KSÜ Mühendislik Bilimleri Dergisi, 28(1), 2025
Araştırma Makalesi

KSU J Eng Sci, 28(1), 2025
Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 16.12.2024
Kabul Tarihi : 26.01.2025

Received Date : 16.12.2024
Accepted Date : 26.01.2025

A COMPARATIVE STUDY OF DOUBLE-STEP DEEP LEARNING FRAMEWORK FOR BURNED AREA IDENTIFICATION AND SEVERITY ASSESSMENT IN WILDFIRES

YANGINLARDA YANMIŞ ALANLARIN BELİRLENMESİ VE ŞİDDET DEĞERLENDİRMESİ İÇİN ÇİFT ADIMLI DERİN ÖĞRENME ÇERÇEVESİNİN KARŞILAŞTIRMALI BİR ÇALIŞMASI

Murat Mert YURDAKUL[1*] (ORCID: 0009-0000-7285-7802)
Bülent BAYRAM[2] (ORCID: 0000-0002-4248-116X)
Tolga BAKIRMAN[2] (ORCID: 0000-0001-7828-9666)
Hamza Osman İLHAN[1] (ORCID: 0000-0002-1753-2703)

[1] Yıldız Technical University, Department of Computer Engineering, İstanbul, Türkiye
[2] Yıldız Technical University, Department of Geomatic Engineering, İstanbul, Türkiye

*Sorumlu Yazar / Corresponding Author: Murat Mert YURDAKUL, mert.yurdakul@std.yildiz.edu.tr

ABSTRACT

As wildfires become more frequent and intense, it is essential to develop sophisticated techniques for precise detection and damage evaluation. This research examines a Double-Step Deep Learning Framework using several U-Net models, including MultiResUNet, to identify burned areas and estimate severity. Using satellite images, the study explores the effect of different severity levels within mask output, focusing on both 4 and 5 level severity classifications. Additionally, the Mask R-CNN model was evaluated independently for image segmentation, revealing challenges due to its reliance on pretrained weights and limited spectral input. The comparative analysis illustrates how changes in the granularity of severity intervals influence model performance, providing insights into the benefits of more nuanced severity segmentation for wildfire assessment. This approach has the potential to improve the precision of damage assessments and support more informed decision-making in the management and response of wildfires.

Keywords: Wildfire severity prediction; deep neural networks; double-step architecture; satellite imagery analysis.

ÖZET

Orman yangınlarının daha sık ve yoğun hale gelmesiyle birlikte, doğru tespit ve hasar değerlendirmesi için gelişmiş tekniklerin geliştirilmesi büyük önem taşımaktadır. Bu araştırma, yanmış alanları belirlemek ve yangın şiddetini tahmin etmek için MultiResUNet dahil olmak üzere çeşitli U-Net modellerini kullanan Çift Aşamalı Derin Öğrenme Çerçevesi'ni incelemektedir. Uydu görüntülerinden elde edilen maske çıktıları üzerinde, özellikle 4 ve 5 şiddet seviyelerine odaklanılarak, farklı şiddet seviyelerinin etkileri detaylı bir şekilde incelenmiştir. Ayrıca, Mask R-CNN modeli, önceden eğitilmiş ağırlıklar ve sınırlı spektral girdiler nedeniyle görüntü segmentasyonunda yaşanan zorlukları göstermek için bağımsız olarak incelenmiştir. Yapılan analizler, şiddet aralıklarının granülerliğindeki değişikliklerin model performansını nasıl etkilediğini göstererek, yangın değerlendirmesi için daha ayrıntılı şiddet segmentasyonunun faydalarına dair önemli bilgiler sağlamaktadır. Bu yaklaşım, hasar değerlendirmelerinin doğruluğunu artırma ve yangın yönetimi ile müdahalesinde daha bilinçli kararlar alınmasını destekleme potansiyeline sahiptir.

Anahtar Kelimeler: Yangın şiddeti tahmini; derin sinir ağları; çift adımlı mimari; uydu görüntü analizi.

ToCite: YURDAKUL, M.M., BAYRAM, B., BAKIRMAN, T. & İLHAN, H.O. (2025). A COMPARATIVE STUDY OF DOUBLE-STEP DEEP LEARNING FRAMEWORK FOR BURNED AREA IDENTIFICATION AND SEVERITY ASSESSMENT IN WILDFIRES. *Kahramanmaraş Sütçü İmam Üniversitesi Mühendislik Bilimleri Dergisi*İ 28(1), 513-523.

INTRODUCTION

The growing threat posed by wildfires in recent years has increased the need for advanced technologies to assess fire damage accurately and efficiently. The destruction caused by these fires affects ecosystems, economies, and communities worldwide, making it essential to have reliable tools for fire detection and damage assessment. High-resolution satellite imagery from missions such as Sentinel-2 offers a critical resource for monitoring large-scale environmental events such as wildfires.

This paper explores the Double-Step Deep Learning Framework (DSF) with derivative models of U-Net, as used in the paper prepared by Monaco et al. (Monaco et al., 2021), for wildfire detection and severity estimation. The framework incorporates a binary segmentation model to detect burned areas and a regression model to estimate the severity of the damage.

This study introduces MultiResUNet (Ibtehaz and Rahman, 2020), a U-Net variant not previously evaluated in the DSF, to assess its performance in comparison to other U-Net-based models such as Attention U-Net, Nested U-Net, and SegU-Net. This study also investigates the impact of adjusting severity levels by merging levels 4 and 3 into a 4-level classification, addressing data imbalance issues and improving model training stability. Additionally, a pretrained Mask R-CNN model is evaluated independently from the DSF to benchmark its performance on burned area detection. This comparison highlights the strengths and limitations of the DSF framework relative to state-of-the-art object detection models.

