Wildfire Detection with MobileNetV2 and Mixed Attention from Satellite Imagery

Shubham Acharekar [23CSM2R22]
M.Tech I Year CSIS
National Institute of Technology, Warangal
Telangana, India 506004

***Abstract*— Wildfires pose significant threats to ecosystems, infrastructure, and human lives worldwide. Traditional methods for wildfire detection in satellite imagery have relied heavily on Convolutional Neural Networks (CNNs) without specific emphasis on channel importance or region specificity, leading to suboptimal performance. In this study, we propose a novel approach for wildfire detection using a combination of MobileNetV2 architecture and attention mechanisms applied to satellite imagery. The integration of attention mechanisms allows the model to focus on crucial regions and channels, enhancing its ability to discriminate between wildfire and non-wildfire scenes. Additionally, modifications to the classification head further improve the model's performance. Experimental results on benchmark datasets demonstrate the effectiveness of the proposed method, achieving superior wildfire detection accuracy compared to conventional CNN-based approaches. This research contributes to the advancement of wildfire monitoring and mitigation efforts, offering a promising solution for timely and accurate detection of wildfire events from satellite imagery.**

I. INTRODUCTION

Forest fires, caused by a combination of environmental factors and human activities, are an enormous challenge with far-reaching ecological, economic and social consequences. As climate change intensifies, wildfires will continue to increase in frequency, duration and severity, putting ecosystems, infrastructure and communities at risk around the world. In response to this growing threat , innovative approaches to wildfire detection and monitoring are urgently needed to facilitate timely intervention and mitigation . Ground-based systems , including individual sensors and networks of ground-based sensors , are also used to detect wildfires . These systems use optical and infrared cameras to record data related to flame and smoke characteristics . Although effective , ground-based systems face challenges such as limited coverage and susceptibility to false alarms . Space-based (satellite) systems play a vital role in wildfire monitoring , as they provide comprehensive insight into fire dynamics over large geographic areas . Satellites operating in various orbits , including geostationary , low-Earth orbit , and polar sun-synchronous orbit , offer different spatial and temporal resolutions , each with unique advantages for wildfire detection and monitoring . Satellite imagery has become an important tool for wildfire monitoring due to its wide coverage , high spatial resolution , and frequent revisits . However , traditional forest fire detection methods in satellite images have mostly relied on convolutional neural networks (CNNs) , which treat all channels and regions of the image equally . This approach ignores the inherent complexity and spatiotemporal dynamics associated with wildfires , leading to suboptimal detection . To address these limitations , this study proposes a new framework for wildfire detection using the MobileNetV2 architecture and alerting mechanisms applied to satellite imagery . The MobileNetV2 architecture , known for its efficiency and effectiveness in image classification , is the backbone of our recognition model . By integrating attention mechanisms into the network , we enable the model to dynamically highlight the most important regions and channels in the input images , enabling more effective discrimination between forest fire scenes and non-forest fire scenes . Additionally , we are making changes to the classification end of the network to improve its ability to capture the complex spatial and spectral patterns characteristic of wildfires . These changes aim to improve the model's ability to generalize across different environmental conditions and wildfire scenarios , thereby improving overall detection accuracy and reliability . The research stems from the critical need for advanced wildfire detection systems that can provide early warning and situational awareness to support wildfire management and response . Using the power of deep learning and alerting mechanisms , we aim to advance the state-of-the-art in wildfire detection technology and provide stakeholders with actionable insights and decision support tools to mitigate wildfire impacts on ecosystems and communities . In this paper , we provide a comprehensive overview of the proposed wildfire detection framework , including detailed descriptions of the MobileNetV2 architecture , alerting mechanisms , and classification head changes . Through extensive experiments and evaluation of benchmark datasets , we demonstrate the efficiency and robustness of our approach compared to traditional CNN-based methods . Our results highlight the potential of the proposed framework to transform wildfire detection and control practices , paving the way for more sustainable and adaptive wildfire management strategies in the face of increasing fire risk .

