***I.J. Intelligent Systems and Applications*, 2024, 5, 53-67**
Published Online on October 8, 2024 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2024.05.05

Reduction of Misclassifications in Wildfire Detection: A Weighted Ensemble Deep Learning Approach

Lawrence Owusu*
North Carolina Agricultural and Techincal State University/Department of Computational Data Science and Engineering, Greensboro, NC, 27411, USA
E-mail: lowusu@aggies.ncat.edu
ORCID iD: https://orcid.org/0009-0005-8150-2970
*Corresponding Author

Robert B Eshun
North Carolina Agricultural and Techincal State University/ Department of Computational Data Science and Engineering, Greensboro, NC, 27411, USA
E-mail: rbeshun@aggies.ncat.edu
ORCID iD: https://orcid.org/0009-0003-1568-2622

Leila Hashemi-Beni
North Carolina Agricultural and Techincal State University/Department of Built Environment, Greensboro, NC, 27411, USA
E-mail: lhashemibeni@ncat.edu
ORCID iD: https://orcid.org/0000-0003-1026-4555

Ali AlQahtani
North Carolina Agricultural and Techincal State University/Computer System Technology, Greensboro, NC, 27411, USA
E-mail: aalqahtani@ncat.edu
ORCID iD: https://orcid.org/0000-0002-8225-4180

Masud R Rashel
University of Evora/Institute of Earth Sciences, Evora, 7002-554, Portugal
E-mail: mrashel@uevora.pt
ORCID iD: https://orcid.org/0000-1367-1383

AKM K. Islam
North Carolina Agricultural and Techincal State University/Department of Computational Data Science and Engineering, Greensboro, NC, 27411, USA
E-mail: akislam@ncat.edu
ORCID iD: https://orcid.org/0000-0003-0267-7982

Received: 26 March 2024; Revised: 17 May 2024; Accepted: 25 July 2024; Published: 08 October 2024

Abstract: Governments worldwide are increasingly prioritizing early wildfire detection to safeguard lives, property, and the environment. Although CNN-based models have demonstrated exceptional performance in various computer vision applications, the evolving nature of wildfire images poses significant challenges for a single CNN-based model in wildfire detection. In this study, we addressed this issue by integrating and weighting the differential learning capabilities of three individual transfer learning models: InceptionV3, ResNet50, and VGG16. Experimental results show that the ensemble deep learning models significantly outperformed all single classifiers across all performance metrics. Both the ensemble and weighted ensemble deep learning models achieved 99.7% accuracy, 99.5% precision, 100% recall, 99.8% F1-score, 0.5%false positive rate, 0.0% false negative rate and 0.3% error rate. Additionally, these models reduced the error rate by 98%, 91%, and 40% compared to the error rates of ResNet50, InceptionV3, and VGG16 respectively. A false negative rate of 0% indicates that our proposed ensemble deep learning models identified and predicted all the wildfire instances present in the test set correctly without a single misclassification. This positions our proposed ensemble deep learning models as superior choices for reducing misclassifications in wildfire detection.

Index Terms: Ensemble Learning, Wildfire, Deep Learning, Forest Fire, Confusion Matrix, Classification.

1. Introduction

Wildfires are becoming more frequent and intense due to climate change, with rising temperatures, extreme weather events, and vegetation desiccation. These fires can spread rapidly, causing severe damage to human lives, infrastructure, and ecosystems, and in 2018 alone, wildfires resulted in an estimated $1.6 billion in economic losses in the United States and Australia. Wildfires disrupt the ecological balance, alter local weather patterns, and threaten the extinction of vulnerable species. To mitigate these impacts, governments are actively seeking advanced methods for early detection of wildfires [1-3].

Traditionally, wildfires are detected using networks of cameras, fire watchtowers, or satellite images of areas of interest. However, the detection of wildfire with these systems is problematic, challenging, and error-prone because of the involvement of human beings in the monitoring of the data streams emitted from these wildfire detection systems [4, 5]. For these reasons, researchers have leveraged the advancement in computer vision technology to vigorously research efficient and effective techniques of automated fire detection systems [4-7].

Although deep learning models have achieved state-of-the-art results in a variety of computer vision use cases, the major and common drawback is that the Convolutional Neural Network (CNN) requires large data samples and computational resources to train. The reality is that the common wildfire datasets are small, repetitive, and collected as images or video samples from the Internet and experiments, and there is no standard dataset to adopt in the training mode and performance evaluation task [8].

Transfer learning technique fine-tunes a pre-trained neural network architecture on a smaller, wildfire-specific dataset. This allows for smaller datasets and less computational complexity without compromising model detective performance accuracy. Transfer learning is crucial in wildfire detection due to its ability to overcome data limitations, improve feature extraction, generalize findings, accelerate model convergence, and adapt to diverse environmental conditions [9]. Transfer learning is a machine learning technique that leverages the knowledge gained from the first task to improve learning efficiency and performance on the second task, especially where labeled data are scarce [10].

Despite advancements in the algorithms, individual models struggle with the complex and changing nature of wildfires as each algorithm has its own set of strengths and weaknesses. Additionally, the evolving nature of wildfires makes wildfire detection more complex and challenging, rendering it impractical for individual models to identify wildfires across diverse conditions. Furthermore, weather conditions like fog and sunset, which mimic or have similarities in pixel values with actual forest fires and smoke, further complicate their detection [11].

With ensemble deep learning, different individual learners work synergistically by integrating the strengths of multiple algorithms to detect diverse forms of forest fires under various scenarios [12]. This study proposes a weighted ensemble deep learning approach that integrates and assigns differential weights based on the learning capabilities of three pre-trained models: InceptionV3, VGG16, and ResNet50. By assigning weights proportionally based on their predictive and learning capabilities, this approach will enhance the overall accuracy of wildfire detection under diverse environmental conditions.

