

A novel deep Siamese framework for burned area mapping Leveraging mixture of experts

Seyd Teymoor Seydi^a, Mahdi Hasanlou^{a,*}, Jocelyn Chanussot^{b,c}

^a School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran

^b Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China

^c Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000, Grenoble, France



ARTICLE INFO

Keywords:

Wildfire mapping
Sentinel-2 imagery
Change detection
Siamese network
Mixture of experts

ABSTRACT

Due to the complexity of the areas and the diversity of the objects, traditional Burned Area Mapping (BAM) methods cannot provide promising results. Moreover, these methods focus on additional processing to improve their results, which is time-consuming and complex. Therefore, an advanced framework is needed to achieve accurate results in burned area mapping. In this context, this study proposes a novel Siamese-based mixture of expert networks for burned area mapping, which takes advantage of a mixture of experts (MoE) and position and channel attention mechanisms for deep feature generation. The proposed framework has two deep feature extractor channels to account for the bi-temporal multispectral pre- and post-fire input datasets. To evaluate the performance of the SMoE model, three multispectral Sentinel-2 were used in different countries. The results were compared with other advanced machine and deep learning models, including Light Gradient Boosting Machine (LGBM), Extreme Gradient Boosting (XGBoost), 2D-Siamese Convolutional Neural Network (2D-SCNN), and 3D-Siamese Convolutional Neural Network (3D-SCNN). The result of the burned mapping shows that the proposed model has a high effectiveness compared to other models, as it provides an average accuracy of more than 98% and 0.91 b y overall accuracy (OA) and kappa coefficient (KC) indices, respectively.

1. Introduction

Natural hazards such as wildfires play an important role in the evolution of ecosystems worldwide, affecting economies, global biochemical cycles, biodiversity, carbon sequestration, soil properties, water supplies, and human well-being (Fernandez-Carrillo et al., 2019; Sannigrahi et al., 2020). Wildfires can be caused by a variety of factors, including natural conditions such as topography, lightning, and climatic conditions, as well as human actions such as smoking and campfires (Febriandhika et al., 2018). In recent years, the world has experienced many wildfires, which have led to an increase in fire intensity, less water (drier summers), more greenhouse gas emissions, and the loss of many animal and plant species (Philipp and Levick, 2020; Tanase et al., 2010). In addition, wildfires produce various pollutants, including gases that contribute to global warming (Chowdhury et al., 2021).

BAM allows authorities to assess the extent of areas affected by fire, which is critical to understanding the impacts on ecosystems, biodiversity, and human settlements (Bar-Massada et al., 2014; Cochrane, 2009). BAM helps ensure optimal resource allocation and deployment

by providing accurate data on the size and location of burned areas (Griffith et al., 2017). BAM also helps monitor post-fire ecosystem recovery, enabling informed decisions about when and where to intervene to support vegetation and habitat regeneration (Bousquet et al., 2022). Historical BAM data help identify regions prone to frequent wildfires, informing land-use planning and risk-mitigation strategies to reduce community and ecosystem vulnerability (Meng et al., 2015). BAM also contributes to a better understanding of the role of wildfires in the carbon cycle and climate change by estimating carbon emissions from burned areas (Liu and Yang, 2020). The data guide the formulation of fire management policies and strategies. Insights into patterns of fire occurrence and severity are used for more effective prevention, preparedness, and response plans (Ager et al., 2015; Syphard et al., 2013).

The primary objectives of post-fire forest management are to determine the total area affected, the severity of damage, and acceptable measures for vegetation restoration (Elhag et al., 2020). Spatiotemporal mapping of burned areas is critical for assessing biological and ecological impacts, calculating economic losses, and, more recently, predicting the climatic effects of biomass burning (Han et al., 2022). Consequently,

* Corresponding author.

E-mail address: hasanlou@ut.ac.ir (M. Hasanlou).

fire management organizations and provincial governments need cost-effective and reliable fire mapping systems, especially as the frequency and severity of wildfires increase. It is therefore essential to identify burned areas accurately in space and time, for which satellite Earth observation has become increasingly popular in recent decades, especially at coarse spatial resolution. Globally, satellite imagery provides the most important information about the Earth at the lowest cost and in the shortest time. Because of their extensive coverage and excellent spatial and temporal resolution, satellite sensors are an excellent tool for generating accurate maps of burned areas (Abdikan et al., 2022; Zhang et al., 2021).

Satellite-based burned area mapping has long been a hot research topic, and satellite data are used to monitor fires by identifying actively burning fires and mapping the burned area (Palandjian et al., 2009). Most burned area products are produced using polar-orbiting satellites because they have better global coverage and temporal resolution (1–3 days). Recently, numerous global burned area products have been developed using Visible Infrared Imaging Radiometer Suite (VIIRS) and Moderate Resolution Imaging Spectroradiometer (MODIS) data with spatial resolutions of 375–500 (m) (Chuvieco et al., 2018; Lasko, 2021; Pinto et al., 2020). Despite their coarse resolution, these products provide daily coverage, allowing for near real-time applications and greater resilience in the presence of clouds. However, the coarse resolution is a limitation for applications that require small details of burned areas or to properly estimate the total burned area at the end of the season, a fundamental measure that can still show significant variations depending on the methods used (Bastarrika et al., 2011; Roteta et al., 2019). Furthermore, the accuracy of these products has been evaluated and many studies have reported an overall accuracy between 65 and 77% (Boschetti et al., 2010; Franquesa et al., 2022).

To obtain better resolution results, the BAM can be carried out based on high-resolution sensors (10–30 (m)), including synthetic aperture radar (SAR) (e.g. Sentinel-1, RCM), hyperspectral (e.g. PRISMA, AVIRIS) and multispectral (Landsat-9, Sentinel-2) sensors. The previously described systematic burned area products, several regional databases have consolidated burned area delineations from different sources to facilitate the monitoring of fire activities in the corresponding administrative regions.

SAR systems are less affected by weather conditions and their signals can penetrate clouds. In addition, SAR sensors can operate both day and night. These advantages of SAR sensors lead to the use of SAR data sets for BAM in many studies. For example, Lestari et al. (2021) evaluated the performance of Sentinel-1 SAR images for BAM using convolutional neural network (CNN), multi-layer perceptron (MLP), and random forest (RF) models. They extracted some parameters such as radar burn ratio, radar burn difference, and gray level co-occurrence matrix texture features as input for multi-layer perceptron and random forest models. In addition, Engelbrecht et al. (2017) consider a simple normalized difference procedure for BAM based on the C-band RADARSAT-2 dataset. For this purpose, they used a polarimetric decomposition to determine α -angles for pre- and post-fire datasets. Finally, the normalized difference α -angle is measured to detect vegetation changes between bi-temporal images (Seydi and Hasanlou, 2021). Seydi et al. (2022) proposed a deep learning-based BAM framework based on time-series Sentinel-1 imagery. They generated three coherence maps based on pre/pre-fire, pre/post-fire, and post/post-fire datasets. Next, the coherence maps are stacked and fed into the deep learning model for classification. Although SAR has been used effectively for BAM, it has several limitations, such as relatively low temporal resolution. Furthermore, the SAR sensors acquire the dataset with limited polarization channels (i.e. HH, VV, HV, VH), which causes the detection of the burned area change into a big challenge.

Hyperspectral imagery has a high content of spectral information, which allows it to detect similar objects with high complexity (Shah et al., 2021; Yadav et al., 2022). The advantages of hyperspectral imagery have motivated researchers to propose several burned area

Table 1

The BAM frameworks are based on different satellite imagery.

Dataset	Sensor	Method	Reference
SAR	Sentinel-1	CNN, RF, MLP	Lestari et al. (2021)
	RADARSAT-2	Polarimetric change detection	Engelbrecht et al. (2017)
	Sentinel-1	CNN	Seydi et al. (2022)
Hyperspectral	DESI	Analysis of spectral bands	Nolde et al. (2021)
	PRISMA	Deep learning	Seydi et al. (2021)
	AVIRIS	Analysis of spectral bands	van Gerrevink and Veraverbeke (2021)
Multispectral	Sentinel-2, Landsat-8	Thresholding of spectral indices	Achour et al. (2022)
	Sentinel-2	Multi-criteria approximate reasoning procedure	Sali et al. (2021)
	Landsat-8	RF	Gholamrezaie et al. (2022)
	PlanetScope	U-Net	Martins et al. (2022)
	Sentinel-2, Landsat-8	Logistic regression and Markov Random Field	Negri et al. (2022)

detection methods, for example, Nolde et al. (2021) have analyzed the suitability of red and near-infrared (NIR) spectral bands for BAM based on the DLR Earth Sensing Imaging Spectrometer (DESI) sensor. They evaluated 232 different band combinations in terms of accuracy and their results show that the spectral bands have a great influence on the detection of burned areas. Seydi et al. (2021) developed a deep learning-based burned area detection framework based on change detection through hyperspectral and multispectral datasets. They used the pre-fire Setinel-2 and post-fire hyperspectral datasets to detect the burned areas using a Deep Siamese Morphological Neural Network (DSMNN-Net). They reported that using a hyperspectral dataset instead of just a multispectral dataset improved the mapping result. van Gerrevink and Veraverbeke (2021) considered the spectral sensitivity of spectral bands for burn severity based on the hyperspectral dataset. In this regard, they analyzed different spectral bands of the hyperspectral dataset for a differenced normalized burn ratio index to identify the optimal bands for burn severity mapping. They used the AVIRIS hyperspectral sensor dataset with 224 spectral bands for burn severity mapping. Although these studies show the effeteness of hyperspectral bands in BAM, the hyperspectral datasets have several limitations: (1) hyperspectral satellite sensors have a poor temporal resolution, (2) the spectral information in the hyperspectral dataset is highly redundant leading to the demand for additional pre-processing (i.e. dimensional reduction), (3) the complexity and high dimensional of hyperspectral dataset cause to process this dataset for large scale area more challenging.

The multispectral satellite datasets are the most conventional source datasets (Table 1) that have been widely used in many applications (Rogan and Franklin, 2001). The free availability, global coverage, and good temporal and spatial resolution of these datasets led to the use of multispectral datasets for BAM in many studies. For example, Achour et al. (2022) compared the performance of Sentinel-2 and Landsat-8 sensors for BAM. First, several spectral indices were generated to predict burned areas, and then the final burned map was generated by thresholding these indices. They reported that the Sentinel-2 dataset was more efficient than Landsat-8 imagery. Sali et al. (2021) proposed an automatic burned area detection framework based on bi-temporal Sentinel-2 images and a multi-criteria approximate reasoning procedure. For this purpose, they used the region growth method and fuzzy set theory in the BAM framework. Gholamrezaie et al. (2022) proposed an automated BAM framework based on time-series Landsat-8 imagery. They combined spectral indices and original spectral bands to identify burned areas. The genetic algorithm and neural network were used to select informative features. Finally, the RF and neural network

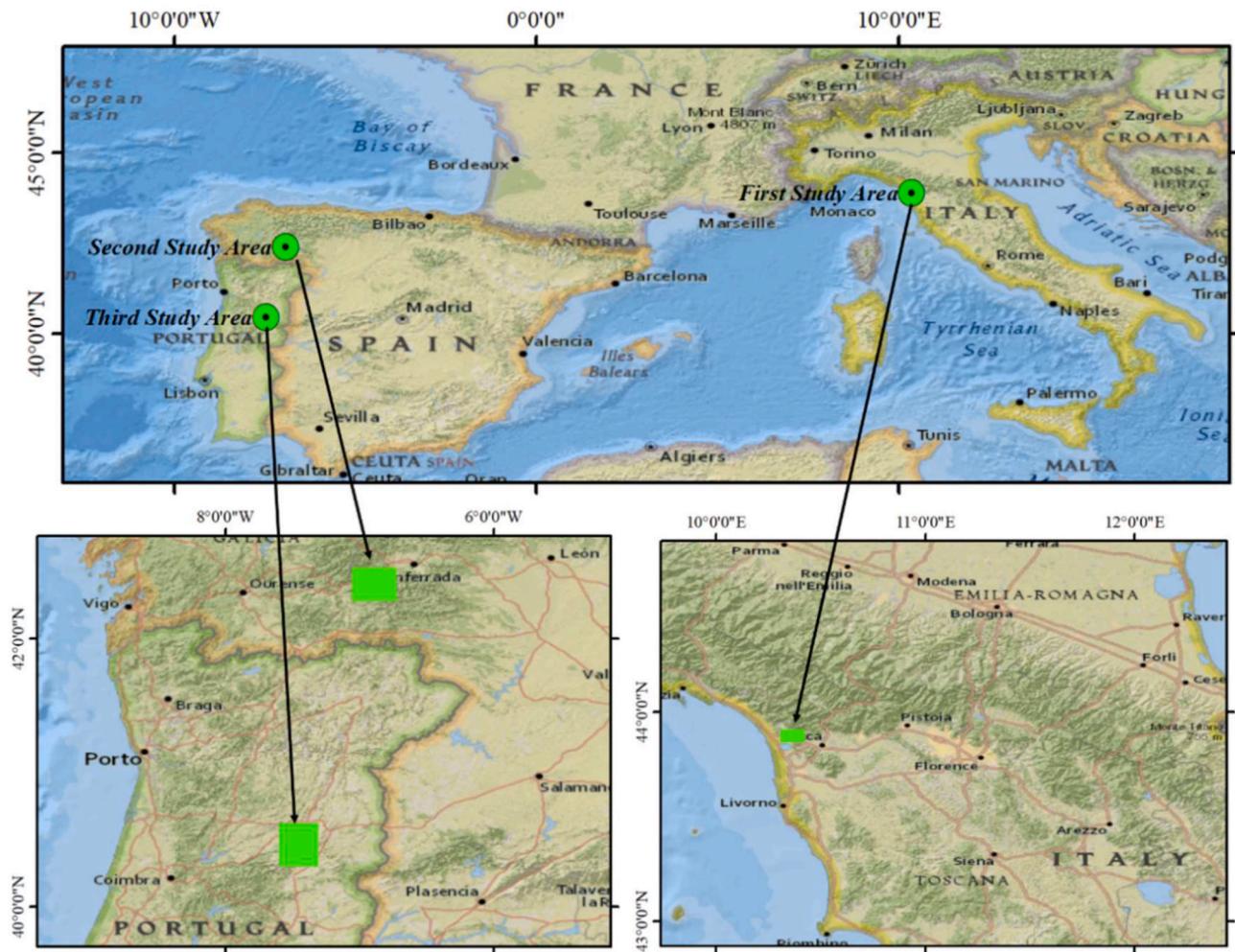


Fig. 1. The location of three case study areas in this paper. The first study area is located in Italy, the second study area is located in Spain, and the third study area is located in Portugal.

classifiers are used to classify the burned areas. Martins et al. (2022) developed a deep learning-based framework for burned area identification using PlanetScope and transfer learning. They trained the U-Net model based on the pre-existing Landsat-8 derived burned area reference dataset. The trained model is then used to predict burned areas using the PlanetScope dataset. Negri et al. (2022) presented a unified data-driven BAM framework using multi-temporal Landsat-8, Terra, and Sentinel-2 imagery. This framework is implemented in four steps: (1) multi-temporal normalized burn ratio calculation, (2) spatio-temporal representation for difference normalized burn ratio, (3) deviation analysis, (4) classification based on logistic regression and Markov Random Field.

