Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 10 No. 3 (2026) 735 - 744 e-ISSN: 2580-0760

Burned Area Segmentation Using Random Band Selection and Ensemble Encoder

Annisaa Nurul Ramadhani Novelika[1], Laksmita Rahadianti[2]
[1,2]Faculty of Computer Science, Universitas Indonesia, Depok, Indonesia

[*]*annisaa.nurul@ui.ac.id*

Abstract

Wildfires are increasingly recognized as major environmental hazards that endanger ecosystems, human health, and economic stability. In practice, various monitoring efforts, such as hotspot detection, burned-area indices, and field inspections still depend on reliable spatial information to quantify fire impacts over large regions. Accurate segmentation of burned areas from satellite imagery plays a vital role in supporting post-disaster response and sustainable land management. This paper proposes a novel framework for burned area segmentation using multispectral imagery sampled by Random Subspace Band Selection (RSBS) with a U-Net architecture with an ensemble encoder. The RSBS module generates multiple 3-band subsets from Landsat-8 data, incorporating spectrally informative bands such as Near-Infrared (NIR) or Shortwave Infrared (SWIR). These subsets are used to train U-Net models with an ensemble of encoder backbones ResNet34, ResNet50, DenseNet121, MobileNetV2 and InceptionV4 for spectral and architectural diversity. The final predictions are aggregated using majority voting to enhance robustness and generalization. The framework is evaluated on a publicly available Indonesian burned area dataset encompassing diverse land cover types. Experimental results demonstrate that the proposed ensemble model achieves up to 0.7673 IoU, 0.8734 mIoU, and 0.8683 F1 Score, outperforming the best single-backbone model. These findings confirm that the proposed framework offers a scalable, accurate, and adaptable solution for wildfire damage assessment using satellite data.

Keywords: burned area segmentation; deep learning; ensemble learning; multispectral imagery; random subspace; remote sensing; U-Net

How to Cite: A. N. R. Novelika and L. Rahadianti, "Burned Area Segmentation Using Random Band Selection and Ensemble Encoder ", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 10, no. 3, pp. 735 - 744, Jun. 2026.
Permalink/DOI: [https://doi.org/10.29207/resti.v10i3.7370

Received: December 5, 2025
Accepted: May 2, 2026
Available Online: June 21, 2026

This is an open-access article under the CC BY 4.0 License
Published by Ikatan Ahli Informatika Indonesia

1. Introduction

Wildfires are among the most destructive natural hazards, causing ecological degradation, biodiversity loss, and economic damage. Accurate and timely burned area mapping is therefore crucial for assessing fire impacts, supporting post disaster recovery, and informing environmental management. With the increasing availability of satellite data, multispectral sensors such as Sentinel 2 and Landsat 8 have become essential tools for monitoring burned areas due to their accessible coverage, consistent spatial resolution, and rich spectral information [1-3]. However, traditional index based approaches such as the Normalized Burn Ratio (NBR) and the Normalized Difference Vegetation Index (NDVI) often struggle under heterogeneous land cover conditions, complex terrain, and variable illumination [1], [2]. These limitations highlight the need for more adaptive, data driven methods that can capture the diverse spectral and spatial characteristics of fire affected landscapes.

Recent advances in deep learning have enabled models to learn feature representations directly from satellite imagery [3-6]. Convolutional Neural Networks (CNNs) have shown strong capabilities in modeling hierarchical spatial spectral patterns and often outperform index based or thresholding approaches [5-7]. Among these models, U-Net is widely used for semantic segmentation because its encoder decoder design and skip connections help preserve important spatial details [8], [9].

A range of U-Net variants has also been developed to address different challenges in burned area mapping. Examples include TV U-Net for SAR based mapping [10], Bia-Net for bi temporal Sentinel 2 data [11], and

residual attention U-Net models for forest fire delineation [12]. Other studies focus on rapid or high-resolution mapping, such as Basnet [13], SAFE [14], Burnt Net [15], and hybrid convolutional transformer approaches using Landsat-9 imagery [16]. Research on cross resolution mapping highlights the benefit of combining medium and high-resolution imagery [17], while optical and SAR fusion has proven useful for dealing with clouds or smoke [18]. Lightweight architectures such as attention enhanced DeepLabV3+ [19] and machine learning based approaches including Extreme Gradient Boosting and U-Net hybrids [20], [21] further expand the available methods.

Several reviews and regional case studies also provide insights into current challenges and opportunities. Surveys on fire detection emphasize the importance of scalable models, multisensory integration, and uncertainty estimation for real operational use [22], [23]. Recent work on wildfire monitoring pipelines highlights the growing use of cloud-based processing platforms and new satellite missions [24]. Studies from South Korea [2], [25] and research using residual attention U-Nets with Sentinel 2 time series [6], [26] show that adapting models to local vegetation, terrain, and regional fire patterns is often essential. Ensemble and meta learning approaches in related segmentation problems, such as road scene understanding and hyperspectral analysis, also encourage the exploration of hybrid ensemble strategies for burned area mapping [27], [28].

Although many advancements have been made, several challenges still need to be addressed. Multispectral imagery often contains redundant or uninformative bands, and the most effective band combinations can vary across different regions, fire events, and atmospheric conditions. Deep learning models that rely on a fixed set of input bands or a single encoder may therefore struggle to generalize well in diverse and highly dynamic environments. Ensemble based methods offer promising improvements, but most existing studies focus only on architectural ensembles or spectral subspace selection. Only a small number of works attempt to integrate both aspects into one unified framework [28], [29]. Research involving tropical wildfire datasets, including those from Indonesia, also remains limited, even though these environments pose unique challenges such as dense vegetation, rapid land cover changes, and interference from smoke [30].

Motivated by these gaps, this study proposes a hybrid burned area mapping framework that integrates Random Subspace Band Selection (RSBS) with an ensemble of U-Net models using heterogeneous encoder backbones [31]. RSBS introduces spectral diversity by sampling multiple three band subsets from Landsat 8 imagery, while ensuring that each subset includes at least one fire sensitive band such as Red, NIR, or SWIR. Each subset is processed by a dedicated U-Net model with a different encoder, including Resnet [32], Dense Net [33], MobileNet [34], and Inception [35]. This diversity allows the ensemble to capture complementary spatial and spectral information. Final predictions are generated through pixel level majority voting, enabling the strengths of each model to be combined effectively.

