Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

Deep Learning–Based Forest Fire Classification Using MobileNetV3, ResNet50, and YOLOv8

Djarot Hindarto

[1)] Informatika, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional
[1]djarot.hindarto@civitas.unas.ac.id

ABSTRACT

Forest and land fires pose serious environmental, economic, and public health threats, particularly in regions with extensive forest areas and prolonged dry seasons. Early and accurate detection is essential to reduce damage and support rapid response. This study proposes a deep learning-based framework for forest fire image classification using MobileNetV3, ResNet50, and YOLOv8. The experiments were conducted using a dataset compiled from public sources, including The Wildfire Dataset, consisting of 5,000 images evenly distributed into five classes: cloud, fire, normal, risky, and smoke, with 1,000 images per class. The dataset was divided into training, validation, and testing sets using an 80:10:10 ratio, corresponding to 4,000, 500, and 500 images, respectively. Prior to training, all images were preprocessed through resizing, normalization, and data augmentation to improve robustness and generalization under varying environmental conditions. This standardized data preparation ensured a fair comparison among the three architectures. Model performance was evaluated using accuracy, precision, recall, F1-score, Matthews Correlation Coefficient, and Cohen's Kappa. The results show that YOLOv8 achieved the best overall performance, with 95.2% accuracy, 0.9566 precision, 0.952 recall, 0.9519 F1-score, 0.9412 MCC, and 0.9400 Kappa. ResNet50 achieved 94.0% accuracy, slightly outperforming MobileNetV3 at 93.8%. Overall, the study highlights the potential of deep learning for reliable forest fire monitoring and early warning systems.

Keywords: Forest Fire; Image Classification; Deep Learning; YOLOv8; MobileNetV3; ResNet50; Computer Vision; Disaster Monitoring

INTRODUCTION

Forest and land fires are among the most severe environmental disasters worldwide, particularly in tropical regions where prolonged dry seasons, climate variability, and anthropogenic activities significantly increase fire vulnerability. These incidents result in extensive ecological damage, including deforestation, loss of biodiversity, soil degradation, and disruption of ecosystem stability. Moreover, forest fires release large quantities of smoke, particulate matter, and greenhouse gases, contributing to air pollution, climate change, and serious public health problems. The economic impact is also considerable, affecting agriculture, transportation systems, tourism, and regional development. As the frequency and intensity of fire events continue to rise globally, the development of effective early detection and monitoring systems has become increasingly critical for mitigating environmental and socio-economic losses (Barmpoutis et al. 2020).

Conventional fire detection approaches typically rely on manual surveillance, ground patrols, observation towers, or community reporting. Although these methods have been employed for decades, they suffer from limited spatial coverage, delayed response times, and dependence on human vigilance. Satellite-based monitoring systems provide wider coverage but often lack sufficient temporal resolution and are susceptible to cloud cover and atmospheric interference, which can delay fire identification. Consequently, fires may spread uncontrollably before detection, reducing the effectiveness of emergency response measures. These limitations highlight the urgent need for automated, accurate, and real-time detection systems capable of processing visual data captured from cameras, drones, or satellites (Zhao et al. 2018).

Recent advances in artificial intelligence, particularly deep learning, have significantly transformed computer vision applications, enabling automated analysis of complex visual data. Convolutional Neural Networks (CNNs) have demonstrated superior performance in image classification, object detection, and scene understanding due to

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

their ability to learn hierarchical features directly from raw images. In forest fire monitoring, deep learning models can detect distinctive visual characteristics of flames and smoke, including color intensity, texture, shape, and spatial patterns, even under challenging environmental conditions. However, early-stage fire detection remains problematic because flames and smoke may appear small, partially occluded, or visually like clouds, fog, or sunlight reflections. Variations in illumination, background complexity, and atmospheric disturbances often lead to misclassification and high false alarm rates, reducing system reliability. These challenges highlight the need for more robust automated solutions capable of accurate fire identification across diverse real-world scenarios (Muhammad et al. 2018; Frizzi et al. 2016).

To address these issues, this study employs three complementary deep learning models: MobileNetV3, ResNet50, and YOLOv8. MobileNetV3 is selected for its lightweight architecture and computational efficiency, making it suitable for deployment on edge devices or embedded systems with limited resources (Howard et al. 2019). ResNet50, a deep residual network, is utilized for its strong feature extraction capability, enabling accurate classification in complex visual environments (He et al. 2016). Meanwhile, YOLOv8 is incorporated through its classification variant, YOLOv8n-cls, to perform image-level multi-class classification while benefiting from the efficient feature learning capability of the YOLO architecture. By combining lightweight and deep classification-oriented approaches, this study aims to capture both efficiency and accuracy aspects essential for practical fire monitoring applications (Redmon et al. 2016; Huang et al. 2023).

The primary objective of this research is to develop and evaluate a comparative deep learning framework for forest fire image classification using the three models under identical experimental conditions. Unlike many previous studies that evaluate models independently using different datasets or evaluation protocols, this research employs a unified dataset containing diverse fire and non-fire scenes captured under varying environmental conditions. Model performance is assessed using comprehensive metrics, including accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and Cohen's Kappa, to provide a robust evaluation of predictive reliability and class agreement. Such an evaluation framework enables a fair comparison of model effectiveness in terms of accuracy, generalization capability, and suitability for real-world deployment (El-Madafri et al. 2023).

The main contribution and novelty of this study lie in the systematic benchmarking of heterogeneous deep learning architectures—lightweight classification (MobileNetV3), deep residual classification (ResNet50), and classification-based YOLOv8n-cls within a single, standardized experimental framework. This integrated evaluation not only identifies the strengths and limitations of each model but also provides practical guidance for selecting appropriate architecture based on application requirements, whether prioritizing accuracy, computational efficiency, or real-time performance. By addressing the lack of comprehensive comparative studies across different model paradigms, this research contributes to the development of reliable, scalable, and deployable forest fire monitoring systems that can support early warning mechanisms, disaster mitigation, and environmental protection efforts (Guede-Fernández et al. 2021).