The environments in which wildfires occur are often complex, with a wide variety of objects. This complexity makes it difficult to distinguish between burned and non-burned areas when a wildfire occurs. As a result, traditional methods for identifying burned areas are not very effective and do not provide a high level of accuracy. In general, many burn area detection frameworks have been developed by different studies, they generally have one of the following drawbacks: (1) Most frameworks use manual feature generation using spectral features (i.e. normalized burn ratio index) rather than taking advantage of spatial features. In addition, several methods have focused on thresholding the spectral of the indices, the selection of a global threshold does not support a large region, (2) There is a tendency to use conventional classifiers (e.g., support vector machines) when advanced classifiers (e.g., CNN) have proven to be more reliable, (3) The effectiveness of

machine learning models depends only on the input features, and the extraction of informative features is a challenge. In addition, (4) some models used different algorithms and additional processing (e.g., change detection or feature extraction) to produce a burned map, which is a time-consuming process.

The deep learning-based frameworks have provided robust results in many remote sensing applications (Gun and Chen, 2023) such as change detection (Ma et al., 2023), classification (Meher, 2020), flood mapping (Macharyulu et al., 2022), fire detection (Khan et al., 2022), and agricultural crop mapping (Saranya et al., 2023). The complexity of the background causes conventional deep learning models to fail to provide promising results. Furthermore, deep learning-based CNN frameworks have a fundamental limitation in that the entire network must be run for all input feature data (Wang et al., 2019). As a result, the size of the network is limited due to this computational burden (Eigen et al., 2014). By increasing the number of parameters and hidden units, but using only a portion of the network for each input, the computational cost can be kept low. After mapping the inputs to appropriate portions, the network can then be computed inexpensively. The MoE combines the output of multiple experts by training the gating network to produce a final output. The traditional structure of the MoE was introduced by (Jacobs et al., 1991), which greatly increases the capacity of the model (Chen et al., 2022).

The main objective of this paper is to propose a new deep learning-based model for burned area mapping based on bi-temporal Sentinel-2 datasets to minimize the above challenges. The main contributions of

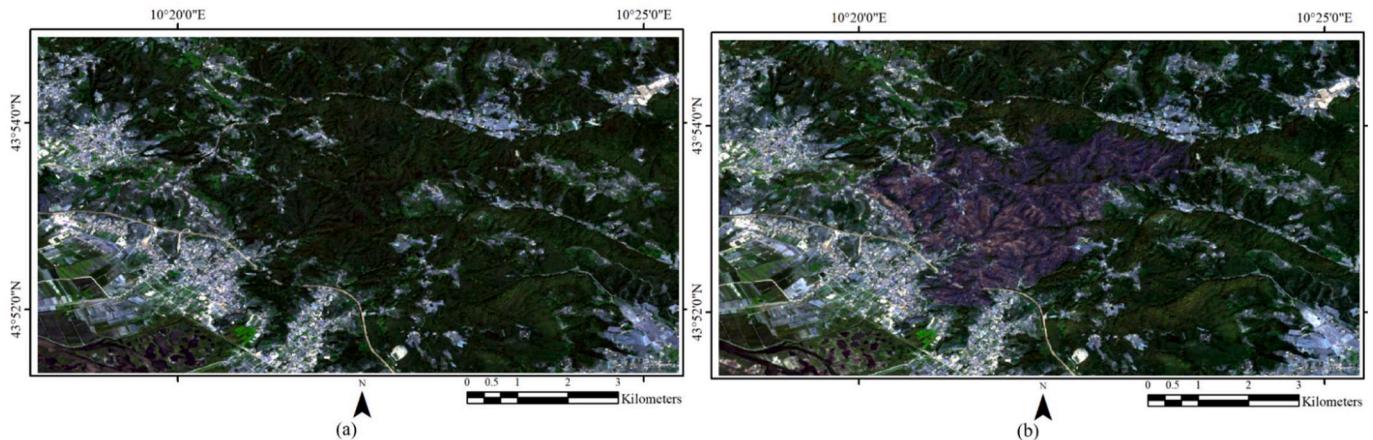


Fig. 2. The incorporated dataset for the first study area. (a) Pre-Fire, and (b) Post-Fire.

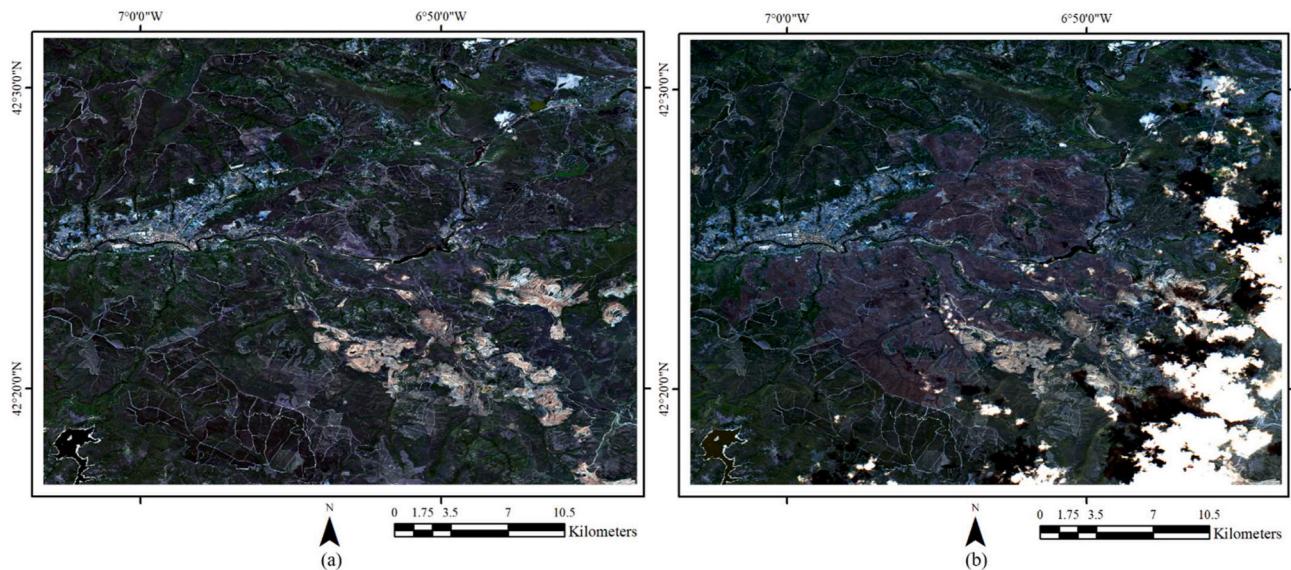


Fig. 3. The incorporated dataset in the second study area. (a) Pre-Fire, and (b) Post-Fire.

the current study are summarized as follows: (1) a novel BAM structure is proposed by using the MoE layer for the first time; (2) the SMoE takes advantage of the channel and position attention modules for BAM; (3) we compared the performance of the proposed architecture with other advanced machine and deep learning frameworks for BAM.

2. Study areas and datasets

Sentinel-2 satellite imagery, with a spatial resolution of up to 10 m and a revisit time of 5 days, takes into account Sentinel-2A and Sentinel-2B, the latter of which has been operational since early 2017 (Drusch et al., 2012). Sentinel-2's improved spatial resolution and better temporal coverage have made it an excellent option for mapping burned regions since 2017, as shown by several studies (Gonçalves et al., 2023; Saisawad and Ninsawat, 2023). Three real bi-temporal Sentinel-2 datasets were selected to evaluate the performance of the SMoE model, both quantitatively and qualitatively. Fig. 1 shows the geographical location of the study areas. The study areas are located in different countries (Spain, Italy, and Portugal) with complex background diversity. In this study, the Sentinel-2 dataset with spatial resolutions of 10 and 20 (m) was used. Thus, the input dataset for the model included 10 spectral bands, and other spectral bands with a spatial resolution of 60 (m) were removed.

2.1. The first study area

Since the afternoon of July 18, 2022, large fires have been burning in the forests surrounding the village of Massarosa, in the province of Lucca, Italy. The average humidity and temperature in this area are 12% and 30 °C respectively. This fire lasted for two days, as about 60 ha of forest and Mediterranean scrub were destroyed by the fire front, which is several kilometers long. Fig. 1 shows the geographical location of the first study area. Fig. 2 shows the pre- and post-fire Sentinel 2 data sets. This dataset covers the area with boundaries containing 1277 × 668 samples.

2.2. The second study area

The provinces of Lugo and Ourense (Galicia - NW Spain) were affected by major forest fires on July 14, 2022 and July 15, 2022. Folgoso do Courel, O Barco de Valdeorras, and Carballeda de Valdeorras are the main areas affected by these fires. The fires not only last for a few days but also affect populated areas due to unfavorable weather conditions. The average humidity and temperature in this study area are 1% and 27.5 °C, respectively. As can be seen, Fig. 3 illustrates the incorporated data sets for this study area, which covers 3670 × 2755 samples.

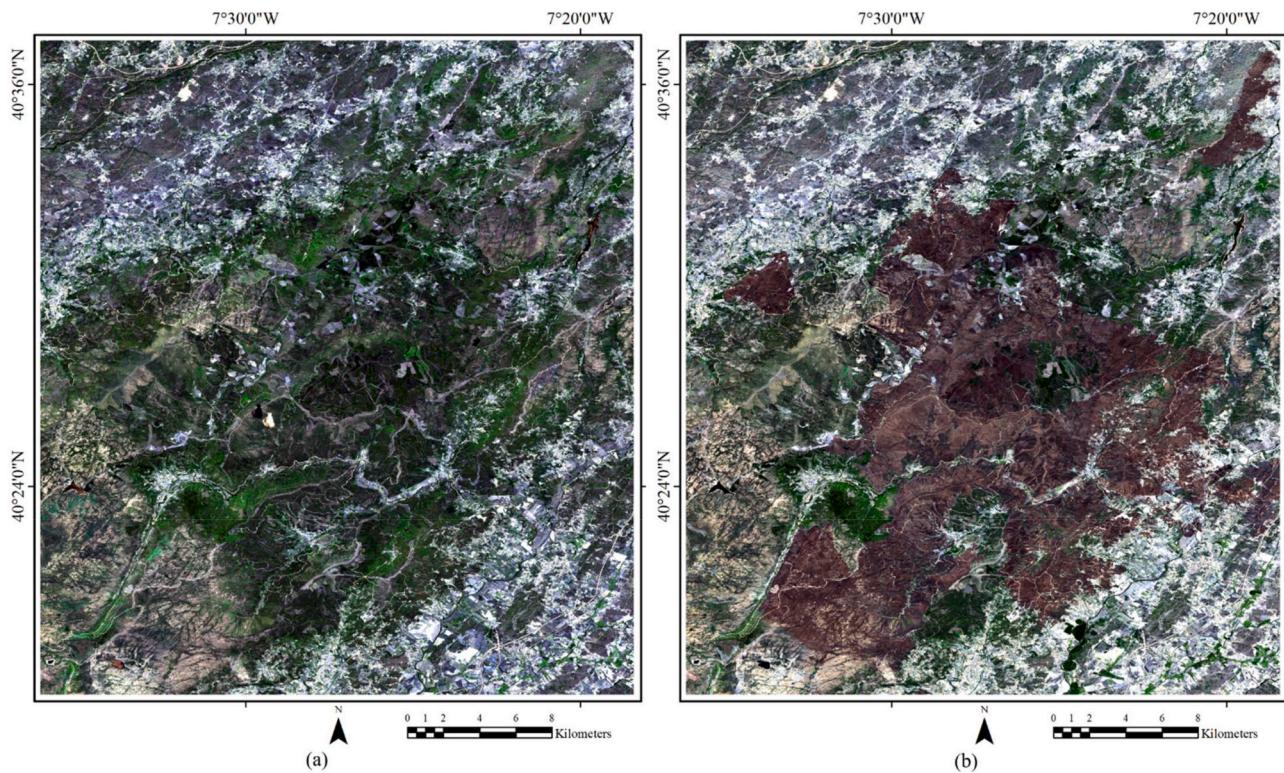


Fig. 4. The incorporated dataset in the third study area. (a) Pre-Fire, and (b) Post-Fire.