The main contributions of this study are threefold. First, we propose a novel framework that integrates spectral subspace sampling with heterogeneous encoder ensemble learning within a U-Net based architecture to improve burned-area segmentation. This approach leverages diverse spectral representations while capturing complementary features from multiple encoders. Second, the method is evaluated on a publicly available Indonesian wildfire dataset that reflects complex tropical conditions, including high variability in vegetation, land cover, illumination, and burn severity, ensuring realistic and challenging assessment. Third, this study provides an in-depth analysis of band selection strategies, encoder diversity, and ensemble configurations, highlighting their impact on segmentation accuracy and robustness.

This work bridges spectral subspace exploration and deep ensemble modelling, improving performance while reducing sensitivity to spectral redundancy and environmental variability. The proposed framework offers a scalable, robust, and transferable solution for burned-area mapping, enabling reliable detection across regions and satellite sensors and supporting adaptive wildfire monitoring.

2. Methods

2.1. Dataset Description

This study employs the Dataset of Deep Learning from Landsat-8 Satellite Images for Estimating Burned Areas in Indonesia [30]. This dataset was collected by National Institute of Aeronautics and Space (LAPAN), now known National Research and Innovation Agency (BRIN) in Indonesia, through the Innovative and Productive Research (RISPRO) program through the National Research Priority (PRN) scheme. The dataset is composed of three major components: image subsets, burned area masks, and quicklook images.

The image subsets consist of 227 image patches extracted from Landsat-8 scenes captured between 2019 and 2021. Each patch has a spatial resolution of 512 × 512 pixels and includes eight multispectral bands. Bands 1 through 7 follow the original Landsat-8 sequence, while Band 8 corresponds to Band 9 (cirrus). All images are stored in GeoTIFF format using the geographic coordinate system (WGS 1984 datum) with a spatial resolution of 0.00025 degrees. The images are encoded as 16-bit unsigned integers, with pixel values ranging from 0 to 65,535. These subsets depict burned areas in diverse ecological contexts such as forests, shrublands, grasslands, open areas, water bodies, settlements, clouds, and shadows. Some subsets also contain smoke from active fires. Spatial overlaps exist between some patches to ensure full coverage of

extensive burn areas. The Landsat-8 spectral configuration used in this study, including wavelength ranges and angular resolution for all bands, is summarized in Table 1.

Table 1. Landsat-8 band set

Band Names	Wavelength [μm]	Resolution [°]
Band 1—Coastal/Aerosol	0.43–0.45	0.000 25
Band 2—Blue	0.45–0.51	0.000 25
Band 3—Green	0.53–0.59	0.000 25
Band 4—Red	0.64–0.67	0.000 25
Band 5—Near Infrared (NIR)	0.85–0.88	0.000 25
Band 6—Short Wave Infrared (SWIR-1)	1.57–1.65	0.000 25
Band 7—Short Wave Infrared (SWIR-2)	2.11–2.29	0.000 25
Band 8—Cirrus	1.36–1.38	0.000 25

The burned area masks are binary annotation images indicating burned regions (value = 1) and unburned background (value = 0). Each mask is saved in 8-bit unsigned integer format. The masks were refined to exclude interference from thick clouds, shadows, and vegetation. Of the 227 images, 206 contain labeled burned areas, while 21 consist solely of background. In most cases, the burned areas cover less than 10% of the image surface. The quicklook images provide false color composite visualizations using Bands 7 (SWIR-2), 5 (NIR), and 4 (Red). These images, also sized at 512 × 512 pixels, are stored in GeoTIFF format with 8-bit unsigned integer values. A contrast stretching process was applied to enhance visual clarity. Quicklook images serve both as rapid visual references and as alternative inputs for model development and evaluation.

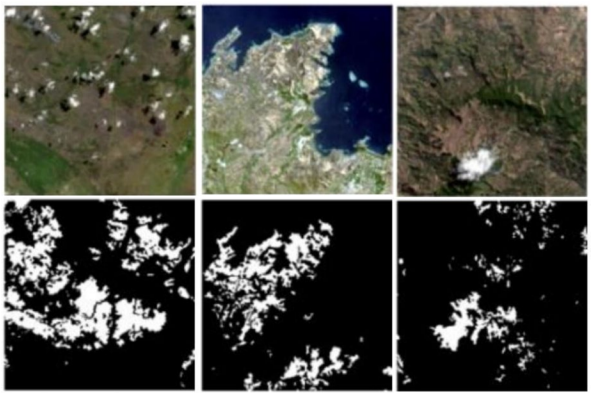


Figure 1. Representative sample images and ground truth masks from the dataset

As shown in Figure 1, the dataset presents representative sample images. Overall, it provides a comprehensive, high-resolution, and well-annotated resource suitable for training and validating deep learning models aimed at detecting burned areas from satellite imagery. Its structure supports various applications in remote sensing, environmental monitoring, and wildfire management.

2.2. Data Preprocessing

The image dataset and its corresponding burned area mask undergo a series of preprocessing stages prior to model training. First, a random subspace band selection technique is applied, where 3 out of the available 8 spectral bands are randomly chosen to reduce dimensionality and enhance model generalization [31]. Subsequently, the selected image bands are normalized to a pixel value range of [0, 1] in float32 format, enabling stable gradient flow during training, following common practice in remote-sensing segmentation pipelines [1], [5].

The burned area mask, originally stored in 8-bit format with values of 0 and 255, is binarized into {0, 1} to match the expected binary output of the segmentation model [5]. Following normalization and binarization, both image and mask pairs are patched into smaller regions of 256 × 256 pixels using a sliding-window approach from the original 512 × 512 resolution, which increases the number of training samples and allows the model to focus on local spatial patterns [5], [8]. At the end of preprocessing, each training sample consists of a 256 × 256 image with 3 channels and its corresponding 256 × 256 binary masks indicating the burned area.