The Flow of the Study

The introduction section sets the context for the study, while the methodology section describes the dataset and outlines the approaches for developing the ensemble deep learning models. The subsequent section presents discusses and compares the performance of the proposed ensemble deep learning models against individual models and other models trained on the same benchmark DeepFire dataset. Finally, the conclusion section provides a summary of the key findings along with recommendations for future studies.

2. Related Works

Due to the parallel computation power offered by graphical processing units (GPUs), deep learning architectures like convolutional neural networks (CNNs) have become increasingly popular for detecting forest fires [8]. Since Krizhevsky and his team published their pioneering work, training a deep CNN on 1.2 million high-resolution images for the ImageNet LSVRC-2010 contest, the field of computer vision has been harnessed to develop models capable of automatically detecting forest fires [13]. These models excel in extracting and learning complex feature representations and deeper semantic meanings [14].

Muhammad *et al.* created a CNN model inspired by the GoogleNet architecture to detect fires in surveillance videos [15]. Kim *et al* proposed a method that uses video sequences as input, employing a region-based CNN (R-CNN) to learn spatial features and classify areas as either potential fire or non-fire. They further used Long Short Term Memory (LSTM) networks to verify the reliability of fire alarms [16].

Although transfer learning has achieved state-of-the-art results with a limited dataset, using a single pre-trained model for wildfire detection is often unreliable due to the complexity and evolving nature of wildfire under different environmental conditions. As a result, many researchers have turned to ensemble learning for more reliable forest fire detection [11].

For instance, Renjie and his colleagues combined two object detectors, Yolov5 and EfficientDet, with a transfer learning model, EfficientNet, to enhance wildfire detection accuracy. Yolov5 and EfficientDet were used in parallel to detect forest fires, but they primarily focused on fire-like objects rather than the entire image. To address this, the EfficientNet image classifier was incorporated, enabling the model to utilize global information and ensuring that fire-like objects did not disproportionately affect detection results. Tests showed that this ensemble learning method (combining Yolov5, EfficientDet, and EfficientNet) increased detection accuracy by 2.5% to 10.9% and reduced false positives by 51.3% without adding extra latency [12].

Similarly, Bahhar *et al* [17] proposed and evaluated a system for detecting wildfires and smoke by integrating a CNN-based model with YOLO as the object detection algorithm. The proposed method operated through a two-stage process. Initially, the CNN-based model detected anomalies in the frame, indicating the presence of fire or smoke. Upon detection, the YOLO model localized these anomalies, providing precise locations. During evaluations, the CNN-based classification model achieved impressive results, with an accuracy of 0.99 and an F1-score of 0.95. The smoke detection model attained a mean average precision of 0.85 despite challenges such as the limited availability of high-quality real-world UAV-captured fire and smoke images. This ensemble deep learning approach effectively addressed both classification and detection tasks [17].

In response to the rising incidence of forest fires driven by global warming and human activities, Abdusalomov et al. leveraged deep learning techniques to develop an improved method for detecting forest fires using the Detectron2 model. Initially, they trained their proposed model on a custom dataset comprising 5,200 images of fire and non-fire, taken both day and night under various scenarios. To enhance the model's representativeness, the dataset was augmented to 348,600 images. Testing their model in real-time on a Raspberry Pi 3B+ system, they found that their Detectron2-based model demonstrated superior performance, reaching an accuracy of 99.3% [18].

Furthermore, Almasoud leveraged an Integrated Sensor System (ISS) and deep learning techniques to develop a reliable method for detecting and alerting wildfires. The proposed system, an intelligent system for detecting and alarming wild forest fires using advanced deep learning techniques (IWFFDA-DL), aimed to identify forest fires at their early stages using an ISS that combined multiple sensors to gather input data for wildfire forecasting. Central to the system were the Attention-based Convolutional Neural Network with Bidirectional Long Short-Term Memory (ACNN-BLSTM) model for data processing and fire detection and the Bacterial Foraging Optimization (BFO) algorithm for performance optimization. After extensive simulations to evaluate its effectiveness, the proposed IWFFDA-DL system demonstrated its capability to classify different fire statuses accurately and showed significant improvements in various performance metrics [19].

To enhance the detection capabilities of deep learning-based wildfire detection systems under varying environmental conditions, Seydi and his team developed a deep learning framework called Fire-Net. This model was trained on Landsat-8 imagery, utilizing visible and thermal data to improve detection accuracy. Fire-Net employed a two-stream feature extraction method that leveraged residual convolution and separable convolution blocks, allowing for the extraction of relevant features from coarse datasets. To demonstrate the model's robustness and transferability across different environmental conditions and forest densities, Fire-Net was tested in various geographical areas, including Australia, Central Africa, the Amazon Rainforest, Chernobyl, and North America. The study revealed that Fire-Net was highly accurate, achieving an overall accuracy rate of 97.35%. It was particularly effective at detecting small active fires, which is crucial for early intervention. This marked a significant advancement in remote sensing for forest fire detection, offering a reliable and efficient tool for monitoring fires in diverse regions [20].

To highlight the potential of integrating UAVs with AI techniques for enhancing real-time forest fire detection and response, Khan *et al* [21] proposed an innovative approach combining a UAV-based system with a VGG19-based transfer learning model. After training and evaluation, they compared the performance of their model against several traditional machine learning algorithms, including logistic regression, random forest, naïve Bayes, support vector machine, and k-nearest neighbors. The experimental results demonstrated that their approach outperformed all traditional machine learning models across all performance metrics, achieving a 95% prediction accuracy, 95.7 precision, and 94.2% recall [21].