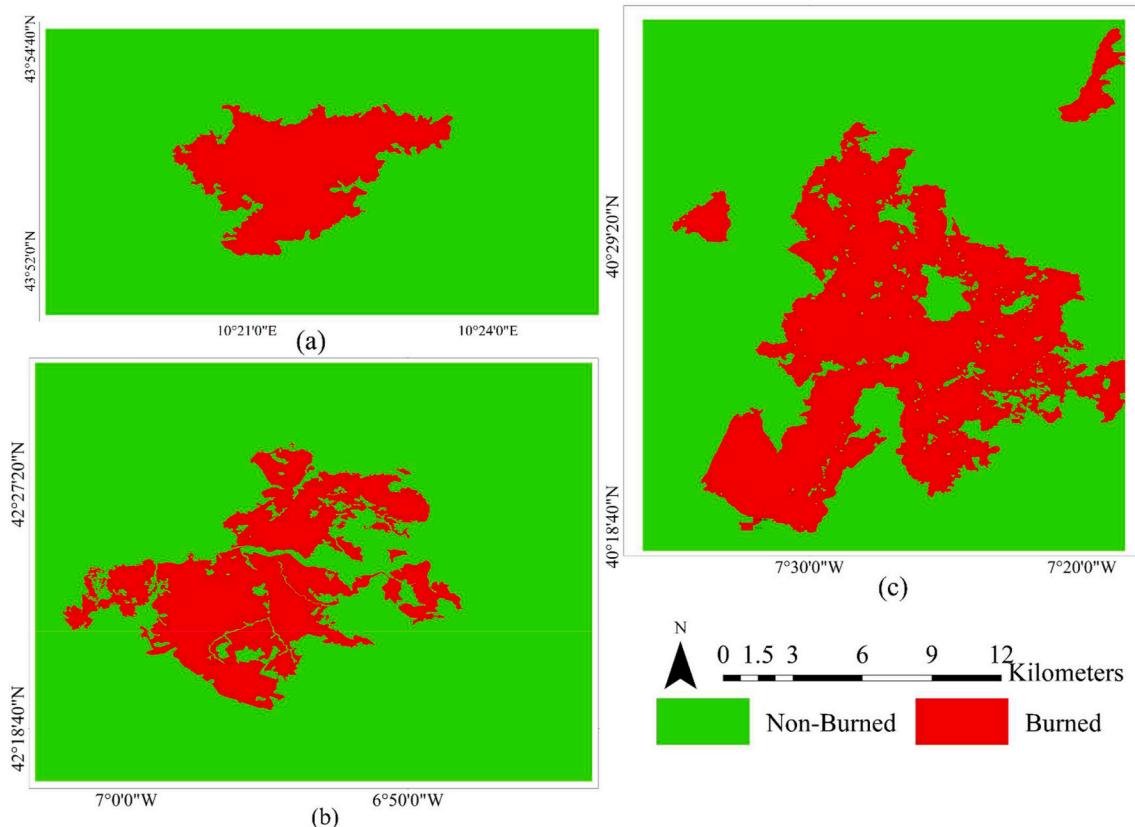
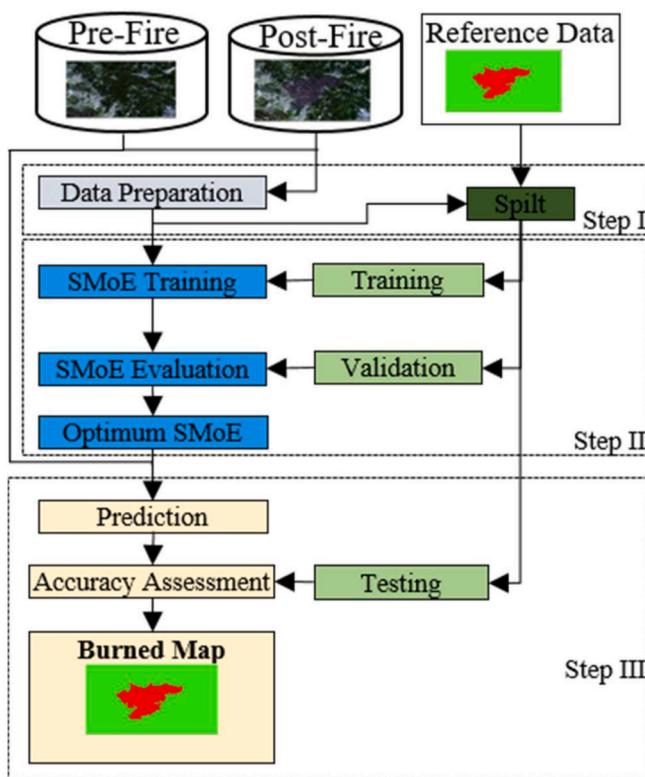


Fig. 5. The data inventory for three study areas. (a) Reference map for the first dataset, (b) Reference map for the second dataset, and (c) Reference map for the third dataset.

Table 2

The statistical details of the reference data samples.

Study Area	Class	All Samples	Training	Validation	Test
The First Study Area	Non-Burned	734220	3121	550	730549
	Burned	118815	505	89	118221
	Total	853036	3626	639	848770
The Second Study Area	Non-Burned	8561980	36387	6422	8519171
	Burned	1548870	6583	1162	1541125
	Total	10110850	42970	7584	10060296
The Third Study Area	Non-Burned	8441192	32613	5756	8402823
	Burned	3474520	13425	2369	3458726
	Total	11915712	46038	8125	11861549

**Fig. 6.** Workflow of the general framework for BAM using the SMoE model.

2.3. The third study area

On August 6, 2022, a forest fire started in the Serra de Estrela Natural Park in Covilhá, Portugal. In this study area, the sky is clear and the average humidity and temperature are 1% and 29.5 °C, respectively. During the three burning days, the fire burned more than 4000 ha of cork oak and pine forest. This dataset contains 3288×3624 samples, of which 8441192 and 3474520 samples belong to the non-burned and burned classes, respectively (Fig. 4).

2.4. Data inventory

The reliability of the reference map is crucial for the comparison of BAM's results. For this purpose, three reference datasets were used in this study, the ground truth of which is available at [<https://emergency.copernicus.eu/>]. The reference map of the three study areas is shown in Fig. 5. In addition, 0.5% of the samples from the reference maps of all

datasets were selected as training sets. It is worth noting that the collection of sample data sets is difficult and time-consuming. In fact, a large amount of sample data set can't be accessed in the practical application. Therefore, the use of a 0.5% sample data set is the simulation of the real condition in practical application. Table 2 shows the details of the sample data sets included for all the study areas.

3. Methodology

A high-level overview of the proposed burn area detection method is shown in Fig. 6. Based on this, the BAM is applied by the proposed framework through three main steps: (1) data preparation, the pre-/post-fire datasets are overlapped with reference datasets and then patched by window size 11×11 with a shift of one pixel, (2) model training, the model parameters are tuned by an iterative manner based on learning using the training dataset and evaluating the model by loss function on the validation dataset, and (3) prediction, the dataset is fed to the predictive model to generate the final burned map, also, the performance model is evaluated based on the test dataset by quantitative measurement indices and visual analysis.

3.1. Deep SMoE

In this study, the proposed SMoE network uses a bi-temporal dataset for BAM. However, some models stack bi-temporal datasets and generate the BAM based on a single deep feature stream, while the proposed model deep feature is based on a double stream channel to improve the effectiveness of the model. On the other hand, the limitation of detection of burned areas has been mentioned above. Therefore, it is necessary to design a model that minimizes the above challenges and has high effectiveness. To this end, this study presents a novel double-stream framework for BAM. The general structure of the proposed SMoE architecture is shown in Fig. 7. As shown, the proposed SMoE framework provides two bi-temporal patch datasets. Then, these patch datasets are fed into the deep feature extractor part by MoE, position attention module, channel attention module, max-pooling, and global average pooling layers separately. The extracted deep features are transferred to the classification head and concatenated. The dense MoE layer is used to represent deep features with more abstraction. Finally, the softmax layer is used to decide on the input patch datasets.

The MoE layer has two main components, which include a set of experts (convolution layer) and a gating network, which is typically linear (Shazeer et al., 2017). For the experts to be complementary, each must learn a distribution that is specialized to a stochastic subspace of the entire problem space (Fu et al., 2018; Jacobs et al., 1991). Gating networks learn weights for experts based on their local efficiency. The main structure of the MoE layer is shown in Fig. 8. Since the same input is fed to a set of experts, different solutions are generated for the entire problem space. In addition, the gating networks receive the same input, and the weight parameters are often normalized.

The weights in the gate network are dynamically determined by the inputs. The output of the MoE layer for input χ with expert networks with weight parameters and a gating network with parameter \hat{W} can be expressed as follows (Eq. 0.1) (Chen et al., 2022):

$$\mathcal{H}(\chi; \hat{W}, W) = \sum_{n \in \Psi_\chi} \mathcal{F}_n(\chi; \hat{W}) h_n(\chi; W) \quad (1)$$

where $\Psi_\chi \in N$ is a set of selected indices, h_n denotes the output of the n^{th} expert network and $\mathcal{F}_n(\chi; \hat{W})$'s are route gate values given by (Eq. 0.2) (Chen et al., 2022).

$$\mathcal{F}_n(\chi; \hat{W}) = \frac{e^{\varphi_n(\chi; \hat{W})}}{\sum_n e^{\varphi_n(\chi; \hat{W})}}, \forall n \in [N] \quad (2)$$

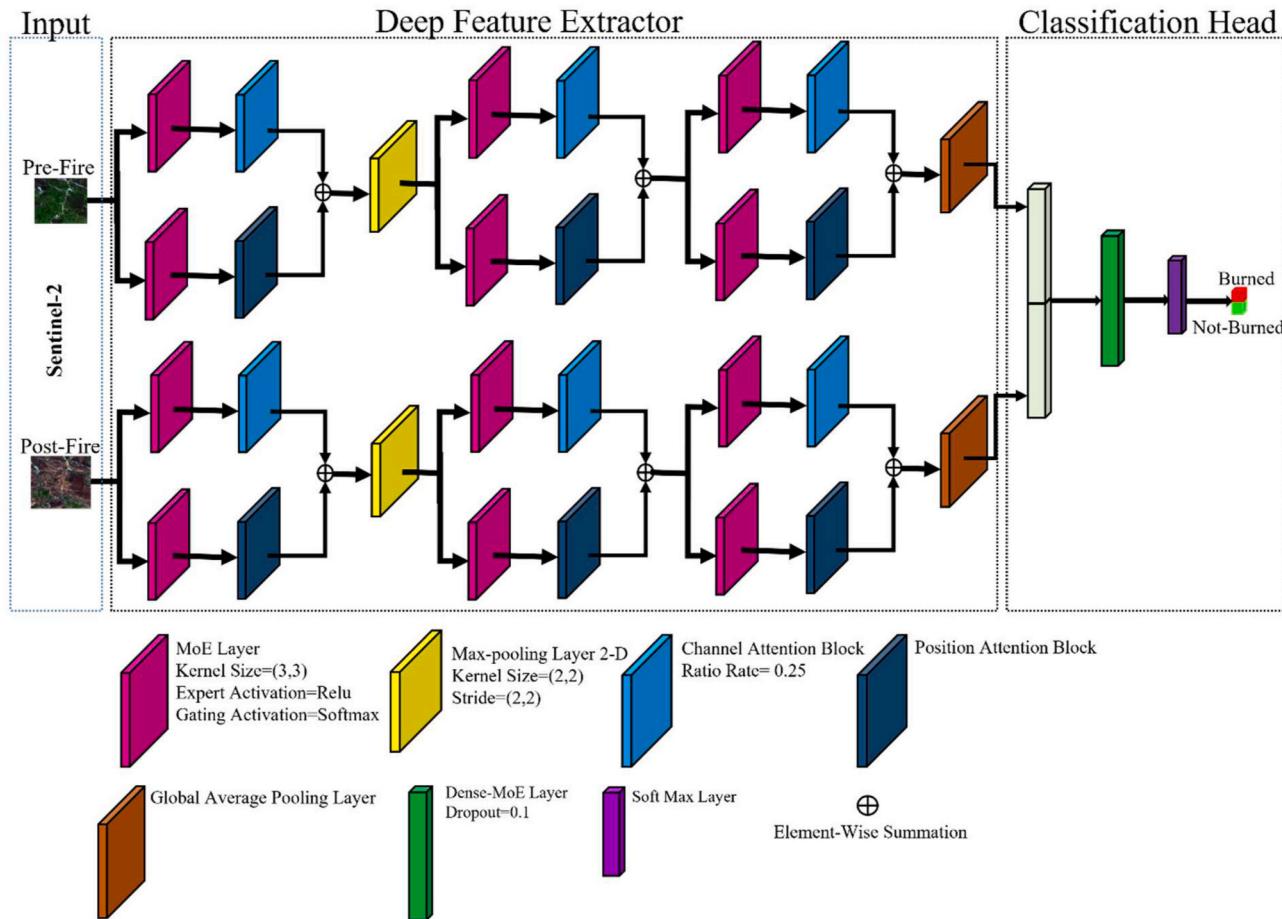


Fig. 7. Structure of the proposed SMoE model.

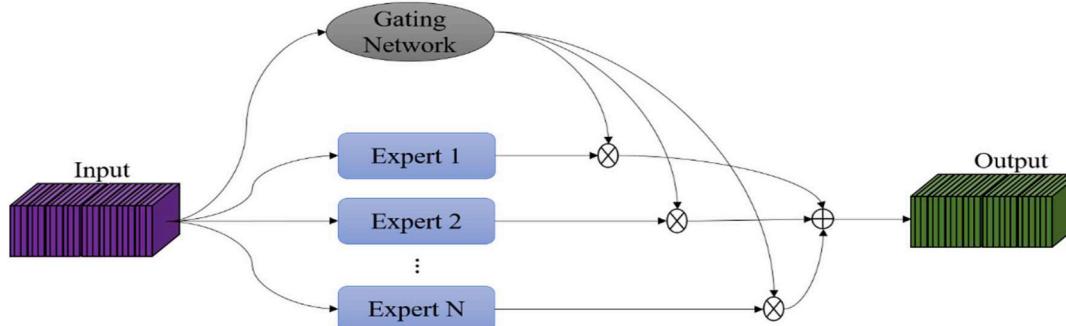


Fig. 8. The general framework of an MoE layer.

where $\varphi(\chi; \hat{W}) = \sum \hat{W}^T \chi$ is the output of getting the network. Furthermore, the h_n can be considered as a convolutional neural network as follows (Eq 0.3) (Chen et al., 2022).

$$h_n(\chi; W) = \sum_{k \in K} \sum g(\langle W_{n,k}, \chi \rangle) \quad (3)$$

where K is the number of filters for n^{th} expert network, g is a nonlinear activation function.

The position attention module introduced by Fu et al. (2018) enables rich contextual relationships over local features. This module encodes a wider range of contextual information into local features, improving their representation. Fig. 9 shows the structure of the position attention module.