2.3. Implementation Details

The framework integrates Random Subspace Band Selection (RSBS) with an ensemble of U-Net encoders to segment burned areas from Landsat-8 imagery [31], [32]. For each training run, a random subset of three spectral bands is sampled from all available channels, constrained to include at least one spectrally informative band (NIR or SWIR). Five U-Net models are trained independently, each using a different encoder backbone (ResNet50, ResNet34, DenseNet121, MobileNetV2, InceptionV4) and a distinct 3-band subset produced by RSBS. All encoders are initialized with ImageNet pretrained weights [36]. The training configuration shared across all backbones is summarized in Table 2.

Table 2. Training configuration (all backbones)

Parameters	Values
Number of epochs	50
Batch size	8
Optimizer	AdamW
Initial learning rate	3×10^{-4}
Weight decay	1×10^{-4}
LR scheduler	Cosine annealing
Minimum learning rate	1×10^{-6}
Loss function	Dice + Focal
Early stopping patience	7 epochs

At inference time, the five predicted masks are aggregated via hard majority voting at the pixel level (a pixel is labeled as burned if at least 3 of 5 models predict burned). Ensembling is applied only during evaluation, following findings that ensemble U-Net variants can improve robustness in remote sensing segmentation tasks [27-29]. Performance is reported using IoU, mIoU, F1, Recall, and Precision. The design is modular and transferable to other sensors or regions with minimal adjustments. An overview of the proposed framework, including its integration of Random Subspace Band Selection (RSBS), heterogeneous encoder backbones, and pixel-level ensemble aggregation, is illustrated in Figure 2.

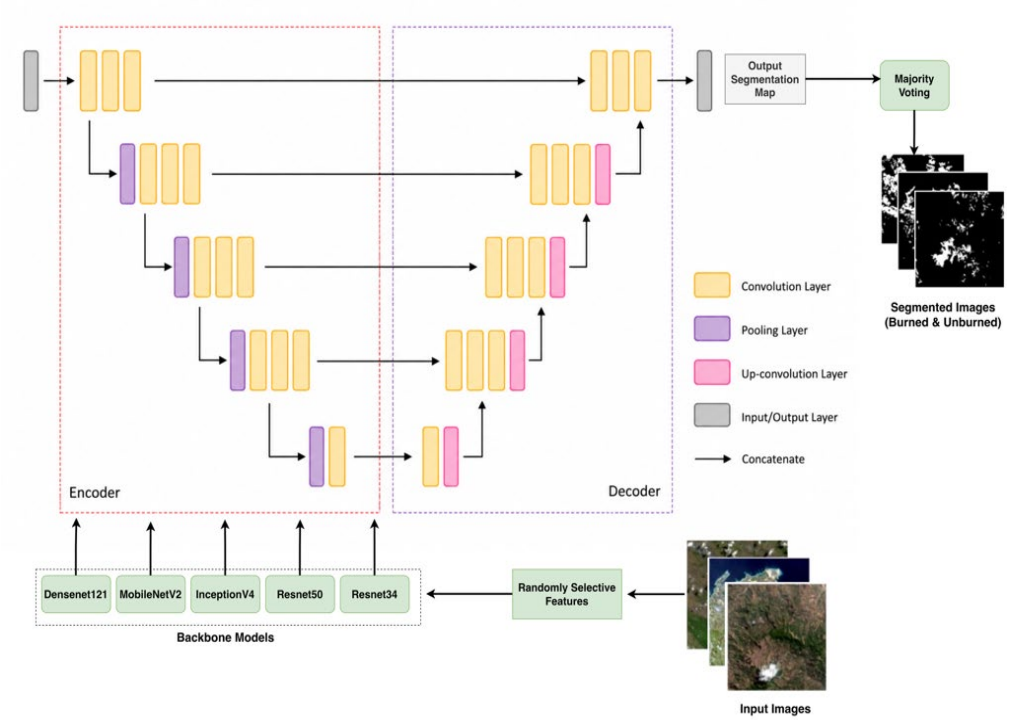


Figure 2. The proposed U-Net with an ensemble encoder architecture and Random Subspace Band Selection (RSBS) for burned area segmentation.

As shown in Figure 2, the architecture retains the traditional U-Net encoder–decoder structure while introducing spectral and architectural diversity through multiple parallel encoder streams, each trained on a distinct spectral subspace. This configuration allows the model to capture complementary spatial–spectral patterns from different band combinations, ultimately enhancing segmentation robustness against spectral redundancy, illumination variation, and land-cover heterogeneity commonly observed in wildfire environments.

2.4. Backbone Models

To enhance the feature extraction capability of U-Net, five encoder backbones are integrated into separate U-Net variants: ResNet34, ResNet50, DenseNet121, MobileNetV2, and InceptionV4. All backbones are pretrained on ImageNet [33] and inserted into the encoder path of U-Net [32] while preserving the standard decoder to ensure architectural consistency and stable optimization during training.

These architectures were selected to introduce complementary representational strengths and to better accommodate the spectral diversity produced by the Random Subspace strategy. ResNet34 provides an efficient residual framework suitable for moderate model complexity, whereas ResNet50 offers deeper bottleneck blocks for richer hierarchical features [32]. DenseNet121 promotes strong gradient flow and feature reuse, which is advantageous for subtle spectral–spatial distinctions [35]. MobileNetV2 contributes a lightweight option with inverted residuals and depthwise separable convolutions [36], and InceptionV4 incorporates multi-scale convolutions that capture both fine and coarse spatial structures [37]. Each backbone is paired with a separate U-Net model trained on the same 3-band input subset, and their predictions are later fused through the ensemble mechanism. This backbone-level heterogeneity complements the spectral diversity from the Random Subspace strategy, ultimately improving robustness, reducing sensitivity to individual subspace limitations, and enhancing overall segmentation performance across varied burned- area conditions [27-29].

2.5. Random Subspace Band Selection (RSBS)

The Random Subspace Method (RSM) is employed as a lightweight feature selection and ensemble strategy that increases spectral diversity without incurring substantial computational cost [37]. Instead of training a single model on the full multispectral stack, RSBS exposes each model to a different subset of Landsat-8 bands, allowing the ensemble to learn complementary spectral cues associated with burn severity, vegetation moisture loss, and background land-cover variability [1], [2], [3]. Input features use eight bands, preserving consistency with the preprocessing described earlier and enabling additional reflectance diversity beneficial for challenging post-fire conditions.

