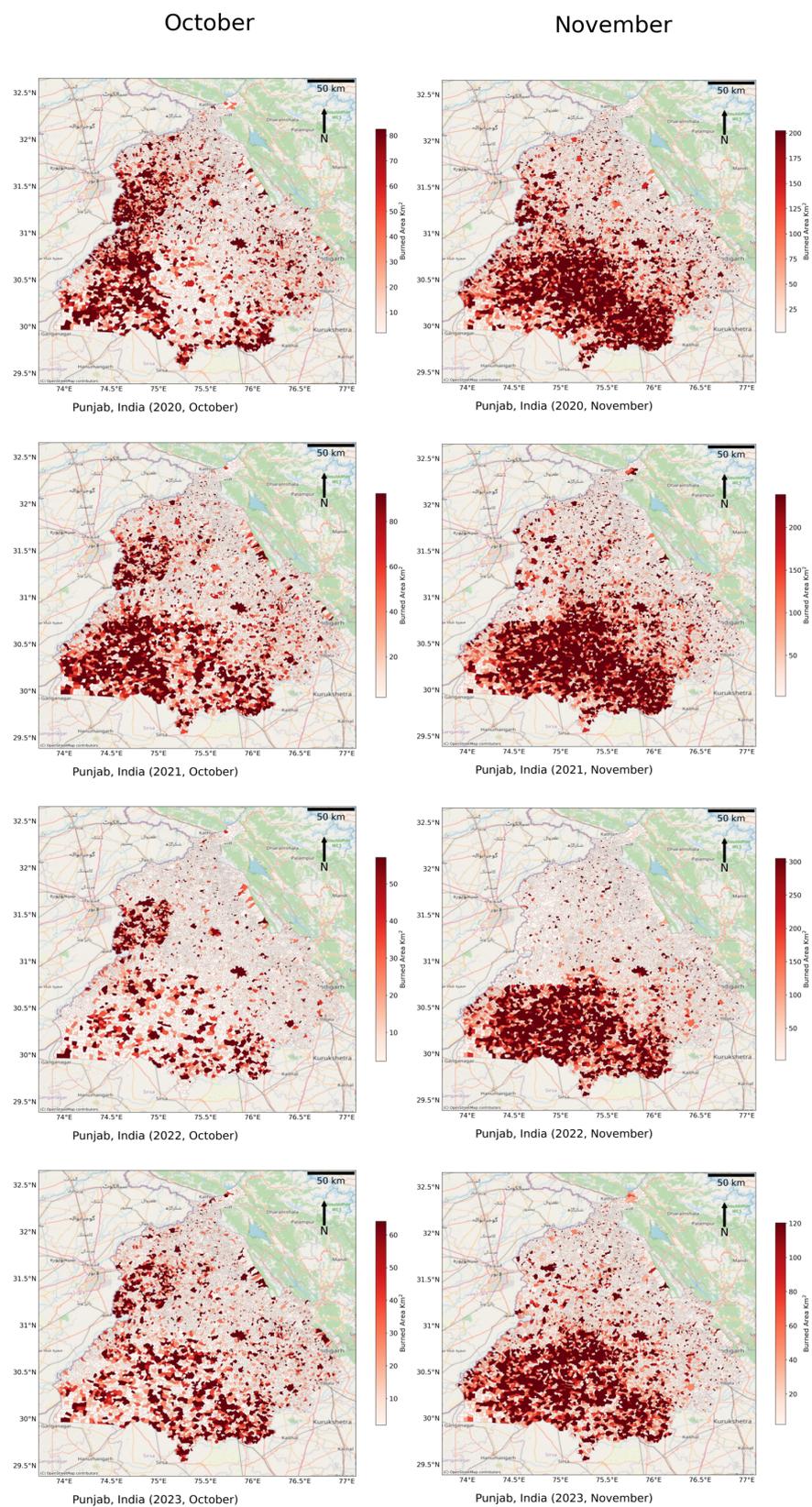
### 3.5. Spatio-Temporal Distribution of Monthly Fire Activity for Post Monsoon Burning Season (2020–2023)

In this section, we analyse the spatial distribution of burn areas across Punjab during the post-monsoon burning season, which extends from October through November. Our study examines the

