Deep Learning Framework for Early Fire and Smoke Detection

1st Sireesha Moturi
Assoc. Professor, Dept of CSE
Narasaraopeta Engineering College
Narasaraopet, Palnadu, India
sireeshamoturi@gmail.com

2nd Kata Ramya
Student, Dept of CSE
Narasaraopeta Engineering College
Narasaraopet, Palnadu, India
kataramya29@gmail.com

3rd Kota Supriya
Student, Dept of CSE
Narasaraopeta Engineering College
Narasaraopet, Palnadu, India
supriyakotareddy@gmail.com

4th Chennupati Jyothika
Student, Dept of CSE
Narasaraopeta Engineering College
Narasaraopet, Palnadu, India
jyothikachennupati19@gmail.com

5th Kondabathini Vydurya
Student, Dept of CSE
Narasaraopeta Engineering College
Narasaraopet, Palnadu, India
18471A05K3@nrtec.in

6th Mothe Suneetha
Asst. Professor, Dept of CSE
Narasaraopeta Engineering College
Narasaraopet, Palnadu, India
msuneetha973@gmail.com

***Abstract*—Most of the forest fires pose threats to ecological balance and human life, maximizing losses with respect to degrading the environment. The traditional techniques for smoke detection are less efficient because they use sensor systems, which lead to delayed fire detection. It is based on the suggestion of this research related to the idea of deep learning with pre-trained convolutional neural networks (VGG16, InceptionV3, and Xception) for early fire and smoke detection. The models are tested on two datasets: one is public, and the other is contributed by the authors in this paper. Substantial improvements were found in experimental results, as well as the best detection accuracy reached 94% during feature extraction and 89% after fine-tuning for InceptionV3. Such a finding proves the full capability of CNN-based approaches to provide real-time fire detection solutions, with their ability to make fire detection systems more efficient and to better enhance early response capabilities toward reducing the severe impact of forest fires.**

***Index Terms*—Forest fire detection, Smoke detection, Pre-trained models, Hyper parameter optimization.**

I. Introduction

Wildfires significantly contribute to environmental, economic, and social concerns globally. They also destroy ecosystems and reduce diversity in nature. They emit vast amounts of carbon dioxide into the air. There is often a threat to human life and property. Recently, countries like the USA, Australia, Indonesia, and India have experienced different types of wildfires in the recent past. Contributing factors include climate changes and adverse land use [1]. This rising threat necessitates the urgent need for reliable fire detection systems that will offer real-time operations to avert all the destruction associated with wild fires.

Traditional methods applied in fire detection include satellite-based thermal imaging, sensor networks, and observer-based approaches. However, such approaches are known to be prone to false alarms, high latency, and low accuracy under adversarial conditions, such as dense smoke or if the fires are small in size [2]. In addition, sensor-based solutions create a heavy infrastructure setup in remote forest areas. Based on the shortcomings of conventional methods, it is essential to look for a smarter, more intuitive solution that can detect the presence of fires with high accuracy and at an early stage.

Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have had a significant effect on image-based fire detection systems. CNNs automatically learn features from images, which is most useful in the detection of complex patterns such as flames and smoke. Extremely high success in classification tasks based on images has been achieved with architectures like VGG16, InceptionV3, and Xception by exploiting transfer learning [3]. These architectures enable models to apply knowledge acquired from large-scale datasets, so that fire and smoke detection becomes even possible in unseen environments. It further reduces the computational cost and time required to train but also enhances the accuracy of detection.

This research further enhances fire and smoke detection through transfer learning by applying a pre-trained DenseNet model. DenseNet has been used because it maintains such a high accuracy level with efficient computation via feature reuse across layers [4]. Wavelet transform techniques have also been utilized in order to enhance the quality of the images and extract features in a better way so that there can be analysis in seven different frequency bands. This is because finer features in an image would be captured which include the slight gradations between the smoke, fire, and non-fire elements within forest environments.

The motivation behind this work is due to the alarming frequency and severity of wildfires occurring around the world. Our approach is anticipated to improve the detection speed and accuracy, the paramount factors in timely intervention in forest fire scenarios. It therefore focuses on fire-prone regions such as India, Indonesia, and the USA,[5] which have caused extreme damage to their ecology and economy.