All 3-band combinations were first enumerated ($\binom{8}{3} = 56$) after which 15 triplets were randomly selected under an informativeness constraint requiring that each triplet include at least one of Red (B4), NIR (B5), or SWIR-1 (B6). These bands are repeatedly identified as among the most responsive to fire-induced reflectance changes across burned-area literature [1], [4], [24], [25]. Each

triplet was then used to train five U-Net models with different encoder backbones [32], [35], [37],

introducing diversity at both spectral-input and architectural levels.

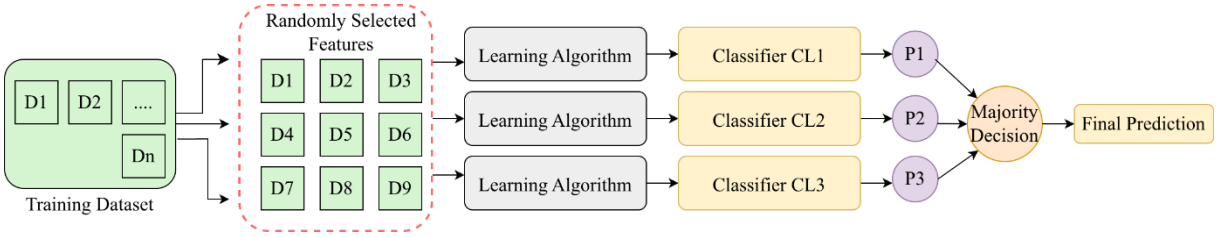


Figure 3. The proposed Random Subspace Band Selection (RSBS) framework for burned-area segmentation

As illustrated in Figure 3, the workflow begins with the training dataset, from which different spectral subspaces (3-band combinations) are randomly generated. Each selected band subset is independently processed by a learning module consisting of a U-Net model with a specific encoder backbone. The predictions produced by these parallel models are then passed to a majority decision module that aggregates the outputs. The final prediction is obtained by combining these independent segmentation results, producing a more stable and spatially consistent burned-area map.

2.6. Evaluation Metrics

To evaluate the performance of the burned-area segmentation model, we used five commonly adopted metrics in remote sensing segmentation tasks [1, 3, 5]: Intersection over Union (IoU), Mean IoU (mIoU), F1 Score, Precision, and Recall.

The Intersection over Union (IoU), also called the Jaccard Index, measures the overlap between the predicted mask and the ground-truth mask It is defined as Equation 1.

$$\text{IoU} = \frac{\text{TP}}{\text{TP+FP+FN}} \tag{1}$$

Here, TP denotes the number of true positive pixels, FP denotes false positive pixels, and FN denotes false negative pixels.

The Mean IoU (mIoU) is obtained by averaging the IoU values across C classes. In binary burned-area segmentation (burned vs. background), mIoU helps assess performance on both majority and minority classes. It is computed as Equation 2.

$$\text{mIoU} = \frac{1}{C}\sum_{i=1}^{C}\frac{\text{TP}_i}{\text{TP}_i+\text{FP}_i+\text{FN}_i} \tag{2}$$

The F1 Score represents the harmonic mean between Precision and Recall and emphasizes the balance between omission and commission errors. It is defined as Equation 3.

$$\text{F1} = \frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision+Recall}} \tag{3}$$

Precision measures the proportion of predicted burned pixels that are correct, while Recall indicates the proportion of actual burned pixels successfully detected by the model. All predictions were thresholded at 0.5 after sigmoid activation, and computations were performed using the scikit-learn library. These metrics provide a comprehensive evaluation of segmentation accuracy, class sensitivity, and error behavior.

2.7. Joint Loss Function

Burned-area segmentation is inherently an imbalanced classification problem, as most pixels in satellite imagery belong to the background class. This imbalance can bias the model toward the majority class and reduce its ability to accurately identify burned areas. To address this issue, a combined loss function consisting of Dice Loss and Focal Loss was employed, as both have shown effectiveness in handling class imbalance in semantic segmentation tasks.

Dice Loss directly evaluates the overlap between predicted and reference burned-area regions and is formulated as Equation 4.

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2\sum_i p_i r_i}{\sum_i p_i + \sum_i r_i + \epsilon} \tag{4}$$

In this expression, p_i represents the predicted probability for pixel i, r_i is the corresponding ground-truth label (0 or 1), and ϵ is a small constant used to avoid division by zero.

Focal Loss reduces the dominance of easy negative examples by focusing more on difficult-to-classify pixels. It is defined as Equation 5.

$$\mathcal{L}_{\text{Focal}} = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{5}$$

Here, p_t denotes the predicted probability for the true class, γ is the focusing parameter (commonly set to 2), and α_t is a balancing coefficient.

The final loss used during training is the sum of both components, as defined in Equation 6.

$$\mathcal{L}_{\text{Joint}} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{Focal}} \tag{6}$$

This joint formulation effectively balances region overlap optimization and class-imbalance handling, resulting in more stable and accurate segmentation performance.

3. Results and Discussions

This section presents the experimental setup and results of the proposed framework. We begin by describing the dataset and implementation details, followed by the evaluation metrics used. We then analyze the performance of individual band subsets and discuss the

impact of ensemble integration through both quantitative and qualitative results.

3.1 Band Subset Evaluation

To assess how different spectral band combinations affect segmentation performance, we adopted a Random Subspace strategy by selecting multiple three-band combinations from the eight-band Landsat-8 imagery. A total of fifteen random subsets were evaluated. This approach was designed to simulate practical situations in which only a limited number of spectral channels are available while also evaluating the relative contribution of different spectral regions to burned-area detection. By using multiple band subsets, the study aims to identify the most informative spectral combinations for distinguishing burned and unburned surfaces.

Each band subset was used as input to an identical U-Net architecture. All training settings, hyperparameters, and optimization strategies were kept constant throughout the experiments. Table 3 presents the Dice scores achieved by the 15 subsets. Subsets 5 [2, 4, 5] and 2 [1, 4, 6] performed the best, with Dice scores of 0.5622 and 0.5172 respectively. These subsets included NIR and SWIR bands, which are known to enhance burned area discrimination. In contrast, subsets 1 [0, 1, 5] and 12 [1, 2, 3] were among the worst, yielding Dice scores of 0.0858 and 0.0000, indicating limited spectral.