The paper by Jonnalagadda and Hashim addressed the pressing challenge of enhancing processing times and detection capabilities in Unmanned Aerial Vehicle (UAV)/drone imagery. This method, called SegNet, uses a segmented deep learning-based Convolutional Neural Network (CNN) to enhance the detection and monitoring of wildfires using UAV/drone imagery. By focusing on reducing irrelevant features, SegNet was able to reduce the number of feature maps in images, thereby increasing both processing speed and detection accuracy. Extensive evaluations of the SegNet model using drone flight datasets showed that the SegNet model performed better in terms of accuracy and processing speed, achieving an accuracy of 98.2% compared to other state-of-the-art algorithms like AlexNet and GoogleNet [22].

Somaiya et al introduced the FFireNet framework, an innovative deep-learning model. They built a binary classifier on the pre-trained convolutional base of the MobileNetV2 model to detect wildfires using the benchmark DeepFire dataset. The FFireNet model performed exceptionally well across all metrics, achieving an accuracy of 98.42%, an error rate of 1.58%, a recall of 99.47%, and a precision of 97.42% [23].

Over the years, researchers have made significant progress in wildfire detection by leveraging deep learning techniques to improve accuracy and reliability. Despite these advancements, a comprehensive literature review revealed a gap: none of the existing studies has considered assigning weight to individual models' learning and predictive capabilities when developing an ensemble deep learning model. This study addresses the limitations with two main contributions.

Key Contributions of the Study

- First, we propose an ensemble deep learning approach that combines the learning and predictive capabilities of InceptionV3, ResNet50, and VGG16 to reduce misclassification.
- Second, we will assign differential weights proportional to the predictive contributions of InceptionV3, ResNet50, and VGG16 to build a weighted ensemble deep learning model to further reduce misclassification.

3. Methodology

3.1. Description of the DeepFire Dataset

In this study, we used the DeepFire dataset, taken from Kaggle. The DeepFire dataset was curated meticulously by Khan and colleagues to serve as a benchmark to aid researchers in developing binary classification models for wildfire detection. Images were gathered from various online resources using keywords such as forest fire, mountain fire, forest, and mountains. The dataset is divided into two fire and no-fire classes. The fire class includes images of forests and mountains with visible fire flames or smoke clouds, while the no-fire class comprises images of forests and green mountains without any signs of fire. The images in the DeepFire dataset have multiple angles of view and a diversified range of scenery to better train the model for robustness and transferability. To enhance model training, the dataset was preprocessed by cropping irrelevant objects, such as extinguishing machinery and people, and resizing them to a uniform resolution of 250 x 250 pixels. This ensured the model could efficiently learn relevant features for classification. The DeepFire dataset consists of 1900 images, with 950 images in the no-fire class and 950 in the fire class [21].


(a) Forest with fire (b) Forest Without Fire

Fig.1. Different Classes in the DeepFire Dataset with Forest Fire (a) and without Fire (b).

Table 1. Class distribution of the dataset and the number of instances in the train, validation and test sets

Class	Train Set	Validation Set	Test Set
Forest Fire	642	107	190
No Forest Fire	658	113	190
Total	1300	220	380

3.2. Data Preprocessing and Flow Diagram of the Study

We conducted data preprocessing to prepare the dataset into the required format by the deep learning algorithms to identify and extract features effectively [24]. The images were resized to 224 by 224 pixels with 3 color channels (RGB) and normalized. Normalizing the data facilitated the backpropagation process, speeding up convergence and improving performance [25, 26].

3.3. Pre-trained Models

Transfer learning is a machine learning technique that leverages knowledge from a source domain, like ImageNet, for a new task with significantly fewer training examples. Given our limited wildfire dataset, three pre-trained CNN models: VGGNet, InceptionV3, and ResNet50 were selected to construct the robust ensemble and weighted ensemble models for wildfire detection [9].

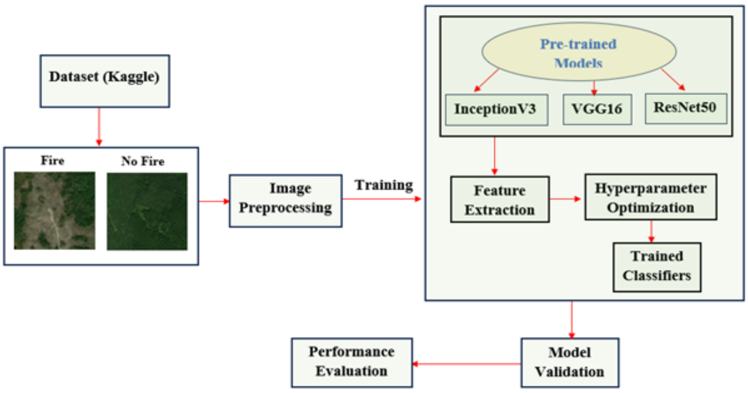


Fig.2. Flow diagram for the study

3.4. *Residual Network*

Microsoft researchers introduced ResNet50 in 2015 to address the vanishing or exploding gradient problem in deep neural networks. It uses residual block and skip connections to minimize training and test error, preserve knowledge gain, and boost model performance. The residual blocks are stacked to form the residual Network. The underlying principle behind the Residual Network is that the stacked layers were made to fit the residual mapping instead of these layers learning the underlying mapping. It has a max-pooling layer that comes after a convolutional layer that is 7x7 and has 64 filters that reduce the spatial dimensions. ResNet50 has residual blocks organized into four stages with 16 blocks each, each with a bottleneck structure consisting of three convolutional layers and two 1x1 and 3x3 convolution operations. It incorporates skip connections to enhance gradient flow throughout the training process and uses global average pooling (GAP) as its final layer to reduce spatial dimensions and the output layer is composed of a fully connected layer with SoftMax activation function [27, 28].