Contributions: New Detection Method: It proposes a new learning transfer from DenseNet to modified version of DenseNet by using wavelet transform for the enhanced extraction of features. High Accurateness and Speed: Our method provides speedier detection. The attainment of accuracy is heightened to be of an earlier intervention with a reduction in false alarms. Application Scenario: It focuses on the application of the model in the forest environment due to high vulnerability to wildfires and provides the scalability of the solution by adapting it towards real-time systems.

Key Contributions: We clarified why existing approaches were insufficient and how deep learning is suited to this problem Objectives: We define what the research aims to achieve. Contributions: Clearly stated the main contributions of the paper and distinguished it from prior work. Paper Organisation: Added a brief outline that lets the reader know in what order the different parts of the paper will follow and, therefore, get an idea of how this paper has been structured. This extended introduction provides better setting for the rest of your paper, and the feedback on contribution, motivation and paper structure.

II. Literature Survey

The increasing prevalence of wildfires has led to the development of various detection systems leveraging deep learning techniques. Recent studies highlight the advantages of using multimodal datasets, such as the combination of RGB and thermal images, to enhance fire detection accuracy. For instance, Chen et al. [6] demonstrated that their dataset improves the detection accuracy of fire-and-smoke scenes, outperforming traditional single-channel systems. A significant advancement in wildfire monitoring is the integration of CCTV images with weather data, proposed by Tran et al. [7]. Their forest fire response system utilizes deep learning strategies to enhance firefighting efforts, enabling real-time data analysis for quicker response times and more effective resource allocation, which is crucial for efficient wildfire management.

The CICLOPE project exemplifies the application of deep learning in wildfire detection using tower-mounted cameras to monitor vast areas in Portugal.l. [8] addressed initial false alarm issues in their smoke detection system by introducing a Dual-Channel Convolutional Neural Network that combines DenseNet with a Detail-selective network, resulting in improved accuracy. In the realm of remote sensing,[9] employed multisensor satellite imagery and deep semantic segmentation using CNNs to detect fire-affected areas, showcasing significant improvements even in challenging conditions, such as cloud cover.

Educational initiatives incorporating wildfire detection into machine vision curricula have also emerged. Wang et al. [10] introduced a two-step approach involving wildfire image classification with a Reduce-VGGNet model, followed by spatio-temporal feature analysis for region detection, achieving high accuracy levels. Another promising direction in flame detection is the improved Faster R-CNN model proposed by Chaoxia et al. [11]. This model enhances detection capabilities through a color-guided anchor strategy and a global information approach, which significantly boosts accuracy and efficiency.

Furthermore, innovative methods combining traditional feature extraction with deep learning techniques have been explored. For instance, Wang et al. [12] presented a system that employs the SSD network alongside ViBe background modeling, effectively reducing false alarms and enhancing detection accuracy in complex outdoor environments. The Fusion-Based Deep Learning (AFFD-FDL) Model discussed by Duhayyim et al. [13] integrates classical and modern detection methods using HOG, SqueezeNet, and Inception v3, optimized through Glowworm Swarm Optimization to improve overall detection capabilities.

Lastly, the UAVs-FFDB dataset, designed to support UAV-based fire detection and monitoring, includes over 15,000 images representing various fire conditions [14]. The development of the MHCNNFD model from this dataset achieved an impressive accuracy of 99.81%, underscoring the dataset's value for advancing UAV technology in wildfire detection.

III. Proposed Methodology

We had provided a new concept in our research through which RGB cameras are mounted on drones to capture images of real-time fires. Those images captured by the drones are classified into four classes - fire, non-fire, smoke, and fire-smoke using state-of-art deep learning models. The overall process is described in the following flowchart:

The proposed methodology is illustrated in Figure 1, which outlines the key steps involved in data collection, preprocessing, and model training.