LITERATURE REVIEW

Forest fire monitoring based on image data has developed rapidly due to the widespread availability of surveillance cameras, unmanned aerial vehicles (UAVs), and satellite imagery. These technologies support large-scale visual monitoring and enable earlier identification of fire events. However, accurate recognition of flames and smoke remains challenging, especially in the early stage of fire, when visual patterns are still weak and often resemble clouds, fog, or lighting reflections. Environmental factors such as complex backgrounds, dense vegetation, and atmospheric disturbance further increase the risk of false alarms. Deep learning methods, particularly convolutional neural networks, have therefore become widely used because of their ability to automatically extract discriminative visual features from complex scenes (Muhammad et al. 2018; Saleh et al. 2024).

Several previous studies have addressed this problem using different model paradigms. Muhammad et al. (2018) applied convolutional neural networks for early fire detection in surveillance imagery and showed that CNN-based approaches are effective for extracting fire-related visual characteristics automatically. Frizzi et al. (2016) also demonstrated the feasibility of CNN models for fire and smoke detection in video data, emphasizing the importance of temporal visual patterns in challenging scenes. In contrast, Zhao et al. (2018) focused on UAV imagery and combined saliency detection with deep learning to improve wildfire identification from aerial views, showing that viewpoint and image source strongly influence model behavior. These studies confirm the potential of deep learning, but they differ substantially in data characteristics, target tasks, and evaluation settings.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

Recent research has also explored architectures optimized for practical deployment. Zheng et al. (2023) investigated MobileNetV3-based fire detection on small, embedded devices and showed that lightweight models are attractive for real-time applications with limited computational resources. On the other hand, (Guede-Fernández et al. 2021) highlighted the benefit of object identification frameworks for forest fire detection, where spatial localization of relevant regions improves recognition capability. In addition, (El-Madafri et al. 2023) introduced The Wildfire Dataset to improve representativeness and support more robust deep learning evaluation, emphasizing that dataset diversity is crucial for reducing false detections and improving generalization. These studies indicate that lightweight classifiers, deep residual networks, and modern deep classification frameworks each offer different advantages depending on deployment objectives, scene complexity, and computational constraints.

Despite these advances, most prior studies evaluate only one model family or focus on a specific task such as classification, detection, or video-based analysis. As a result, there is still limited evidence regarding how lightweight classification models, deep residual classification networks, and modern classification architectures compare under the same dataset, preprocessing pipeline, and evaluation metrics. This lack of direct comparison makes it difficult to identify which architecture provides the best balance between classification accuracy, robustness, and deployment efficiency. Therefore, this study addresses that gap by conducting a standardized comparative evaluation of MobileNetV3, ResNet50, and YOLOv8 using the same dataset and experimental settings. This comparison is expected to provide clearer guidance for selecting suitable deep learning architectures for practical forest fire monitoring systems.

Compared with previous studies, this research differs in three important aspects. First, it evaluates three heterogeneous architectures representing lightweight classification, deep residual classification, and YOLO-based image classification within a single framework. Second, all models are trained and tested using the same dataset composition, preprocessing strategy, and evaluation criteria to ensure fairness. Third, the study emphasizes not only predictive accuracy but also reliability-oriented metrics such as MCC and Cohen's Kappa, which are rarely discussed together in previous forest fire classification studies.

To provide a clearer positioning of this study with respect to prior works, Table 1 summarizes several representative studies on deep learning-based forest fire detection and classification. The comparison highlights differences in model type, data scenario, research focus, and remaining gaps, thereby clarifying the novelty of the present study.

Table 1
Comparison of Previous Studies on Deep Learning-Based Forest Fire Detection and Classification

No.	Study	Model / Approach	Data / Scenario	Main Findings	Limitation / Gap
1	(Muhammad et al. 2018)	CNN-based early fire detection	Surveillance imagery	CNN effectively learned fire-related visual features and improved early fire detection performance	Focused on CNN only; no comparison with lightweight and detection-based modern architectures
2	(Frizzi et al. 2016)	CNN for video fire and smoke detection	Video-based fire and smoke scenes	Demonstrated that deep learning can detect fire and smoke patterns in video sequences	Focused on video data; not directly comparable to image-based multi-model benchmarking
3	(Zhao et al. 2018)	Saliency detection + deep learning	UAV imagery	Improved wildfire identification from aerial images and highlighted the importance of viewpoint	Focused on UAV scenarios and hybrid preprocessing; no standardized comparison across architectures
4	(Zheng et al. 2023)	MobileNetV3 and YOLOv4	Small embedded devices / real-time fire detection	Showed that lightweight and real-time models are suitable for resource-constrained environments	Emphasized deployment efficiency rather than a broad comparison with deep residual networks
5	(Guede-Fernández et al. 2021)	Deep learning-based object identification	Forest fire detection scenes	Spatial localization improves identification of fire-related regions	Focused on object identification; limited comparison with classification-oriented models
6	El-Madafri et al. (2023)	Wildfire dataset and multi-task	Diverse open-source wildfire dataset	Highlighted the importance of dataset diversity and	Focused more on dataset design and representativeness

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

No.	Study	Model / Approach	Data / Scenario	Main Findings	Limitation / Gap
		learning perspective		representativeness for robust wildfire analysis	than direct benchmarking of multiple architectures
7	This study	MobileNetV3, ResNet50, and YOLOv8	Image-based forest fire classification using the same dataset and preprocessing pipeline	Provides a direct comparison of lightweight, deep residual, and YOLO-based classification models using identical settings and multiple evaluation metrics.	Addresses the lack of standardized benchmarking across heterogeneous architectures

As shown in Table 1, previous studies have demonstrated the effectiveness of deep learning for forest fire analysis under different scenarios, including surveillance images, UAV imagery, video data, and embedded deployment. However, most of them focus on a single model family or a specific application setting. In contrast, the present study performs a standardized comparison of MobileNetV3, ResNet50, and YOLOv8 using the same dataset composition, preprocessing procedures, and evaluation metrics, thereby providing a fairer benchmark for model selection in practical forest fire monitoring systems.