This paper also reviews related works that focus on wildfire detection and burned area analysis, with an emphasis on studies that predominantly utilize deep learning models, some of which are variants of the U-Net architecture. It details the materials and methodology, including a description of the employed dataset and the architecture of the proposed MultiResUNet approach. The experimental results are presented and discussed, comparing the performance of different U-Net architectures on the dataset. Finally, the paper concludes with insights and suggestions for future research.

RELATED WORK

Various U-Net models and their extensions have been widely applied in forest fire-related tasks. Adaptations such as attention mechanisms, residual blocks, and integration of temporal and environmental data have enhanced the effectiveness of U-Net in wildfire detection, smoke segmentation, and severity assessment.

Farasin et al. introduced a methodology for assessing wildfire damage severity using Sentinel-2 satellite imagery through a deep learning model named Double-Step U-Net (Farasin et al., 2020). In their study, they also proposed and compared the performance of Single, Parallel, and Double-Step U-Net frameworks, evaluating their effectiveness in burned area detection and severity estimation.

Monaco et al. utilized an extension of the Double-Step U-Net (Farasin et al., 2020). to predict wildfire severity using a multichannel deep learning framework for post-fire satellite imagery analysis. Their approach, referred to as the Double-Step Deep Learning Framework (Monaco et al., 2021), incorporates an attention mechanism into the U-Net. The framework compares its results with derivatives of the U-Net model to evaluate its effectiveness. The Double-Step Framework operates in two steps: the first performs binary segmentation to identify burned regions, while the second assigns severity levels to each pixel within the identified burned areas. Leveraging Sentinel-2 data with twelve channels, the Double-Step Framework captures diverse terrain features and applies multi-channel attention analysis to assess the importance of each spectral band in predicting burn severity.

MultiResUNet model, proposed by Ibtehaz and Rahman (Ibtehaz and Rahman, 2020), is an enhancement of the original U-Net architecture designed specifically for improved performance in medical image segmentation. While traditional U-Net consists of an encoder-decoder structure with simple convolutional layers, MultiResUNet introduces MultiRes blocks. Furthermore, Ibtehaz and Rahman replaced the standard U-Net skip connections with Res paths. These paths apply additional convolutional layers to reduce the semantic gap between encoder and decoder features.

Colomba et al. introduced a new data set specifically designed to delineate the burned area and estimate severity from satellite imagery (Colomba et al., 2022). The authors used Sentinel-1 and Sentinel-2 satellite data to categorize burn severity on a scale from 0 (undamaged) to 4 (completely destroyed).

Han et al. proposed a transformer-based change detection model to accurately map burned areas and assess the severity of the burn (Han et al., 2024). Their approach leverages multispectral Landsat-8 imagery and auxiliary environmental data such as vegetation indices to enhance model performance and accuracy in identifying fire-affected areas. The model compares pre-fire and post-fire images to detect changes in vegetation and land surface features. This work demonstrates the potential of combining transformer-based methods with environmental data to improve wildfire impact assessments and aid in post-fire recovery planning (Han et al., 2024).

Pinto et al. developed a deep learning model called BA-Net, which is tailored for identifying and timing burned regions by utilizing a temporal series of satellite images from the VIIRS sensor (Pinto et al., 2020). This approach addresses key limitations of traditional burned area mapping techniques by eliminating the need for cloud masking and complex preprocessing steps. The model utilizes a U-Net architecture modified with a Long-Short-Term Memory (LSTM) layer to capture temporal dependencies (Pinto et al., 2020).

Navarro et al. analyzed 2016 forest fires on Madeira Island using Sentinel-2A imagery to evaluate the severity of the burn (Navarro et al., 2017). They focused on the application of spectral indices such as the Normalized Difference Vegetation Index (NDVI), Green Normalized Difference Vegetation Index (GNDVI), and Normalized Burn Ratio (NBR). By comparing these indices with data from the Copernicus Emergency Management Service, the authors confirmed that red-edge spectral indices were highly effective in identifying levels of burn severity (Navarro et al., 2017).

Li et al. developed an early forest fire segmentation algorithm named F-Unet, designed to aid fire rescue operations (Li et al., 2021). F-Unet introduces a contraction path, a feature fusion layer, and an expansion path to enhance the precision of the segmentation. The contraction path leverages the first 13 layers of VGG16 to capture multiscale feature maps. The feature fusion network integrates these features to improve segmentation accuracy. The results from the experimental evaluation of the FLAME dataset (Shamsoshoara et al. 2020) indicate that F-Unet substantially improves the accuracy of fire segmentation (Li et al., 2021).

Khennou and Akhloufi presented FU-NetCastV2, a deep learning convolutional neural network developed to predict fire spread and map burned areas (Khennou and Akhloufi, 2023). FU-NetCastV2 identifies areas around wildfires that are at high risk of future spread by integrating satellite imagery with topographical and weather data. This approach leverages consecutive forest wildfire perimeters, satellite images, Digital Elevation Model maps, aspect, slope, and weather variables (Khennou and Akhloufi, 2023).

Wang et al. proposed an enhanced approach to early forest fire smoke detection through the development of Smoke-Unet, an improved U-Net model incorporating attention mechanisms and residual blocks (Wang et al., 2022). Their method leverages the RGB, SWIR2, and AOD bands to improve recognition accuracy, demonstrating a 3.1% improvement over standard U-Net performance on a diverse dataset encompassing multiple seasons, regions, and types of land cover.

Zou et al. propose an attention-based deep learning model (Zou et al., 2023) for wildfire spread prediction, integrating spatial and channel attention modules with a CNN. Trained on fire-tracking satellite data and environmental factors, it outperforms benchmark models like FARSITE (Finney, 1998) in predicting fire progression and final perimeters, offering valuable applications for fire risk management.