A. Data Collection

We gather images from different sources using Google and Kaggle [**?**] along with our proprietary dataset BoWFire for building up a strong dataset. Therefore, we have images left 1,104. We decided to use 80% data for training purposes. The remaining balance 20% would be kept for validation and testing. This would make our model learn on the basis of examples and is capable of cross-verifying with appropriate amount of data. Sample images from the BoWFire dataset are shown in Figure 2, depicting different classes such as fire, no fire, smoke, and fire smoke.

B. Data Preprocessing

The pre-processing of our data increases the diversity of our dataset, and this allows for generalizability in the model. This is why we applied random rotations to our images, which we then adjusted to be between -15 and +15 degrees. It will simulate different angles at which fire might be viewed, emphasizing a robust model to detect fire and smoke in many situations.

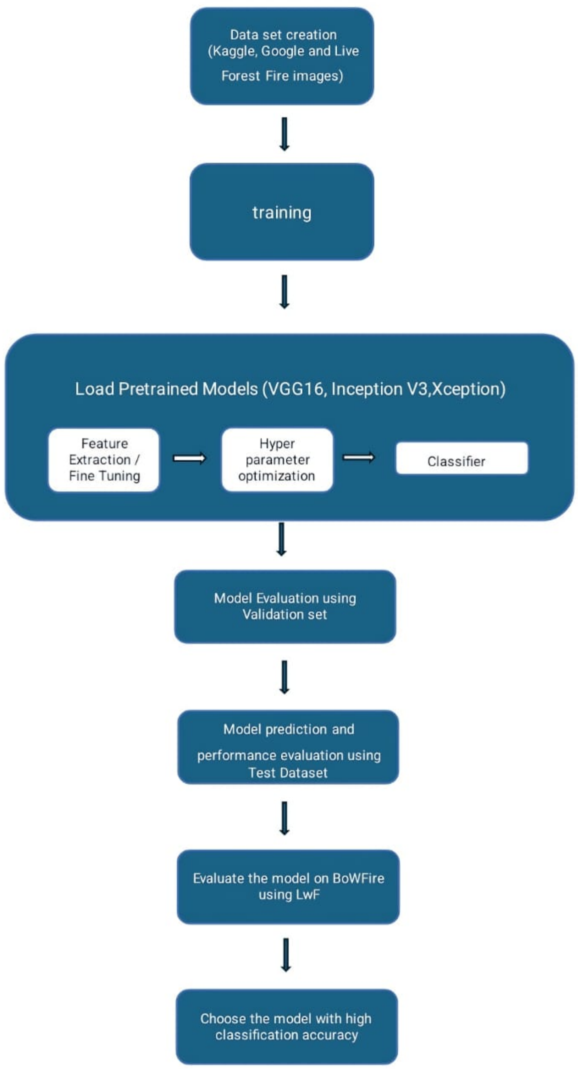

Fig. 1. Flowchart for Proposed Methodology

TABLE I
DATASET DESCRIPTION

Dataset	Fire	No Fire	Smoke	Smoke Fire	Total
Train	609	195	5	74	883
Validation	76	24	1	9	110
Test	76	24	1	10	111

C. Data Splitting

Our dataset has 1,104 samples divided into three subcategories:

- Fire: 755 images
- Fire Smoke: 100 images
- Non-Fire: 244 images

We have balanced it at 80-10-10 for splitting. 80 percent of the data is trained on, and the remaining 20 percent is again divided equally into validation and testing. Let's see how that looks in detail:

Training Set

Fig. 2. Sample Images from BoWFire Dataset: (a) Fire (b) No Fire (c) Smoke Fire (d) Smoke

- Fire: 609 images
- Fire Smoke: 74 images
- Non-Fire: 195 images
- Smoke: 5 images

Validation Set:

- Fire: 76 images
- Fire Smoke: 9 images
- Non-Fire: 24 images
- Smoke: 1 image

Test Set:

- Fire: 76 images
- Fire Smoke: 10 images
- Non-Fire: 24 images
- Smoke: 1 image

The carefully split ratios ensure that our model trains on a balanced dataset. In addition, the validation and testing datasets are unbiased. This is particularly important since we experience class imbalances that may bias the model in its learning and generalization in new data.