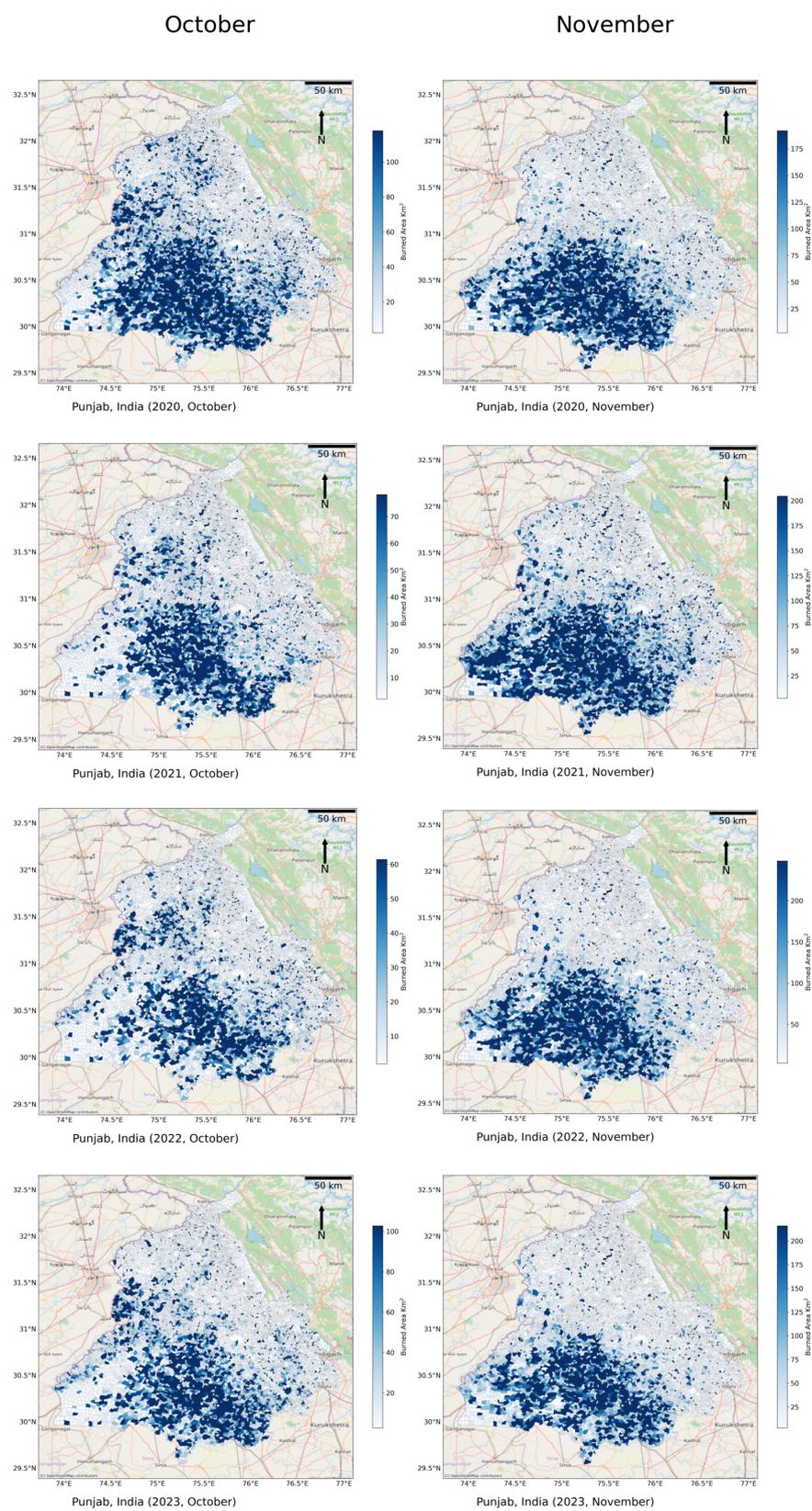
spatial distribution of burn areas over four consecutive years (2020–2023), particularly focusing on village-level burn area patterns. We evaluated the DL-model predictions by calculating the total burn area within each village in Punjab. For comparison purposes, we compared our findings against the MODIS-based MCD64A1 product [31]. MCD64A1 is a global burn area product, integral to the Global Fire Emission Database (GFED, ver. 4), which combines 500 m resolution observations with 1 km MODIS active fire data. The burn area mapping algorithm uses a burn-sensitive vegetation index (VI) derived from shortwave infrared bands 5 and 7 [68]. Detailed information related to the MCD64A1 algorithm can be found elsewhere in the literature [69]. Although MCD64A1 operates at a coarser resolution of 500 m, this comparison provides a meaningful benchmark against an established dataset, by allowing us to demonstrate the importance of Hi-resolution sentinel observation in capturing the small and fragmented burn areas. We would also like to clarify that this compares the cumulative extent of burn areas on the village level rather than the specific time or capture of active fire events.

Punjab is a vast region, making it challenging to pinpoint emission sources and their transport dynamics without detailed spatial analysis. Figure 14(a) portrays the burn area distribution at the village level as detected using our DL model. Both the DL model and MCD64A1 show a high proportion of burn areas in the villages of northwest Punjab (Amritsar, Taran Taran, Ferozpur) and the southern central part of Punjab (Ludhiana, Barnala, Sangrur, Maler Kotla, Bathinda) during October. In 2022 and 2023, less burning activity was detected by both the DL model and MCD64A1. This trend is consistent with the lower MODIS and VIIRS fire counts for the same years [70].

The peak burning activity can be observed from late October to mid-November. During this period, the southern part of Punjab (mainly Sangrur, Barnala, Mansa, Moga, and Bathinda) shows high number of burn areas during this period. These observations are also consistent with high fire detection counts in southern districts reported by the [70] during an intensive observation campaign. Recently, another study currently under review based on the Fire Detection Counts (FDC) by VIIRS S-NPP and PM 2.5 observations, [71] reported a continuous spike in FDC in Punjab from 2022 - 2021 and then observed a decline in FDC from 2022 - 2023. Both the DL model and MCD64A1 follow this trend of increasing FDC and show higher burning activity in 2020 and 2021. This increase can be attributed to the introduction of the Farmer's Bill (Indian Agricultural Act), which led to excess crop residue burning by farmers. The decline in FDC from 2022 - 2023, is also evidently observed by both MCD64A1 and DL model. This reduction in FDC can be attributed to the strict regulation and monitoring of the burning activity by the authorities and improvement in alternative residue management practices [72]. However, the DL model has surprisingly even lower burning activity, especially for the year 2022 (October), the burning activity observed by the DL model shows fewer burning activity. Several factors such as cloud cover and temporal gaps can be the possible reason. Additionally, for the year 2022 (October), the total number of sentinel tiles processed with less than 20% cloud cover was significantly lower. Applying robust cloud and cloud shadow masking techniques such as [73] in the data preparation can be useful to overcome this issue. Our analysis reveals marked spatial and temporal variations in burn area activity across Punjab, highlighting the importance of using high-resolution data and advanced models to capture small and fragmented burn areas. The DL model shows superiority against the MCD64A1 when detecting smaller burn areas, particularly during periods with lower overall activity.