METHOD

This study proposes a deep learning framework for forest fire image classification using three architectures MobileNetV3, ResNet50, and YOLOv8 (Hindarto et al. 2023) selected for their complementary strengths in lightweight classification, deep feature extraction, and efficient image-level classification (Krizhevsky et al. 2017). The models are evaluated using the same dataset, identical preprocessing procedures, uniform training protocols, and consistent evaluation metrics to ensure a fair and unbiased comparison, allowing performance differences to reflect the intrinsic characteristics of each architecture (Hindarto et al. 2024). The workflow includes data acquisition, preprocessing and augmentation to enhance data quality, model training under identical conditions, and performance evaluation using multiple metrics that assess accuracy, precision, recall, and reliability. Through this standardized approach, the study aims to determine the most effective model for accurate, efficient, and deployable forest fire image classification systems suitable for early warning, disaster mitigation, and environmental monitoring applications (Ramadhani et al. 2025).

A. Dataset Acquisition

The dataset used in this study was compiled from public sources, including The Wildfire Dataset, and consists of 5,000 original images distributed evenly across five classes: cloud, fire, normal, risky, and smoke, with 1,000 images per class. The dataset was designed to represent diverse environmental conditions, including variations in illumination, smoke density, vegetation background, and viewing angle, in order to improve model generalization. Unlike binary fire detection datasets that only distinguish between fire and non-fire images, this study adopts a multi-class image classification setting to better capture the visual complexity of real-world forest fire monitoring scenarios. The dataset was divided into training, validation, and testing sets using an 80:10:10 ratio, corresponding to 4,000, 500, and 500 images, respectively. Each subset maintained a balanced class distribution to ensure fair model training and evaluation.

B. Data Preprocessing and Augmentation

All images were pre-processed before model training to ensure consistency across the three evaluated architectures. MobileNetV3 and ResNet50 used an input resolution of 224 × 224 pixels, while YOLOv8 employed an input resolution adjusted to its architectural requirement. Pixel values were normalized to improve training stability and convergence. To enhance robustness and reduce overfitting, data augmentation was applied during the training phase using random rotation, horizontal flipping, scaling, and brightness adjustment. These augmentation strategies were intended to simulate real-world variations in environmental conditions, such as changes in lighting, smoke intensity, and camera perspective. By using the same preprocessing and augmentation strategy across all models, this study aimed to maintain a fair experimental setting and improve the generalization capability of the trained models.

C. Deep Learning Models

This study evaluates three deep learning architectures with complementary characteristics, namely MobileNetV3, ResNet50, and YOLOv8, for multi-class forest fire image classification. These models were selected to represent

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

different architectural strengths in terms of computational efficiency, deep feature extraction capability, and modern classification performance.

1) MobileNetV3

MobileNetV3 is a lightweight convolutional neural network designed for efficient deployment on mobile and embedded devices. Its architecture combines depthwise separable convolution and lightweight attention mechanisms to reduce computational cost while maintaining competitive classification performance. In this study, MobileNetV3 was used as a multi-class image classifier to evaluate its effectiveness in recognizing five categories of forest fire-related images under resource-constrained conditions.

2) ResNet50

ResNet50 is a deep residual neural network that employs shortcut connections to overcome degradation problems in very deep architectures. This design enables stable training and strong feature extraction from complex visual patterns. In this study, ResNet50 was used as a multi-class image classifier to capture more discriminative features from images belonging to the cloud, fire, normal, risky, and smoke classes.

3) YOLOv8

In this study, YOLOv8 was implemented using the YOLOv8 classification variant (YOLOv8n-cls) rather than the object detection variant. The model was employed for image-level multi-class classification, where each input image was assigned to one of the five target classes: cloud, fire, normal, risky, and smoke. By using the classification version of YOLOv8, the model can be fairly compared with MobileNetV3 and ResNet50 under the same classification framework while still benefiting from the efficient feature learning capability of the YOLO architecture.

D. Model Training

All models were trained under the same experimental protocol to ensure a fair comparison. The training set was used for model learning, the validation set was used to monitor convergence during training, and the testing set was used for final performance evaluation. Since the task in this study is multi-class image classification, each model was optimized to predict one of the five predefined classes: cloud, fire, normal, risky, and smoke.

MobileNetV3 and ResNet50 were trained using the Adam optimizer with an initial learning rate of 0.0001, a batch size of 32, and 50 epochs. To improve generalization and reduce overfitting, a weight decay of 0.0005 was applied to both models. For YOLOv8, the YOLOv8n-cls model was trained using the Adam optimizer with an initial learning rate of 0.0001, an input image size of 224 × 224 pixels, a batch size of 32, and 50 epochs. No early stopping strategy was applied during training.

This standardized training configuration was intended to ensure that differences in model performance were primarily influenced by architectural characteristics rather than inconsistencies in optimization settings. By applying a consistent training procedure across all evaluated models, the comparison results become more reliable and reproducible for forest fire image classification research.

In this study, the main task is defined as multi-class image classification, where each input image is assigned to one of five classes: cloud, fire, normal, risky, or smoke. Although YOLOv8 originates from the YOLO family, the model used in this research is the classification variant (YOLOv8n-cls). Therefore, the term classification is used consistently throughout this manuscript as the final prediction task for all evaluated models.