II. RELATED WORK

The increasing frequency and severity of wildfires has attracted considerable research interest in recent years due to the urgent need for effective fire detection and mitigation strategies . In 2019 alone , the number of wildfires in the Brazilian Amazon increased by 80% compared to the previous year , highlighting the threat of such incidents . In response to this growing challenge , researchers have explored different approaches to wildfire detection and suppression , leveraging advances in machine learning and computer vision . One promising approach is transfer learning , a machine learning technique that efficiently guides pre-trained models to new tasks . By transferring information from models trained on large amounts of data to target tasks where data is limited , transfer learning enables the rapid development and deployment of effective solutions . In the field of fire detection , recent studies have demonstrated the effectiveness of transfer learning using pre-trained convolutional neural network (CNN) models to extract features from images . Deep learning-based methods , especially those using CNNs , have become powerful tools for fire detection and classification . Spiking neural networks and CNNs have been used to achieve high accuracy in forest fire detection . Maximum reported accuracy is 90.91% and 89.92% . These approaches exploit the ability of CNNs to learn characteristics of visual data , which allows them to distinguish between fire and non-fire images with high accuracy . Pre-trained CNN models such as Google Inception , Microsoft ResNet50 and Oxford VGG have become invaluable resources for image-related tasks due to their ability to extract meaningful features from various datasets . These models are typically trained on large datasets such as ImageNet , which contains thousands of annotated images organized according to a hierarchical structure . In the context of fire detection , researchers have investigated the use of pre-trained CNNs such as VGG16 , VGG19 , ResNet-50 and InceptionV3 to extract features from fire-related images . These models trained on ImageNet show high accuracy in capturing relevant features , which facilitates accurate classification of fire and non-fire image . In addition, research has highlighted the importance of optimizing network architectures to improve performance . For example , researchers studying food classification have shown that improvements in the architecture of transfer learning models such as Inception-ResNet and Inception-V3 have led to improved classification accuracy of foods based on their nutritional value . Similarly , advances in CNN architectures such as the VGG and ResNet series have shown improved performance in a variety of image-related tasks , including image classification , object detection , and face recognition . The VGG network architecture , characterized by its simplicity and stacked convolutional layers , has been widely used in image classification tasks . VGG16 and VGG19 with the number of printing layers showed competitive performance in image classification tasks , although the error rate is different . In contrast , ResNet-50 introduces a new architectural design with residual connectivity that allows training very deep networks mitigating the vanishing gradient problem . Skipping connections in ResNet-50 facilitates gradient flow and improves learning by enabling model-learning identity functions , contributing to its excellent performance in image recognition tasks . In general , research in the field of fire detection emphasizes the importance of transfer learning and deep learning approaches , especially approaches that use pre-trained CNN models , to solve the challenges presented by wildfires . By exploiting the characteristics of these models and optimizing the network architecture , researchers aim to develop robust and efficient systems for early forest fire detection and mitigation , ultimately improving safety and environmental protection . Despite their effectiveness , these approaches are not without their limitations . One of the main challenges is the need for large , annotated datasets to efficiently tune pre-trained models . Although models such as VGG16 , VGG19 , ResNet-50 , and InceptionV3 are trained using large data sets such as ImageNet , the specificities of fire-related features may require additional data addition or domain adaptation for optimal performance . In addition , the computational resources required to train and infer deep CNNs can be significant , creating practical challenges for deployment in resource-constrained environments or real-time applications . Furthermore , the interpretability of deep learning models remains a concern , as understanding the decision-making process of complex neural networks is often difficult , especially in safety-critical areas such as fire detection . Addressing these issues is critical to advancing the field of fire detection and realizing the full potential of deep learning approaches to wildfire mitigation . Future research may focus on developing effective data augmentation techniques , optimizing network architecture for resource-constrained environments , and improving the interpretability of deep learning models to promote trust and transparency in fire detection systems . By overcoming these challenges , researchers can pave the way for more effective and scalable solutions to protect lives , protect ecosystems , and mitigate the devastating effects of wildfires.

III. PROPOSED MODEL

A. Transfer learning with MobileNetV2 :

Base model selection and configuration : The MobileNetV2 architecture , pre-trained on the ImageNet dataset , is chosen as the basis of the fire detection model due to its proven effectiveness in image-related tasks . MobileNetV2 offers a balance between computational efficiency and effectiveness , making it suitable for use in resource-constrained environments .