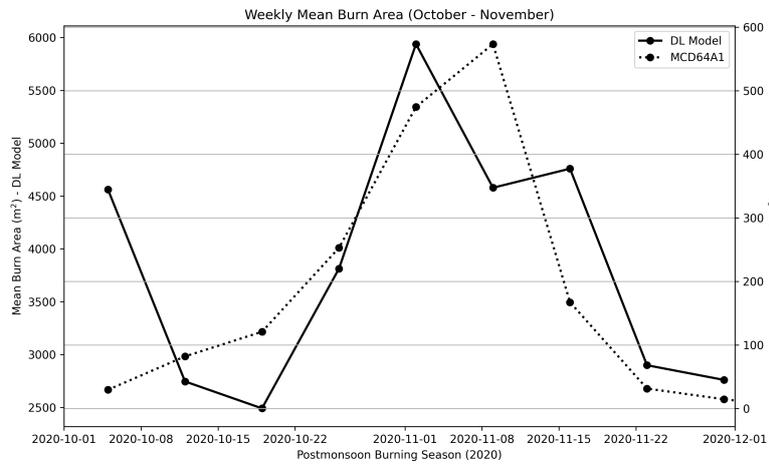


**Figure 14.** Village areas level spatial distribution of burn areas during October–November by DL-Model.

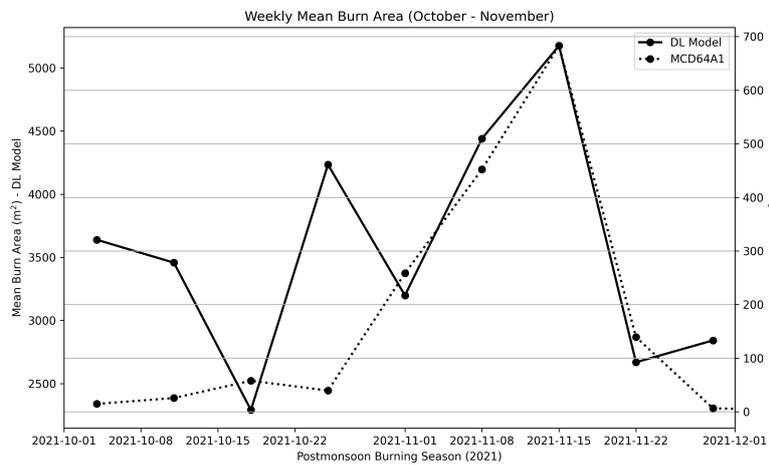


**Figure 15.** Village areas level spatial distribution of burn areas during October–November by MCD64A1.

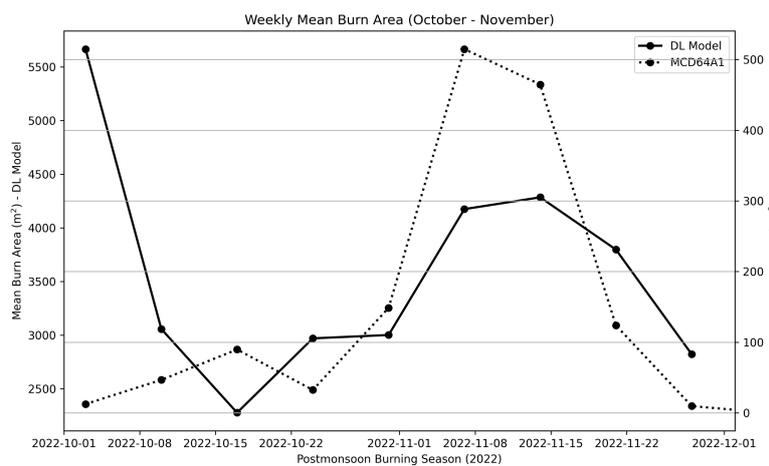
### 3.6. Comparison with MCD64A1 Burn Area Product



(a) 2020

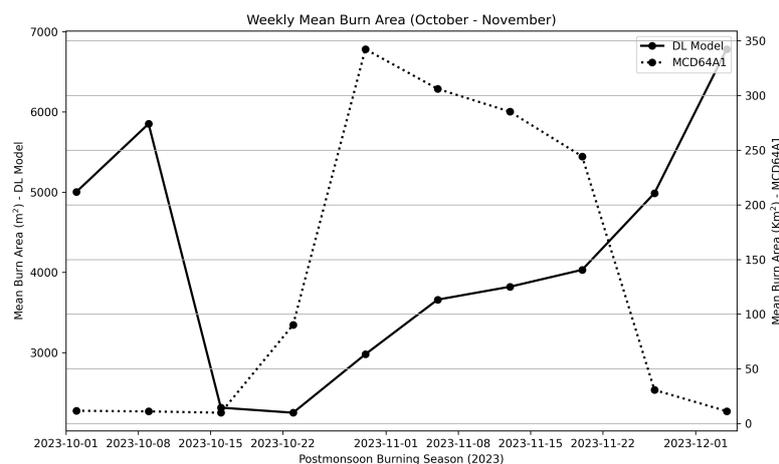


(b) 2021



(c) 2022

Figure 16. Cont.



(d) 2023

**Figure 16.** Time series of burn area distribution by size for DL-Model and MCD64A1

As described in this section, we compare the time series of burn area data from MCD64A1 and our DL) model for the post-monsoon burning seasons during October–November over four consecutive years (2020–2023). Despite considerably large differences in the scales of burn areas captured by MCD64A1 and the DL model, both datasets generally follow the conventional trend of burning, which peaks gradually from mid-October to mid-November. The primary objective of this comparison is to understand how the trends and patterns of small fires change throughout the burning season. The MCD64A1 dataset can capture fires up to  $1000\text{ m}^2$  under regular conditions and up to  $50\text{ m}^2$  under very clear, cloud-free conditions. However, MCD64A1 has been reported to have low capability for detecting smaller fires ( $<250\text{ m}^2$ ), especially those related to crop burning [11]. For 2020, the peak for small fires captured by the DL model coincides with the peak observed in the MCD64A1 dataset. It remains inconclusive why the peaks differed in 2023, comparison with active fire counts might shed light on this point.

Given the fragmented and small nature of agricultural fields, burn areas smaller than  $50\text{ m}^2$  are often overlooked by MCD64A1. This likely explains the high peaks observed in early October and late November from the smaller fires detected using the DL model, which are not clearly captured by MCD64A1. A recent study by [74] analysed the spatial accuracy of several MODIS-based burn area products, including MCD64A1 and FireCCI51, against fire size. They reported that fires smaller than  $250\text{ ha}$  ( $2.5\text{ m}^2$ ) had Dice scores of 0.22 for MCD64A1 and 0.24 for FireCCI51. In contrast, the DL model showed overall accuracy with a Dice score of 0.64 for fire sizes up to  $250\text{ m}^2$ . A summary of recent burn area algorithms and their respective accuracy can be found in Table 4. [11] also highlighted the challenges in mapping crop fires using MODIS-based burn area products because of the coarse resolution of MODIS and the short duration of fires and subsequent land changes. The high accuracy of burn area mapping by the DL model demonstrates its significant potential to overcome these existing limitations.