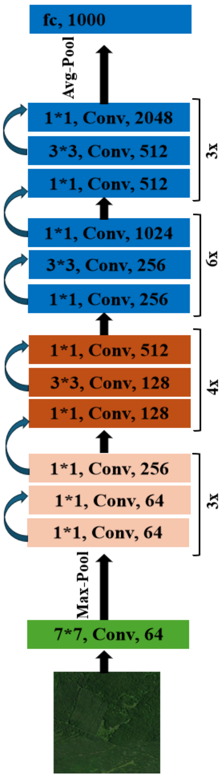


Fig.3. Architecture of ResNet50 [27]

3.5. *InceptionV3*

InceptionV3 is a CNN-based architecture from the Inception family, designed to enhance training speed and convergence. It leverages techniques such as factorized convolutions, regularization, spatial factorization into asynchronous convolutions, dimension reduction, auxiliary classifiers, and parallel computations to create a computationally efficient network with fewer parameters. The architecture comprises three key components: a convolutional block for feature extraction, an improved inception module for parallel multi-scale convolutions to capture diverse and complex patterns in data, and an auxiliary classifier for better convergence [29, 30].

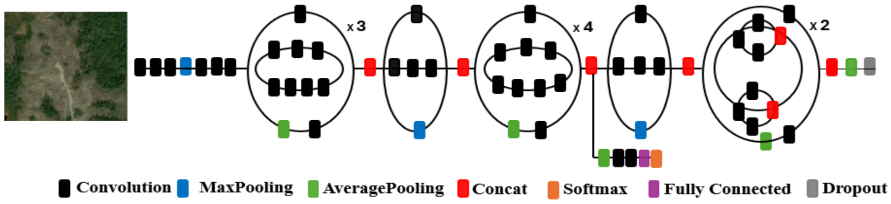


Fig.4. Architecture of InceptionV3 [31]

3.6. *VGG16*

VGG16 is a CNN-based architecture proposed by K. Simonyan and A. Zisserman in their paper "Very Deep Convolutional Networks for Large-Scale Image Recognition". It made tremendous improvements over the AlexNet architecture by replacing the large kernel-sized filters with multiple 3x3 kernel filters. Architecturally, the VGG16 model consists of 13 convolutional layers, 2 fully connected layers, and 1 SoftMax classifier. The first and second convolutional layers comprise 64 feature kernel filters, each of size 3x3, and the resulting feature map is passed to a max pooling layer with a stride of 2. The third and fourth convolutional layers consist of 124 feature kernel filters, each of size 3x3, and these two layers are followed by a max pooling layer with a stride of 2. The fifth, sixth, and seventh layers are convolutional layers with kernel filters of size 3x3, and these layers are followed by max pooling with a stride of 2. From the eighth to the thirteenth are two sets of convolutional layers with 512 filters, each of kernel size 3x3, and the resulting feature map is passed to a max pooling layer with a stride of 1 [30, 32].

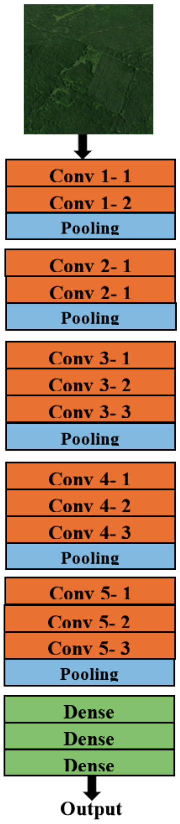


Fig.5. Architecture of VG [30]

3.7. *Model Training and Evaluation*

The ResNet50 architecture was chosen for its small parameter size, high performance, and control of vanishing and explosion gradient problems [33]. The InceptionV3 architecture was chosen for its computational efficiency, faster speed, and better optimization techniques [34]. The VGG16 was chosen for its ability to learn and extract complex features from input images [30]. We preserved the convolutional base of the pre-trained models and customized them to extract useful features from the DeepFire dataset. A new classifier was constructed on top of the pre-trained models to leverage feature mapping from the previous ImageNet dataset for wildfire detection. The models were trained over 50 epochs using the Adam optimizer, binary cross-entropy loss function, and the Rectified Linear Unit (ReLU) activation function to introduce non-linearity into the network and facilitate learning complex functions [24]. We used a batch size of 32 and a learning rate of 0.001. For the output layer, we used the sigmoid activation function. The sigmoid function is a differentiable squashing function that outputs the probability of prediction between 0 to 1, where x represents the input value to a neuron. The binary cross-entropy loss function is applied in binary classification tasks, where the objective is to assign instances to one of two classes. It is specifically used with the sigmoid activation function. The y in eq (1) represents the class and p(y) is the predicted probability of the input belonging to class 1 or 0 [24, 35].

$$Softmax\ (x) = \frac{1}{1+e^{-x}} \tag{1}$$

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

$$Loss = -\frac{1}{N}\sum_{i=1}^{N} y_i \log\big(p(y_i)\big) + (1-y)\log\big(1-p(y_i)\big) \tag{3}$$