Retraining settings : The convolutional basis of the model is preserved while the classification head is dropped . This allows us to take advantage of ImageNet's learned features that capture common visual patterns when adapting the model for fire detection .

Selection of intermediate layers : 12 intermediate layers are selected from the different layers of MobileNetV2 to capture complex and abstract features at a higher level of representation . In particular , Block-12-expand-relu , with its

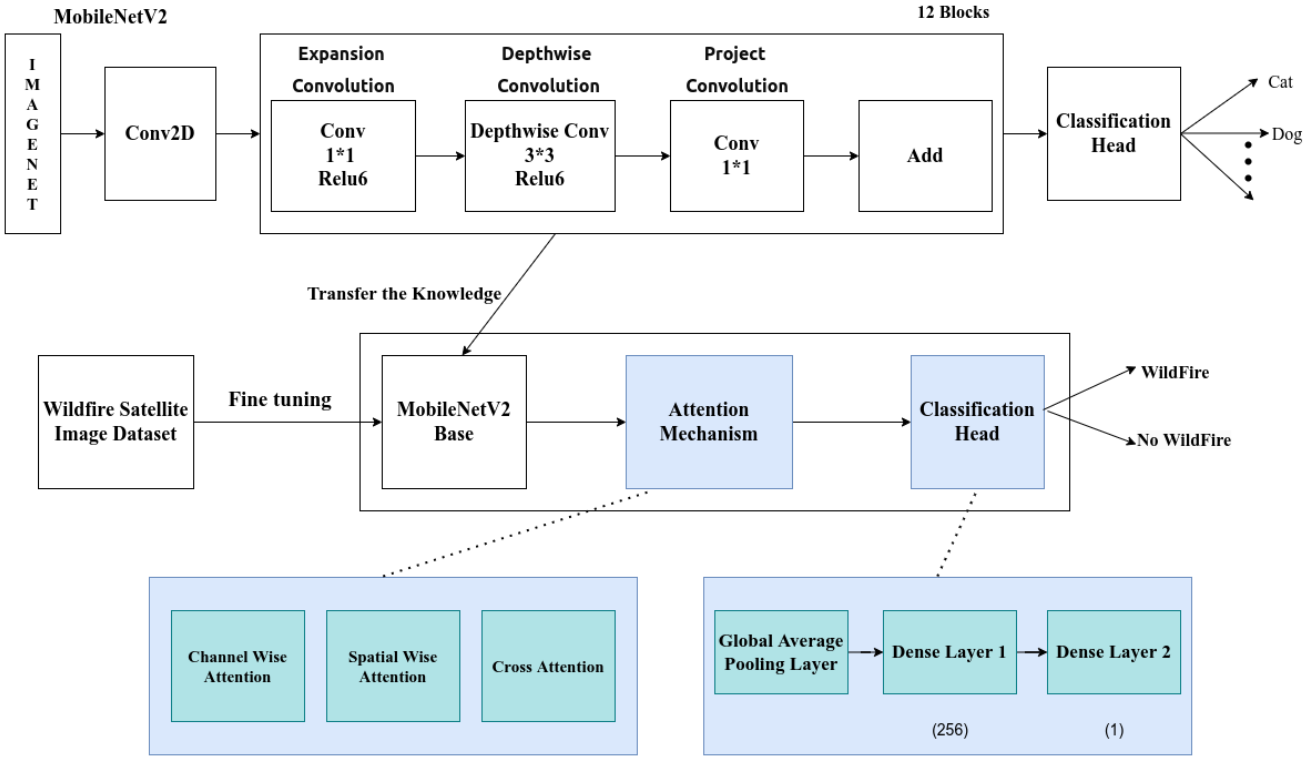

Fig. 1: Architecture

many convolutional operations and nonlinear activations , is particularly good at capturing the complex features required for forest fire detection .