MATERIAL AND METHOD

Dataset Information

The dataset includes 73 distinct regions, each typically represented by four images: pre-fire and post-fire Sentinel-1 and Sentinel-2 images, along with a mask image indicating the burned area. The mask images feature five pixel intensity levels, ranging from black (level 0 - unburned) to white (level 4 - severely burned). The time frame between pre-fire and post-fire images is standardized to two months (Colomba et al., 2022). The data set is divided into training, validation, and test sets based on predefined folds as shown in Table 1 from the article (Colomba et al., 2022):

Table 1. The Organization Schema of The Validation and Test Sets In the Experiments

Sets	Folds						
Test	Purple	Coral	Pink	Grey	Cyan	Lime	Magenta
Validation	Coral	Cyan	Coral	Coral	Coral	Coral	Coral

The fold colors represent sets grouped by their similar morphological features, primarily corresponding to regions in closely situated cities throughout Europe. As shown in Table 1, if the test set is 'purple', the validation set will be 'coral'. The remaining folds form the training set. The majority of the dataset consists of unburned areas as shown by the grey bar in Figure 1. Only a small portion of the dataset represents burned areas. This disparity highlights the difficulty of training models effectively to predict burned areas, given that fire-affected regions are relatively less represented than unburned zones.

Data Preprocessing

This research exclusively utilized post-fire Sentinel-2 imagery for the experiments. During the selection process, images with excessive cloud cover were filtered out by applying a cloud cover threshold. The resolution of the images is excessively high for straightforward use in training neural network models. Consequently, the images were split into square tiles, each measuring 480x480 pixels.

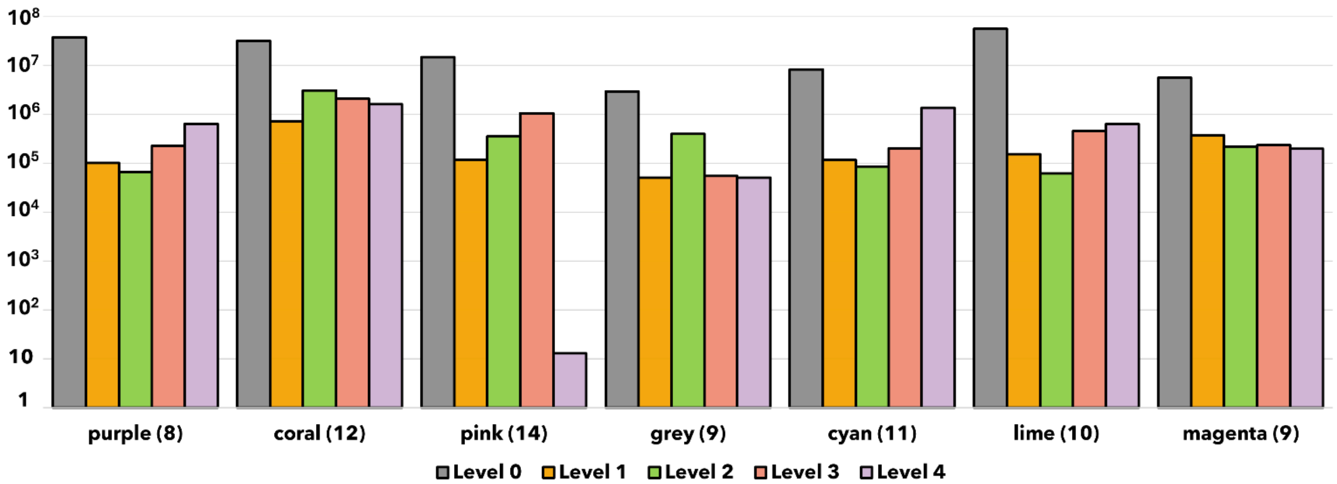


Figure 1. Logarithmic Representation of Pixels from Mask Images Based on Folds. From Left to Right: Purple, Coral, Pink, Grey, Cyan, Lime, and Magenta. The Number of Images in Each Fold Set Is Indicated in Parentheses, With Purple Containing 8, Coral 12, Pink 14, Grey 9, Cyan 11, Lime 10, and Magenta 9 Images, Totaling 73 Images Across All Folds.

Data Augmentation

During training, random data augmentation techniques were applied to improve model performance, ensure robustness, address data imbalances, and prevent overfitting. The methods employed include rotating images by a maximum of 50 degrees, cropping images by as much as 20 degrees, and flipping images both horizontally and vertically. Each of these techniques was executed with a 50% probability for each batch of images. Furthermore, all images were normalized using a mean and standard deviation of 0.5.

Only the training dataset is enhanced using data augmentation methods like rotation and flipping. The performances are assessed using Intersection over Union (IOU) for binary predictions and Root Mean Square Error (RMSE) for regression tasks. IoU is a metric calculated by dividing the area of overlap between the predicted and ground truth regions by the area of their union. RMSE measures the average magnitude of the error between predicted and ground truth values, calculated as the square root of the mean of the squared differences. IoU is used to evaluate the binary segmentation performance in the first step of the framework, while RMSE is employed in the second step to assess the accuracy of severity estimation.

Mask R-CNN

Mask R-CNN (He et al., 2017) is an advanced deep learning architecture designed for instance segmentation tasks. It extends Faster R-CNN (Ren et al., 2015) by adding a branch for predicting segmentation masks on each detected object. The architecture combines object detection and pixel-level segmentation, making it highly effective for applications requiring precise localization, such as satellite image analysis and medical imaging. Mask R-CNN uses a backbone network (e.g., ResNet) (He et al., 2016) for feature extraction and a Region Proposal Network (RPN) (Ren et al., 2015) to identify candidate object regions.