The input feature data ($\varphi \in R^{\gamma \times \alpha \times \beta}$) is fed into a convolution layer to extract three new feature maps ($\varphi_1, \varphi_2, \varphi_3 \in R^{\gamma \times \alpha \times \beta}$). Next, these new feature maps are reshaped to a new dimension $\gamma \times N$, where N is the number of pixels ($N = \alpha \times \beta$). Then, matrix multiplication is applied between the transpositions of φ_1 and φ_2 , and a softmax layer is performed to compute the spatial attention map (ψ) (Eq 0.4) (Fu et al., 2018).

$$\psi_{i,j} = \frac{e^{(\varphi_{1i}, \varphi_{2j})}}{\sum_{i=1}^N e^{(\varphi_{1i}, \varphi_{2j})}} \quad (4)$$

Then, a matrix multiplication is applied between the transpose of ψ and φ_3 and is reshaped the result to $R^{\gamma \times \alpha \times \beta}$. Finally, the final output

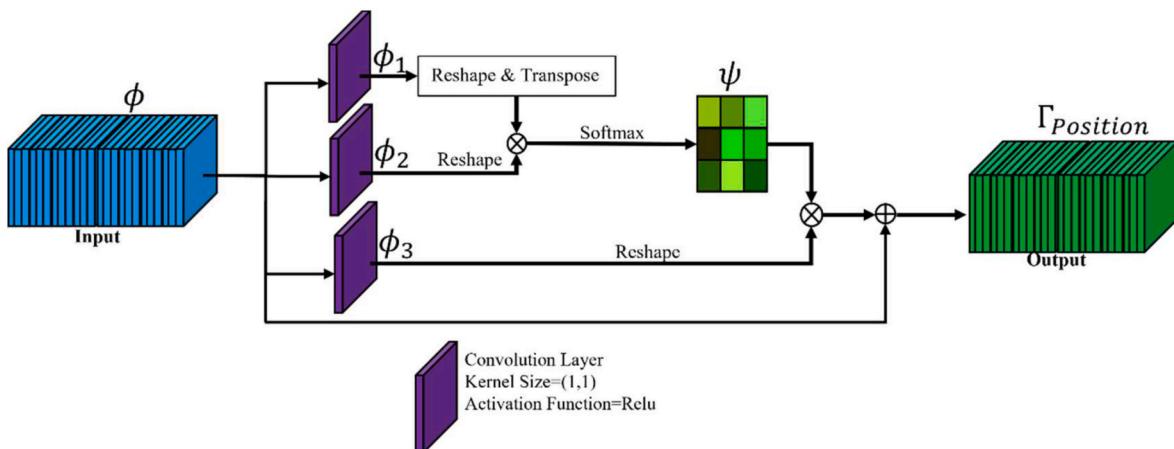


Fig. 9. The general framework of a Position Attention Module.

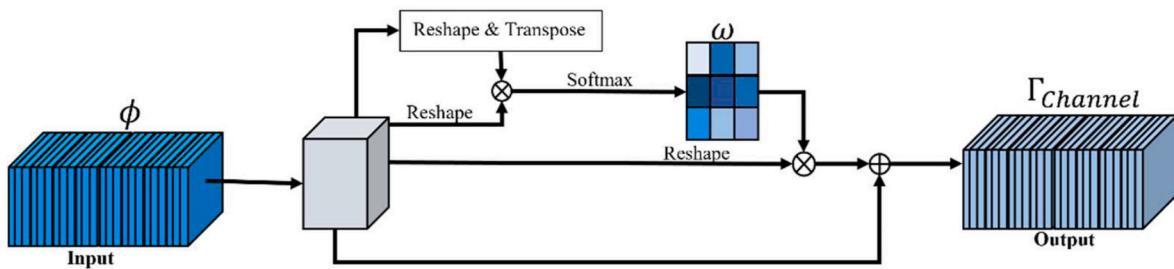


Fig. 10. The general framework of a Channel Attention Module.

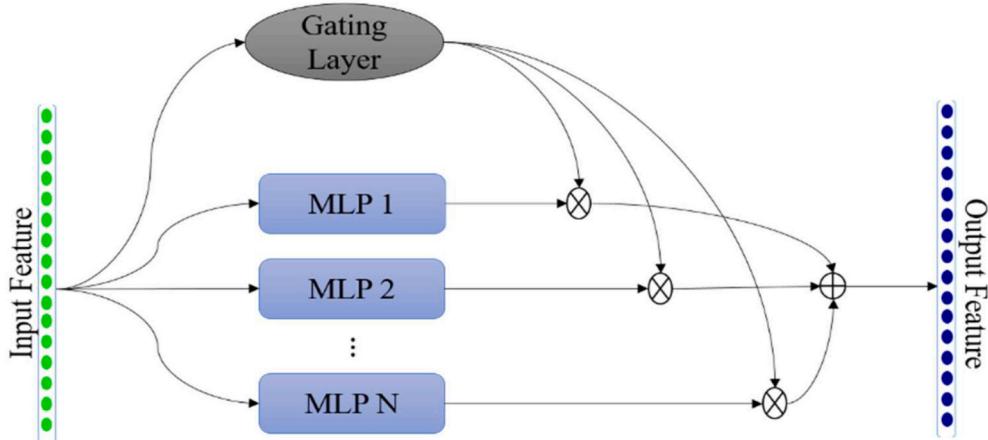


Fig. 11. Overview of Dense-MoE layer.

$(\Gamma_{Position})$ is obtained by multiplying a scale parameter (τ), and then element-wise summing with φ (Eq 0.5) (Fu et al., 2018).

$$\Gamma_{Position_j} = \tau \sum_{i=1}^N (\psi_{ij} \varphi_{3i}) + \varphi_j \quad (5)$$

Utilizing interdependencies between channel maps allowed us to emphasize interdependent feature maps and improve semantic feature representation (Fu et al., 2018). To this end, a channel attention module is built to model channel interdependencies explicitly. The mechanism of the channel attention module is close to the position attention module (Fig. 10). The main difference is that the input feature is directly employed to generate the channel attention map ($\omega \in R^{r \times r}$) (Eq 0.6). Furthermore, the scale parameters (ρ) are employed to calculate the

final output $\Gamma_{Channel}$ (Eq 0.7). To this end, the input feature dataset is reshaped to a new dimension $\gamma \times N$ (Fu et al., 2018).

$$\omega_{ij} = \frac{e^{(\varphi_i \varphi_j)}}{\sum_{i=1}^{\gamma} e^{(\varphi_i \varphi_j)}} \quad (6)$$

$$\Gamma_{Channel_j} = \rho \sum_{i=1}^{\gamma} (\omega_{ij} \varphi_i) + \varphi_j \quad (7)$$

The dense-MoE layer is employed in the final part of the proposed SMoE model which is built by several MLP layers and gating layers. The general structure of the dense-MoE layer is shown in Fig. 11. As can be seen, the dense-MoE layer is similar to the MoE layer but it uses MLP

Table 3

The optimum values of the hyper-parameters of different models.

Model	Parameters
XGboost	The number of estimators is 105, the learning rate is 0.1, the maximum depth is 20, and the subsample ratio is 0.7.
LGBM	The number of estimators is 150, the learning rate is 0.1, the regularization parameter is 0.9, the number of leaves is 150, and the maximum depth is 9.
Deep Learning	patch-size 15×15 , weight-initializer He-Normal, learning rate 0.001, dropout rate 0.1, batch-size 355, number of epochs 150, the number of neurons in the hidden layer 256, shuffle, and loss function binary-cross-entropy.

layers instead of convolution layers. In addition, these layers receive the flattened deep features from the MoE layers as an input dataset.

3.2. Contribution of study

This research presents an innovative deep-learning framework using bi-temporal Sentinel-2 imagery. The primary essence of our work can be encapsulated in the following key contributions: (1) Proposing a novel double-stream framework for informative deep feature generation, (2) Leveraging the MoE in burned area mapping for the first time, (3) Combining MoE layers and position and channel attention mechanisms for BAM for the first time, and (4) Analysing the sensitivity of hyper-parameters of our framework in BAM.

3.3. Accuracy assessment

The evaluation of the accuracy is carried out on the basis of a comparison of reference data with the output of the model. This comparison can be done based on visual interpretation and quality measurement indices. For this purpose, this research employs seven quality

measurement indices (Overall Accuracy (OA), F1-Score, Precision, Kappa Coefficient (KC), Precision (PCC), Recall, Intersection over Union (IOU), and Balance Accuracy (BA)), which are estimated from the confusion matrix. In addition, the normalized confusion matrix is used to effectively consider the performance of the models. To evaluate the performance of the SMoE model, we implemented six other deep learning and advanced machine learning models. These models include extreme gradient boosting (XGBoost), light gradient boosting machine (LGBM), 2D-Siamse convolutional neural network (2D-SCNN), 3D-Siamse convolutional neural network (3D-SCNN), multi-layer perceptron mixer (MLP-Mixer), and Hierarchical Vision Transformer using Shifted Windows (Swin-T). We stacked the pre- and post-fire bi-temporal data sets as input for four LGBM, XGBoost, MLP-Mixer, and Swin-T models. Moreover, both deep learning models have a similar architecture of SMoE, but they use only convolutional layers instead of MoE layer and attention mechanism modules.

4. Experiments and results

4.1. Parameter setting

It is important to determine the optimal value of the tuning parameters of the proposed method and other classification methods to achieve the highest possible classification accuracy. As shown in Table 3, the tuning parameters for each classifier and their optimal values were determined by trial and error. The deep learning models were trained on a PC equipped with an Intel i7 CPU, 32 GB RAM, and an NVIDIA GTX GPU.

4.2. Results

This section examines the results of burned area mapping for three

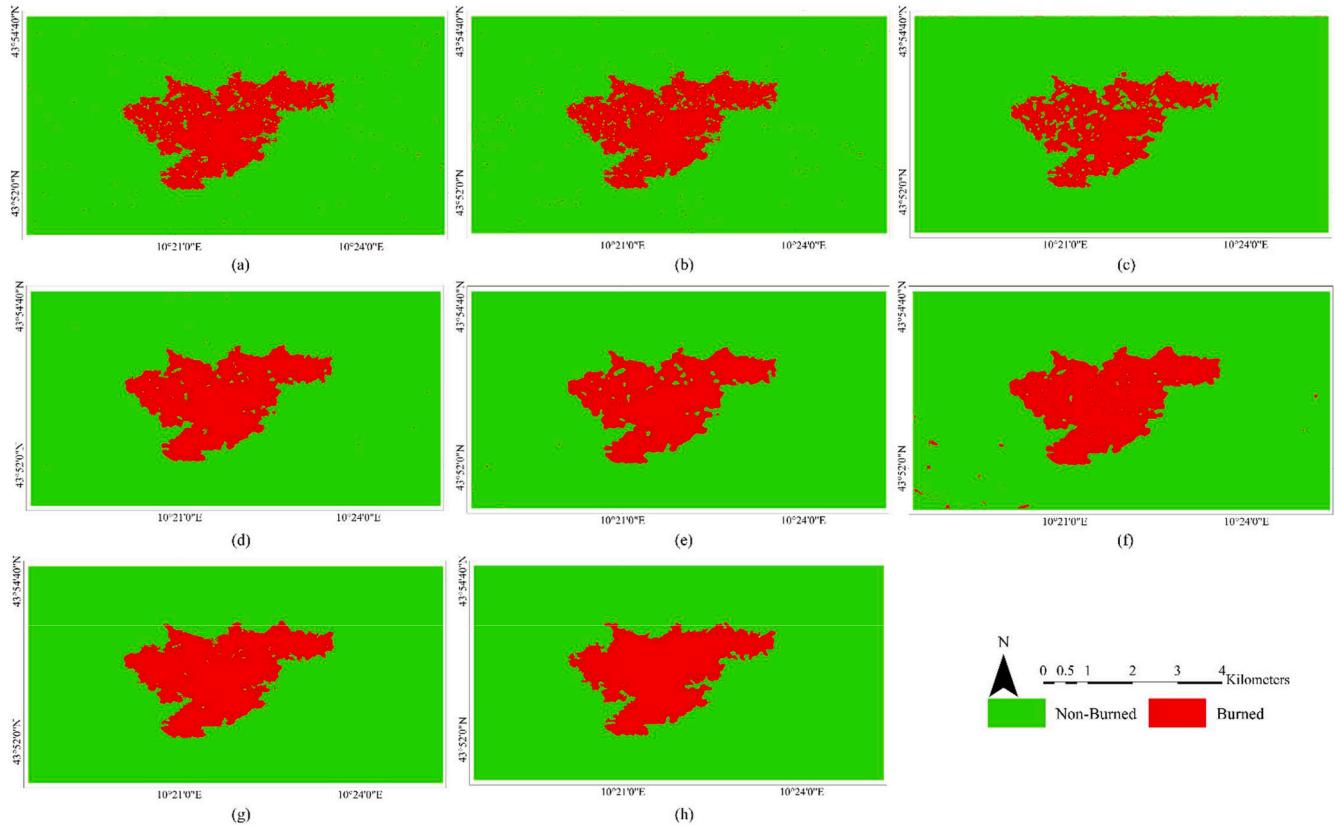


Fig. 12. Visual comparison of BAM results in the first study area: (a) XGBoost, (b) LGBM, (c) 2D-SCNN, (d) 3D-SCNN, (e) MLP-Mixer, (f) Swin-T, (g) SMoE, and (h) ground truth map.

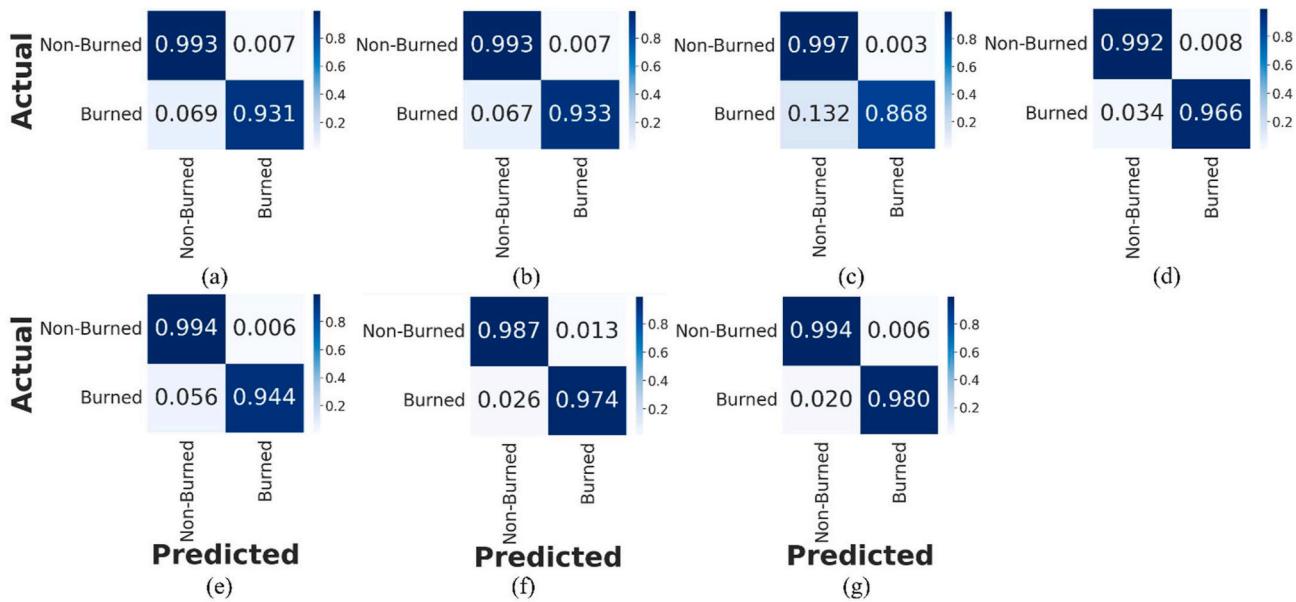


Fig. 13. Comparison of confusion matrix all methods: (a) XGBoost, (b) LGBM, (c) 2D-SCNN, (d) 3D-SCNN, (e) MLP-Mixer, (f) Swin-T, and (g) SMoE.