These findings confirm that NIR and SWIR bands are sensitive to changes caused by fire and can help distinguish burned areas from unburned areas. The results also show that selecting the right band combination is important, especially when using satellite images with limited spectral bands, because it can affect the model performance and segmentation results.

Table 3. Dice scores of 15 random band subsets

Subset ID	Selected Bands	Mapped Bands	Dice Score
1	[0, 1, 5]	[1, 2, 6]	0.0858
2	[1, 4, 6]	[2, 5, 7]	0.5172
3	[0, 5, 6]	[1, 6, 7]	0.3253
4	[4, 5, 7]	[5, 6, 8]	0.4698
5	[2, 4, 5]	[3, 5, 6]	0.5622
6	[0, 3, 6]	[1, 4, 7]	0.3005
7	[4, 6, 7]	[5, 7, 8]	0.4513
8	[1, 4, 7]	[2, 5, 8]	0.3299
9	[2, 3, 5]	[3, 4, 6]	0.3586
10	[3, 4, 7]	[4, 5, 8]	0.3884
11	[2, 3, 4]	[3, 4, 5]	0.3435
12	[1, 2, 3]	[2, 3, 4]	0.0000
13	[1, 3, 5]	[2, 4, 6]	0.1100
14	[0, 4, 5]	[1, 5, 6]	0.4823
15	[1, 3, 7]	[2, 4, 8]	0.1231

3.2. Performance Comparison Across Band Subsets and Encoders

To investigate the impact of spectral band selection on burned area segmentation performance, we conducted a detailed evaluation using the two best performing and two worst performing three-band subsets identified during preliminary experiments. Each subset was evaluated using five pretrained encoder backbones within the U-Net framework, along with an ensemble model that integrates their outputs. The evaluation includes both quantitative metric analysis and qualitative visual inspection of sample predictions.

A comparative evaluation of the five individual backbones and the ensemble model was performed across four selected spectral band combinations. Table 4 presents the segmentation results for subset 5 [2,4,5], Table 5 reports the results for subset 2 [1,4,6], Table 6 shows the performance for subset 12 [1,2,3], and Table 7 presents the results for subset 1 [0,1,5]. Standard metrics (IoU, mIoU, F1, Recall, Precision) are reported to evaluate how backbone choice and spectral subspace jointly influence performance under post-fire conditions. This evaluation highlights the robustness and consistency of each backbone across different spectral inputs, showing how band combinations affect segmentation performance.

Table 4. Results on subset 5 [2,4,5] (best performing)

Model	IoU	mIoU	F1	Recall	Precision
ResNet50	0.7491	0.8635	0.8565	0.8445	0.8689
Resnet34	0.7536	0.8657	0.8595	0.8716	0.8478
DenseNet121	0.7488	0.8633	0.8564	0.8504	0.8624
MobileNetV2	0.7497	0.8636	0.8570	0.8663	0.8479
InceptionV4	0.6922	0.8315	0.8181	0.8425	0.7951
Ensemble	0.7673	0.8734	0.8683	0.8708	0.8659

Table 5. Results on subset 2 [1,4,6] (second best)

Model	IoU	mIoU	F1	Recall	Precision
ResNet50	0.7724	0.8759	0.8716	0.8952	0.8492
ResNet34	0.7748	0.8774	0.8731	0.8837	0.8628
DenseNet121	0.7771	0.8785	0.8746	0.8984	0.8519
MobileNetV2	0.7645	0.8716	0.8665	0.8843	0.8494
InceptionV4	0.6271	0.7972	0.7708	0.7173	0.8330
Ensemble	0.7846	0.8828	0.8793	0.8882	0.8705

Table 6. Results on subset 12 [1,2,3] (poor performing)

Model	IoU	mIoU	F1	Recall	Precision
ResNet50	0.6099	0.7862	0.7577	0.7604	0.7550
ResNet34	0.5679	0.7588	0.7244	0.8490	0.6317
DenseNet121	0.6362	0.8020	0.7777	0.7317	0.8298
MobileNetV2	0.6307	0.7958	0.7735	0.8588	0.7037
InceptionV4	0.2144	0.5674	0.3530	0.2915	0.4475
Ensemble	0.6611	0.8148	0.7960	0.7930	0.7990

Table 7. Results on subset 1 [0,1,5] (worst performing)

Model	IoU	mIoU	F1	Recall	Precision
ResNet50	0.2430	0.5627	0.3910	0.4986	0.3216
ResNet34	0.1872	0.5226	0.3153	0.4338	0.2477
DenseNet121	0.2586	0.5898	0.4110	0.3683	0.4649
MobileNetV2	0.1354	0.5122	0.2384	0.2341	0.2429
InceptionV4	0.1368	0.5120	0.2407	0.2411	0.2404
Ensemble	0.2336	0.5756	0.3787	0.3358	0.4342

The best performing ensemble configuration was obtained from subset 5 [2,4,5], as shown in Table 4, which corresponds to bands [3,5,6] and includes Green, NIR (near infrared), and SWIR-1 (short-wave infrared). This ensemble achieved an IoU of 0.7673, mIoU of 0.8734, and F1 score of 0.8683, indicating a strong

ability to delineate burned areas, particularly because NIR and SWIR bands are highly sensitive to vegetation condition and moisture. A close second was subset 2 [1,4,6], as shown in Table 5, which produced the highest mIoU (0.8828) and F1 score (0.8793). Although its IoU was slightly lower, this configuration demonstrated excellent generalization, especially in capturing small-scale burned regions and reducing class imbalance, as reflected in its higher Recall.