Double-Step Deep Learning Framework

Double-Step Deep Learning Framework is designed to predict the severity of wildfires, dividing the task into two steps: detection of the wildfire and estimation of the severity. Figure 2 outlines the general structure of the Double-Step Framework.

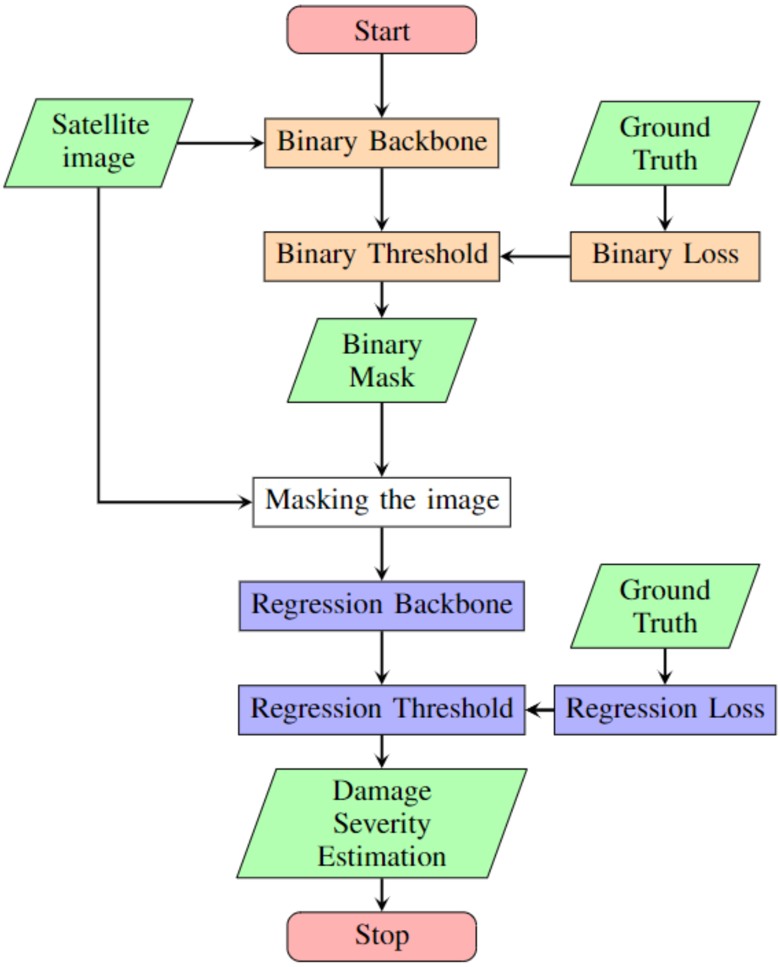


Figure 2. Double-Step Framework Architecture. Green Shapes: Inputs/Outputs. Orange Shapes: Wildfire Detection. Blue Shapes: Severity Estimation.

The initial phase of the Double-Step Framework, known as the Wildfire Detection Task (Step One), aims to identify if a particular region has been burned. In this stage, a binary-class backbone was utilized that analyzes satellite images and generates a binary probability map. Each pixel represents the likelihood of being burned (values between 0 and 1). The initial backbone training uses a binary loss function to compare the predictions with the ground-truth binary masks. After generating the probability map, the values are thresholded to create a final binary mask (burned/unburned regions). This research utilized a threshold value of 0.5. Applying the resulting binary mask to the original image effectively conceals the unburned regions, ensuring that only the burned areas are visible in preparation for the model's subsequent step.

In the second step, the model uses masked satellite images and predicts the severity of fire damage using a regression backbone. The result is a map assigning each pixel a severity score from 0 to 4, indicating varying levels of burn damage. The second step involves posing it as a regression problem where the model is trained to reduce the disparity between the predicted and actual severity levels. If the binary mask produced initially contains inaccuracies, it will negatively impact the performance of the second model. Thus, ensuring an accurate binary mask enhances the overall effectiveness of severity estimation.

DSF approach supports the implementation of various neural network architectures for both binary detection and severity assessment processes. In this study, several deep learning-based segmentation networks have been deployed. Each step uses the same deep learning model architecture. The subsequent subtitles provide a summary of the models employed.

U-Net

U-Net (Ronneberger et al., 2015) is a convolutional neural network architecture specifically designed for image segmentation tasks. The architecture consists of an encoder that captures context through convolutions and pooling operations and a decoder that uses upsampling. This structure allows U-Net to excel in segmenting complex structures within images, making it highly suitable for applications like medical imaging, satellite image analysis, and more.

Attention U-Net

It is an advanced version of the U-Net, incorporating attention mechanisms to enhance segmentation accuracy by focusing on relevant image features. Attention U-Net (Oktay et al., 2018) applies attention gates in these connections. This approach enables the model to selectively emphasize significant regions while suppressing less important areas. The architecture retains the same encoder-decoder structure as U-Net. Attention gates ensure that only the most relevant features are passed to the decoding layers.

Nested U-Net

Also known as UNet++ (Zhou et al., 2018), is an enhancement of the original U-Net architecture designed to improve segmentation performance by adding multiple nested and dense skip pathways between encoder and decoder. This architecture introduces a series of intermediate layers called nested dense skip connections.

SegU-Net

It is a segmentation model (Kamal et al., 2020) that combines the strengths of the U-Net architecture with a pretrained VGG backbone for feature extraction. The encoder path leverages the VGG network to capture detailed spatial and contextual information from the input image.