IV. PROPOSED DEEP LEARNING MODEL ARCHITECTURE

We test three state-of-the-art deep learning models: VGG16, InceptionV3, and Xception, specifically designed for early fire and smoke detection. For classifying the images into four major classes—fire, non-fire, fire smoke, and smoke—we used transfer learning along with fine-tuning. The architectures adapted for these experiments are described below.

A. VGG16 Architecture

VGG16 is quite efficient for all tasks of image classification. We downscaled the pre-trained VGG16 by removing its top fully connected layers and then added the following:

- **Global Average Pooling Layer:** Replaces the flattening layer to prevent overfitting.
- **Fully Connected Dense Layer:** A layer with 256 units and ELU as the activation function to enhance the flow of gradients and efficiency during training.

- **Softmax Layer:** The final dense layer consists of 4 output neurons with softmax as the activation function for multi-class classification.

During training, the early layers of VGG16 were frozen, and only the new dense layers were fine-tuned on our dataset.

The convolution operation in VGG16 can be expressed mathematically as:

$$y = f(Wx + b) \tag{1}$$

where:

- y is the result of the convolution,
- W represents the weights (filters or kernels),
- x is the input image or feature map, and
- b is the bias term.

The ReLU activation is applied as:

$$f(x) = \max(0, x) \tag{2}$$

The pooling operation to reduce spatial dimensions can be expressed as:

$$z = \max_{i,j}(x_{ij}) \tag{3}$$

Finally, the softmax function calculates the class probabilities:

$$p_i = \frac{e^{h_i}}{\sum_j e^{h_j}} \tag{4}$$

The convolution process in VGG16 is illustrated in Figure 3, showing how feature maps are generated through filters.

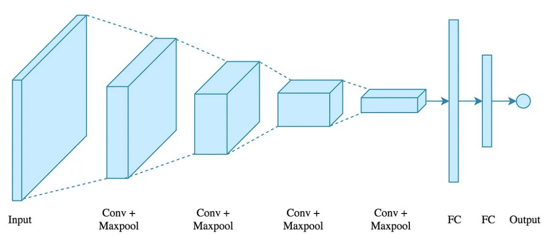

Fig. 3. Convolution operation

The max pooling process is demonstrated in Figure 4, which reduces the size of the feature maps by focusing on the most prominent features.

Fig. 4. Max pooling being applied to a feature map, reducing its size

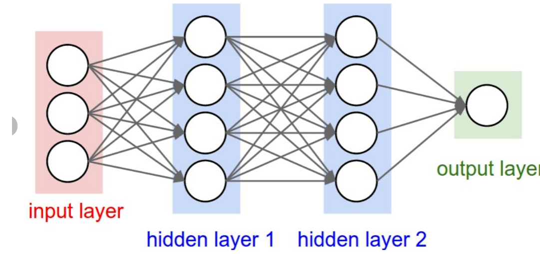

Fig. 5. Fully connected layer with neurons connected to each other

B. InceptionV3 Architecture

InceptionV3 performs well with multi-scale convolutions. We adapted it as follows:

- **Global Average Pooling Layer:** This layer conserves key spatial information without adding any additional parameters to the network.
- **Custom Dense Layers:** Two fully connected layers, each with 512 and 256 units respectively, using the ReLU activation function to introduce non-linearity.
- **Softmax Layer:** An output layer of 4 neurons using softmax for classification.

As shown in Figure 5, fully connected layers in InceptionV3 connect all neurons between layers. As in VGG16, the lower layers of InceptionV3 were frozen, and only the newly added dense layers were fine-tuned.

The concept of factorized convolutions in InceptionV3 can be represented as:

$$Y = f(W_x * x) + f(W_y * x) \tag{5}$$

where Y is the result of applying two 1D convolutions sequentially along the x-dimension W_x and the y-dimension W_y. The outputs of parallel convolutions are combined as:

$$z = \sum_{i=1}^{n} W_i * x \tag{6}$$

C. Xception Architecture

Xception learns efficiently using depthwise separable convolutions. We adopted the pre-trained Xception model as follows:

- **Global Average Pooling Layer:** Replaces the flatten layer to enhance generalization.
- **Custom Dense Layers:** Two dense layers of 512 and 256 units sequentially, followed by ELU activation to improve learning capability.
- **Softmax Layer:** The final softmax layer classifies images into one of four categories.