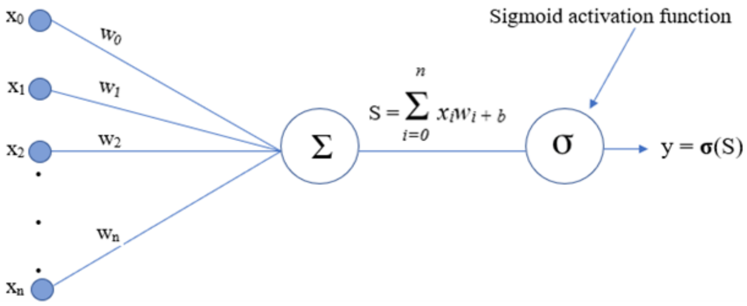


Fig.6. A Sigmoid activation function of a neural network

4. Results and Discussion

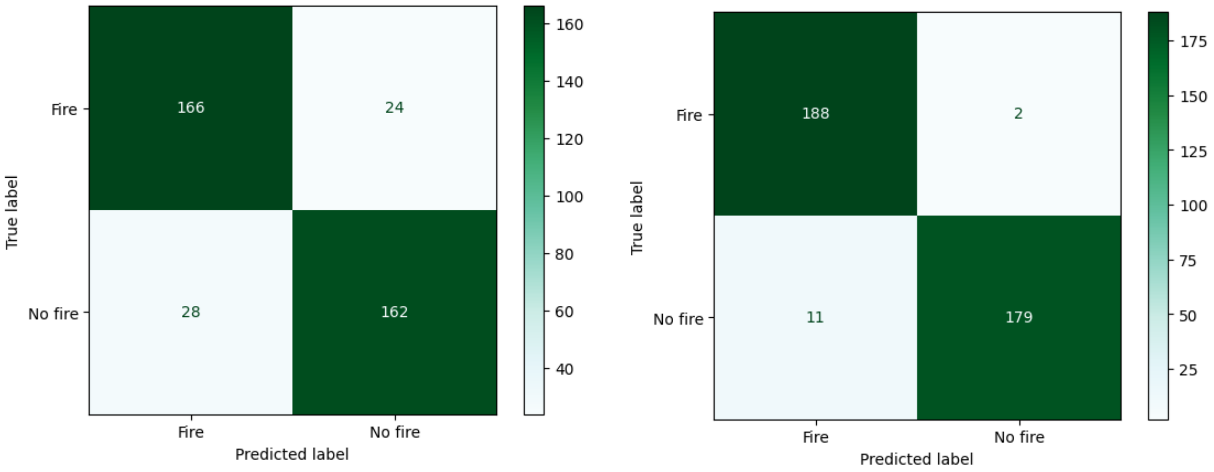
4.1. Experimental Environment

The models were trained in Google Colab, which provides a powerful and convenient cloud-based environment. Specifically, we used the NVIDIA-SMI 535.104.05 TP4 GPU for substantial computational power [36]. The Google Colab environment featured a virtual machine with 7 processors, each with 4 cores (Intel(R) Xeon(R) CPU @ 2.00GHz), and up to 53.4707GB of memory, ensuring robust computational resources without memory bottlenecks. For the software components, we used Python 3.10.12 as the runtime. We used TensorFlow 2.15.0 as the backend and Keras as the high-level API for training the neural network models. Additionally, we used NumPy 1.25.2 and Pandas 2.0.3 for numerical computations and data manipulation respectively, and Seaborn 0.13.1 and Matplotlib 3.7.1 were used for visualization tasks. We also leveraged OpenCV 4.8.0 to preprocess the images into the required shape before passing them to the neural network for training.

4.2. Establishment of Baseline Models and Performance Analysis

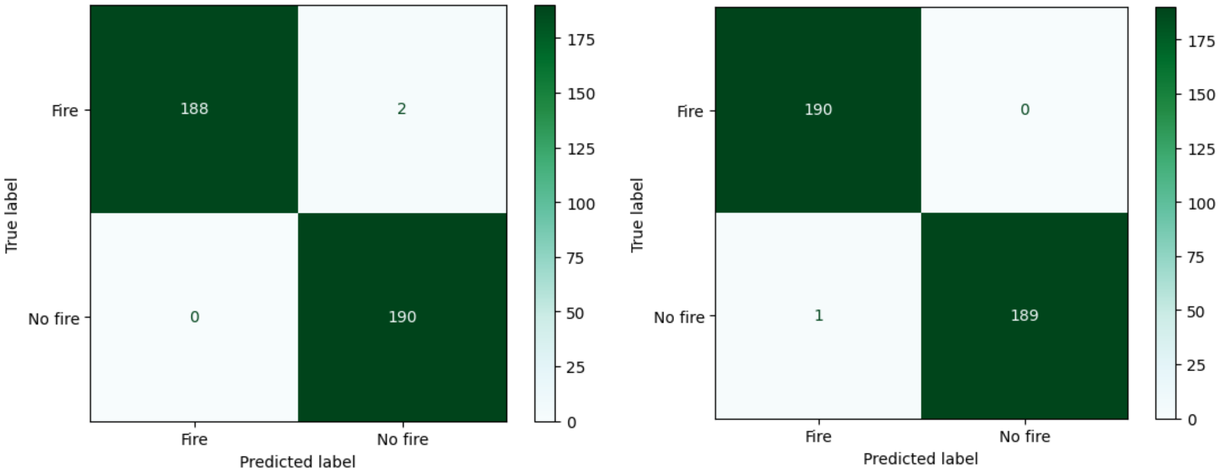
We first trained ResNet50, InceptionV3, and VGG16 on the DeepFire Dataset to establish baseline performance. These models served as benchmarks for assessing the effectiveness of our proposed ensemble and weighted ensemble deep learning methods. We tracked the training and validation accuracies of the three individual classifiers; ResNet50, InceptionV3, and VGG16) over 50 epochs, using a model checkpointing system to save the network weights only when there was an improvement in validation accuracy and the results presented in Fig. 8.

We evaluated the predictive performance of the classifiers on the test set across various metrics, including accuracy, precision, recall, F1-score, ROC-AUC score, false positive rate, false negative rate, and error rate. The confusion matrices for our proposed ensemble deep learning models and the three other individual models are presented in Fig. 7. A confusion matrix is a table that visualizes the performance of a supervised machine learning algorithm when the true labels of the test set are known. Each row of the matrix represents the actual class instances, while each column represents the predicted ones. Diagonal elements indicate correct classifications, while off-diagonal elements indicate misclassifications [24, 35]. From the confusion matrices in Fig. 7, the Y-axis represents the actual classes, while the X-axis represents the predicted classes. From these confusion matrices, we obtained the true positives (instances where the model correctly predicted wildfire), true negatives (situations where there is no wildfire and the model accurately predicts no wildfire), false positives (the model predicts a wildfire, but there is no wildfire), and false negatives (instances where the model predicted no wildfire, but there is an actual wildfire). These values were used to compute the accuracy, precision, recall, F1-score, false positive rate, false negative rate and error rate for each model.