Following are the equations that are used in the MobileNetV2 model while training:

1. Expansion Convolution: Expansion convolution expands the number of channels using a 1×1 convolution.

Let F_{exp} be the input feature map tensor with dimensions $H \times W \times C_{\text{in}}$, where H is the height, W is the width, and C_{in} is the number of input channels.

Let W_{exp} denote the expansion convolutional kernel weights, and b_{exp} denote the corresponding biases.

$$F_{\text{exp}} = \text{Conv2D}(F_{\text{exp}}, W_{\text{exp}}) + b_{\text{exp}} \tag{1}$$

2. Depthwise Convolution : Depthwise convolution applies a separate convolutional kernel to each input channel independently .

Let F_{exp} be the input feature map tensor with dimensions $H \times W \times C_{\text{exp}}$.

Let W_{dw} denote the depthwise convolutional kernel weights, and b_{dw} denote the corresponding biases.

$$F_{\text{dw}} = \text{DepthwiseConv2D}(F_{\text{exp}}, W_{\text{dw}}) + b_{\text{dw}} \tag{2}$$

3. Projection Convolution : Projection convolution reduces the number of channels back to the desired output dimension using a 1×1 convolution .

Let F_{dw} be the input feature map tensor with dimensions $H \times W \times C_{\text{exp}}$.

Let W_{proj} denote the projection convolutional kernel weights, and b_{proj} denote the corresponding biases.

$$F_{\text{proj}} = \text{Conv2D}(F_{\text{dw}}, W_{\text{proj}}) + b_{\text{proj}} \tag{3}$$

B. Integration of attention mechanisms :

Channel-specific attention mechanism : Channel-specific attention mechanism is seamlessly integrated into the model architecture to selectively highlight important channel-specific information extracted from feature maps . This mechanism allows the model to dynamically prioritize significant features , improving its discrimination and ability to capture complex patterns , which are crucial for accurate classification . This is mathematically denoted as,

$$\text{ChannelAttention}(F)_c = \sigma\left(\frac{1}{HW}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{i,j,c}\right) \tag{4}$$

Spatial Attention Mechanism : In addition to channel-specific attention , a spatial attention mechanism is introduced to store important location information of maps of transformed map objects . By focusing on the spatial context of features , the model gains a deeper understanding of the image content , facilitating more accurate classification decisions and improving overall performance . This is mathematically denoted as,

$$\text{SpatialAttention}(F)_{i,j} = \sigma\left(W_{\text{spatial}} * F_{i,j}\right) \tag{5}$$

Cross-Attention mechanism : In addition , a cross-attention mechanism is included in the model architecture to capture dependencies between different regions of the image . By considering not only spatial and channel-specific features , but also the relationships between them , the model gains a comprehensive understanding of the underlying image structure , leading to further improvements in classification accuracy and robustness . This is mathematically denoted as,

$$\text{CrossAttention}(F_1, F_2)_{i,j} = \sigma\left(W_{\text{cross}} * F_1\right)_{i,j} \odot F_2 \quad (6)$$

Mechanisms of strategic attention : Mechanisms of attention are strategically placed after the middle layers of the model architecture . This strategic placement makes it possible to suppress irrelevant noise and highlight important features , ensuring that the model's attention is focused on key parts of the input data . By refining the way features are represented in this way , the model becomes more adept at capturing significant patterns and making accurate classification decisions , improving its overall performance in image analysis tasks .

C. Modifying the classification head:

Global average pooling level : To combine the spatial information of maps of map objects into one value per map , a global average pooling level is applied . This step helps to reduce the dimensionality of feature maps while preserving important spatial information that is essential for accurate classification .

$$\text{GAP}(F)_c = \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} F_{i,j,c} \quad (7)$$

Here:

- $\text{GAP}(F)_c$ represents the output of the Global Average Pooling operation for the c^{th} channel.
- $F_{i,j,c}$ denotes the value of the feature map at spatial location (i, j) and channel c.
- H and W are the height and width of the feature map, respectively.
- HW represents the total number of spatial locations in the feature map.