E. Performance Evaluation

Model performance was evaluated using a confusion matrix and several standard classification metrics, namely Accuracy, Precision, Recall, F1-Score, Matthews Correlation Coefficient (MCC), and Cohen's Kappa. These metrics were selected to provide a comprehensive assessment of classification performance in terms of correctness, class discrimination, and prediction reliability.

Accuracy measures the proportion of correctly classified samples among all evaluated samples and is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

Precision measures the proportion of positive predictions that are actually correct and is defined as:

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall measures the proportion of actual positive samples that are correctly identified and is defined as:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

F1-Score represents the harmonic mean of Precision and Recall and is defined as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Matthews Correlation Coefficient (MCC) provides a balanced evaluation by considering all elements of the confusion matrix and is defined as:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5)$$

Cohen's Kappa measures the agreement between predicted and true labels while accounting for agreement occurring by chance, and is defined as:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where p_o denotes the observed agreement and p_e denotes the expected agreement by chance.

For the multi-class classification setting used in this study, these metrics were computed based on the overall confusion matrix to assess the comparative performance of MobileNetV3, ResNet50, and YOLOv8. By combining general accuracy measures with reliability-oriented metrics such as MCC and Cohen's Kappa, this evaluation framework provides a more comprehensive interpretation of model effectiveness.

F. Research Methodology Flowchart

This subsection presents the structured workflow of the proposed research methodology, outlining each stage involved in developing and evaluating the deep learning–based forest fire image classification system. The flowchart provides a clear and systematic representation of the research process, starting from data acquisition and preprocessing to model training, evaluation, and final model selection. By visualizing the sequential steps and their interconnections, the diagram enhances the transparency and reproducibility of the experimental design. Furthermore, it illustrates how multiple deep learning architectures are integrated within a unified framework to ensure a fair and consistent comparative analysis.

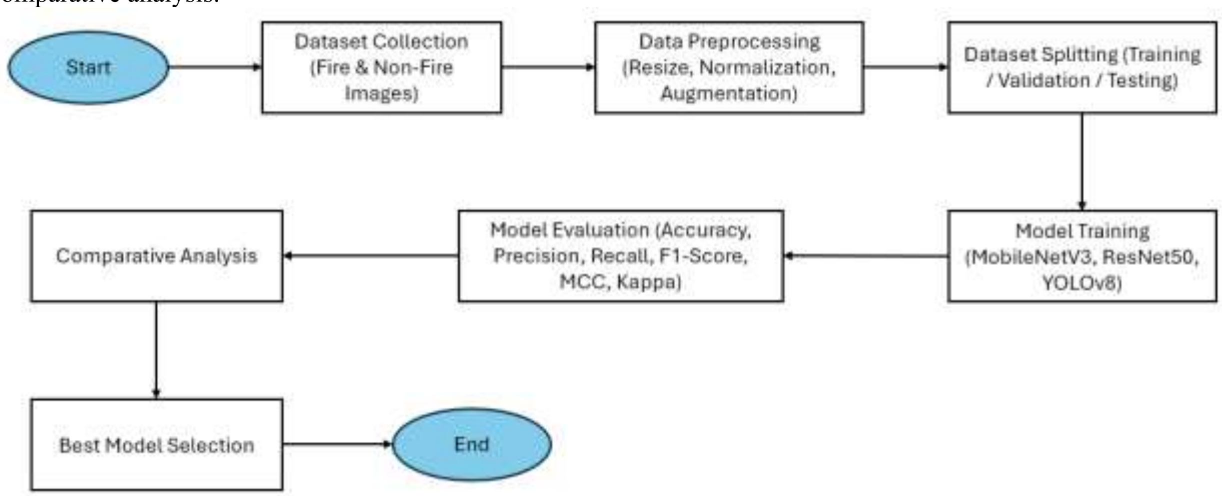

Fig. 1 Overall Research Workflow for Forest Fire Image Classification

Figure 1 illustrates the complete methodological framework adopted in this study for forest fire image classification. The workflow begins with dataset collection consisting of fire and non-fire images, followed by data

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

preprocessing steps such as resizing, normalization, and augmentation to enhance data quality and variability. The dataset is then divided into training, validation, and testing subsets to ensure a structured and unbiased evaluation process. Subsequently, the training phase involves three deep learning architectures—MobileNetV3, ResNet50, and YOLOv8—applied under identical experimental settings to maintain consistency in comparison.

After model training, performance evaluation is conducted using multiple quantitative metrics, including accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and Cohen's Kappa, providing a comprehensive assessment of predictive capability and reliability. The results obtained from each model are then subjected to comparative analysis to identify differences in learning behavior, generalization performance, and computational characteristics. Finally, the best-performing model is selected based on overall evaluation outcomes. This structured workflow ensures transparency, reproducibility, and fairness in assessing the effectiveness of different deep learning approaches for forest fire monitoring applications.

Dataset

The dataset used in this study consists of 5,000 original images evenly distributed across five classes: cloud, fire, normal, risky, and smoke, with 1,000 images per class. The data were partitioned using a standard machine learning split to ensure objective and reliable model evaluation. Specifically, 80% of the data (4,000 images) were allocated for training, while 10% (500 images) were used for validation and the remaining 10% (500 images) for testing. Each subset maintains a balanced class distribution, comprising 800 images per class for training and 100 images per class for both validation and testing. This balanced design helps prevent model bias toward any particular class and enhances the model's ability to generalize to unseen data. With a well-structured and evenly distributed dataset, the models are expected to effectively learn distinctive visual characteristics associated with each category.