MultiResUNet

It is an advanced version of the U-Net architecture specifically designed for medical image segmentation. Unlike the classical U-Net, MultiResUNet (Ibtehaz and Rahman, 2020) incorporates MultiRes blocks to capture image features at multiple resolutions. This design allows it to more effectively identify structures at varying scales within medical images. Additionally, the model replaces U-Net's standard skip connections with "Res paths", helping to bridge the gap between feature representations at different levels. These enhancements make MultiResUNet more robust in segmenting challenging images. Given its ability to handle multi-resolution features and bridge encoder-decoder disparities, MultiResUNet was considered for wildfire segmentation tasks, where complex and varied spatial patterns in satellite imagery pose significant challenges. Its unique architectural features align well with the need for precise detection and segmentation of burned areas across diverse geographic regions.

EXPERIMENTAL RESULTS

In this study, DSF is used to assess the effectiveness of several variations of U-Net for the detection and segmentation of wildfire severity. In the experiments, U-Net, Attention U-Net, Nested U-Net, SegU-Net, and MultiResUNet models were tested. The first four models are integral components of the Double-Step Framework proposed by (Monaco et al., 2021). MultiResUNet, however, is included primarily for comparison purposes with minimal customization. Additionally, this study employs two distinct severity scales for the second step of DSF: one with five levels and another with four levels. The 5-level classification allows for more detailed differentiation of severity, useful for ecological studies requiring precise segmentation of fire impacts. In contrast, the 4-level classification simplifies the analysis by addressing data imbalance and grouping levels 4 and 3.

In DSF model training, the binary cross-entropy loss was used for the binary segmentation criterion, which measures the difference between predicted probabilities and actual binary labels (0 or 1) by calculating the negative log likelihood of the correct classification. Table 2 shows the rest of the parameters utilized in this study.

Table 2. Hyperparameters

Hyperparameter	Value
Batch size	8
Input channels	12
Epochs	50
Seed	1
Patience	5
Tolerance	0.01
Criterion	nn.BCEWithLogitsLoss
Regr_criterion	nn.MSELoss
Optimizer	optim.Adam (lr=0.0001)
Trainer - Precision	16
Trainer - Gradient_clip_val	5.0

The first step of the Double-Step Framework focuses on binary detection, distinguishing burned from unburned areas. IoU results, as shown in Table 3, include models within the Double-Step Framework as well as an evaluation of Mask R-CNN as an external benchmark for binary segmentation. MultiResUNet demonstrated notable performance, achieving an IoU score of 0.84 on the Coral and Pink dataset. However, it is important to note that the model's predicted mask images often exhibited predicted lines close to the edge of the images, likely due to the minimal customization applied to the model. Attention U-Net also performed well in this step. It achieved an IoU of 0.82 in the Magenta data set. Nested U-Net exhibited the most consistent performance overall with its highest IoU score of 0.85 on the Magenta and Pink dataset. The results of Step 1 suggest that MultiResUNet and Nested U-Net may be better suited for binary detection, as they achieved high IoU scores across different levels and datasets.

Table 3. Step 1 Result (Wildfire Detection Phase) Results

Binary - IoU						
	Double-Step Framework					**Single Step**
Fold	**U-Net**	**Attention U-Net**	**Nested U-Net**	**SegU-Net**	**MultiResUNet**	**Mask R-CNN**
Coral	0.83	0.74	0.84	0.75	0.84	0.31
Cyan	0.66	0.68	0.68	0.59	0.66	0.27
Grey	0.48	0.71	0.78	0.70	0.64	0.22
Lime	0.56	0.60	0.57	0.57	0.58	0.38
Magenta	0.84	0.82	0.85	0.82	0.82	0.17
Pink	0.81	0.79	0.85	0.82	0.84	0.28
Purple	0.82	0.81	0.81	0.81	0.65	0.26

Mask R-CNN was evaluated independently from the DSF to observe its performance on binary segmentation tasks. The model was trained using 3-channel satellite images as input and mask images for binary segmentation of burned areas. The results were extremely poor, with IoU values across different test folds shown in Table 3, suggesting that the model failed to learn effectively. This could be attributed to the reliance on pretrained weights, which are likely optimized for general-purpose object detection tasks rather than wildfire-specific segmentation. The model's inability to generalize may also stem from the unique spectral and geographical characteristics of wildfire datasets, which require domain-specific training. Training the Mask R-CNN model from scratch, without using pretrained weights, could better adapt it to the wildfire dataset and improve its performance. Additionally, modifying the model to accept 12-channel Sentinel-2 images, which capture a broader range of spectral information, could further enhance its segmentation accuracy for wildfire datasets.

The second step of the Double-Step Framework is designed to perform a severity estimation on the detected burned areas, evaluating the models' ability to predict severity levels accurately. Table 4 shows the RMSE results for each dataset and their corresponding levels. MultiResUNet achieved the lowest RMSE value of 0.17 in the Lime data set with the 4-level configuration. However, it is important to note that this RMSE result is at level 0, which represents unburned areas. Predicting unburned areas should be evaluated differently from other levels since it largely reflects the model's binary detection accuracy rather than its capability for detailed severity classification. Nested U-Net displayed higher variability in RMSE values across different datasets, especially in the Coral dataset. The standard U-Net model maintained relatively stable RMSE results. In Figure 3 examples of predicted images are shown.