Table 4

A comparison between the BAM accuracies obtained by each model on the first dataset.

Index	XGBoost	LGBM	2D-SCNN	3D-SCNN	MLP-Mixer	Swin-T	SMoE
OA (%)	98.50	98.50	97.90	98.85	98.70	98.51	99.18
Precision (%)	95.77	95.60	97.59	95.04	96.04	92.16	96.09
Recall (%)	93.14	93.34	86.78	96.59	94.36	97.36	97.97
F1-Score (%)	94.43	94.46	91.87	95.81	95.20	94.69	97.02
BA (%)	96.24	96.33	93.22	97.90	96.87	98.03	98.67
KC	0.936	0.936	0.907	0.951	0.944	0.938	0.965
IOU	0.895	0.895	0.850	0.920	0.908	0.899	0.942

*The best accuracy is marked in bold.

study areas. All models are trained and tested separately for each study area. It is worth noting that the accuracy evaluation of the burned area mapping results is applied to the test dataset.

4.2.1. The first study area

The BAM results for the first study area are shown in Fig. 12. Visually, the burned mapping results appear to be relatively accurate in detecting non-burned and burned classes. Fig. 12-a-b shows the results of the XGboost and LGBM models, which can be seen in some noisy labeled pixels. In addition, there are many mis-detected pixels by 2D-SCNN and MLP-Mixer results (Fig. 12-c and 12-e). The Swin-T has some false detection pixels in some areas (Fig. 12-f). The proposed method has provided the best performance in detecting both non-burned and burned classes compared to other methods.

As shown in Fig. 13, the confusion matrices represent the overall performance of the burned mapping models. As can be seen, all models have provided an accuracy of more than 99.2% for the non-burned class, which is a considerable result. The performance models for the burned class are different as they cover an accuracy between 86 and 98%. In addition, the non-diagonal elements of the normalized confusion matrix show the Miss Detection (MD) and False Alarm (FA) rates. The FA rates of the models are between 0.3 and 0.8%, while the MD rates cover 2–13%. It is clear that the proposed model has outperformed other models in both classes. However, the 2D-SCNN model outperformed the proposed model and other models in the non-burned class, it misses

performances in detecting burned areas. As it has provided a miss-detection rate of more than 13%. The proposed model has achieved robust results in both classes as it has provided an accuracy of more than 98% and an error rate of less than 2%.

The numerical results of the burned mapping models are shown in Table 4. The quantitative results confirmed the results provided by the normalized confusion matrix as; our model outperformed other models by metrics of OA, Recall, F1-Score, BA, KC, and IOU. The SMoE model has provided an accuracy of more than 99.18% and 0.942 by OA and IOU indices. Furthermore, this performance is better than 0.68%, 0.68%, 1.28%, 0.33%, 0.48%, and 0.68% of XGBoost, LGBM, 2D-SCNN, 3D-SCNN, MLP-Mixer, and Swin-T by OA index, respectively. It is worth noting that 2D-SCNN has provided better performance than SMoE models, but has missed the effectiveness in the burned mapping by other metrics.

4.2.2. The second study area

Fig. 14 shows the visual results of the burned area map obtained from different machine and deep learning models. The analysis of the burned area maps (Fig. 14) showed that two machine learning models (LGBM and XGBoost) provided some noisy labeled pixels (Fig. 14-a-b). However, the deep learning models have provided promising results, the proposed model has presented an adequate performance for this dataset in comparison with other models (Fig. 14-e).

The normalized confusion matrices in Fig. 15 show a comparison of normalized confusion matrices for different BAM models. Most of the models have provided significant results in the non-burned class, as the accuracy in the non-burned class is more than 98%. In the burned class, the models have significantly different performance (88%–93%). Among these models, the SMoE has provided the best performance compared to other models (Fig. 15-g) in the detection burned class.

The quantitative evaluation of the BAM models is shown in Table 5. The deep learning-based models have provided better accuracy than the machine learning model, as the accuracy of machine learning models is less than 97%, while deep learning methods are more than 97% by OA index. Based on the provided results, the SMoE has achieved the highest accuracy compared to other models. However, 2D-SCNN and Swin-T have provided better performance than SMoE by OA and Precision but have missed their performance by other metrics (i.e. Precision and IOU).

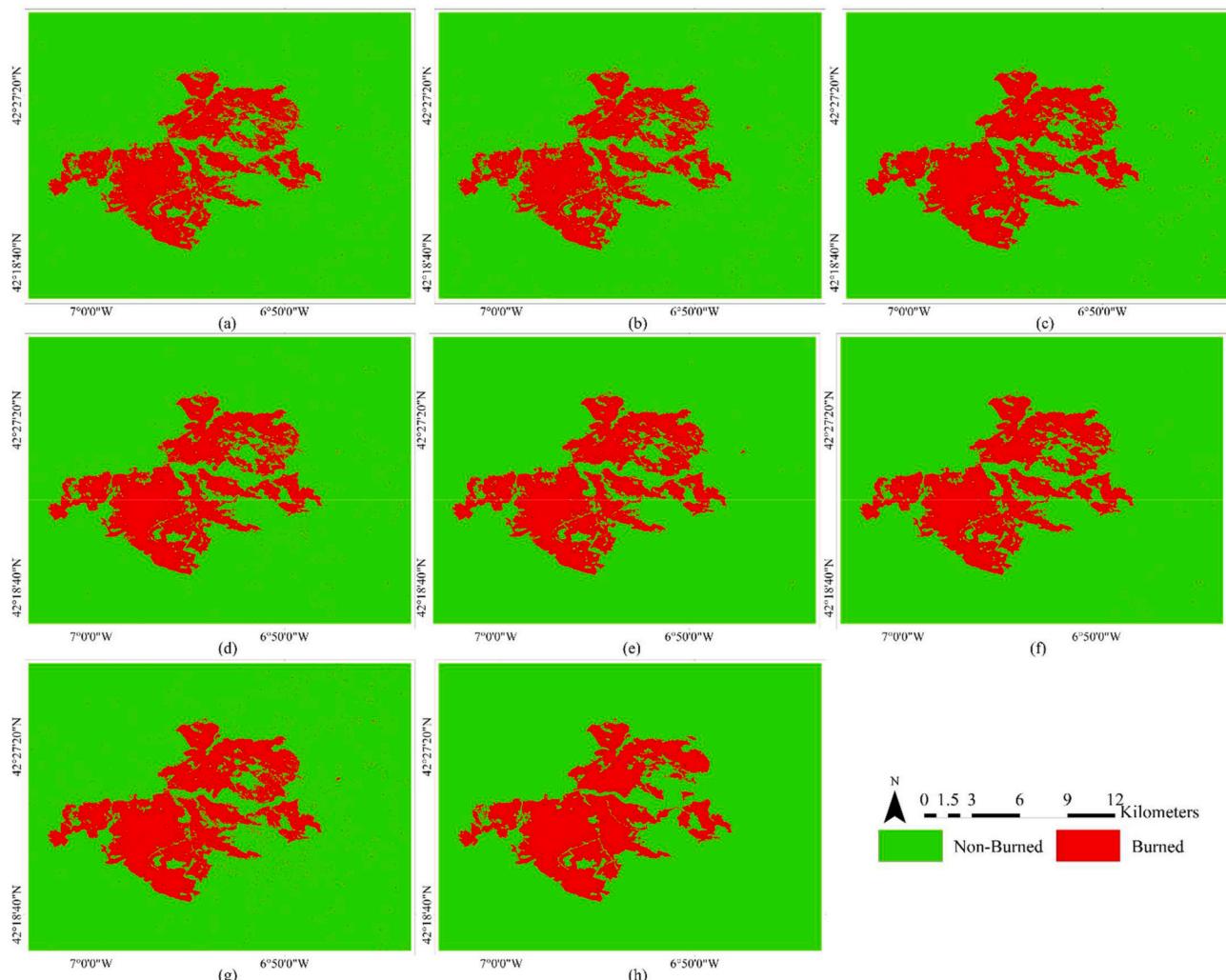


Fig. 14. Visual comparison of BAM results in the second study area: (a) XGBoost, (b) LGBM, (c) 2D-SCNN, (d) 3D-SCNN, (e) MLP-Mixer, (f) Swin-T, (g) SMoE, and (h) ground truth map.

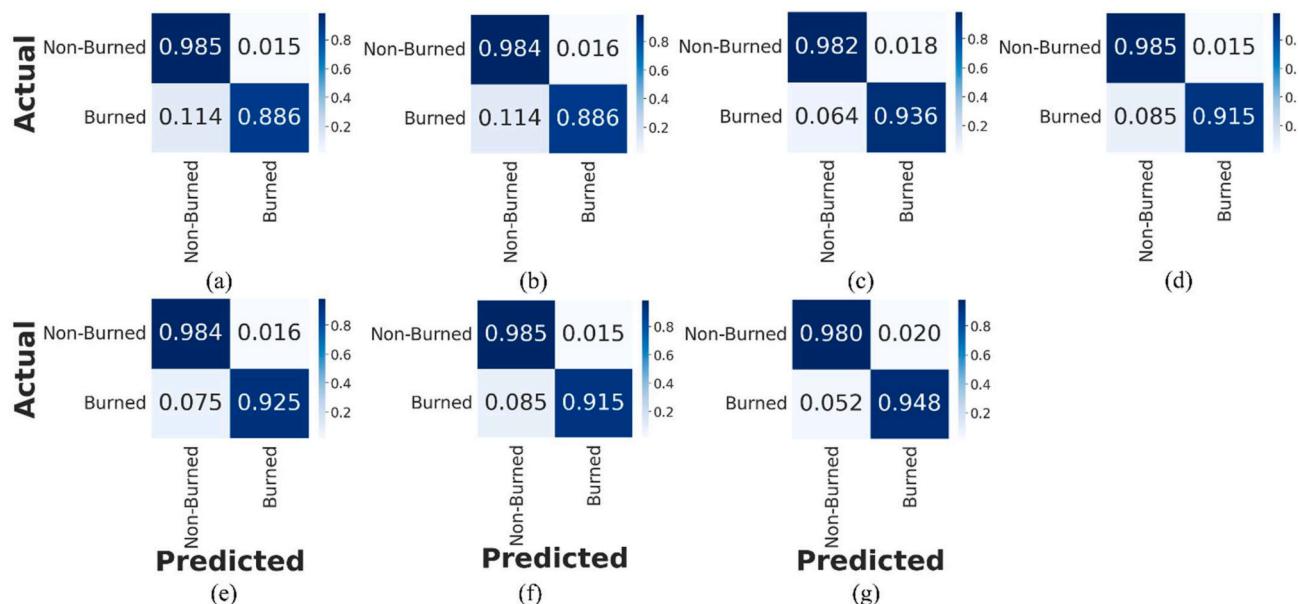


Fig. 15. Comparison of the confusion matrix of all methods for the second study area: (a) XGBoost, (b) LGBM, (c) 2D-SCNN, (d) 3D-SCNN, and (e) MLP-Mixer, (f) Swin-T, and (g) SmoE.

Table 5

The BAM accuracies of different methods in the second study area.

Index	XGBoost	LGBM	2D-SCNN	3D-SCNN	MLP-Mixer	Swin-T	SMoE
OA (%)	96.99	96.94	97.51	97.43	97.46	97.43	97.49
Precision (%)	91.46	91.12	90.47	91.73	91.10	91.71	89.50
Recall (%)	88.60	88.65	93.60	91.46	92.48	91.48	94.76
F1-Score (%)	90.01	89.87	92.01	91.60	91.79	91.59	92.05
BA (%)	93.55	93.54	95.91	94.98	95.42	94.99	96.37
KC	0.882	0.881	0.905	0.901	0.903	0.901	0.906
IOU	0.818	0.816	0.852	0.845	0.848	0.845	0.853

*The best accuracy is marked in bold.

4.2.3. The third study area

Fig. 16 illustrates the visual burned mapping plots obtained using different models on the third dataset. As shown in Fig. 16, the burned mapping result obtained with the SMoE model was clearer and closer to the reference map than those obtained with the other methods. Based on the comparison of the proposed SMoE approach with other methods, it was found that most of the samples could be accurately predicted by the SMoE framework.

The effectiveness of the burned detection models is represented by the normalized confusion matrices in Fig. 17. As can be seen, the performance of all models in this dataset is good, as those that have provided an accuracy of more than 99% and 96% in non-burned and burned classes, respectively. In general, the proposed model has provided the accuracy and error rate in both classes compared to other models. As it has provided an accuracy of 99% and an error detection rate below 0.004 for burned and non-burned classes.

The numerical experiment of the different models for the third study area can be observed in Table 6. As can be seen, the proposed framework performed better on the burned mapping in terms of several indices, with the highest OA equal to 99.42%. According to the LGBM model, the lowest OA value was reported at 98.61%. Furthermore, the effectiveness of the SMoE model was confirmed by other indices limited that outperformed other models (i.e. F1-Score, IOU, and KC).