During the training phase, data augmentation was applied on-the-fly using the *torchvision.transforms* library to increase data diversity without physically expanding the dataset. This approach enables the model to encounter different variations of the same images at each epoch, such as random rotations, horizontal flips, scaling, and brightness adjustments, thereby improving robustness and reducing overfitting. Because augmentation is performed dynamically during training, the model effectively learns from a broader data distribution compared to relying solely on static original images. This strategy also enhances the model's resilience to real-world variations, including changes in lighting conditions, camera angles, and environmental noise. Consequently, although the number of original training images remains 4,000, the model is exposed to virtually unlimited variations throughout the learning process. Such dynamic augmentation is essential for developing reliable image-based classification systems for forest fire monitoring applications in real-world environments.

RESULT

This section presents the outcomes obtained from the implementation of the proposed deep learning framework for forest fire image classification using MobileNetV3, ResNet50, and YOLOv8. The evaluation was conducted on the test dataset consisting of previously unseen images to measure the generalization capability of each model. Performance was assessed using several quantitative metrics, including Accuracy, Precision, Recall, F1-Score, Matthews Correlation Coefficient (MCC), and Cohen's Kappa. These metrics provide a comprehensive assessment of classification effectiveness, considering both predictive correctness and agreement with ground truth labels.

The experimental results indicate that all three models achieved high classification performance, with noticeable differences in effectiveness across architectures. MobileNetV3 demonstrated strong performance while maintaining computational efficiency, making it suitable for resource-constrained environments. ResNet50 achieved slightly higher performance than MobileNetV3, reflecting its deeper architecture and stronger feature extraction capability. YOLOv8 produced the highest performance among the evaluated models, indicating its effectiveness in capturing both spatial and semantic information relevant to fire-related features in images.

Table 2
Comparison of Model Performance Based on Accuracy, Precision, Recall, F1-Score, MCC, and Kappa

Model	Accuracy	Precision	Recall	F1-Score	MCC	Kappa
MobileNetV3	0.938	0.941045	0.938	0.938096	0.923244	0.9225
ResNet50	0.940	0.942291	0.940	0.940079	0.925560	0.9250
YOLOv8	0.952	0.956604	0.952	0.951916	0.941210	0.9400

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

As shown in Table 2, YOLOv8 achieved the highest accuracy of 95.2%, followed by ResNet50 at 94.0% and MobileNetV3 at 93.8%. Similar trends are observed across Precision, Recall, and F1-Score, indicating consistent performance differences among the models. The MCC and Kappa values also confirm strong agreement between predictions and actual labels for all models, with YOLOv8 exhibiting the highest reliability.

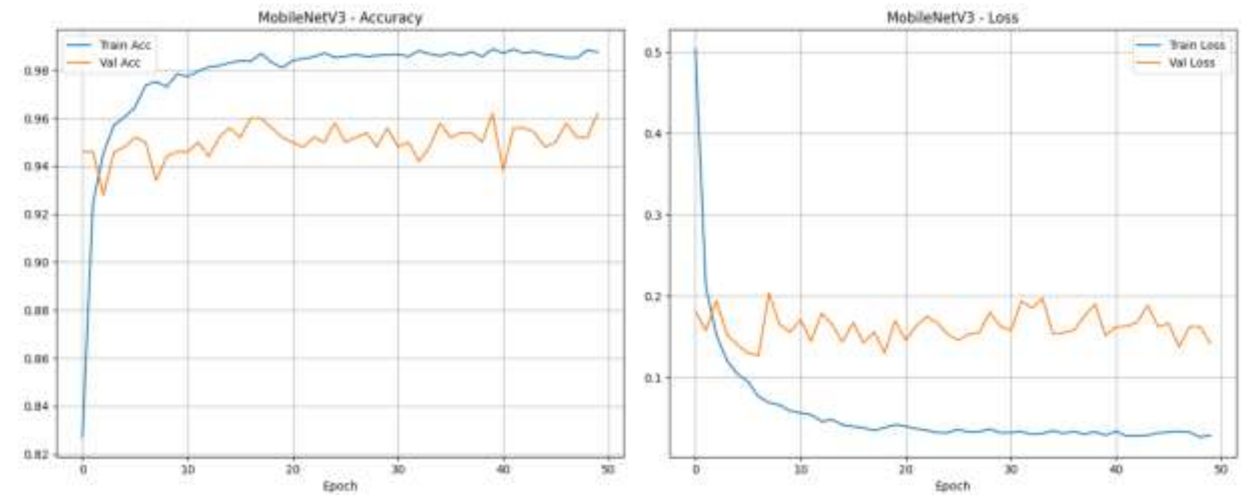

Fig. 2 Training and Validation Accuracy and Loss Curves of MobileNetV3 Model

Figure 2 illustrates the training and validation performance of the MobileNetV3 model across 50 epochs, presenting both accuracy and loss curves. The accuracy graph shows a rapid improvement during the initial epochs, with training accuracy increasing sharply and stabilizing above 98%, while validation accuracy consistently remains around 94–96%, indicating strong generalization performance. Meanwhile, the loss curve demonstrates a significant decrease in training loss during the early stages, followed by gradual stabilization at a low value, reflecting effective convergence of the model. Although minor fluctuations are observed in the validation loss, no significant divergence between training and validation trends is evident, suggesting that overfitting is minimal. Overall, the curves indicate that MobileNetV3 achieves stable learning behavior and satisfactory convergence within the given training epochs.