**Table 3.** Sentinel-2 spectral bands characteristics (adapted from [46]).

Band	Description	Bandwidth (nm)	Central wavelength (nm)	Spatial resolution (m)
1	Aerosol	20	443	60
2	Blue	65	490	10
3	Green	35	560	10
4	Red	30	665	10
5	Vegetation edge	15	705	20
6	Vegetation edge	15	740	20
7	Vegetation edge	20	783	20
8a	NIR	115	842	10
8b	Narrow NIR	20	865	20
9	Water vapor	20	945	60
10	Circus	30	1380	60
11	SWIR1	90	1610	20
12	SWIR2	180	2190	20

**Table 4.** Accuracy comparison with other burn area detection algorithms.

Model	Fire size	Dice score	Source
MCD64A1	2.5 km <sup>2</sup>	0.22	[74]
FireCCI51	2.5 km <sup>2</sup>	0.24	[74]
BA-Net	>5 km <sup>2</sup>	0.90	[41]
MCD64A1C6	>5 km <sup>2</sup>	0.90	[41,69]
<b>DL-Model</b>	<b>&gt;250 m<sup>2</sup></b>	<b>0.64</b>	

### 3.7. Limitations

The fine-tuned model demonstrates great capability for delineating small agricultural burn areas and consistently achieves higher accuracy than the baseline model based on conventional methods, but it does have some limitations. One important limitation is the model's specificity to agricultural burn areas. Because it is deliberately tuned for this purpose, its performance might not be as effective when capturing other fire types such as grassland or peat fires, which have different features. Another major shortcoming is the model's sensitivity to cloud and cloud shadow presence. Cloud interference can strongly affect the model's ability to detect burns because the burning efficiency is reduced. To address this point, we plan to apply robust cloud masking techniques in future analyses to achieve more consistent burn area mapping. Additionally, the model's temporal coverage is limited by the Sentinel-2 satellite's five-day revisit period. Given the unpredictable nature of burning practices during the post-harvest period, it is likely that some fire events are missed within this interval. Integrating observations from other satellites such as Landsat might help improve the maximum capture of burn areas.

## 4. Conclusion and Future Scope

This study demonstrated the effectiveness of high-resolution Sentinel-2 observations, DL models, and transfer learning approaches for accurately mapping small agricultural burn areas. Initially trained on data from Portugal, the model showed reliable capability in capturing large fires and delineating burn areas. After fine-tuning with data from Punjab, the model's ability to detect small burn patches in agricultural areas improved considerably. This enhanced accuracy is crucially important for quantifying the contribution of small fires to overall emissions. Additionally, the DL model exhibited superior burn area mapping efficiency compared to the baseline NBR index method. One salient benefit of this study is its promising methodology for assessing burn areas and their effects in regions with limited ground truth data. Furthermore, the model's adaptability allows it to be applied to detect burn areas in other similar agricultural regions worldwide. This study specifically addressed a single deep learning architecture, but future research might benefit from a comparative analysis

of different neural networks in burn area mapping. Exploring advanced neural networks such as U-Net++, Diffusion models, and one-shot learning might be particularly useful in scenarios with scarce data. Given the model's accommodation of multiple datasets, integrating higher-resolution datasets such as Planetscope might provide even more precise burn area mapping. Additionally, modifying the network architecture, especially during the transfer learning step, by adjusting the final layers of the neural network might enhance overall model accuracy further. In summary, this study not only highlights the methodological advancements in using DL models for detecting small agricultural burn areas: it also paves the way for additional research and application in various agricultural regions worldwide.

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

**Funding:** This research was partially supported by Research Institute for Humanity and Nature (RIHN: a constituent member of NIHU) Project No. 14200133. This research was conducted using the computational resources of Fujitsu PRIMEHPC FX1000 and FUJITSU PRIMERGY GX2570 (Wisteria/BDEC-01) at the Information Technology Center, The University of Tokyo.

**Acknowledgments:** This research was partially supported by the Research Institute for Humanity and Nature (RIHN: a constituent member of NIHU) Project No. 14200133. We are especially grateful to Dr. Tanbir Singh, Prof. Kamal Vatta, and Prof. Kanako Muramatsu for providing the on-site images from the AAKASH project campaign. We express heartfelt thanks to Ms. Lisa Knopp for her expertise and encouragement, which have been greatly appreciated. This research was conducted using the computational resources of Fujitsu PRIMEHPC FX1000 and FUJITSU PRIMERGY GX2570 (Wisteria/BDEC-01) at the Information Technology Center, The University of Tokyo. We have used the Land Use Land Cover information on our research work from the Natural Resources Census Project of the National Remote Sensing Centre (NRSC), ISRO, Hyderabad, India. Digital Database: Digital Database Bhuvan-Thematic Services, LULC503/MAP/PB.jpg, NRSC/ISRO - India.

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

## Abbreviations

The following abbreviations are used in this manuscript:

DL	Deep Learning
CNN	Convolutional Neural Networks
NBR	Normalized Burn Ratio
MODIS	Moderate Resolution Imaging Spectroradiometer
VIIRS	Visible Infrared Imaging Radiometer Suite
NIR	Near Infrared
SWIR	Shortwave Infrared
ICNF	Institute for Nature Conservation and Forests
MSI	Multispectral Instrument
TOA	Top of Atmosphere
IoU	Intersection over Union
Ano1	Annotation 1
Ano2	Annotation 2
FDC	Fire Detection Counts