Table 4. Double-Step Framework Step 2 (Wildfire Severity Classification) Results

		Regression - RMSE									
	Level	**U-Net**		**Attention U-Net**		**Nested U-Net**		**SegU-Net**		**MultiResUNet**	
Fold		**5 Lvl**	**4 Lvl**	**5 Lvl**	**4 Lvl**	**5 Lvl**	**4 Lvl**	**5 Lvl**	**4 Lvl**	**5 Lvl**	**4 Lvl**
Coral	0	0.71	0.61	0.66	0.49	0.56	0.54	0.42	0.44	0.48	0.47
	1	0.93	0.84	0.97	0.60	0.79	0.77	0.56	0.67	0.65	0.65
	2	1.04	0.60	1.01	0.86	0.88	0.53	0.83	0.79	0.54	0.63
	3	0.74	0.51	0.52	1.07	0.35	0.55	1.10	1.19	0.98	1.05
	4	0.83		1.79		1.52		2.48		2.42	
	ALL	0.79	0.60	0.86	0.69	0.72	0.55	0.90	0.70	0.86	0.63
Cyan	0	0.45	0.54	0.52	0.47	0.46	0.49	0.41	0.49	0.56	0.41
	1	1.05	1.08	1.09	1.02	1.04	0.93	0.81	0.65	1.17	0.90
	2	0.99	0.83	0.90	0.90	0.93	0.85	0.98	0.96	0.94	1.04
	3	0.71	1.36	0.68	1.44	0.74	1.39	1.14	1.80	0.67	1.62
	4	2.30		2.21		2.31		2.67		2.33	
	ALL	1.03	0.78	1.02	0.76	1.03	0.75	1.15	0.89	1.07	0.79
Grey	0	0.38	0.26	0.30	0.24	0.31	0.29	0.63	0.27	0.29	0.29
	1	1.14	0.95	1.15	0.72	0.91	0.88	0.70	0.75	0.89	0.93
	2	1.25	1.37	1.06	0.89	0.66	1.07	0.93	0.65	0.81	0.94
	3	1.23	1.47	1.16	1.55	1.21	1.27	1.55	1.44	1.47	1.30
	4	2.37		1.91		2.02		2.64		2.49	
	ALL	0.57	0.52	0.49	0.42	0.44	0.46	0.72	0.40	0.47	0.44
Lime	0	0.28	0.32	0.26	0.23	0.21	0.26	0.19	0.28	0.23	0.17
	1	0.76	0.63	0.75	0.73	0.71	0.65	0.82	0.80	0.77	0.78
	2	1.00	1.06	1.05	1.06	1.16	1.02	1.32	1.26	1.13	1.51
	3	1.60	1.48	1.85	1.61	2.03	1.48	2.30	1.90	1.95	2.27
	4	1.91		2.11		2.18		2.58		2.23	
	ALL	0.47	0.44	0.50	0.41	0.50	0.41	0.56	0.49	0.51	0.52
Magenta	0	0.38	0.42	0.39	0.45	0.34	0.39	0.34	0.26	0.29	0.48
	1	1.28	1.34	1.30	1.41	1.15	1.41	1.10	0.82	0.74	1.57
	2	0.84	0.84	0.96	0.80	0.79	0.84	0.84	0.76	0.76	0.87
	3	1.05	0.97	1.18	0.83	1.17	0.91	1.37	1.29	1.54	0.95
	4	1.24		1.27		1.42		1.72		2.12	
	ALL	0.63	0.62	0.67	0.61	0.63	0.60	0.70	0.57	0.74	0.68
Pink	0	0.32	0.23	0.28	0.25	0.23	0.26	0.26	0.24	0.18	0.21
	1	0.94	0.80	0.87	0.84	0.95	0.93	0.88	0.64	0.56	0.54
	2	0.56	0.79	0.92	0.80	0.66	0.60	0.78	0.55	0.93	0.70
	3	0.91	1.14	1.36	1.09	1.05	1.05	1.25	1.52	1.78	1.56
	4	1.69		2.55		1.66		1.78		2.26	
	ALL	0.45	0.48	0.57	0.48	0.45	0.46	0.52	0.57	0.65	0.58
Purple	0	0.20	0.21	0.28	0.20	0.20	0.26	0.24	0.17	0.22	0.34
	1	1.11	0.93	1.08	0.88	1.00	1.03	0.95	0.73	0.83	0.84
	2	0.54	0.51	0.71	0.44	0.48	0.57	0.36	0.50	0.69	0.56
	3	1.45	1.51	1.39	1.18	1.25	1.07	1.43	1.34	1.89	1.23
	4	1.74		1.67		1.85		2.03		2.24	
	ALL	0.57	0.41	0.59	0.43	0.60	0.43	0.65	0.46	0.71	0.51

The experiments reveal that merging levels 4 and 3 into a 4-level classification consistently reduces RMSE values across most folds and models compared to the 5-level classification. This improvement is particularly notable in folds with significant data imbalance, such as Coral and Grey. For example, in the Coral fold, the RMSE for U-Net decreases from 0.79 (5-level) to 0.60 (4-level), and MultiResUNet exhibits similar trends with RMSE decreasing from 0.86 to 0.63. However, the Lime fold showed less consistent improvement, highlighting that the effectiveness of this adjustment may vary depending on the dataset's severity level distribution.

In Table 5, the performance benchmarks of various models and methods employing the Double-Step Framework, including related works and the results from this study, are presented. The table includes metrics such as the Intersection over Union, Root Mean Squared Error for all levels (RMSE All), and Root Mean Squared Error for individual severity levels (RMSE Level). RMSE Level does not include Level 0 (unburned pixels), as unburned pixels dominate the dataset and could disproportionately influence the metric. By excluding Level 0, RMSE Level focuses on burned severity levels, providing a more meaningful comparison of the models' performance in assessing burned areas. The results allow for a comparative analysis of different approaches, highlighting the effectiveness of the Double-Step Framework and related methods.