4.3. Sensitivity analysis

The deep learning models have several hyperparameters that need to be set. To this end, the sensitivity of the hyperparameters of the SMoE in mapping the burned area was analyzed. As can be seen, Fig. 18 shows the effect of the number of mixture experts on the performance of the SMoE for the first dataset. Based on this figure, the number of mixture experts significantly influenced the result of BAM. The lower value of the higher value of the mixture of experts reduces the performance of the model. The best value of a mixture of experts is 10. The high value of a mixture of experts significantly increased the parameters of the model and resulted in overfitting the model. On the contrary, the low value of this parameter results in the SMoE model not capturing sufficient BAM patterns.

The loss function plays a key role in the training of deep learning models. In this regard, the impact of three of the most commonly used loss functions in BAM was compared. These three loss functions are Kullback-Leibler divergence, Poisson, and binary cross entropy. The results of the BAM based on the different loss functions show that there are some differences between the different loss functions (Fig. 19). Furthermore, Kullback-Leibler divergence and binary cross-entropy have the least difference in the result of the BAM.

The activation function guides the deep learning models to learn complex patterns (Jagtap et al., 2020; Jagtap and Karniadakis, 2022). There are many different kinds of activation functions which the most common activation functions without external parameters are included Relu, hyperbolic tangent (Tanh), exponential linear unit (Elu), scaled exponential linear unit (Selu), and Gaussian error linear unit (Gelu). The numerical results of utilizing different activation functions in the SMoE model are presented in Fig. 20. As seen, the activation function has an

impact on the performance SMoE model for BAM. Basis this figure, the Tanh activation function has provided the lowest performance. Furthermore, Gelu and Relu have the highest influence on the effectiveness of the model in BAM.

4.4. Ablation analysis

Ablation studies typically involve removing a component of the model and observing how this affects performance. To this end, this study conducts the ablation analysis in three main scenarios: (1) without the dense mixture of expert (S#1), (2) without the position attention module (S#2), and (3) without the channel attention module (S#3). As seen, Table 7 reports the result of the ablation analysis of the proposed SOME model for the first dataset. From the ablation analysis results reported in Tables 7 and it can be seen that the SMoE model in S#1 has the lowest performance. By specializing in different aspects of the data, combining expert knowledge, adapting to changing conditions, improving generalization, and increasing model robustness, the mixture of expert layers improves BAM. This results in burned area mapping predictions that are more accurate and resilient. Thus, the dense mixture of experts plays a key role in the BAM. In addition, the channel attention module has the lowest influence on the performance of the proposed SMoE model.

5. Discussion

5.1. Accuracy

In this study, we present a novel deep learning-based framework based on a mixture of experts. We evaluated the performance of the proposed SMoE model on three datasets and also compared it with other machine learning and deep learning models. It can be observed from Figs. 12–17 and Tables 4–6 that the proposed model outperforms other models in detecting both non-burned and burned classes. The OA shows the overall performance of the model in the detection of both burned and non-burned classes. The KC considers the performance of the model in an imbalanced dataset. The IOU metric measures the spatial agreement between predicted and real pixels. Precision refers to the performance of the model in the detection of non-burned pixels while Recall refers to the permeance of the model in the detection of burned pixels. Furthermore, the F1-score is the average of two metrics. There is a trade-off that exists in the detection of non-burned and burned classes. This trade-off often involves optimizing the model to excel in one class while potentially sacrificing performance in the other due to the inherent challenges and imbalances in the dataset. However, the 2D-SCNN has provided accuracy more than the SMoE model by Recall and BA metrics; it missed the performance by other metrics. Furthermore, this theme can be observed as the result of Swin-T in some metrics. The average accuracy of all models for the three datasets is shown in Fig. 21. From the average accuracy results shown in Fig. 18, it can be seen that the deep learning models outperformed other machine learning models (XGBoost and LGBM). Furthermore, the proposed SMoE model has achieved high performance compared to other deep learning models (2D-SCNN, 3D-SCNN, MLP-Mixer, and Swin-T). This is more evident in the F1-Score

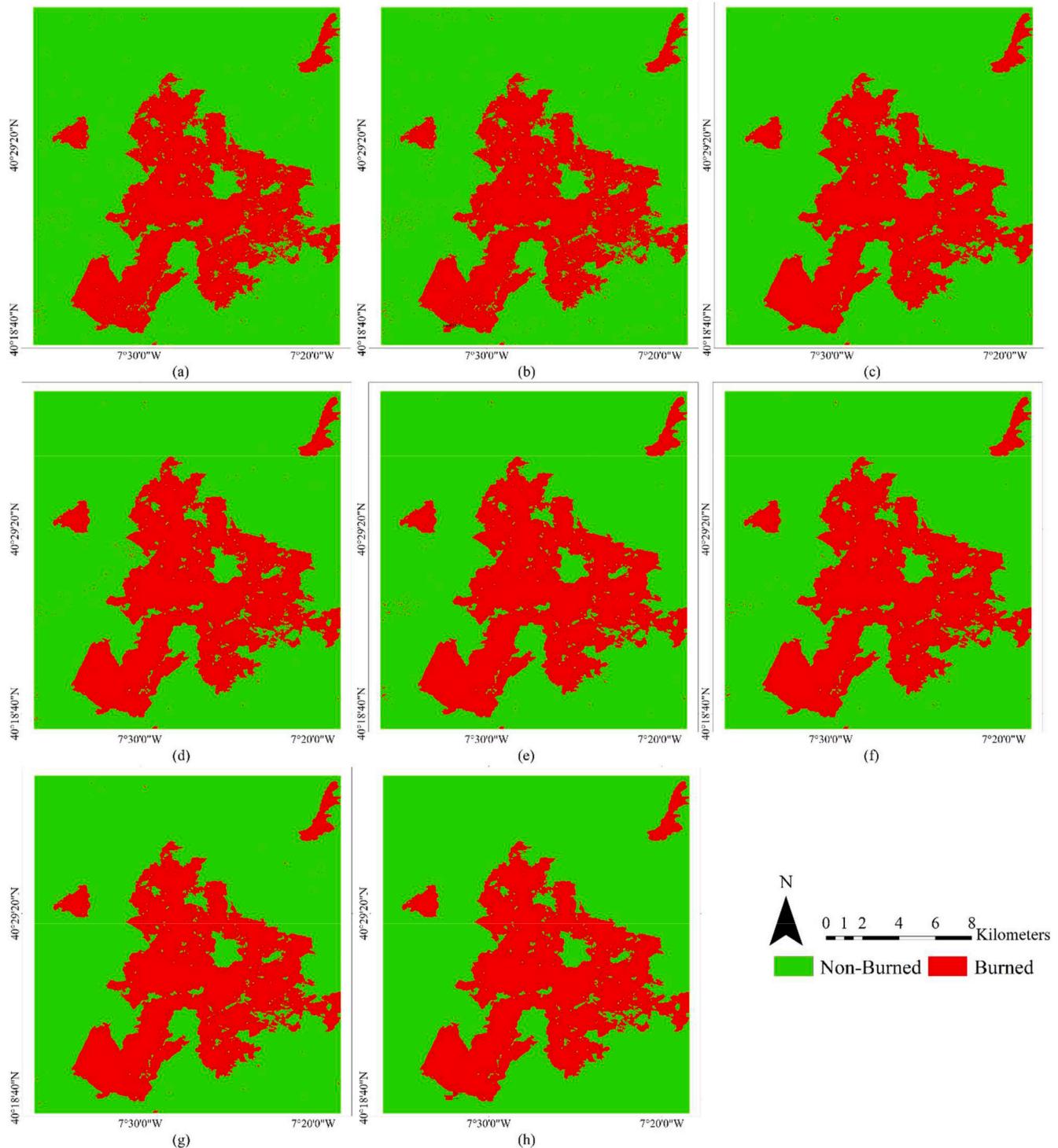


Fig. 16. Visual comparison of BAM results in the second study area: (a) XGBoost, (b) LGBM, (c) 2D-SCNN, (d) 3D-SCNN, (e) MLP-Mixer, (f) Swin-T, (g) SMoE, and (h) ground truth map.

and BA metrics, which passed 96% and 98%, respectively.

5.2. Significant test

The McNemar test is an established statistical technique for evaluating the statistical significance of differences in the effectiveness of the model. McNemar's test is tailored for use with a 2×2 contingency table. Within this table, the cells capture the counts of samples classified correctly and incorrectly by both methods, as well as those classified

correctly by only one of the methods. Table 8 shows the results of the McNemar test for different models with statistical significance at the 0.05 level. Based on this table, the p-value for the models is lower than 0.05, which means that the difference between the SMoE and the other models is significant.

5.3. Feature extraction

Recently, many machine learning-based burn area detection

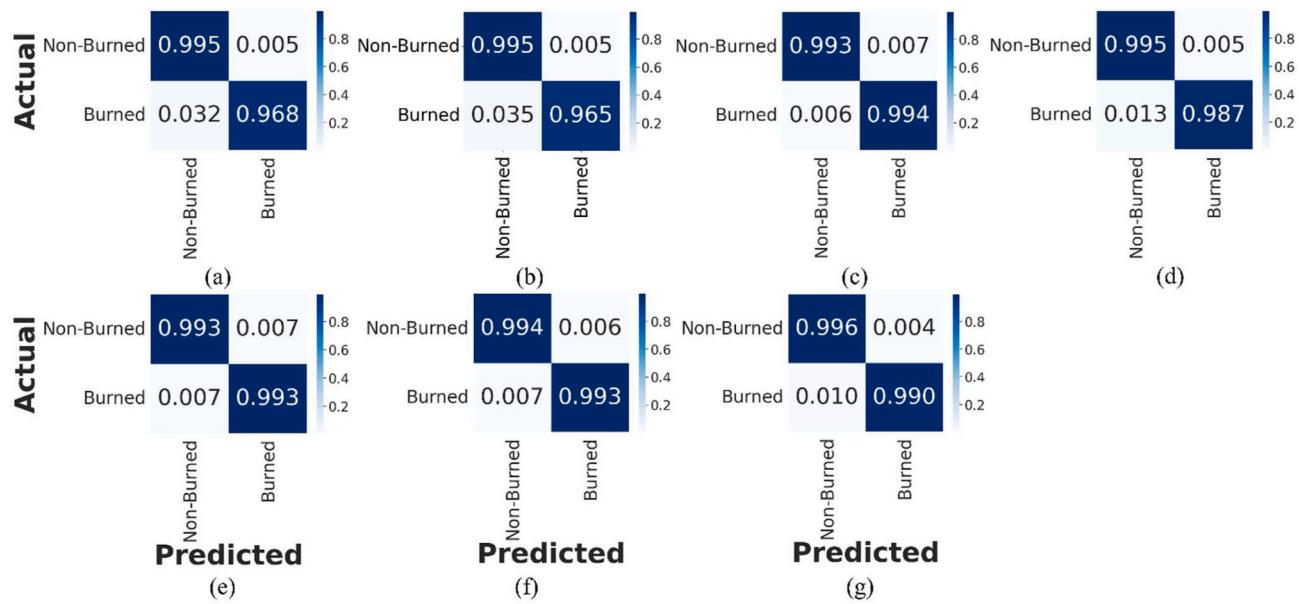


Fig. 17. Comparison of confusion matrix all methods on the third dataset: (a) XGBoost, (b) LGBM, (c) 2D-SCNN, (d) 3D-SCNN, (e) MLP-Mixer, (f) Swin-T, and (g) SMoE.

Table 6

A comparison between the BAM accuracies obtained by different models on the third dataset.

Index	XGBoost	LGBM	2D-SCNN	3D-SCNN	MLP-Mixer	Swin-T	SMoE
OA (%)	98.73	98.61	99.34	99.27	99.29	99.35	99.42
Precision (%)	98.81	98.71	98.37	98.81	98.31	98.51	98.97
Recall (%)	96.81	96.50	99.40	98.70	99.25	99.26	99.04
F1-Score (%)	97.80	97.59	98.88	98.75	98.78	98.88	99.00
BA (%)	98.17	97.99	99.36	99.10	99.28	99.32	99.31
KC	0.969	0.966	0.984	0.982	0.983	0.984	0.986
IOU	0.957	0.953	0.978	0.975	0.976	0.978	0.980

*The best accuracy is marked in bold.

frameworks have been proposed in many studies. These frameworks mainly focused on manual feature engineering based on a spatial feature (i.e., texture analysis using the gray-level-co-occurrence-matrix) and a spectral feature (i.e., normalized burn ratio index). In other words, machine learning models need to generate spatial and spectral features to achieve promising results. The extraction of informative features is a

time-consuming pre-processing, especially for such applications (BAM) on a large scale. The DL methods need not only to extract manual features but also to extract deep features automatically. The results of BAM show that deep learning-based frameworks have provided noisy labeled pixels lower than machine learning models. Thus, this problem can be attributed to deep learning-based models that use deep features, which are a combination of both spatial and spectral features.

The main difference between deep learning models (i.e., 2D SCNN and 3D SCNN) is the extraction of deep features by convolutional layers. Therefore, the design of an appropriate architecture for deep feature extraction is crucial. In this study, a novel deep feature generation framework based on mixture-of-expert layers is employed. The results of BAM for three datasets prove that the proposed SMoE model has high efficiency for BAM in comparison with the other two deep learning frameworks.

5.4. Single post-fire dataset

The proposed BAM in this study is based on a Siamese network, which simultaneously considers pre-fire and post-fire data sets. However, the use of the bi-temporal datasets increases the processing time and computational cost, it improves the performance of BAM. To this

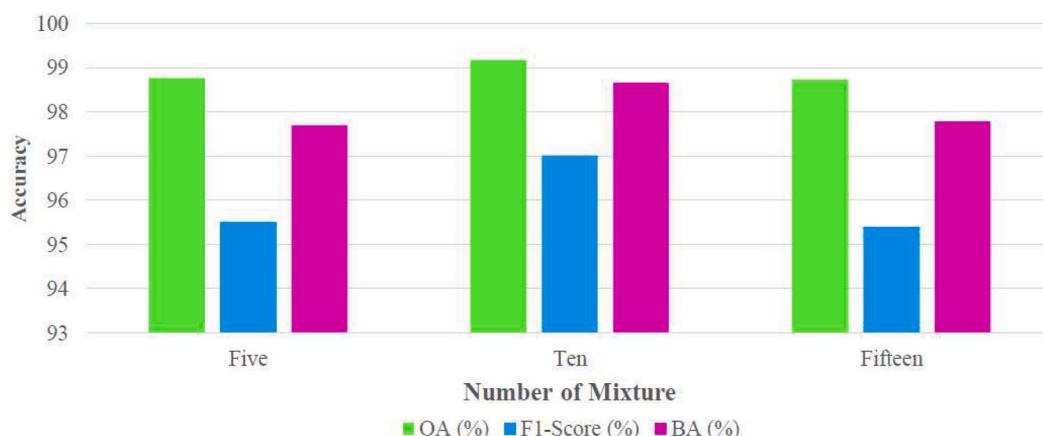


Fig. 18. The effect of the number of mixture experts on the effectiveness of the SMoE.