Adding Dense Layers : After the global average aggregation layer , two dense layers are added to the model architecture . The first dense layer transforms the condensed features resulting from global averaging into a higher-dimensional representation , facilitating richer feature representations . Next , a second dense layer allows the model to learn more complex patterns and relationships within the extracted features , further improving its discrimination ability . Let x denote the input vector to the dense layer, W represent the weight matrix, and b represent the bias vector. The output of the dense layer, denoted as y, can be calculated as follows:

$$z = Wx + b$$

$$y = \text{Activation}(z)$$

Here:

- W is a matrix of shape (m, n), where m is the number of neurons in the current layer and n is the number of neurons in the previous layer.
- x is a vector of shape $(n, 1)$.
- b is a vector of shape $(m, 1)$.
- z represents the affine transformation result, obtained by multiplying the input vector x with the weight matrix W and adding the bias term b.
- Activation is the activation function applied to the output z, introducing non-linearity into the network.

D. Wildfire dataset fine-tuning :

Dataset preparation : a comprehensive Wildfire dataset containing uncured satellite images marked as wildfires . Special attention is given to ensuring a balanced distribution of classes in the dataset to avoid bias during model training .

Data Augmentation Techniques : Various data augmentation techniques including rotation , zooming , scaling , etc . are used to augment the training data set . By introducing variations in the training samples , data augmentation helps improve model reliability and generalization .

Fine-tuning process : A pre-trained MobileNetV2 model augmented with alerting mechanisms and a modified classification head is fine-tuned on the wildfire dataset . During fine-tuning , the model parameters are adjusted to adapt to the specific characteristics of forest fire detection , while using ImageNet pre-trained features to ensure efficient data transfer .

Validation and Testing : The fine-tuned model undergoes rigorous validation and testing procedures using a separate validation dataset . Performance measures such as precision , accuracy , recall , and F1 score are calculated to evaluate the effectiveness of the model in accurately classifying satellite images into forest fires and non-fires .

IV. DATASET

The dataset used in this project comes from Canadian fire data available on the Open Government Portal and is licensed under the Quebec Creative Commons 4.0 Attribution (CC-BY) license . It consists of 35,850 satellite images , each measuring 350 x 350 pixels , divided into two categories : Wildfire and No Wildfire . The dataset is further divided into training (70%) , testing (15%) and validation (15%) to facilitate model training , evaluation and validation .

Wildfires Category : This category contains 22,710 satellite images depicting scenes affected by fires . These images capture the different stages and intensities of wildfires, including active fire fronts, burned areas and smoke plumes. No Fires Category : This category contains 20,140 satellite images depicting forests , countryside , waterways , urban areas and natural landscapes unaffected by wildfires . The dataset provides a comprehensive overview of forest fires and non-forest fires , enabling the development and evaluation of forest fire detection algorithms using satellite imagery . The dataset has been carefully augmented to enrich its

Fig. 2: Dataset Description

diversity and size , which increases the efficiency of forest fire detection algorithms using satellite images . Through the strategic application of extension methods , we have carefully expanded the breadth and depth of the dataset , ensuring a more complete representation of real scenarios . These upscaling strategies include :

Shear Range Transformation : The use of andom shear transformations dynamically changed the spatial orientation of images , promoting a wider perspective of the data set . This high-scale technique facilitates the model's ability to detect wildfire-related features from different angles and locations , improving its adaptability to different environmental conditions .

Zoom Range Expansion : By introducing random zoom adjustments , the dataset is scalable and full of resolution , enabling complex details and contextual nuances to be captured at different magnifications . This high-scale method allows the model to distinguish subtle cues indicating the presence of forest fire at multiple spatial resolutions , increasing its robustness in wildfire detection amid heterogeneous landscapes .

Improvement of horizontal translation : The inclusion of horizontal translation created mirror images in the dataset , diversifying the perspective of the dataset . This high-scale strategy helps to better understand the patterns and structures associated with wildfires from different perspectives , enhancing the model's ability to generalize to different environmental conditions . When these high-scale techniques are carefully integrated , the dataset has greater variability and richness , giving the model a more comprehensive and nuanced understanding of wildfire dynamics . Thus, the model is ready to demonstrate the effectiveness and generalizability of forest fire detection in multiple environmental conditions , which will contribute to forest fire detection and mitigation through satellite image analysis .