Table 5. Performance Benchmark of Related Works

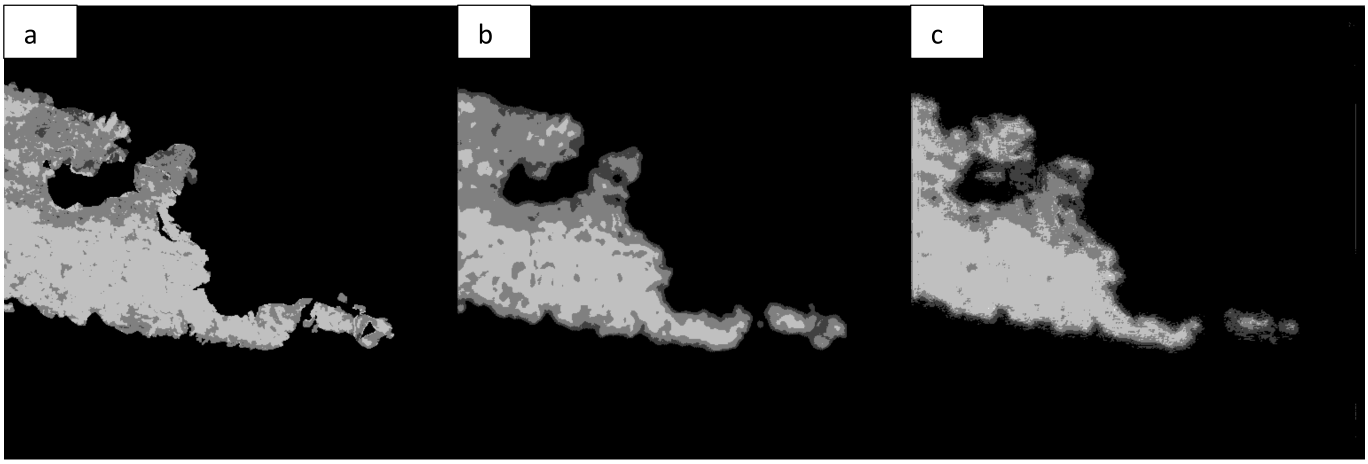
Model/Method	IoU	RMSE All	RMSE Level	Reference
DS-UNet	0.75	1.32	-	Monaco et al.,2021
DS-UNet++	0.74	1.41	-	
DS-SegU	0.65	1.66	-	
DS-AttU	0.72	1.38	-	
dNBR	-	0.91	0.48	Farasin et al.,2020
Single U-Net	-	0.95	0.35	
Parallel U-Net	-	0.97	0.35	
Double-Step U-Net	-	0.76	0.39	
MultiResUNet	0.84	0.44	0.54	This Study

CONCLUSION

This paper presents the implementation and evaluation of a Double-Step Deep Learning Framework for the detection and severity assessment of burned areas using Sentinel-2 satellite imagery. The Double-Step Framework leverages various U-Net model variations like Attention U-Net, Nested U-Net, SegU-Net, and MultiResUNet to assess wildfire impact across multiple severity levels. The first step in this framework successfully utilizes binary detection to distinguish burned from unburned regions, with MultiResUNet and Nested U-Net demonstrating strong performance in terms of IoU. Although MultiResUNet achieved high accuracy in the Coral dataset, predicted mask images exhibited lines close to the edge of the image, suggesting potential limitations due to the model's minimal customization.

The results of the 4-level and 5-level severity classifications demonstrate that merging levels 4 and 3 helps address data imbalance and reduces RMSE for underrepresented high-severity categories. This adjustment simplifies the classification task and often leads to better overall model performance, as seen in the lower RMSE values across most folds and models. However, the variability in performance across certain folds, such as Lime, underscores the need for further investigation into the impact of severity level distribution on model generalization.

The Mask R-CNN model was evaluated independently to benchmark its binary segmentation performance. Its results were extremely poor, likely due to the reliance on pretrained weights and the use of 3-channel input images instead of the full spectral range provided by Sentinel-2. Enhancing Mask R-CNN through domain-specific training and modifying its architecture to accept 12-channel Sentinel-2 imagery could improve its ability to process wildfire datasets effectively. Incorporating Mask R-CNN into the severity estimation step of the DSF represents a promising area for future exploration.



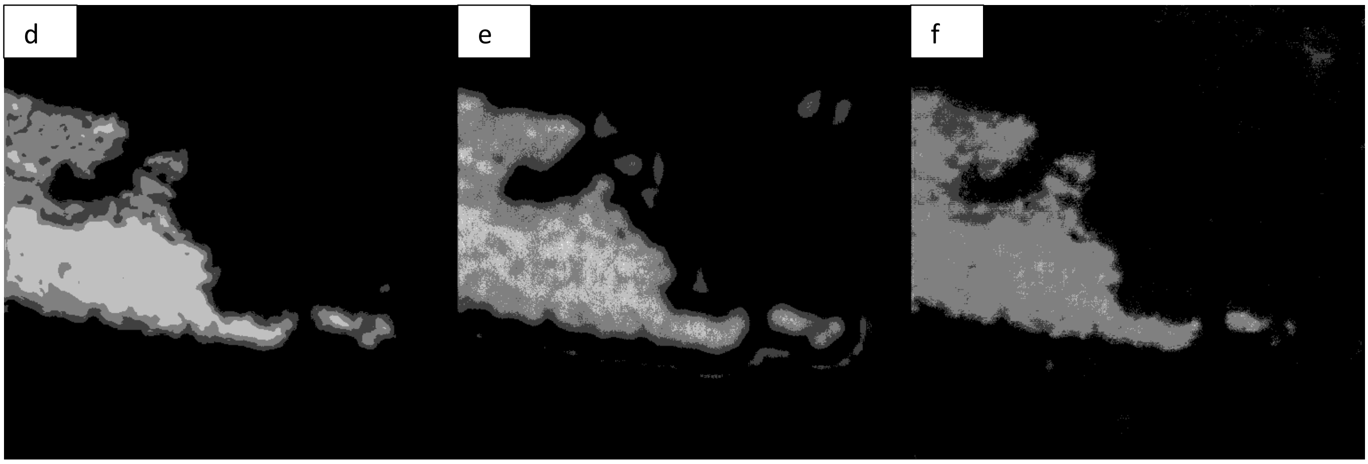


Figure 3. a. Ground Truth **b.** U-Net Predicted Mask Image **c.** Attention U-Net Predicted Mask Image **d.** Nested U-Net Predicted Mask Image **e.** SegU-Net Predicted Mask Image **f.** MultiResUNet Predicted Mask Image

The experimental results from the second step show that Nested U-Net showed greater variability in RMSE across different folds, with the Coral dataset posing particular challenges. The Pink fold achieved the best overall performance, suggesting that the distribution of severity levels in the data set can impact the accuracy of the model. By contrast, the Lime fold exhibited the poorest detection results, likely due to the unique geographical characteristics of volcanic areas, which may introduce greater complexity in accurately detecting burned regions.