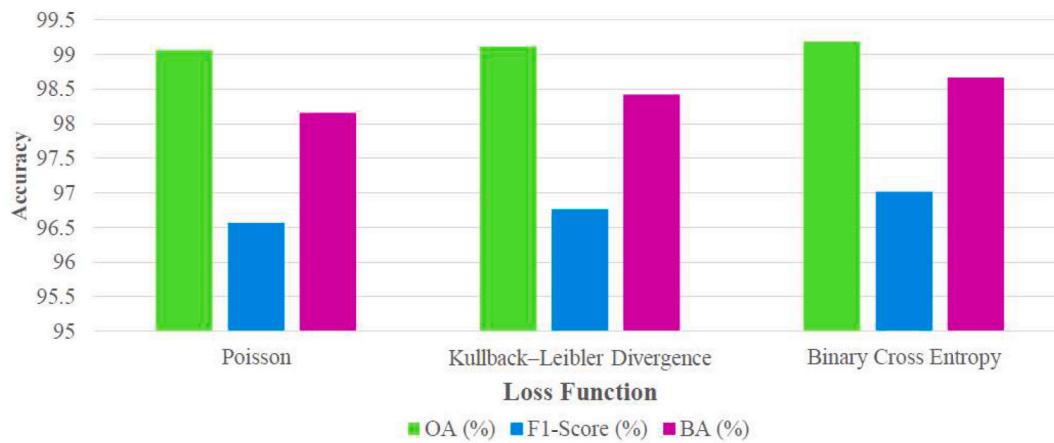


Fig. 19. Comparison of the influence of different loss functions in the training of the SMoE.

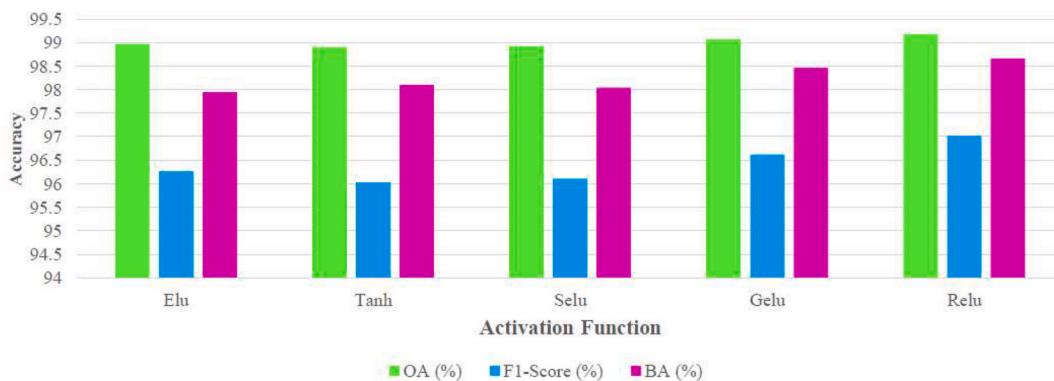


Fig. 20. Comparison of the influence of different activation functions in BAM by the SMoE.

Table 7
Ablation analysis of the SMoE model for BAM in the first dataset.

Index	S#1	S#2	S#3	SMoE
OA (%)	98.70	98.74	98.78	99.18
Precision (%)	95.07	94.53	95.31	96.09
Recall (%)	95.42	96.33	95.76	97.97
F1-Score (%)	95.24	95.42	95.53	97.02
BA (%)	97.32	97.72	97.51	98.67
KC	0.945	0.947	0.948	0.965
IOU	0.909	0.912	0.914	0.942

*The best accuracy is marked in bold.

end, we evaluated the performance of the proposed SMoE model in two scenarios. The first scenario is the BAM only based on the post-fire dataset, and the second uses the pre/post-fire datasets. The results of the performance of the SMoE model in both scenarios for the first dataset are shown in Table 9 and Fig. 22. As can be seen, the results of the BAM by SMoE model for the first study area show that the proposed method can provide high performance in both scenarios as the OA is more than

Table 8

Result of McNemar's test to compare the performance of the models.

Index	XGBoost	LGBM	2D-SCNN	3D-SCNN	MLP-Mixer	Swin-T
p-value	ϵ	ϵ	10^{-290}	ϵ	ϵ	ϵ

ϵ is lower than 10^{-350} .

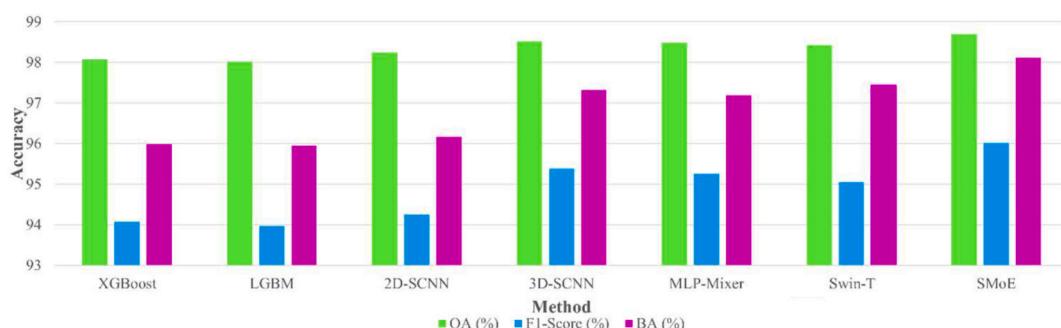


Fig. 21. A comparison of the average of the OA, F1-Score, and BA metrics between the different models.

Table 9

Comparison of the performance of the SMoE model based on utilizing only post-fire and bi-temporal datasets.

Dataset	OA (%)	Precision (%)	Recall (%)	F1-Score (%)	BA (%)	KC	IOU
Only Post-Fire	98.62	94.76	95.18	94.97	97.17	0.942	0.904
Pre/Post-Fire	99.18	96.09	97.97	97.02	98.67	0.965	0.942

*The best accuracy is marked in bold.

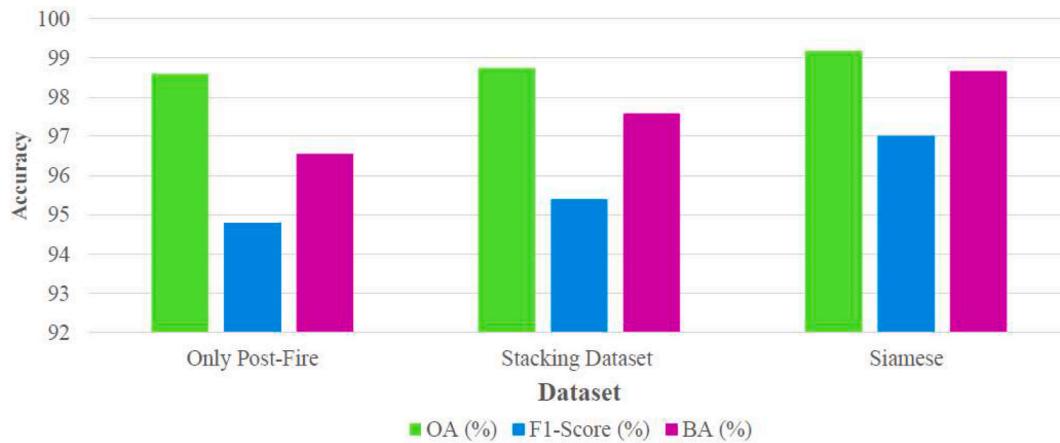


Fig. 22. Comparison of the performance of the SMoE model using post-fire only, stacking bi-temporal datasets, and bi-temporal datasets.

Table 10

Generalization of the SMoE model in the first and second datasets using the trained model from the third dataset.

Dataset	OA (%)	Precision (%)	Recall (%)	F1-Score (%)	BA (%)	KC	IOU
First dataset	97.55	94.81	86.85	90.66	93.05	0.892	0.829
Second dataset	91.25	64.78	94.05	76.71	92.40	0.715	0.622

*The best accuracy is marked in bold.

98.6%. In addition, the use of bi-temporal datasets increases the accuracy of BAM by more than 2%, 2.5%, and 0.02 in terms of F1 score, recall, and KC indices, respectively.

5.5. Sample dataset

The acquisition of sample datasets for supervised learning methods is a time-consuming and expensive process. The deep learning-based semantic segmentation-based BAM methods have been used in many studies (Bo et al., 2022; Brand and Manandhar, 2021; Lee et al., 2020). These models require a large number of sample datasets to be compared with CNN models. In addition, these models need to label all samples of the input patch dataset. Although they have provided promising results for BAM, providing a large amount of labeled sample data is a big challenge and very expensive. Our SMoE framework uses only 0.5% of the whole sample dataset, which is much lower than semantic segmentation models.

5.6. Model generalization

A critical aspect of wildfire mapping is the generalization of deep learning models. To this end, we evaluated the generalization of our proposed model based on a trained model using the third dataset and predicted the first and second datasets. Table 10 shows the numerical result of some model generalizations. As seen, the proposed model provided a performance of more than 97% and 91% for the first and second datasets, respectively.

Table 11
Comparison of the computational cost of the deep learning models.

Model	2D-SCNN	3D-SCNN	MLP-Mixer	Swin-T	SMoE
Time (s)	30.91	89.61	72.41	185.59	35.01
Number of Parameters (million)	0.9	4.3	0.8	1.6	9.7

5.7. Complexity analysis

The computational cost of deep learning models is an important factor. To evaluate the computation of the SMoE model, we measure the training time of the models for the first dataset. Table 11 shows the training time of the deep learning model under the same conditions. Based on this, the proposed model needed 35 s for training, while the 3D-SCNN needed 85 s. Also, the 2D-SCNN model required the lowest processing time (30 s). The deep feature generation by the 2D convolution layer is faster than the 3D convolution, which leads to the SMoE model being faster than the 3D SCNN model. However, the SMoE model has high parameter numbers compared with the other four models but it outperformed the other two models. There is a trade-off between the computational cost and the effectiveness of the model. For example, a higher number of parameters may allow a model to learn more complex patterns, but it may also require more computation.

6. Conclusions and future work

Accurate mapping of burned areas is important for post-fire management and subsequent activities. To achieve this, this study proposed

a novel BAM framework using bi-temporal pre-/post-fire Sentinel-2 imagery. In addition, we evaluated the performance of the proposed SMoE model on three real-world burned datasets in different areas based on visual comparison and accuracy assessment by several metrics. The experimental results show that all models have provided promising results, as those have achieved an accuracy of more than 91% by the OA index. The results of BAM show that the deep learning-based models have provided high performance in BAM comparison with advanced machine learning models (XGBoost and LGBM) as there are low noisy labeled pixels. The proposed SMoE model outperformed other deep learning-based models (2D-SCNN and 3D-SCNN) in three datasets. The proposed SMoE model can provide high performance only for the post-fire dataset, as it provides an accuracy of more than 96% by OA index. The use of a bi-temporal dataset improves the result of BAM. The generalization of the SMoE model was evaluated in the first and second study areas using the trained model from the third study area dataset. The result of generalization shows that the proposed model can provide high effectiveness for the unseen dataset. Generally, the proposed SMoE model has several comparisons with other models: (1) high accuracy in BAM, (2) low miss detection and false alarm rates, (3) doesn't require additional processing such as feature extraction, (4) high generalization for the unseen dataset, and (5) adaptive with mono-temporal and bi-temporal datasets for BAM.

Since the activation function influences the performance of the SMoE model, for future work we analyzed the effectiveness of advanced activation functions such as adaptive activation function (Jagtap et al., 2020, 2021; Jagtap and Karniadakis, 2020) in the BAM. Furthermore, the use of the Segment Anything Model (SAM) model can help to reduce the sample dataset and model generalization. Due to the potential of the SMoE model in the BAM, the focus for future work is on the multi-hazard mapping (i.e. active fire, flood) based on a multi-modal dataset. Furthermore, we analyzed the performance of the model with a large sample dataset utilizing SMoE for semantic segmentation tasks and compared it with state-of-the-art models in the huge dataset.

CRediT authorship contribution statement

Seyd Teymoor Seydi: Conceptualization, Methodology, Software, Visualization, Investigation. **Mahdi Hasanlou:** Data curation, Writing – original draft, Supervision. **Jocelyn Chanussot:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Mahdi Hasanlou reports administrative support, statistical analysis, and writing assistance were provided by School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran. Mahdi Hasanlou reports a relationship with School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran that includes: employment.

Data availability

Data will be made available on request.