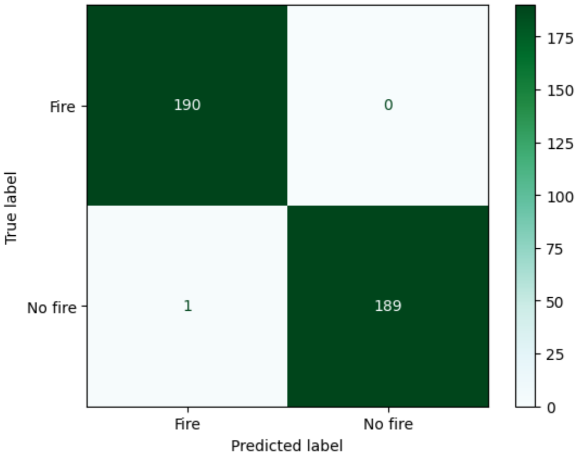
(a) Resnet50

(b) InceptionV3



(c) VGG16

(d) Ensemble deep learning model



(e) Weighted ensemble deep learning model

Fig.7. Confusion matrices of the models; resnet50 (a), inceptionV3 (b), VGG16 (c), ensemble deep learning model and weighted ensemble deep learning model

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

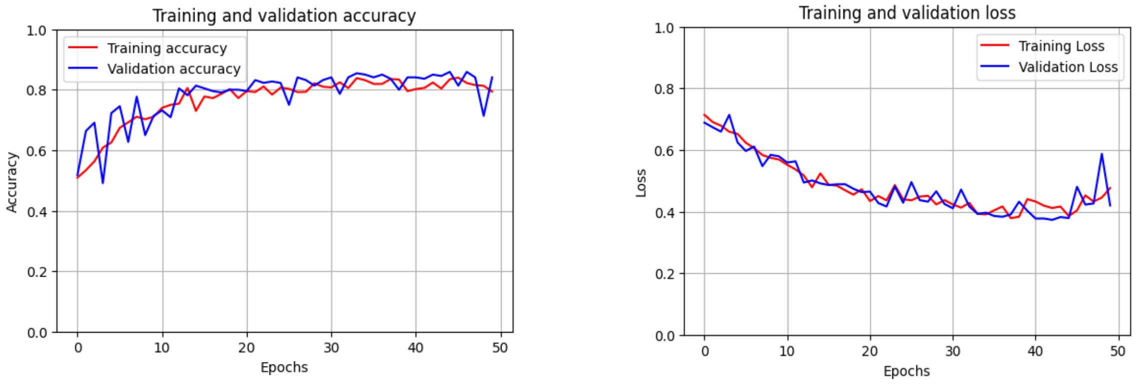
$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$F1 - Score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \tag{7}$$

$$False\ Positive\ Rate\ = \frac{FP}{FP + TN} \tag{8}$$

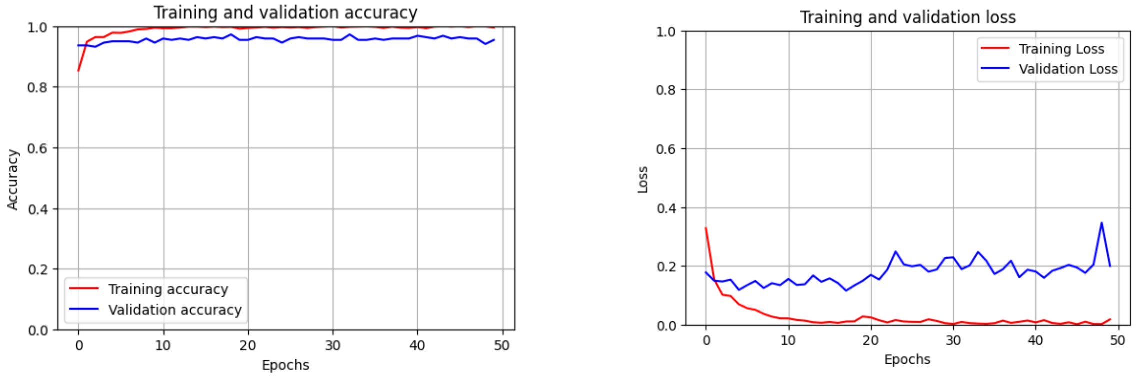
$$False\ Negative\ Rate\ = \frac{FN}{FN + TP} \tag{9}$$