This study demonstrates the potential of integrating multi-sensor deep learning models, leveraging the rich spectral information provided by Sentinel-2 imagery, for wildfire detection and severity assessment. Among the models tested, MultiResUNet with minimal customization showed similar results to other U-Net variants, particularly for burned area detection. This suggests that the first step of the Double-Step Framework could be optimized specifically for U-Net-based architectures. Furthermore, a potential area of future research could involve leveraging different model families, such as applying U-Net-based models for binary detection and architectures like Mask R-CNN for severity estimation. Additionally, employing more balanced satellite datasets could enhance the Double-Step Framework's effectiveness across a wider range of geographical features and severity levels.

ACKNOWLEDGMENT

This study was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) ARDEB 2549 (NCBR) Grant No: 122N254.

REFERENCES

Colomba, L., Farasin, A., Monaco, S., Greco, S., Garza, P., Apiletti, D., Baralis, E., & Cerquitelli, T. (2022). A dataset for burned area delineation and severity estimation from satellite imagery. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM ’22)* (pp. 3893–3897). Association for Computing Machinery. https://doi.org/10.1145/3511808.3557528

Farasin, A., Colomba, L., & Garza, P. (2020). Double-Step U-Net: A deep learning-based approach for the estimation of wildfire damage severity through Sentinel-2 satellite data. *Applied Sciences*, *10*(12), 4332. https://doi.org/10.3390/app10124332

Finney, M. A. (1998). *FARSITE: Fire Area Simulator—Model Development and Evaluation* (Research Paper RMRS-RP-4, Revised 2004). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Han, Y., Zheng, C., Liu, X., Tian, Y., & Dong, Z. (2024). Burned area and burn severity mapping with a transformer-based change detection model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 17*, 13866–13880. https://doi.org/10.1109/JSTARS.2024.3435857

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2961–2969. https://doi.org/10.1109/ICCV.2017.322

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Ibtehaz, N., & Rahman, M. S. (2020). Multiresunet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks, 121*, 74–87. https://doi.org/10.1016/j.neunet.2019.08.025

Kamal, U., Tonmoy, T. I., Das, S., & Hasan, M. K. (2020). Automatic traffic sign detection and recognition using SegU-Net and a modified Tversky loss function with L1-constraint. *IEEE Transactions on Intelligent Transportation Systems, 21*(4), 1467–1479. https://doi.org/10.1109/TITS.2019.2911727

Khennou, F., & Akhloufi, M. A. (2023). Improving wildland fire spread prediction using deep U-Nets. *Science of Remote Sensing, 8,* 100101. https://doi.org/10.1016/j.srs.2023.100101

Li, M., Zhang, Y., Xin, J., Mu, L., Yu, Z., Liu, H., Xie, G., Jiao, S., & Yi, Y. (2021). Early forest fire segmentation based on deep learning. In *2021 CAA Symposium on Fault Detection, Supervision, and Safety for Technical Processes (SAFEPROCESS)* (pp. 1–5). IEEE. https://doi.org/10.1109/SAFEPROCESS52771.2021.9693660

Monaco, S., Greco, S., Farasin, A., Colomba, L., Apiletti, D., Garza, P., Cerquitelli, T., & Baralis, E. (2021). Attention to fires: Multi-channel deep learning models for wildfire severity prediction. *Applied Sciences, 11*(22), 11060. https://doi.org/10.3390/app112211060

Navarro, G., Caballero, I., Silva, G., Parra, P.-C., Vázquez, Á., & Caldeira, R. (2017). Evaluation of forest fire on Madeira Island using Sentinel-2A MSI imagery. *International Journal of Applied Earth Observation and Geoinformation, 58,* 97–106. https://doi.org/10.1016/j.jag.2017.02.003

Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. *arXiv preprint*, arXiv:1804.03999

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 Journal of Photogrammetry and Remote Sensing, 160,* 260-274. https://doi.org/10.1016/j.isprsjprs.2019.12.014

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems* (Vol. 28).

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P. Z., & Blasch, E. (2020). *Aerial imagery pile burn detection using deep learning: The FLAME dataset*. arXiv. https://arxiv.org/abs/2012.14036

Wang, Z., Yang, P., Liang, H., Zheng, C., Yin, J., Tian, Y., & Cui, W. (2022). Semantic Segmentation and Analysis on Sensitive Parameters of Forest Fire Smoke Using Smoke-Unet and Landsat-8 Imagery. *Remote Sensing*, *14*(1), 45. https://doi.org/10.3390/rs14010045

Zou, Y., Sadeghi, M., Liu, Y., Puchko, A., Le, S., Chen, Y., Andela, N., & Gentine, P. (2023). Attention-Based Wildland Fire Spread Modeling Using Fire-Tracking Satellite Observations. *Fire*, *6*(8), 289. https://doi.org/10.3390/fire6080289

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4* (pp. 3-11). Springer International Publishing. https://doi.org/10.1007/978-3-030-00889-5_1