The effectiveness of the ensemble models was strongly influenced by the composition of the spectral subspace. Subsets that included SWIR and NIR consistently outperformed those dominated by visible bands. The Random Subspace strategy helped promote model diversity through varied spectral inputs, reduced overfitting by training on different feature perspectives, and allowed explicit testing of ensemble robustness against weak or redundant input bands. However, this approach is not inherently robust, when the selected subspace contains less informative bands, such as in subset 1 [0,1,5], as shown in Table 7, which combines Coastal Aerosol, Blue, and SWIR-1, ensemble performance dropped substantially, with an IoU of only 0.2336 and F1 score of 0.3787. This result confirms that purely random selection without spectral consideration does not guarantee effective segmentation.

Interestingly, even in the poor-performing subset 12 [1,2,3], as shown in Table 6, the ensemble still produced substantially better results than most individual backbones. InceptionV4 alone yielded an IoU of just 0.2144 and F1 score of 0.3530, whereas the ensemble reached an IoU of 0.6611 and F1 score of 0.7960, nearly doubling the performance of the weakest model. This behavior illustrates an important advantage of ensemble methods, as they can compensate for the weaknesses of poorly performing models and stabilize overall predictions. In this context, the ensemble acts as a robust aggregator of heterogeneous learners, some of which may be suboptimal in isolation, in line with the ensemble learning principle that multiple weak or moderately strong learners can outperform a single complex model. By smoothing out errors from individual networks, the ensemble produces more reliable segmentation, even when the underlying spectral subspace is not optimal.

Performance trends across encoder backbones show that ResNet50 consistently achieved strong IoU and F1 scores, supported by its depth and ability to capture rich spatial features. ResNet34 delivered well-balanced results across all metrics, making it a stable and reliable choice. DenseNet121 obtained the highest Recall in nearly all settings, indicating strong sensitivity to burned regions. MobileNetV2 was less effective as a standalone model but contributed positively when combined with other encoders in the ensemble. In contrast, InceptionV4 repeatedly produced the lowest scores across all subsets, suggesting limited suitability for this segmentation task.

To build robust burned area segmentation pipelines, it is therefore advisable to construct encoder ensembles that include DenseNet121, ResNet34, and ResNet50, while excluding InceptionV4 due to its consistently weak performance. Spectral subspaces should prioritize combinations that involve Red, NIR, and SWIR bands to exploit their strong contrast between burned and unburned surfaces. The experimental results show that combining informed spectral selection with architectural diversity through ensemble modeling not only improves accuracy but also enhances robustness across varying spectral conditions and environmental settings, leading to more reliable and generalizable burned area segmentation.

The qualitative results for each selected subset, including the input image, ground truth mask, and predictions from ResNet50, ResNet34, DenseNet121, MobileNetV2, InceptionV4, and the ensemble model, are presented to facilitate comprehensive visual evaluation. These visualizations enable side-by-side comparison and highlight the strengths and weaknesses of each encoder.

The visual comparison highlights differences in how each encoder captures spatial patterns and boundary details. The ensemble produces more consistent predictions by combining the strengths of individual encoders.

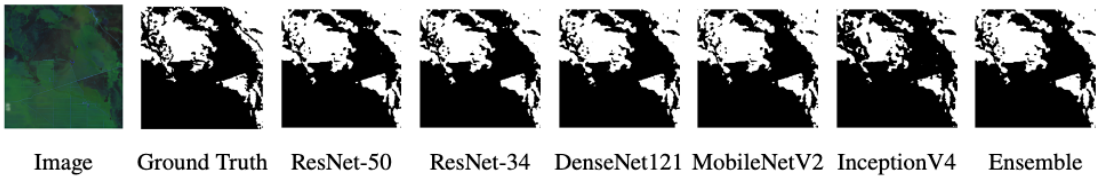


Figure 4. Segmentation results on Subset 5 [2,4,5] — best performing

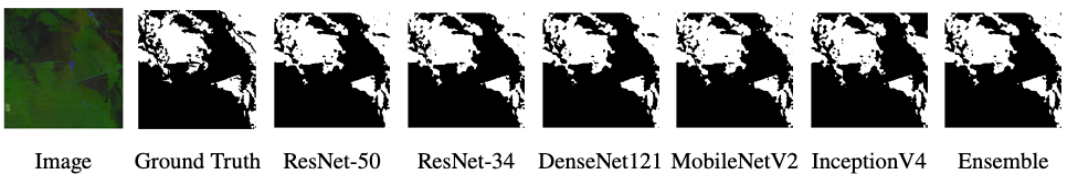


Figure 5. Segmentation results on Subset 2 [1,4,6] — second best

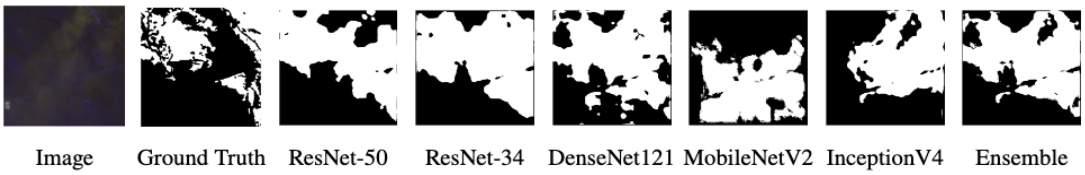


Figure 6. Segmentation results on Subset 1 [0,1,5] — worst performing

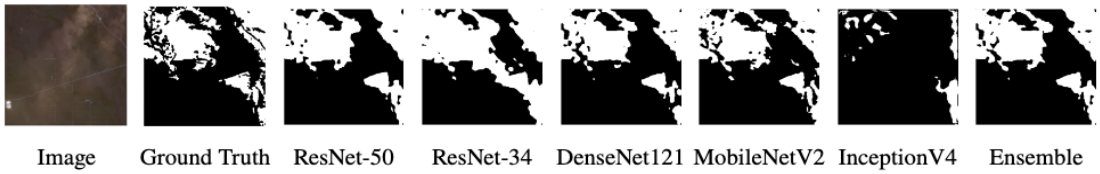


Figure 7. Segmentation results on Subset 12 [1,2,3] — poor performing

For subset 5 [2,4,5], as shown in Figure 4, the input combines Green, NIR, and SWIR1 bands, which provide strong spectral contrast between burned and unburned areas. Visually, the predictions from ResNet34, ResNet50, and DenseNet121 are very close to the ground truth, with sharp boundaries and little noise. MobileNetV2 also performs well, although it tends to slightly over segment thin burn edges, while InceptionV4 produces oversmoothed masks with noticeable false negatives. The ensemble output offers the most balanced result, capturing fine burn patterns while suppressing noise, consistent with this subset's highest F1 score and IoU in the quantitative evaluation.