## Appendix A

### Appendix A.1

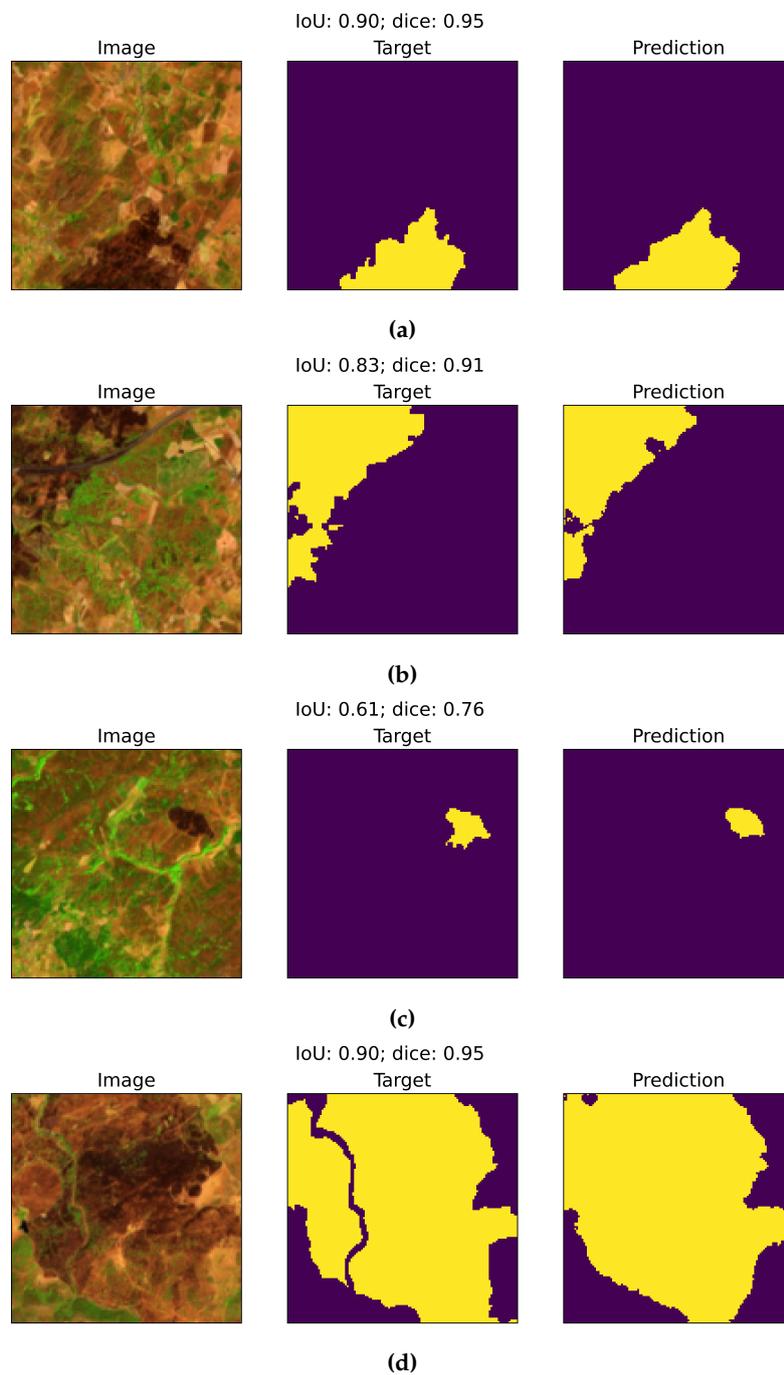
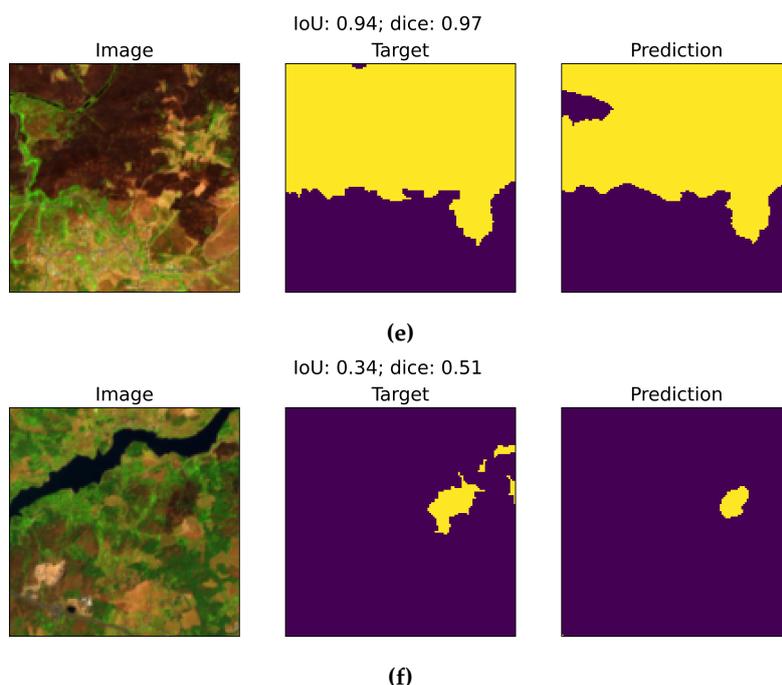


Figure A1. Cont.



**Figure A1.** Left panel shows Sentinel-2 false-color images (Green, NIR, and SWIR) over Portugal. Middle panels display ground truth data from ICNF. Right panels present model predictions. IoU and Dice scores quantify similarities between ground truth and the model output. (a), (b), and (d): demonstrate high similarity and the model's capability to capture the full extent of large fires. (c): shows the detection of a small fire with high similarity in unburned areas. (e): highlights a potential error in the ground truth data, but the model accurately identifies unburned areas. (f): represents the model's difficulty in capturing some small burn patches, leading to lower IoU and Dice scores, although it detects large fires with some fragmentation.

## References

1. Korontzi, S.; McCarty, J.; Loboda, T.; Kumar, S.; Justice, C. Global distribution of agricultural fires in croplands from 3 years of Moderate Resolution Imaging Spectroradiometer (MODIS) data. *Global Biogeochemical Cycles - GLOBAL BIOGEOCHEM CYCLE* **2006**, *20*. <https://doi.org/10.1029/2005GB002529>.
2. Cassou, E. Field Burning. Technical report, World Bank, Washington, DC, 2018. <https://doi.org/10.1596/29504>.
3. Kumar, P.; Kumar, S.; Joshi, L., The Extent and Management of Crop Stubble. In *Socioeconomic and Environmental Implications of Agricultural Residue Burning: A Case Study of Punjab, India*; Springer India: New Delhi, 2015; pp. 13–34. [https://doi.org/10.1007/978-81-322-2014-5\\_2](https://doi.org/10.1007/978-81-322-2014-5_2).
4. Jethva, H.; Torres, O.; Field, R.D.; Lyapustin, A.; Gautam, R.; Kayetha, V. Connecting Crop Productivity, Residue Fires, and Air Quality over Northern India. *Scientific Reports* **2019**, *9*. <https://doi.org/10.1038/s41598-019-52799-x>.
5. Deshpande, M.V.; Kumar, N.; Pillai, D.; Krishna, V.V.; Jain, M. Greenhouse gas emissions from agricultural residue burning have increased by 75% since 2011 across India. *Science of The Total Environment* **2023**, *904*, 166944. <https://doi.org/10.1016/j.scitotenv.2023.166944>.
6. van der Werf, G.R.; Randerson, J.T.; Giglio, L.; Collatz, G.J.; Mu, M.; Kasibhatla, P.S.; Morton, D.C.; DeFries, R.S.; Jin, Y.; van Leeuwen, T.T. Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmospheric Chemistry and Physics* **2010**, *10*, 11707–11735. <https://doi.org/10.5194/acp-10-11707-2010>.
7. Singh, R.; Sinha, B.; Hakkim, H.; Sinha, V. Source apportionment of volatile organic compounds during paddy-residue burning season in north-west India reveals large pool of photochemically formed air toxics. *Environmental Pollution* **2023**, *338*, 122656.