The convolutional layers of Xception were frozen, and the newly added layers were fine-tuned similarly to the other architectures.

The depthwise separable convolution can be expressed as:

$$y_p = \sum_{k=1}^{K} y(k)_d * W(k)_p + b \tag{7}$$

where $y(k)_d$ represents the k-th channel output from the depth-wise convolution, and $W(k)_p$ is the point-wise convolution filter applied to the output.

In conclusion, these models leverage state-of-the-art architectures, adapted and fine-tuned to achieve high accuracy in early fire and smoke detection.

V. Experimental Results and Analysis

In this work, the hyperparameters of deep learning models are optimized to attain nearly optimal performance in early fire and smoke detection. The learning rate is set to 0.0001 and optimized by the Adam optimizer to enable good convergence during training. This value was determined through preliminary experiments to ensure that the model learns efficiently without overshooting optimal solutions. An appropriate batch size of 32 was selected to optimize memory usage and computing efficiency.

A. *Training Time*

For VGG16, InceptionV3, and Xception models, training was conducted for 70 epochs. After this, an additional 10 epochs for fine-tuning were performed, optimizing accuracy and yielding the best classification results.

1) A. Results of VGG16: For VGG16, the model showed an increase in both training and validation accuracy until around 85% by the 4th epoch. The loss was low, indicating minimal overfitting. However, the validation accuracy plateaued around 78%, with a significant jump in validation loss after the second epoch.

To alleviate overfitting, the following strategies are recommended:

- **Data Augmentation:** Techniques such as flipping, rotation, and noise addition.
- **Regularization Techniques:** Implementing dropout and L2 regularization.
- **Model Simplification:** Reducing the complexity of the architecture.

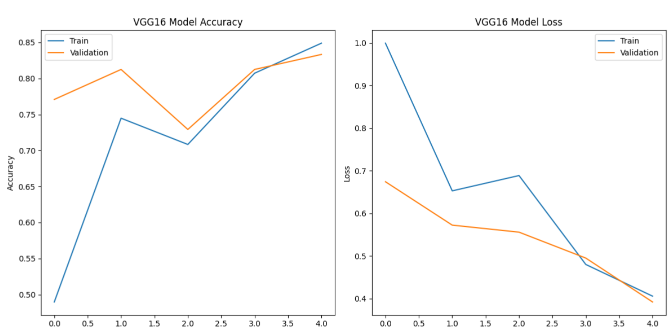

Fig. 6. Accuracy and Loss Curves for vgg

2) B. Results of InceptionV3: The InceptionV3 model achieved training accuracy of almost 95% by the 4th epoch, while validation accuracy remained around 70%. The training loss decreased to approximately 0.1%, but an increase in validation loss indicated potential overfitting.

TABLE II
Performance Metrics of VGG16, InceptionV3, and Xception on BoWFire Dataset

Model	Class	Precision	Recall	F1-score
VGG16	Fire	1.00	0.94	0.97
	Normal	0.94	0.75	0.83
	Smoke	0.92	0.93	0.93
InceptionV3	Fire	0.93	0.88	0.90
	Normal	0.88	0.88	0.88
	Smoke	0.79	0.78	0.89
Xception	Fire	1.00	0.94	0.97
	Normal	0.94	0.94	0.94
	Smoke	0.89	0.92	0.93

To counteract this, the following measures are suggested:

- **Early Stopping:** Monitoring validation loss to stop training when it begins to rise.
- **Data Variability:** Enhancing data through augmentation techniques.