$$Error\ Rate\ = \frac{FN + FP}{TP + TN + FN + FP} \tag{10}$$



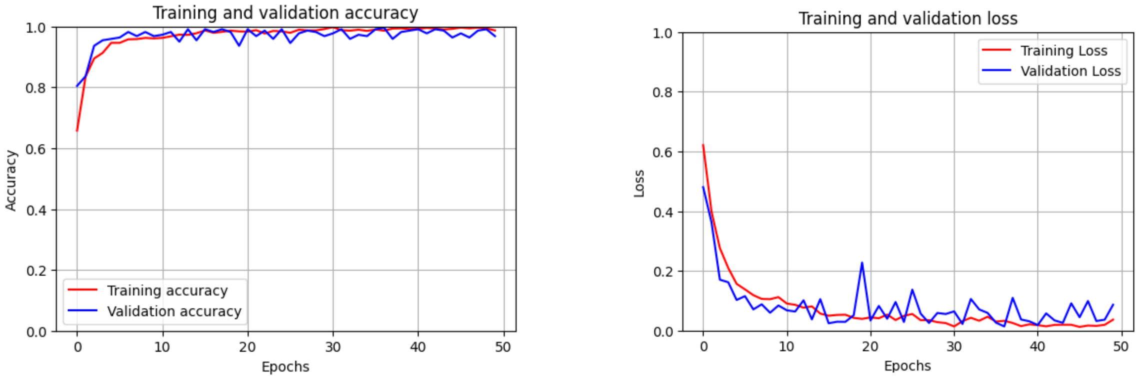
(a) ResNet50

(b) Resnet50



(c) InceptionV3

(d) InceptionV3



(e) VGG16

(f) VGG16

Fig.8. Training and validation accuracy and loss of resnet50 (a and b), inceptionV3 (c and d) and VGG16 (e and f) models

The performance of the trained models, both the three single models and the ensemble and weighted deep learning models are presented in Table 2. The experimental results demonstrate that our proposed ensemble deep learning approaches outperformed the individual baseline models across all the relevant performance metrics, achieving the lowest error rate of 0.3%.

A critical metric in wildfire detection is the false negative rate. A false negative occurs when the model fails to detect an actual wildfire, incorrectly predicting no fire, which can lead to severe consequences. Early detection enables prompt firefighting efforts to contain and mitigate fires before they escalate.

Our analysis revealed that the ResNet50, InceptionV3, and VGG16 models misclassified some wildfire instances as no-wildfire, with false negative rates of 12.6%, 1.1%, and 1.1%, respectively as shown in Fig.9. This indicates that the separate models have limited learning ability and were not able to relevant features from input images to identify and classify wildfire instances in the testing set accurately.

Table 2. Predictive performance of the models on the test set

Metric	ResNet50	InceptionV3	VGG16	Ensemble Deep Learning Model	Weighted Ensemble Deep Learning Model
Accuracy	0.863	0.966	0.995	0.997	0.997
Precision	0.856	0.945	1.000	0.995	0.995
Recall	0.874	0.989	0.989	1.000	1.000
F1-Score	0.865	0.967	0.995	0.998	0.998
ROC-AUC Score	0.863	0.966	0.995	0.997	0.997
Error Rate	0.137	0.034	0.005	0.003	0.003
False Positive Rate	0.147	0.058	0.000	0.005	0.005
False Negative Rate	0.126	0.011	0.011	0.000	0.000

Conversely, our proposed ensemble and weighted ensemble deep learning models outperformed all the single models with a false negative rate of 0.00%. This means both models correctly identified and predicted all the wildfire instances in the test set without a single misclassification. This suggests that integrating the three different individual models allowed them to work synergistically by pooling and combining their differential learning capabilities for better feature representation and perfect detection of all the wildfire instances in the test set.

Furthermore, we leveraged the area under the receiver operating characteristic curve (ROC-AUC) score to evaluate how well the models performed on the testing set. AUC denotes the degree of separability, while ROC is a probability curve. It indicates how well the model can distinguish between classes. The ROC curve plots recall on the y-axis and false positive rate on the x-axis using different probability thresholds, with the ROC-AUC ranging from 0 to 1. The closer the curve hugs the top left corner of the plot, the better the classifier's performance. By analogy, a model with a high AUC score is more efficient at distinguishing images with fire from those without fire [37, 38]. Our experimental results show that our proposed ensemble deep learning models outperformed the single models, achieving a near-perfect ROC-AUC score of 0.997.

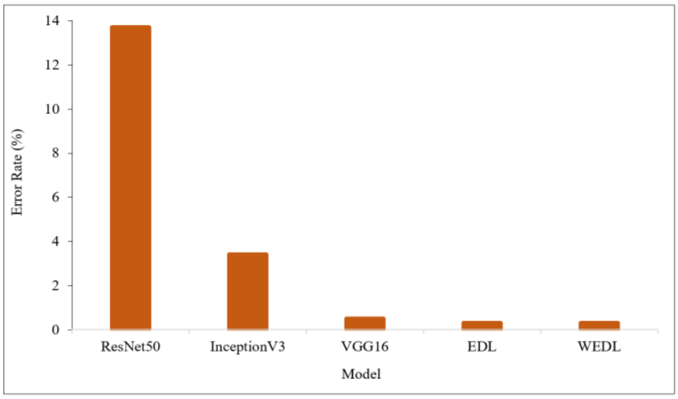


Fig.9. Performance comparison of resnet50, inceptionV3, VGG16, ensemble deep learning (EDL) and weighted ensemble deep learning (WEDL) models using error rates