8. Dhaka, S.K.; Chetna.; Kumar, V.; Panwar, V.; Dimri, A.; Singh, N.; Patra, P.K.; Matsumi, Y.; Takigawa, M.; Nakayama, T.; et al. PM<sub>2.5</sub> diminution and haze events over Delhi during the COVID-19 lockdown period: an interplay between the baseline pollution and meteorology. *Scientific reports* **2020**, *10*, 13442.
9. Hall, J.V.; Loboda, T.V. Quantifying the Potential for Low-Level Transport of Black Carbon Emissions from Cropland Burning in Russia to the Snow-Covered Arctic. *Frontiers in Earth Science* **2017**, *5*. <https://doi.org/10.3389/feart.2017.00109>.
10. Andreae, M.O., Biomass Burning: Its History, Use, and Distribution and Its Impact on Environmental Quality and Global Climate. In *Global Biomass Burning*; The MIT Press, 1991; p. 3–21. <https://doi.org/10.7551/mitpress/3286.003.0003>.
11. Hall, J.V.; Loboda, T.V.; Giglio, L.; McCarty, G.W. A MODIS-based burned area assessment for Russian croplands: Mapping requirements and challenges. *Remote Sensing of Environment* **2016**, *184*, 506–521. <https://doi.org/10.1016/j.rse.2016.07.022>.
12. Seiler, W.; Crutzen, P.J. Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. *Climatic change* **1980**, *2*, 207–247.
13. Ramo, R.; Roteta, E.; Bistinas, I.; Van Wees, D.; Bastarrika, A.; Chuvieco, E.; Van der Werf, G.R. African burned area and fire carbon emissions are strongly impacted by small fires undetected by coarse resolution satellite data. *Proceedings of the National Academy of Sciences* **2021**, *118*, e2011160118.
14. WMO.; UNEP.; ICSU.; IOC-UNESCO. The Global Observing System for Climate: Implementation Needs. Technical Report GCOS -200, World Meteorological Organization (WMO), 2016.
15. Roteta, E.; Bastarrika, A.; Padilla, M.; Storm, T.; Chuvieco, E. Development of a Sentinel-2 burned area algorithm: Generation of a small fire database for sub-Saharan Africa. *Remote Sensing of Environment* **2019**, *222*, 1–17. <https://doi.org/10.1016/j.rse.2018.12.011>.
16. Kaskaoutis, D.; Kumar, S.; Sharma, D.; Singh, R.P.; Kharol, S.; Sharma, M.; Singh, A.; Singh, S.; Singh, A.; Singh, D. Effects of crop residue burning on aerosol properties, plume characteristics, and long-range transport over northern India. *Journal of Geophysical Research: Atmospheres* **2014**, *119*, 5424–5444.
17. Ravindra, K.; Singh, T.; Mor, S. Emissions of air pollutants from primary crop residue burning in India and their mitigation strategies for cleaner emissions. *Journal of Cleaner Production* **2019**, *208*, 261–273. <https://doi.org/10.1016/j.jclepro.2018.10.031>.
18. Chetna, N.; Devi, M.; et al. Forecasting of Pre-harvest Wheat Yield using Discriminant Function Analysis of Meteorological Parameters of Kurukshetra District of Haryana. *Journal of Agriculture Research and Technology* **2022**, *47*, 67–72.
19. Wang, J.; Christopher, S.A.; Nair, U.S.; Reid, J.S.; Prins, E.M.; Szykman, J.; Hand, J.L. Mesoscale modeling of Central American smoke transport to the United States: 1. “Top-down” assessment of emission strength and diurnal variation impacts. *Journal of Geophysical Research: Atmospheres* **2006**, *111*. <https://doi.org/10.1029/2005JD006416>.
20. Sharma, A.R.; Kharol, S.K.; Badarinath, K.V.S.; Singh, D. Impact of agriculture crop residue burning on atmospheric aerosol loading & a study over Punjab State, India. *Annales Geophysicae* **2010**, *28*, 367–379. <https://doi.org/10.5194/angeo-28-367-2010>.
21. Kumar, R.; Barth, M.; Pfister, G.; Nair, V.; Ghude, S.D.; Ojha, N. What controls the seasonal cycle of black carbon aerosols in India? *Journal of Geophysical Research: Atmospheres* **2015**, *120*, 7788–7812.
22. National Remote Sensing Centre (NRSC), ISRO. Land Use Land Cover Digital Database. <https://bhuvan-app1.nrsc.gov.in/2dresources/thematic/LULC503/MAP/PB.jpg>, 2024. Natural Resources Census Project, accessed 2024.
23. Knopp, L.; Wieland, M.; Rättich, M.; Martinis, S. A Deep Learning Approach for Burned Area Segmentation with Sentinel-2 Data. *Remote Sensing* **2020**, *12*. <https://doi.org/10.3390/rs12152422>.
24. 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 Sensing of Environment* **2019**, *225*, 45–64. <https://doi.org/10.1016/j.rse.2019.02.013>.
25. Chang, D.; Song, Y. Estimates of biomass burning emissions in tropical Asia based on satellite-derived data. *Atmospheric Chemistry and Physics* **2010**, *10*, 2335–2351.