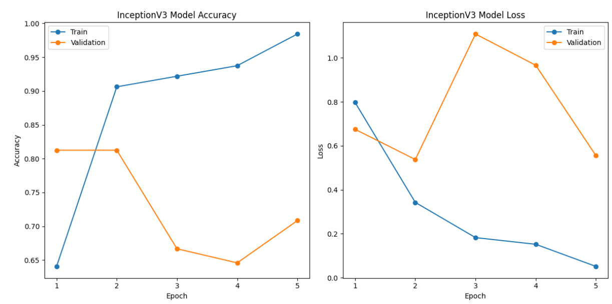

Fig. 7. Accuracy and Loss Curves for Inception

3) C. Results of Xception: The Xception model exhibited the highest training accuracy, approaching 95%. However, the validation accuracy remained around 70% from the 4th epoch onward. Similar to the other models, the validation loss initially decreased but began to rise afterward, highlighting overfitting concerns.

To improve generalization, the following steps are crucial:

- **Regularization Methods:** Using dropout and L2 techniques.
- **Architecture Simplification:** Streamlining the model structure.

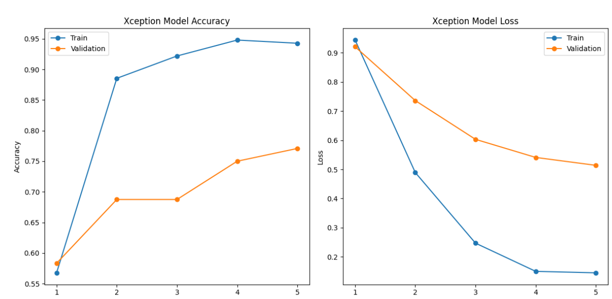

Fig. 8. Accuracy and Loss Curves for Xception

B. Hyperparameter Justification

The choice of hyperparameters was critical for achieving good model performance. The learning rate of 0.0001 was carefully chosen to maintain a balance between convergence speed and stability, avoiding overshooting during optimization. The batch size of 32 effectively managed memory usage while ensuring that the model could learn from sufficient data points in each iteration.

The number of epochs for training was determined based on observed performance trends. VGG16 was trained for 70 epochs, while InceptionV3 and Xception were trained for 50 epochs. The 10 epochs of fine-tuning aimed to refine model accuracy without leading to overfitting.

VI. Other Methods Comparison

Three deep learning models, VGG16, InceptionV3, and Xception, have been used to classify images into the classes namely fire, smoke, smoke fire, and no fire. Of these, the best was InceptionV3, that could demonstrate a validation accuracy of 94% and a test accuracy of 93%. These results demonstrate efficiency with early fire and smoke incidents. The main reason for the excellent performance of InceptionV3 is said to be the sophisticated architecture that it uses for dealing with complex visual patterns in classification.VGG16 could achieve only passable results, with only an accuracy marginally below that of InceptionV3. Xception, although showing a bit better result than VGG16, could not go beyond InceptionV3. Figure 7 captures the comparison of validation and test accuracies, where it is evident that InceptionV3 has outperformed the rest in both accuracy and overall robustness. The feature of InceptionV3 comes across as highly appealing for the domain since the inception modules allow multi-scale feature extraction, which makes it suitable for the application, which includes complex visual patterns with varying kinds of smoke and fire.These results have proved the relevance of the use of complex architectures such as InceptionV3 to early fire-detection applications and thus hold enough promise for improvement in firefighting systems.

Model	Accuracy (Validation, Feature Extraction)	Accuracy (Testing, Feature Extraction)	Accuracy (Validation, Fine-tuning)	Accuracy (Testing, Fine-tuning)
VGG 16	89%	93%	89%	87%
InceptionV3	**94%**	**93%**	85%	89%
Xception	91%	93%	71%	59%

Fig. 9. Model Accuracy Comparisons for Feature Extraction and Fine-tuning

VII. Conclusion

The InceptionV3 model has been shown to be highly effective when it comes to detecting forest fires and smoke; hence, this explains why it acquired the highest accuracy when compared to the other tested models. Its architectural improvements are built for the effective exploitation of complicated visual patterns that would be needed in any task involving these image classification tasks whose level of precision is related.Through transfer learning, it was possible to significantly reduce the computational load which ended up reducing the processing time but with a high performance level. The design will make the model well suitable for real-time applications requiring environmental monitoring where fast and accurate detection is a critical necessity.It would also evolve toward improving its adaptability toward the broader range of conditions, optimizing it to run on low-power devices, and to integrate into remote monitoring systems. This will add further range to its application and effectiveness in the real world.