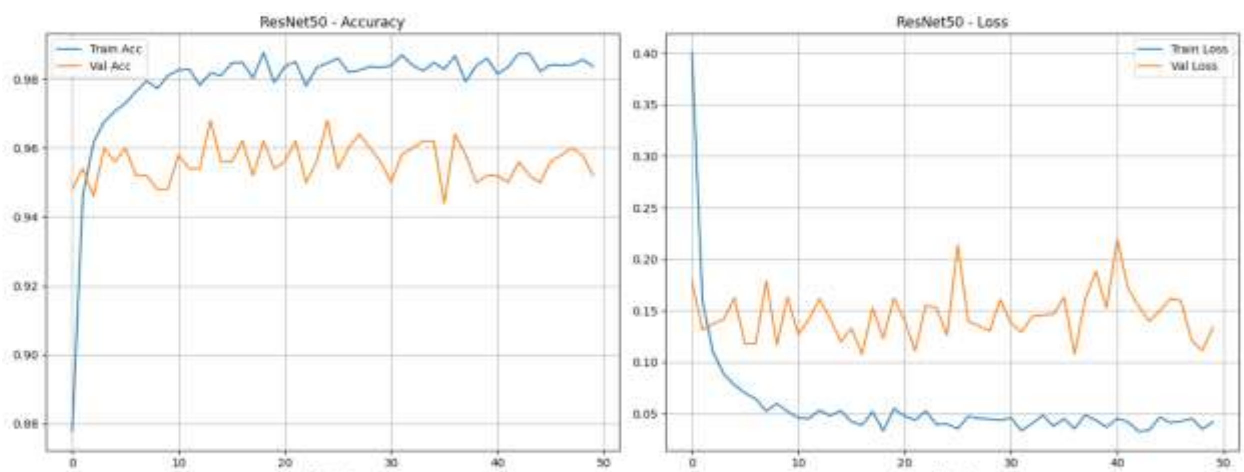

Fig. 3 Training and Validation Accuracy and Loss Curves of the ResNet50 Model

Figure 3 presents the learning behavior of the ResNet50 model over 50 training epochs, showing both accuracy and loss trends for training and validation phases. The accuracy curve indicates a rapid increase during the initial epochs, with training accuracy quickly surpassing 98% and maintaining a stable plateau, while validation accuracy fluctuates slightly around 95–97%, reflecting strong but slightly variable generalization performance. The loss curve demonstrates a sharp decline in training loss at the beginning of the training process, followed by stabilization at a

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

low level, which suggests effective convergence of the model. In contrast, validation loss exhibits moderate oscillations throughout the training period, likely due to variations in the validation samples or model sensitivity to complex patterns in unseen data. Despite these fluctuations, there is no significant divergence between training and validation trends, indicating that overfitting remains limited. Overall, the curves suggest that ResNet50 successfully learns discriminative features for forest fire image classification while maintaining stable performance across epochs.

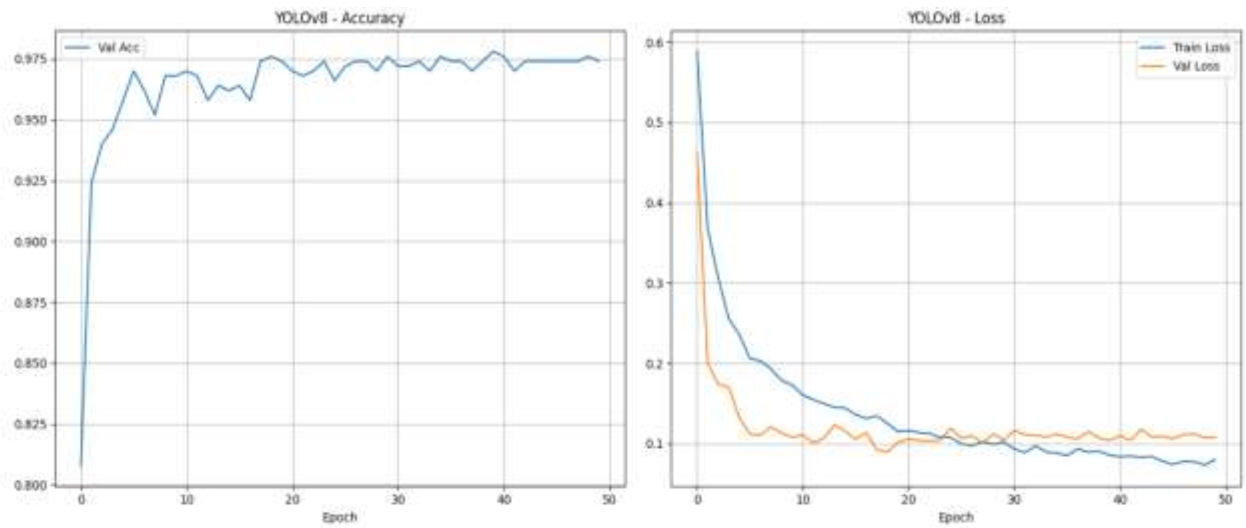

Fig. 4 Training and Validation Accuracy and Loss Curves of the YOLOv8 Model

Figure 4 illustrates the training dynamics of the YOLOv8 model across 50 epochs, showing both accuracy and loss behavior during the learning process. The accuracy curve demonstrates a rapid improvement in the early epochs, with validation accuracy quickly rising above 95% and stabilizing around 97–98%, indicating strong generalization capability even at early training stages. The loss curve shows a steep decline in both training and validation loss during the initial epochs, followed by gradual convergence to low values, which suggests effective optimization and stable learning. Although minor fluctuations are observed in the validation loss, the overall trend remains consistent with the training loss, indicating minimal overfitting. Compared to typical deep classification models, YOLOv8 exhibits faster convergence and maintains stable performance throughout the training process. Overall, these curves confirm that YOLOv8 effectively captures relevant visual features for forest fire image classification while achieving high accuracy and reliable model stability.