V. Experimental Results

In this study , a novel approach for Wildfire Detection for Satellite Imagery using MobileNetV2 along with attention mechanism like channelwise attention , spatial attention and cross attention is proposed . This approach is evaluated using various performance metrics as follows :

1. Precision : Precision is a metric used to evaluate the accuracy of positive predictions made by a model . It represents the ratio of true positive predictions to the total number of positive predictions made by the model . Precision is calculated using the formula :

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

In simpler terms , precision measures the ability of the model to correctly identify relevant instances from all the instances it has classified as positive .

2. Recall : Recall , also known as sensitivity or true positive rate, assesses the ability of a model to correctly identify all relevant instances in the dataset . It is calculated as the ratio of true positive predictions to the total number of actual positive instances in the dataset , including both the correctly and incorrectly classified ones . The formula for recall is :

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Recall provides insights into how effectively a model captures all positive instances in the dataset , regardless of any false positives .

3. F1 Score : The F1 score is a single metric that combines precision and recall into a single value , providing a balanced assessment of a model's performance . It is the harmonic mean of precision and recall and is calculated using the formula :

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1 score considers both false positives and false negatives and is particularly useful when dealing with imbalanced datasets or when there is an uneven cost associated with false positives and false negatives .

4. Accuracy : Accuracy is a basic metric used to evaluate the overall performance of a model . It measures the proportion of correctly classified instances (both true positives and true negatives) out of the total number of instances in the dataset . The formula for accuracy is :

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}}$$

While accuracy provides an overall measure of model performance , it may not be suitable for imbalanced datasets where the classes are disproportionately represent .

Following are the results obtained while evaluating the proposed model :

TABLE I: Performance Metrics

Precision	0.96
Recall	0.98
F1-Score	0.96

Accuracy is comparison done with plotting graph of Training accuracy VS Validation Accuracy. Comparing training and validation accuracies using plotting helps to understand the learning dynamics of models , identify potential overfitting , and optimize model performance during training .

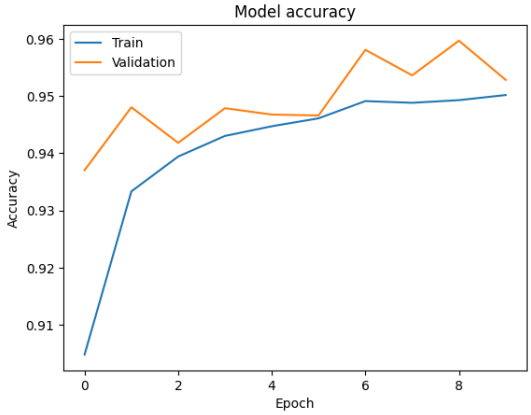

Fig. 3: Training Accuracy vs Validation Accuracy

Estimating test accuracy with a confusion matrix provides a comprehensive breakdown of model performance showing true positives , true negatives , false positives , and false negatives . This distribution helps to assess the strengths and weaknesses of the model in classification tasks . By providing insight into the model's ability to correctly classify cases , it highlights areas for improvement that are critical to improving accuracy and reliability in real-world applications . Understanding these forecasting results is key to improving model performance and ensuring its effectiveness in different scenarios .

In addition , we compared the results of our model with the results of other existing models to confirm our conclusions . This comparative analysis allows us to assess the relative effectiveness and efficiency of our model against established benchmarks . This comparison is summarized in the table below :

TABLE II: Comparision with existing models

Model	**Accuracy**
CNN	0.91
Inception	0.92
Proposed Model	0.96

This comparative analysis allows us to contextualize the performance of our model within the broader context of existing methodologies , helping to identify areas for improvement and potential opportunities for further research and development .