Dataset Used

- Fire and Smoke Dataset on Kaggle

References

[1] Manoj, S., Valliyammai, C., ”Drone network for early warning of forest fire and dynamic fire quenching plan generation,” *J Wireless Com Network*, 2023, 112 (2023). https://doi.org/10.1186/s13638-023-02320w.

[2] Liu, J., Wang, Y., Guo, H., et al., ”Spatial and temporal patterns and driving factors of forest fires based on an optimal parameter-based geographic detector in the Panxi region, Southwest China,” *Fire Ecology* 20, 27 (2024). https://doi.org/10.1186/s42408-024-00257-z.

[3] Mishra, B., Panthi, S., Poudel, S., et al., ”Forest fire pattern and vulnerability mapping using deep learning in Nepal,” *Fire Ecology* 19, 3 (2023). https://doi.org/10.1186/s42408-022-00162-3.

[4] San Martín, R., Ottlé, C., Sörensson, A., ”Fires in the South American Chaco, from dry forests to wetlands: response to climate depends on land cover,” *Fire Ecology* 19, 57 (2023). https://doi.org/10.1186/s42408-023-00212-4.

[5] Tydersen, J. M., Collins, B. M., Coppoletta, M., et al., ”Fuel dynamics and reburn severity following high-severity fire in a Sierra Nevada, USA, mixed-conifer forest,” *Fire Ecology* 15, 43 (2019). https://doi.org/10.1186/s42408-019-0060-x.

[6] Chen, X., et al., ”Wildland Fire Detection and Monitoring Using a Drone Collected RGB/IR Image Dataset,” *IEEE Access*, vol. 10, pp. 121301-121317, 2022, doi: 10.1109/ACCESS.2022.3222805.

[7] Tran, D. Q., Park, M., Jeon, Y., Bak, J., and Park, S., ”Forest-Fire Response System Using Deep-Learning-Based Approaches With CCTV Images and Weather Data,” *IEEE Access*, vol. 10, pp. 66061-66071, 2022, doi: 10.1109/ACCESS.2022.3184707.

[8] Gonçalves, M., Brandão, T., and Ferreira, J. C., ”Wildfire Detection With Deep Learning—A Case Study for the CICLOPE Project,” *IEEE Access*, vol. 12, pp. 82095-82110, 2024, doi: 10.1109/ACCESS.2024.3406215.

[9] Rashkovetsky, D., Mauracher, F., Langer, M., and Schmitt, M., ”Wildfire Detection From Multisensor Satellite Imagery Using Deep Semantic Segmentation,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7001-7016, 2021, doi: 10.1109/JSTARS.2021.3093625.

[10] Wang, L., Zhang, H., Zhang, Y., Hu, K., and An, K., ”A Deep Learning-Based Experiment on Forest Wildfire Detection in Machine Vision Course,” *IEEE Access*, vol. 11, pp. 32671-32681, 2023, doi: 10.1109/ACCESS.2023.3262701.

[11] Chaoxia, C., Shang, W., and Zhang, F., ”Information-Guided Flame Detection Based on Faster R-CNN,” *IEEE Access*, vol. 8, pp. 58923-58932, 2020, doi: 10.1109/ACCESS.2020.2982994.

[12] Wang, G., Li, J., Zheng, Y., Long, Q., and Gu, W., ”Forest smoke detection based on deep learning and background modeling,” 2020 *IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Shenyang, China, 2020, pp. 112-116, doi: 10.1109/ICPICS50287.2020.9202287.

[13] M.A. Duhayyim et al., ”Fusion-Based Deep Learning Model for Auto mated Forest Fire Detection,” Comput. Mater. Contin., vol. 77, no. 1,pp. 1355-1371. 2023. doi:10.32604/cmc.2023.024198.

[14] Wu, X., Lu, X., and Leung, H., ”An adaptive threshold deep learning method for fire and smoke detection,” 2017 *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Banff, AB, Canada, 2017, pp. 1954-1959, doi: 10.1109/SMC.2017.8122904.