For subset 2 [1,4,6], as shown in Figure 5, the band combination again includes SWIR and NIR channels. All models perform well, with DenseNet121 and ResNet50 delineating burn boundaries cleanly. MobileNetV2 and ResNet34 preserve fine details but tend to overpredict burned areas in some regions, whereas InceptionV4 underperforms with fragmented patches and blurred boundaries. The ensemble result shows crisp and coherent segmentation across both small and large burned patches, matching this subset's highest mIoU and F1 score.

Subset 1 [0,1,5], as shown in Figure 6, combines Coastal Aerosol, Blue, and SWIR1 bands, which provide limited spectral information for burn discrimination. All individual backbones produce noisy, fragmented masks with frequent misclassification. DenseNet121 correctly identifies some larger burned patches but with blurred edges, and the outputs from InceptionV4 and MobileNetV2 produce highly fragmented and inconsistent predictions.

Subset 12 [1,2,3], as shown in Figure 7, consists mainly of visible bands (Blue, Green, Red), which leads to weaker spectral separation and suboptimal results. Burned features are harder to distinguish in the input image, and individual models often fail to produce coherent predictions. InceptionV4 and ResNet34 generate many false positives, while MobileNetV2 misses a substantial portion of burned regions. DenseNet121 performs somewhat better by capturing larger burned areas but still loses finer structures. The ensemble output is more organized than any single model but remains insufficiently precise.

The ensemble result is more spatially consistent than the individual models yet still fails to reproduce the true burned extent accurately, reflecting this subset's poor quantitative performance.

4. Conclusions

Building upon the motivation and architecture introduced earlier, this research presents a framework for burned area segmentation by combining Random Subspace Band Selection with an ensemble of U-Net models that use diverse encoder backbones. The proposed framework addresses two critical challenges in remote sensing-based fire mapping, namely redundant or less informative spectral bands in multispectral imagery and limited generalization of single backbone segmentation networks. Experiments show that the combination of Red, NIR, and SWIR bands consistently outperforms other settings, while the ensemble approach achieves better performance than individual models and shows enhanced generalization across varied land cover types. The ensemble model achieved strong performance with an IoU of 0.7673, mIoU of 0.8734, and F1 score of 0.8683. Key contributions include a modular ensemble architecture that integrates spectral and architectural diversity and a practical application of Random Subspace Band Selection for input generation. Future work will focus on extending the generalizability of this framework through cross-region and cross-sensor validation using datasets such as Sentinel-2 and MODIS.

References

[1] X. Hu, Y. Ban, and A. Nascetti, "Uni-temporal multispectral imagery for burned area mapping with deep learning," Remote Sensing, vol. 13, no. 8, Art. no. 1509, 2021, doi: 10.3390/rs13081509.

[2] D. Lee, S. Son, J. Bae, S. Park, J. Seo, D. Seo, Y. Lee, and J. Kim, "Single-temporal Sentinel-2 for analyzing burned area detection methods: A study of 14 cases in Republic of Korea considering land cover," Remote Sensing, vol. 16, no. 5, Art. no. 884, 2024, doi: 10.3390/rs16050884.

[3] Q. Zhang, L. Ge, R. Zhang, G. I. Metternicht, Z. Du, J. Kuang, and M. Xu, "Deep-learning-based burned area mapping using the synergy of Sentinel-1&2 data," Remote Sensing of