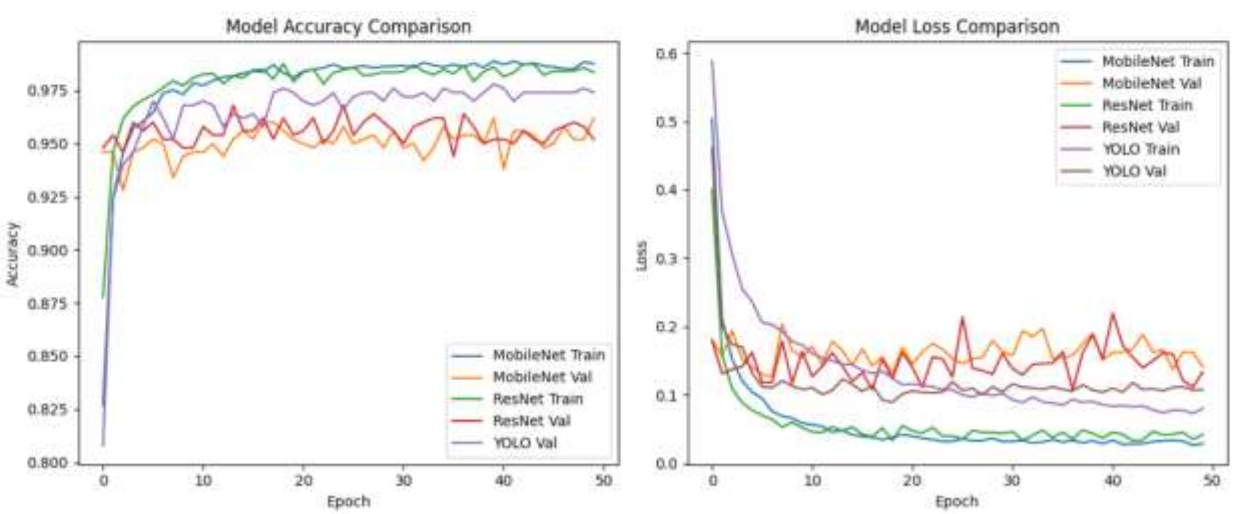

Fig. 5 Comparative Training and Validation Performance of MobileNetV3, ResNet50, and YOLOv8 Models

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

Figure 5 presents a comparative visualization of accuracy and loss trends for MobileNetV3, ResNet50, and YOLOv8 throughout the training process, highlighting differences in convergence behavior and generalization capability among the models. All architectures demonstrate rapid accuracy improvement during the early epochs, followed by stabilization at high performance levels, with training accuracy approaching or exceeding 98% for each model. Validation accuracy remains slightly lower but consistently above 94%, indicating effective learning with limited overfitting. In terms of loss, all models exhibit a sharp decline at the beginning of training, reflecting successful optimization, after which the curves gradually flatten as the models converge. ResNet50 and MobileNetV3 show slightly lower final training loss values, suggesting efficient feature learning, while YOLOv8 maintains competitive performance with stable validation loss across epochs. Minor fluctuations in validation loss are observed for all models, likely due to variations in unseen samples, yet no significant divergence between training and validation curves is evident. Overall, the comparison indicates that all three architectures achieve strong and stable performance, with subtle differences in convergence speed and loss dynamics during the training process.

DISCUSSIONS

The experimental results demonstrate that all three deep learning models—MobileNetV3, ResNet50, and YOLOv8—are capable of accurately classifying forest fire-related images under diverse environmental conditions. However, the performance differences observed among the models highlight the impact of architectural design on classification effectiveness. YOLOv8 achieved the highest overall accuracy and reliability metrics, suggesting that its architecture is highly effective in capturing discriminative spatial and semantic features relevant to forest fire imagery. In contrast, MobileNetV3 showed slightly lower performance but maintained competitive results while offering significant computational efficiency, making it suitable for real-time applications on resource-constrained devices.

A comparison between the lightweight MobileNetV3 and the deeper ResNet50 architecture reveals a trade-off between efficiency and representational power. ResNet50 consistently outperformed MobileNetV3 across most evaluation metrics, indicating that deeper networks are better at extracting complex visual features from challenging scenes such as dense smoke, irregular flame patterns, and cluttered backgrounds. Nevertheless, the relatively small performance gap suggests that MobileNetV3 can still serve as a practical alternative when computational resources are limited. This finding is particularly relevant for deployment scenarios involving edge computing, embedded systems, or mobile monitoring platforms.

The comparative analysis also indicates that YOLOv8n-cls offers advantages beyond conventional CNN-based classification models, especially in environments where fire-related visual cues are subtle or distributed across complex scenes. Its strong features of learning capability appear to contribute to superior performance and faster convergence. Additionally, the training curves show stable behavior with minimal overfitting across all models, which can be attributed to the balanced dataset and on-the-fly data augmentation strategy used during training.

Although the training accuracy of all models exceeded 98% while the test accuracy remained around 94–95%, the gap between training and validation performance did not show severe divergence, suggesting that overfitting was limited rather than substantial. This behavior may be attributed to the balanced dataset composition and the use of on-the-fly data augmentation during training. Nevertheless, the observed gap indicates that further regularization and evaluation on more diverse external datasets would be valuable for confirming model generalization.

Overall, the findings provide valuable insights into selecting appropriate deep learning architectures for automated forest fire image classification. While YOLOv8 offers the highest accuracy and reliability, MobileNetV3 provides a favorable balance between performance and computational cost, and ResNet50 delivers strong feature extraction capabilities for complex visual scenarios. These results contribute to addressing the practical challenge of developing accurate, efficient, and deployable fire monitoring systems, particularly for early warning and disaster mitigation applications. The study also highlights the importance of standardized evaluation frameworks to enable fair comparison among heterogeneous models and to guide future research in intelligent environmental monitoring.

CONCLUSION

This study presented a comprehensive deep learning framework for forest fire image classification using three architectures with distinct characteristics: MobileNetV3, ResNet50, and YOLOv8. Experimental results demonstrated that all models achieved high classification performance, indicating their effectiveness in recognizing fire-related visual patterns under diverse environmental conditions. Among the evaluated models, YOLOv8 obtained the highest accuracy and reliability metrics, suggesting that YOLO-based classification architectures can better capture both spatial and semantic information relevant to fire and smoke patterns. ResNet50 also showed strong performance due

* Corresponding author

Volume 08, Number 2, April 2026
https://doi.org/ 10.47709/cnahpc.v8i2.8112

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

to its deep residual structure, while MobileNetV3 provided competitive results with significantly lower computational requirements.