References

- Abdikan, S., Bayik, C., Sekertekin, A., Bektas Balcik, F., Karimzadeh, S., Matsuoka, M., Balik Sanli, F., 2022. Burned area detection using multi-sensor SAR, optical, and thermal data in mediterranean pine forest. *Forests* 13, 347. <https://doi.org/10.3390/f13020347>.
- Achour, H., Toujani, A., Trabelsi, H., Jaouadi, W., 2022. Evaluation and comparison of Sentinel-2 MSI, Landsat 8 OLI, and EFFIS data for forest fires mapping. Illustrations from the summer 2017 fires in Tunisia. *Geocarto Int.* 37, 7021–7040. <https://doi.org/10.1080/10106049.2021.1980118>.
- Ager, A.A., Kline, J.D., Fischer, A.P., 2015. Coupling the biophysical and social dimensions of wildfire risk to improve wildfire mitigation planning. *Risk Anal.* 35, 1393–1406.
- Bar-Massada, A., Radeloff, V.C., Stewart, S.I., 2014. Biotic and abiotic effects of human settlements in the wildland–urban interface. *Bioscience* 64, 429–437.
- Bastarrika, A., Chuvieco, E., Martín, M.P., 2011. Mapping burned areas from Landsat TM/ETM+ data with a two-phase algorithm: balancing omission and commission errors. *Remote Sens. Environ.* 115, 1003–1012. <https://doi.org/10.1016/j.rse.2010.12.005>.
- Boschetti, L., Roy, D.P., Justice, C.O., Giglio, L., 2010. Global assessment of the temporal reporting accuracy and precision of the MODIS burned area product. *Int. J. Wildland Fire* 19, 705–709. <https://doi.org/10.1071/WF09138>.
- Bousquet, E., Mialon, A., Rodriguez-Fernandez, N., Mermoz, S., Kerr, Y., 2022. Monitoring post-fire recovery of various vegetation biomes using multi-wavelength satellite remote sensing. *Biogeosciences* 19, 3317–3336. <https://doi.org/10.5194/bg-19-3317-2022>.
- Chen, Z., Deng, Y., Wu, Y., Gu, Q., Li, Y., 2022. Towards understanding mixture of experts in deep learning. <https://doi.org/10.48550/arXiv.2208.02813>.
- Chowdhury, S., Zhu, K., Zhang, Y., 2021. Mitigating greenhouse gas emissions through generative adversarial networks based wildfire prediction. <https://doi.org/10.48550/arXiv.2108.08952>.
- Chuvieco, E., Lizundia-Loiola, J., Pettinari, M.L., Ramo, R., Padilla, M., Tansey, K., Mouillot, F., Laurent, P., Storm, T., Heil, A., Plummer, S., 2018. Generation and analysis of a new global burned area product based on MODIS 250 m reflectance bands and thermal anomalies. *Earth Syst. Sci. Data* 10, 2015–2031. <https://doi.org/10.5194/esd-10-2015-2018>.
- Cochrane, M.A., 2009. *Tropical Fire Ecology: Climate Change, Land Use, and Ecosystem Dynamics*. Springer.
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., Meygret, A., Spoto, F., Sy, O., Marchese, F., Bargellini, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote Sens. Environ.*, The Sentinel Missions - New Opportunities for Science 120, 25–36. <https://doi.org/10.1016/j.rse.2011.11.026>.
- Eigen, D., Ranzato, M., Sutskever, I., 2014. Learning factored representations in a deep mixture of experts. <https://doi.org/10.48550/arXiv.1312.4314>.
- Elhag, M., Yimaz, N., Bahrawi, J., Boteva, S., 2020. Evaluation of optical remote sensing data in burned areas mapping of thasos island, Greece. *Earth Syst. Environ.* 4, 813–826. <https://doi.org/10.1007/s41748-020-00195-1>.
- Engelbrecht, J., Theron, A., Vhengani, L., Kemp, J., 2017. A simple normalized difference approach to burn area mapping using multi-polarisation C-band SAR. *Rem. Sens.* 9, 764. <https://doi.org/10.3390/rs9080764>.
- Febrindihka, A.I., Rahman, C.T., Ramdani, F., Saputra, M.C., 2018. Tangible landscape: simulation of estimation of wildfire spread in arjuno mountain taurha R. Soerjo region. In: 2018 4th International Symposium on Geoinformatics (ISyG). Presented at the 2018 4th International Symposium on Geoinformatics (ISyG), pp. 1–8. <https://doi.org/10.1109/ISyG.2018.8611830>.
- Franquesa, M., Stehman, S.V., Chuvieco, E., 2022. Assessment and characterization of sources of error impacting the accuracy of global burned area products. *Remote Sens. Environ.* 280, 113214. <https://doi.org/10.1016/j.rse.2022.113214>.
- Fu, H., Gong, M., Wang, C., Tao, D., 2018. MoE-SPNet: a mixture-of-experts scene parsing network. <https://doi.org/10.48550/arXiv.1806.07049>.
- Gholamrezaie, H., Hasanlou, M., Amani, M., Mirmazloumi, S.M., 2022. Automatic mapping of burned areas using Landsat 8 time-series images in google Earth engine: a case study from Iran. *Rem. Sens.* 14, 6376. <https://doi.org/10.3390/rs14246376>.
- Gonçalves, D.N., Marcato, J., Carrilho, A.C., Acosta, P.R., Ramos, A.P.M., Gomes, F.D.G., Osco, L.P., da Rosa Oliveira, M., Martins, J.A.C., Damasceno, G.A., de Araújo, M.S., Li, J., Roque, F., de Faria Peres, L., Gonçalves, W.N., Libonati, R., 2023. Transformers for mapping burned areas in Brazilian Pantanal and Amazon with PlanetScope imagery. *Int. J. Appl. Earth Obs. Geoinformation* 116, 103151. <https://doi.org/10.1016/j.jag.2022.103151>.
- Griffith, J.D., Kochenderfer, M.J., Moss, R.J., Misic, V.V., Gupta, V., Bertsimas, D., 2017. Automated dynamic resource allocation for wildfire suppression. *Linc. Lab. J.* 22, 38–59.
- Gun, Z., Chen, J., 2023. Novel knowledge graph- and knowledge reasoning-based classification prototype for OBIA using high resolution remote sensing imagery. *Rem. Sens.* 15, 321. <https://doi.org/10.3390/rs15020321>.
- Han, Z., Geng, G., Yan, Z., Chen, X., 2022. Economic loss assessment and spatial-temporal distribution characteristics of forest fires: empirical evidence from China. *Forests* 13, 1988. <https://doi.org/10.3390/f13121988>.
- Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E., 1991. Adaptive mixtures of local experts. *Neural Comput.* 3, 79–87. <https://doi.org/10.1162/neco.1991.3.1.79>.
- Jagtap, A.D., Karniadakis, G.E., 2022. How important are activation functions in regression and classification? A survey, performance comparison, and future directions. <https://doi.org/10.48550/arXiv.2209.02681>.
- Jagtap, A.D., Karniadakis, G.E., 2020. Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *J. Comput. Phys.* 404, 109136. <https://doi.org/10.1016/j.jcp.2019.109136>.
- Jagtap, A.D., Kawaguchi, K., Karniadakis, G.E., 2020. Locally adaptive activation functions with slope recovery term for deep and physics-informed neural networks. *Proc. R. Soc. Math. Phys. Eng. Sci.* 476, 20200334. <https://doi.org/10.1098/rspa.2020.0334>.
- Jagtap, A.D., Shin, Y., Kawaguchi, K., Karniadakis, G.E., 2021. Deep Kronecker neural networks: a general framework for neural networks with adaptive activation functions. <https://doi.org/10.48550/arXiv.2105.09513>.
- Khan, Z.A., Hussain, T., Ullah, F.U.M., Gupta, S.K., Lee, M.Y., Baik, S.W., 2022. Randomly initialized CNN with densely connected stacked autoencoder for efficient

- fire detection. Eng. Appl. Artif. Intell. 116, 105403 <https://doi.org/10.1016/j.engappai.2022.105403>.
- Lasko, K., 2021. Incorporating Sentinel-1 SAR imagery with the MODIS MCD64A1 burned area product to improve burn date estimates and reduce burn date uncertainty in wildland fire mapping. Geocarto Int. 36, 340–360. <https://doi.org/10.1080/10106049.2019.1608592>.
- Lestari, A.I., Rizkinia, M., Sudiana, D., 2021. Evaluation of combining optical and SAR imagery for burned area mapping using machine learning. In: 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC). Presented at the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), pp. 52–59. <https://doi.org/10.1109/CCWC51732.2021.9376117>.
- Liu, M., Yang, L., 2020. Human-caused fires release more carbon than lightning-caused fires in the conterminous United States. Environ. Res. Lett. 16, 014013.
- Ma, C., Weng, L., Xia, M., Lin, H., Qian, M., Zhang, Y., 2023. Dual-branch network for change detection of remote sensing image. Eng. Appl. Artif. Intell. 123, 106324 <https://doi.org/10.1016/j.engappai.2023.106324>.
- Macharyulu, I.S., Satapathy, D.P., Sahoo, A., Samantaray, S., Mohanta, N.R., Ray, A., 2022. Performance evaluation of MLP and CNN models for flood prediction. In: Intelligent System Design: Proceedings of INDIA 2022. Springer, pp. 273–281.
- Martins, V.S., Roy, D.P., Huang, H., Boschetti, L., Zhang, H.K., Yan, L., 2022. Deep learning high resolution burned area mapping by transfer learning from Landsat-8 to PlanetScope. Remote Sens. Environ. 280, 113203 <https://doi.org/10.1016/j.rse.2022.113203>.
- Meher, S.K., 2020. Granular space, knowledge-encoded deep learning architecture and remote sensing image classification. Eng. Appl. Artif. Intell. 92, 103647 <https://doi.org/10.1016/j.engappai.2020.103647>.
- Meng, Y., Deng, Y., Shi, P., 2015. Mapping forest wildfire risk of the world. World Atlas Nat. Disaster Risk 261–275.
- Nolde, M., Plank, S., Riedlinger, T., 2021. Utilization of hyperspectral remote sensing imagery for improving burnt area mapping accuracy. Rem. Sens. 13, 5029. <https://doi.org/10.3390/rs13245029>.
- Palandjian, D., Gitas, I.Z., Wright, R., 2009. Burned area mapping and post-fire impact assessment in the Kassandra peninsula (Greece) using Landsat TM and Quickbird data. Geocarto Int. 24, 193–205. <https://doi.org/10.1080/10106040802488542>.
- Philipp, M.B., Levick, S.R., 2020. Exploring the potential of C-band SAR in contributing to burn severity mapping in tropical savanna. Rem. Sens. 12, 49. <https://doi.org/10.3390/rs12010049>.
- Pinto, M.M., Libonati, R., Trigo, R.M., Trigo, I.F., DaCamara, C.C., 2020. A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. ISPRS J. Photogrammetry Remote Sens. 160, 260–274. <https://doi.org/10.1016/j.isprsjprs.2019.12.014>.
- Roteta, E., Bastarrika, A., Padilla, M., Storm, T., Chuvieco, E., 2019. Development of a Sentinel-2 burned area algorithm: generation of a small fire database for sub-Saharan Africa. Remote Sens. Environ. 222, 1–17. <https://doi.org/10.1016/j.rse.2018.12.011>.
- Saisaward, C., Ninsawat, S., 2023. Evaluation MODIS and sentinel-2 data for detecting crop residue burned area. In: Boonpook, W., Lin, Z., Meksangsouy, P., Wetchayont, P. (Eds.), Applied Geography and Geoinformatics for Sustainable Development. Springer Geography. Springer International Publishing, Cham, pp. 143–158. https://doi.org/10.1007/978-3-031-16217-6_11.
- Sali, M., Piaser, E., Boschetti, M., Brivio, P.A., Sona, G., Bordogna, G., Stroppiana, D., 2021. A burned area mapping algorithm for sentinel-2 data based on approximate reasoning and region growing. Rem. Sens. 13, 2214. <https://doi.org/10.3390/rs13112214>.
- Saranya, T., Deisy, C., Sridevi, S., Anbananthen, K.S.M., 2023. A comparative study of deep learning and Internet of Things for precision agriculture. Eng. Appl. Artif. Intell. 122, 106034 <https://doi.org/10.1016/j.engappai.2023.106034>.
- Seydi, S.T., Hasanlou, M., 2021. A new structure for binary and multiple hyperspectral change detection based on spectral unmixing and convolutional neural network. Measurement 186, 110137. <https://doi.org/10.1016/j.measurement.2021.110137>.
- Seydi, S.T., Hasanlou, M., Chanussot, J., 2022. A quadratic morphological deep neural network fusing radar and optical data for the mapping of burned areas. IEEE J. Sel. Top. Appl. Earth Obs. Rem. Sens. 15, 4194–4216. <https://doi.org/10.1109/JSTARS.2022.3175452>.
- Seydi, S.T., Hasanlou, M., Chanussot, J., 2021. DSMNN-net: a deep siamese morphological neural network model for burned area mapping using multispectral sentinel-2 and hyperspectral PRISMA images. Rem. Sens. 13, 5138. <https://doi.org/10.3390/rs13245138>.
- Shah, S.T.H., Qureshi, S.A., Rehman, A. ul, Shah, S.A.H., Amjad, A., Mir, A.A., Alqahtani, A., Bradley, D.A., Khandaiker, M.U., Faruque, M.R.I., Rafique, M., 2021. A novel hybrid learning system using modified breaking ties algorithm and multinomial logistic regression for classification and segmentation of hyperspectral images. Appl. Sci. 11, 7614. <https://doi.org/10.3390/app11167614>.
- Syphard, A.D., Bar Massada, A., Butsic, V., Keeley, J.E., 2013. Land use planning and wildfire: development policies influence future probability of housing loss. PLoS One 8, e71708.
- Tanase, M.A., Santoro, M., Wegmüller, U., de la Riva, J., Pérez-Cabello, F., 2010. Properties of X-, C- and L-band repeat-pass interferometric SAR coherence in Mediterranean pine forests affected by fires. Remote Sens. Environ. 114, 2182–2194. <https://doi.org/10.1016/j.rse.2010.04.021>.
- van Gerrevink, M.J., Veraverbeke, S., 2021. Evaluating the hyperspectral sensitivity of the differenced normalized burn ratio for assessing fire severity. Rem. Sens. 13, 4611. <https://doi.org/10.3390/rs13224611>.
- Wang, X., Yu, F., Dunlap, L., Ma, Y.-A., Wang, R., Mirhoseini, A., Darrell, T., Gonzalez, J. E., 2019. Deep Mixture of Experts via Shallow Embedding. <https://doi.org/10.48550/arXiv.1806.01531>.
- Yadav, C.S., Pradhan, M.K., Gangadharan, S.M.P., Chaudhary, J.K., Singh, J., Khan, A.A., Haq, M.A., Alhussen, A., Wechtaisong, C., Imran, H., Alzamil, Z.S., Pattanayak, H.S., 2022. Multi-class pixel certainty active learning model for classification of land cover classes using hyperspectral imagery. Electronics 11, 2799. <https://doi.org/10.3390/electronics11172799>.
- Zhang, Q., Ge, L., Zhang, R., Metternicht, G.I., Du, Z., Kuang, J., Xu, M., 2021. Deep-learning-based burned area mapping using the synergy of Sentinel-1&2 data. Remote Sens. Environ. 264, 112575 <https://doi.org/10.1016/j.rse.2021.112575>.