26. Key, C.H.; Benson, N.C. Landscape Assessment: Ground measure of severity, the Composite Burn Index; and Remote sensing of severity, the Normalized Burn Ratio. Technical report, USDA Forest Service, Rocky Mountain Research Station, Ogden, UT, 2006. Report.
27. Roy, D.; Boschetti, L.; Justice, C.; Ju, J. The collection 5 MODIS burned area product — Global evaluation by comparison with the MODIS active fire product. *Remote Sensing of Environment* **2008**, *112*, 3690–3707. <https://doi.org/10.1016/j.rse.2008.05.013>.
28. Giglio, L.; Randerson, J.T.; van der Werf, G.R.; Kasibhatla, P.S.; Collatz, G.J.; Morton, D.C.; DeFries, R.S. Assessing variability and long-term trends in burned area by merging multiple satellite fire products. *Biogeosciences* **2010**, *7*, 1171–1186. <https://doi.org/10.5194/bg-7-1171-2010>.
29. Petropoulos, G.P.; Kontoes, C.; Keramitsoglou, I. Burnt area delineation from a uni-temporal perspective based on Landsat TM imagery classification using Support Vector Machines. *International Journal of Applied Earth Observation and Geoinformation* **2011**, *13*, 70–80. <https://doi.org/10.1016/j.jag.2010.06.008>.
30. Ramo, R.; Chuvieco, E. Developing a Random Forest Algorithm for MODIS Global Burned Area Classification. *Remote Sensing* **2017**, *9*. <https://doi.org/10.3390/rs9111193>.
31. Giglio, L.; Loboda, T.; Roy, D.P.; Quayle, B.; Justice, C.O. An active-fire based burned area mapping algorithm for the MODIS sensor. *Remote Sensing of Environment* **2009**, *113*, 408–420. <https://doi.org/10.1016/j.rse.2008.10.006>.
32. L. Giglio, J.D.K.; Mack, R. A multi-year active fire dataset for the tropics derived from the TRMM VIRS. *International Journal of Remote Sensing* **2003**, *24*, 4505–4525. <https://doi.org/10.1080/0143116031000070283>.
33. McCarty, J.L.; Korontzi, S.; Justice, C.O.; Loboda, T. The spatial and temporal distribution of crop residue burning in the contiguous United States. *Science of the Total Environment* **2009**, *407*, 5701–5712.
34. K. R. Al-Rawi, J.L.C.; Calle, A. Burned area mapping system and fire detection system, based on neural networks and NOAA-AVHRR imagery. *International Journal of Remote Sensing* **2001**, *22*, 2015–2032. <https://doi.org/10.1080/01431160117531>.
35. P. A. Brivio, M. Maggi, E.B.; Gallo, I. Mapping burned surfaces in Sub-Saharan Africa based on multi-temporal neural classification. *International Journal of Remote Sensing* **2003**, *24*, 4003–4016. <https://doi.org/10.1080/0143116031000103835>.
36. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems; Pereira, F.; Burges, C.; Bottou, L.; Weinberger, K., Eds. Curran Associates, Inc., 2012, Vol. 25.
37. Wieland, M.; Li, Y.; Martinis, S. Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. *Remote Sensing of Environment* **2019**, *230*, 111203. <https://doi.org/10.1016/j.rse.2019.05.022>.
38. Wieland, M.; Martinis, S.; Kiefl, R.; Gstaiger, V. Semantic segmentation of water bodies in very high-resolution satellite and aerial images. *Remote Sensing of Environment* **2023**, *287*, 113452. <https://doi.org/10.1016/j.rse.2023.113452>.
39. Maggiori, E.; Tarabalka, Y.; Charpiat, G.; Alliez, P. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 645–657. <https://doi.org/10.1109/TGRS.2016.2612821>.
40. Rustowicz, R.; Cheong, R.; Wang, L.; Ermon, S.; Burke, M.; Lobell, D. Semantic Segmentation of Crop Type in Africa: A Novel Dataset and Analysis of Deep Learning Methods. In Proceedings of the CVPR Workshops, 2019.
41. Pinto, M.M.; Libonati, R.; Trigo, R.M.; Trigo, I.F.; DaCamara, C.C. A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, *160*, 260–274. <https://doi.org/10.1016/j.isprsjprs.2019.12.014>.
42. Pinto, M.M.; Trigo, R.M.; Trigo, I.F.; DaCamara, C.C. A Practical Method for High-Resolution Burned Area Monitoring Using Sentinel-2 and VIIRS. *Remote Sensing* **2021**, *13*. <https://doi.org/10.3390/rs13091608>.
43. Huppertz, R.; Nakalembe, C.; Kerner, H.; Lachyan, R.; Rischard, M. Using transfer learning to study burned area dynamics: A case study of refugee settlements in West Nile, Northern Uganda. *arXiv* **2021**, arXiv:2107.14372.
44. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015; Navab, N.; Hornegger, J.; Wells, W.M.; Frangi, A.F., Eds., Cham, 2015; pp. 234–241.