The findings highlight important trade-offs between accuracy, computational efficiency, and deploy ability. Lightweight models such as MobileNetV3 are well suited for real-time applications on resource-constrained devices, including edge computing platforms and mobile monitoring systems. In contrast, deeper networks like ResNet50 offer stronger feature representation capabilities for complex scenes but require higher processing resources. YOLOv8 demonstrated the ability to balance accuracy and efficiency, making it particularly promising for large-scale monitoring and early warning systems.

Despite the encouraging results, several limitations should be considered. The dataset used in this study, although diverse, may not fully represent all possible environmental conditions, geographic variations, or extreme scenarios encountered in real-world wildfire events. Additionally, the evaluation focused primarily on image-based classification without incorporating temporal information from video data or multimodal inputs such as thermal imagery and meteorological data. These factors may influence system performance in operational environments.

Future research should explore larger and more diverse datasets, integration of multimodal data sources, and deployment strategies on real-time monitoring platforms. Investigating hybrid approaches that combine classification, localization, and temporal analysis may further improve robustness and early warning capability. Overall, the proposed framework provides valuable insights for developing accurate, efficient, and deployable forest fire monitoring systems, with potential applications in early warning, disaster management, environmental protection, and public safety.

REFERENCES

Barmpoutis, P., Papaioannou, P., Dimitropoulos, K., & Grammalidis, N. (2020). A review on early forest fire detection systems using optical remote sensing. *Sensors (Switzerland)*, *20*(22), 1–26. https://doi.org/10.3390/s20226442

El-Madafri, I., Peña, M., & Olmedo-Torre, N. (2023). The Wildfire Dataset: Enhancing Deep Learning-Based Forest Fire Detection with a Diverse Evolving Open-Source Dataset Focused on Data Representativeness and a Novel Multi-Task Learning Approach. In *Forests* (Vol. 14, Issue 9, p. 1697). https://doi.org/10.3390/f14091697

Frizzi, S., Kaabi, R., Bouchouicha, M., Ginoux, J.-M., Moreau, E., & Fnaiech, F. (2016). Convolutional neural network for video fire and smoke detection. *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, 877–882. https://doi.org/10.1109/IECON.2016.7793196

Guede-Fernández, F., Martins, L., de Almeida, R. V, Gamboa, H., & Vieira, P. (2021). A Deep Learning Based Object Identification System for Forest Fire Detection. In *Fire* (Vol. 4, Issue 4, p. 75). https://doi.org/10.3390/fire4040075

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *EEE Computer Society Conference on Computer Vision and Pattern Recognition (2016) 2016-December 770-778*. https://doi.org/10.1109/CVPR.2016.90

Hindarto, D., Amalia, N., & Ningsih, S. (2024). Use of RESNET-50 Neural Network in Diagnosing Diseases Mango Leaves. *Journal of Computer Networks , Architecture and High Performance Computing*, *6*(1), 220–230. https://doi.org/10.47709/cnahpc.v6i1.3308

Hindarto, D., Program, I. S., Technology, I., & Nasional, U. (2023). Exploring YOLOv8 Pretrain for Real-Time Detection of Indonesian Native Fish Species. *Sinkron*, *8*(4), 2776–2785. https://doi.org/10.33395/sinkron.v8i4.13100

Howard, A., Sandler, M., Chen, B., Wang, W., Chen, L.-C., Tan, M., Chu, G., Vasudevan, V., Zhu, Y., Pang, R., Adam, H., & Le, Q. (2019). Searching for MobileNetV3. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 1314–1324. https://doi.org/10.1109/ICCV.2019.00140

Huang, Z., Li, L., Krizek, G. C., & Sun, L. (2023). Research on Traffic Sign Detection Based on Improved YOLOv8. *Journal of Computer and Communications*, *11*(07), 226–232. https://doi.org/10.4236/jcc.2023.117014

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 60, pp. 84–90). Association for Computing Machinery. https://doi.org/10.1145/3065386

Muhammad, K., Ahmad, J., & Baik, S. W. (2018). Early fire detection using convolutional neural networks during surveillance for effective disaster management. *Neurocomputing*, *288*, 30–42.

* Corresponding author

Submitted : 2026-03-03
Accepted : 2026-04-07
Published : 2026-04-09

https://doi.org/10.1016/j.neucom.2017.04.083

Ramadhani, N., Sholihati, I. D., & Hindarto, D. (2025). Integration of Machine Learning and Blockchain for Forest Fire Risk Prediction. *Sinkron*, *9*(4), 3336–3344. https://doi.org/10.33395/sinkron.v9i4.15435

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2016-Decem*, 779–788. https://doi.org/10.1109/CVPR.2016.91

Saleh, A., Zulkifley, M. A., Harun, H. H., Gaudreault, F., Davison, I., & Spraggon, M. (2024). Forest fire surveillance systems: A review of deep learning methods. *Heliyon*, *10*(1), e23127. https://doi.org/10.1016/j.heliyon.2023.e23127

Zhao, Y., Ma, J., Li, X., & Zhang, J. (2018). Saliency detection and deep learning-based wildfire identification in uav imagery. *Sensors (Switzerland)*, *18*(3). https://doi.org/10.3390/s18030712

Zheng, H., Duan, J., Dong, Y., & Liu, Y. (2023). Real-time fire detection algorithms running on small embedded devices based on MobileNetV3 and YOLOv4. *Fire Ecology*, *19*. https://doi.org/10.1186/s42408-023-00189-0

* Corresponding author