VI. CONCLUSIONS

In conclusion , our wildfire detection model is a significant step forward in using advanced technology to effectively control fires . By integrating advanced technologies such as the MobileNetV2 architecture and attention mechanisms , we have developed a model with outstanding accuracy in detecting wildfires from satellite images . An impressive 96% accuracy exceeds the performance of previous systems , which are typically around 91% . This greater accuracy means earlier detection of wildfires , a critical factor in containing and mitigating their impact . Furthermore , the intelligence of our model goes beyond mere detection ; it provides valuable information determining the location and significance of fires detected in the images . This feature improves firefighting operations and resource allocation decisions . Future improvements to the model through parameter optimization and integration of real-time monitoring promise to further improve its performance . These advances could significantly enhance our ability to proactively respond to wildfire threats and ultimately minimize their devastating effects . In conclusion , our wildfire detection model is a significant step forward in using state-of-the-art technology to address the enormous challenge of wildfire management . Its combination of precision , intelligence and adaptability make it a valuable tool for protecting communities and ecosystems from the ravages of wildfires.

REFERENCES

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

[2] M. Sandler AU - A. Howard AU - M. Zhu AU - A. Zhmoginov AU - "MobileNetV2: Inverted Residuals and Linear Bottlenecks" T2 - 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition SP - 4510 EP - 4520 AU - L. -C. Chen PY - 2018 DO - 10.1109/CVPR.2018.00474

[3] P. Agarwal and G. Jha, "Forest Fire Detection Using Classifiers and Transfer Learning," 2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI), Hong Kong, Hong Kong, 2021, pp. 29-33, doi: 10.1109/RAAI52226.2021.9507958.

[4] Sathishkumar, V.E., Cho, J., Subramanian, M. et al. "Forest fire and smoke detection using deep learning-based learning without forgetting. fire" ecol 19, 9 (2023). https://doi.org/10.1186/s42408-022-00165-0

[5] E. Casas, L. Ramos, E. Bendek and F. Rivas-Echeverría, "Assessing the Effectiveness of YOLO Architectures for Smoke and Wildfire Detection," in IEEE Access, vol. 11, pp. 96554-96583, 2023, doi: 10.1109/ACCESS.2023.3312217.

[6] Khan Muhammad, JamilAhmadi, IrfanMehmood, Seungmin Rho and Sung WookBaili, "Convolutional Neural Networks Based Fire Detection in Surveillance Videos", vol6 2018.

[7] Ba, Rui, Chen Chen, Jing Yuan, Weiguo Song, and Siuming Lo. 2019. "SmokeNet: Satellite Smoke Scene Detection Using Convolutional Neural Network with Spatial and Channel-Wise Attention" Remote Sensing 11, no. 14: 1702. https://doi.org/10.3390/rs11141702

[8] https://research.google/blog/mobilenetv2-the-next-generation-of-on-device-computer-vision-networks/

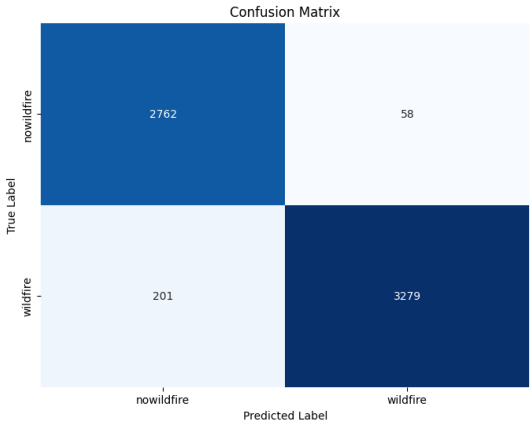

Fig. 4: Confusion Matrix

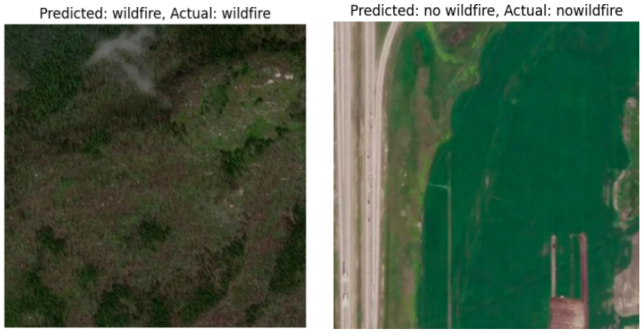

Fig. 5: Prediction Results