45. Main-Knorn, M.; Pflug, B.; Louis, J.; Debaecker, V.; Müller-Wilm, U.; Gascon, F. Sen2Cor for Sentinel-2. In Proceedings of the Image and Signal Processing for Remote Sensing XXIII; Bruzzone, L., Ed. International Society for Optics and Photonics, SPIE, 2017, Vol. 10427, p. 1042704. <https://doi.org/10.1117/12.2278218>.
46. Sentinel-2 ESA's Optical High-Resolution Mission for GMES Operational Services (ESA SP-1322/2 March 2012). Available online: [https://sentinel.esa.int/documents/247904/349490/S2\\_SP-1322\\_2.pdf](https://sentinel.esa.int/documents/247904/349490/S2_SP-1322_2.pdf) (accessed on 4 August 2024).
47. Turco, M.; Jerez, S.; Augusto, S.; Tarín-Carrasco, P.; Ratola, N.; Jiménez-Guerrero, P.; Trigo, R.M. Climate drivers of the 2017 devastating fires in Portugal. *Scientific Reports* **2019**, *9*. <https://doi.org/10.1038/s41598-019-50281-2>.
48. Menezes, L.S.; Russo, A.; Libonati, R.; Trigo, R.M.; Pereira, J.M.; Benali, A.; Ramos, A.M.; Gouveia, C.M.; Morales Rodriguez, C.A.; Deus, R. Lightning-induced fire regime in Portugal based on satellite-derived and in situ data. *Agricultural and Forest Meteorology* **2024**, *355*, 110108. <https://doi.org/10.1016/j.agrformet.2024.110108>.
49. Copernicus Data Space Ecosystem - API - OData. Available online: <https://documentation.dataspace.copernicus.eu/APIs/OData.html> (accessed on 2 August 2024).
50. Institute of Conservation of Nature and Forest. Available online: <https://sig.icnf.pt/portal/home/item.html?id=983c4e6c4d5b4666b258a3ad5f3ea5af#overview> (accessed on 2 August 2024).
51. Vadrevu, K.; Lasko, K. Intercomparison of MODIS AQUA and VIIRS I-Band Fires and Emissions in an Agricultural Landscape—Implications for Air Pollution Research. *Remote Sensing* **2018**, *10*. <https://doi.org/10.3390/rs10070978>.
52. Liu, T.; Marlier, M.E.; Karambelas, A.; Jain, M.; Singh, S.; Singh, M.K.; Gautam, R.; DeFries, R.S. Missing emissions from post-monsoon agricultural fires in northwestern India: regional limitations of MODIS burned area and active fire products. *Environmental Research Communications* **2019**, *1*, 011007. <https://doi.org/10.1088/2515-7620/ab056c>.
53. Key, C.; Benson, N. Measuring and remote sensing of burn severity. In Proceedings of the Joint Fire Science Conference and Workshop Proceedings: 'Crossing the Millennium: Integrating Spatial Technologies and Ecological Principles for a New Age in Fire Management', June 1999.
54. Keeley, J. Fire intensity, fire severity and burn severity: A brief review and suggested usage. *International Journal of Wildland Fire* **2009**, *18*, 116–126. <https://doi.org/10.1071/WF07049>.
55. Rahman, S.; Chang, H.C.; Hehir, W.; Magilli, C.; Tomkins, K. Inter-Comparison of Fire Severity Indices from Moderate (Modis) and Moderate-To-High Spatial Resolution (Landsat 8 & Sentinel-2A) Satellite Sensors. In Proceedings of the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, 2018, pp. 2873–2876. <https://doi.org/10.1109/IGARSS.2018.8518449>.
56. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
57. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6230–6239. <https://doi.org/10.1109/CVPR.2017.660>.
58. Wisteria/BDEC-01 Supercomputer System. Available online: <https://www.cc.u-tokyo.ac.jp/en/supercomputer/wisteria/service/> (accessed on 3 August 2024).
59. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings; Bengio, Y.; LeCun, Y., Eds., 2015.
60. Szpakowski, D.M.; Jensen, J.L.R. A Review of the Applications of Remote Sensing in Fire Ecology. *Remote Sensing* **2019**, *11*. <https://doi.org/10.3390/rs11222638>.
61. Gamper, J.; Koohbanani, N.A.; Benes, K.; Graham, S.; Jahanifar, M.; Khurram, S.A.; Azam, A.; Hewitt, K.; Rajpoot, N. PanNuke Dataset Extension, Insights and Baselines, 2020, [arXiv:eess.IV/2003.10778].
62. Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A Review on Deep Learning Techniques Applied to Semantic Segmentation, 2017, [arXiv:cs.CV/1704.06857].
63. Zou, K.H.; Warfield, S.K.; Bharatha, A.; Tempany, C.M.; Kaus, M.R.; Haker, S.J.; Wells, W.M.; Jolesz, F.A.; Kikinis, R. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic Radiology* **2004**, *11*, 178–189. [https://doi.org/10.1016/S1076-6332\(03\)00671-8](https://doi.org/10.1016/S1076-6332(03)00671-8).

64. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC genomics* **2020**, *21*, 1–13.
65. Biewald, L. Experiment Tracking with Weights and Biases, 2020. Software available from wandb.com.
66. Farhadi, H.; Ebadi, H.; Kiani, A. BADI: A NOVEL BURNED AREA DETECTION INDEX FOR SENTINEL-2 IMAGERY USING GOOGLE EARTH ENGINE PLATFORM. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2023**, X-4/W1-2022, 179–186. <https://doi.org/10.5194/isprs-annals-X-4-W1-2022-179-2023>.
67. Singh, G.; Kant, Y.; Dadhwal, V. Remote sensing of crop residue burning in Punjab (India): a study on burned area estimation using multi-sensor approach. *Geocarto International* **2009**, *24*, 273–292.
68. Roy, D.P.; Landmann, T. Characterizing the surface heterogeneity of fire effects using multi-temporal reflective wavelength data. *International Journal of Remote Sensing* **2005**, *26*, 4197–4218, <https://doi.org/10.1080/01431160500112783>.
69. Giglio, L.; Boschetti, L.; Roy, D.P.; Humber, M.L.; Justice, C.O. The Collection 6 MODIS burned area mapping algorithm and product. *Remote Sensing of Environment* **2018**, *217*, 72–85. <https://doi.org/10.1016/j.rse.2018.08.005>.
70. Daily updates from Intensive observation campaign 2023 Autumn. Available online: <https://aakash-rihn.org/en/campaign2023/> (accessed on 4 August 2024).
71. Mangaraj1, P.; Matsumi, Y.; Nakayama, T.; Biswal, A.; Yamaji, K.; Araki, H.; Yasutomi, N.; Takigawa, M.; Patra, P.K.; Hayashida, S.; et al. Weak coupling of observed surface PM2.5 in Delhi-NCR with rice crop residue burning in Punjab and Haryana. under review.
72. Ministry of Environment, Forest and Climate Change. Paddy harvesting season 2023 comes to an end witnessing significant decrease in stubble burning with efforts made towards management of paddy straw for the current season. <https://pib.gov.in/PressReleaseIframePage.aspx?PRID=1981276>, 2023. Accessed 2024.
73. Wieland, M.; Li, Y.; Martinis, S. Multi-sensor cloud and cloud shadow segmentation with a convolutional neural network. *Remote Sensing of Environment* **2019**, *230*, 111203.
74. Campagnolo, M.; Libonati, R.; Rodrigues, J.; Pereira, J. A comprehensive characterization of MODIS daily burned area mapping accuracy across fire sizes in tropical savannas. *Remote Sensing of Environment* **2021**, *252*, 112115. <https://doi.org/10.1016/j.rse.2020.112115>.

**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.