Environment, vol. 264, Art. no. 112575, 2021, doi: 10.1016/j.rse.2021.112575.
[4] B. Kim, K. Lee, and S. Park, "Burned-area mapping using post-fire PlanetScope images and a convolutional neural network," Remote Sensing, vol. 16, no. 14, Art. no. 2629, 2024, doi: 10.3390/rs16142629.
[5] L. Knopp, M. Wieland, M. Rättich, and S. Martinis, "A deep learning approach for burned area segmentation with Sentinel-2 data," Remote Sensing, vol. 12, no. 15, Art. no. 2422, 2020, doi: 10.3390/rs12152422.
[6] Z. Shirvani, O. Abdi, and R. C. Goodman, "High-resolution semantic segmentation of woodland fires using residual attention UNet and time series of Sentinel-2," Remote Sensing, vol. 15, no. 5, Art. no. 1342, 2023, doi: 10.3390/rs15051342.
[7] D. Alkan and L. Karasaka, "Image segmentation for burned area detection from satellite imagery using the U-Net deep learning model," Bollettino di Geofisica Teorica ed Applicata, vol. 65, no. 4, pp. 649–674, 2024, doi: 10.4430/bgo00472.
[8] I. Mancilla-Wulff, J. Carrasco, C. Pais, A. Miranda, and A. Weintraub, "Two scalable approaches for burned-area mapping using U-Net and Landsat imagery," arXiv preprint arXiv:2311.17368, 2023, doi: 10.48550/arXiv.2311.17368.
[9] E. Arnaudo, L. Barco, M. Merlo, and C. Rossi, "Robust burned area delineation through multitask learning," *Communications in Computer and Information Science*, vol. 2135, pp. 436–447, 2025, doi: 10.1007/978-3-031-74633-8_32.
[10] R. Ghali and M. A. Akhloufi, "Deep learning approaches for wildland fires using satellite remote sensing data: Detection, mapping, and prediction," Fire, vol. 6, no. 5, Art. no. 192, 2023, doi: 10.3390/fire6050192.
[11] P. Zhang, Y. Ban, and A. Nascetti, "Total-variation regularized U-Net for wildfire burned area mapping based on Sentinel-1 C-Band SAR backscattering data," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 203, pp. 301–313, 2023, doi: 10.1016/j.isprsjprs.2023.07.024.
[12] T. Sui, Q. Huang, M. Wu, M. Wu, and Z. Zhang, "BiAU-Net: Wildfire burnt area mapping using bitemporal Sentinel-2 imagery and U-Net with attention mechanism," International Journal of Applied Earth Observation and Geoinformation, vol. 132, Art. no. 104034, 2024, doi: 10.1016/j.jag.2024.104034.
[13] E. Khankeshizadeh, S. Tahermanesh, A. Mohsenifar, A. Moghimi, and A. Mohammadzadeh, "FBA-DPAttResU-Net: Forest burned area detection using a novel end-to-end dual-path attention residual-based U-Net from post-fire Sentinel-1 and Sentinel-2 images," Ecological Indicators, vol. 167, Art. no. 112589, 2024, doi: 10.1016/j.ecolind.2024.112589.
[14] W. Bo, J. Liu, X. Fan, T. Tjahjadi, Q. Ye, and L. Fu, "BASNet: Burned area segmentation network for real-time detection of damage maps in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, Art. no. 5627913, 2022, doi: 10.1109/TGRS.2022.3197647.
[15] S. Liu, Y. Xue, H. Chen, Y. Chen, and T. Zhan, "Segmentation of Any Fire Event (SAFE): A rapid and high-precision approach for burned area extraction using Sentinel-2 imagery," Remote Sensing, vol. 17, no. 1, Art. no. 54, 2025, doi: 10.3390/rs17010054.
[16] S. T. Seydi, M. Hasanlou, and J. Chanussot, "Burnt-Net: Wildfire burned area mapping with single post-fire Sentinel-2 data and deep learning morphological neural network," Ecological Indicators, vol. 140, Art. no. 108999, 2022, doi: 10.1016/j.ecolind.2022.108999.
[17] S. T. Seydi and M. Sadegh, "Improved burned area mapping using monotemporal Landsat-9 imagery and convolutional shift-transformer," Measurement, vol. 216, Art. no. 112961, 2023, doi: 10.1016/j.measurement.2023.112961.
[18] S. Balakavi, V. Vadrevu, and K. Lasko, "Exploring burnt area delineation with cross-resolution mapping: A case study of very high and medium-resolution data," Sensors, vol. 25, no. 10, Art. no. 3009, 2025, doi: 10.3390/s25103009.
[19] X. Xi, M. Kang, L. Dai, Y. Jing, P. Han, and C. Hou, "Identifying forest burned area using a deep learning model based on post-fire optical and SAR remote sensing images," IEEE Access, vol. 12, pp. 188102–188113, 2024, doi: 10.1109/ACCESS.2024.3515205.
[20] L. Liu, Y. Guo, E. Chen, Z. Li, Y. Li, Y. Liu, Q. Zhang, and B. Wang, "A lightweight Deeplab V3+ network integrating deep transitive transfer learning and attention mechanism for burned area identification," *Scientific Reports*, vol. 15, Art. no. 15969, 2025, doi: 10.1038/s41598-024-66060-7.
[21] D. Stavrakoudis and I. Gitas, "Object-based burned area mapping with extreme gradient boosting using Sentinel-2 imagery," Journal of Geographic Information System, vol. 15, pp. 53–72, 2023, doi: 10.4236/jgis.2023.151004.
[22] A. K. Brand and A. Manandhar, "Semantic segmentation of burned areas in satellite images using a U-Net-based convolutional neural network," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLIII-B3-2021, pp. 47–53, 2021, doi: 10.5194/isprs-archives-XLIII-B3-2021-47-2021.
[23] R. N. Vasconcelos et al., "Fire detection with deep learning: A comprehensive review," Land, vol. 13, no. 10, Art. no. 1696, 2024, doi: 10.3390/land13101696.
[24] D. M. Nelson, Y. He, and G. W. K. Moore, "Trends and applications in wildfire burned area mapping: Remote sensing data, cloud geoprocessing platforms, and emerging algorithms," Geomatica, vol. 76, no. 1, Art. no. 100008, 2024, doi: 10.1016/j.geomat.2024.100008.
[25] C. Lee, S. Park, T. Kim, S. Liu, M. N. Md Reba, J. Oh, and Y. Han, "Machine learning-based forest burned area detection with various input variables: A case study of South Korea," Applied Sciences, vol. 12, no. 19, Art. no. 10077, 2022, doi: 10.3390/app121910077.
[26] M. M. Pinto, R. Libonati, R. M. Trigo, I. F. Trigo, and C. C. DaCamara, "A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 160, pp. 260–274, 2020, doi: 10.1016/j.isprsjprs.2019.12.014.
[27] S. Sivanandham and D. B. Gunaseelan, "Development of an ensembled meta-deep learning model for semantic road-scene segmentation in an unstructured environment," Applied Sciences, vol. 12, no. 23, Art. no. 12214, 2022, doi: 10.3390/app122312214.
[28] N. Soucy and S. Y. Sekeh, "CEU-Net: Ensemble semantic segmentation of hyperspectral images using clustering," *Journal of Big Data*, vol. 10, no. 1, Art. no. 43, 2023, doi: 10.1186/s40537-023-00718-3.
[29] I. Dimitrovski, V. Spasev, S. Loshkovska, and I. Kitanovski, "U-Net ensemble for enhanced semantic segmentation in remote sensing imagery," Remote Sensing, vol. 16, no. 12, Art. no. 2077, 2024, doi: 10.3390/rs16122077.
[30] Y. Prabowo, A. D. Sakti, K. A. Pradono, Q. Amriyah, F. H. Rasyidy, I. Bengkulah, K. Ulfa, D. S. Candra, M. T. Imdad, and S. Ali, "Deep learning dataset for estimating burned areas: Case study, Indonesia," Data, vol. 7, no. 6, Art. no. 78, 2022, doi: 10.3390/data7060078.
[31] S. Deegalla, K. Walgama, P. Papapetrou, and H. Boström, "Random subspace and random projection nearest neighbor ensembles for high dimensional data," Expert Systems with Applications, vol. 191, Art. no. 116078, 2022, doi: 10.1016/j.eswa.2021.116078.
[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.
[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proc. CVPR, 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. CVPR, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[35] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. CVPR, 2017, pp. 4700–4708, doi: 10.1109/CVPR.2017.243.
[36] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in Proc. CVPR, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.
[37] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual

connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, pp. 4278–4284, 2017, doi: 10.1609/aaai.v31i1.11231.