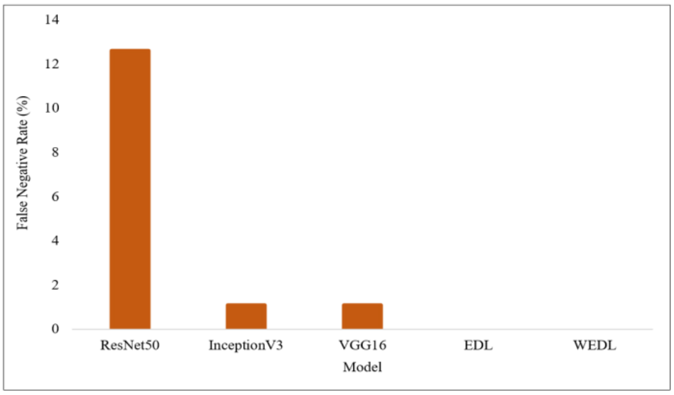
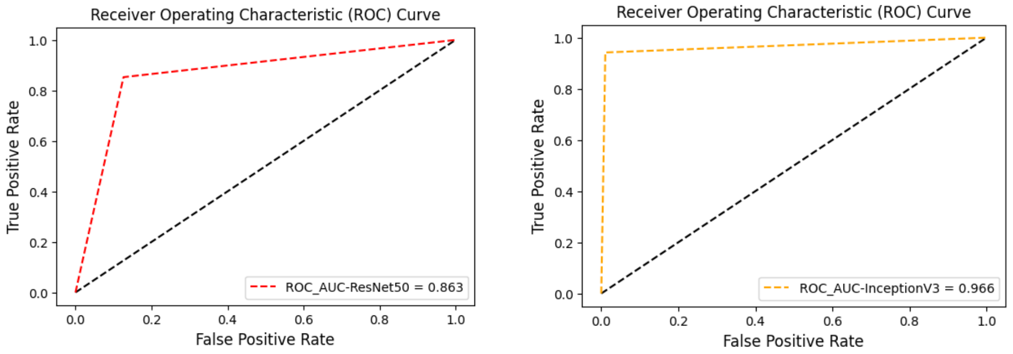
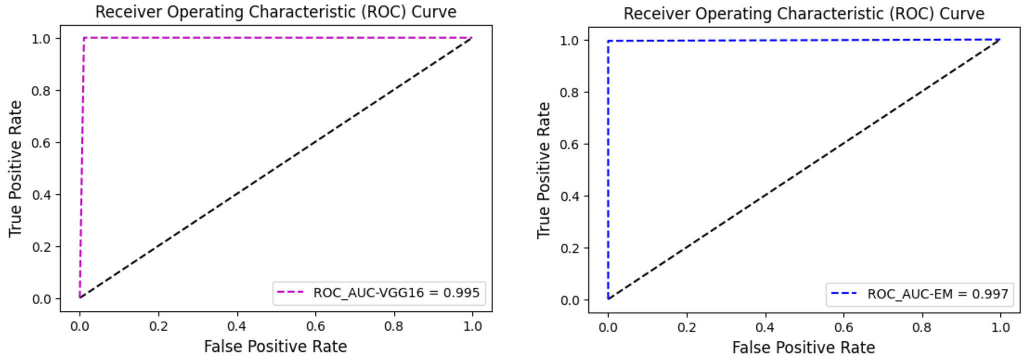


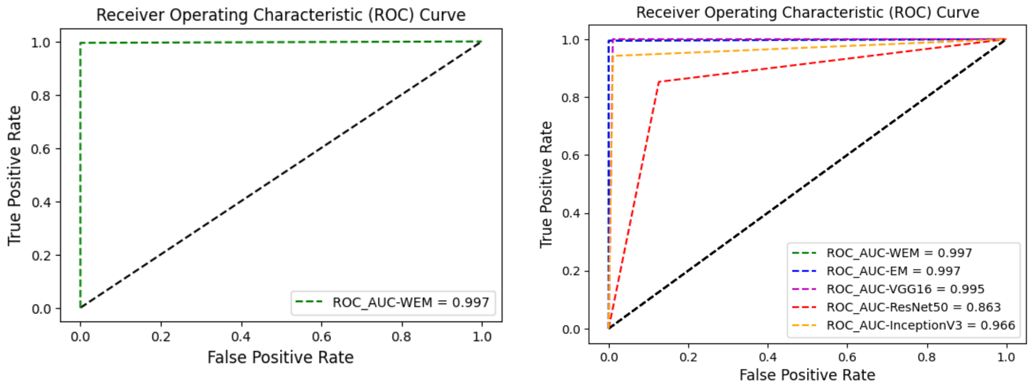
Fig.10. Performance comparison of resnet50, inceptionV3, VGG16, ensemble deep learning (EDL) and weighted ensemble deep learning (WEDL) models using false negative rates



(a) ResNet50 (b) InceptionV3



(c) VGG16 (d) Ensemble deep learning model



(e) Weighted ensemble deep learning model (f) All five models

Fig.11. ROC-AUC curves of resnet50 (a), inceptionV3 (b), VGG16 (c), ensemble deep learning model (d), weighted ensemble deep learning model (e) and all models (f)

4.3. Weighted Ensemble Deep Learning Approach

To develop an optimized weighted ensemble deep learning model, we systematically assigned weights by experimenting with values ranging from 1 to 20, increasing by 1 each time. This process involved iteratively combining different weight values for each model and evaluating the performance of the weighted ensemble model on the testing dataset. By recording the performance for each weight combination, we identified the optimal combination that maximized the chosen performance metrics.

4.4. Effect of Integrating and Weighting the Predictive Capabilities of the Individual Models

The study aimed to compare the performance of ResNet50, InceptionV3, and VGG16 models with our proposed ensemble and weighted ensemble deep learning models in detecting and predicting wildfire instances using the DeepFire dataset, a benchmark for binary wildfire classification. Our comprehensive analysis of the experimental results revealed that the ensemble models, which combine the predictive strengths of ResNet50, InceptionV3, and VGG16, significantly outperformed each separate model, achieving an impressively low error rate of 0.3% as presented in Table 2. This is a significant improvement over the performance of the best individual model, VGG16, which had an error rate of 0.5%. Although the separate individual models produced some misclassifications, they demonstrated differential learning strengths and weaknesses. While some models produced lower false positives and higher false negatives, others produced higher false positives and lower false negatives.

Among the three single models, VGG16 and InceptionV3 consistently demonstrated superior performance over ResNet50 across all the metrics while ResNet50 achieved the lowest performance across all the metrics. Despite ResNet50's ability to address the vanishing or exploding gradient problem, its performance was hampered by the intricate nature of wildfire images. As a result, it could not learn and extract the relevant features from the images, resulting in poor performance across all metrics [27]. Central to the architectural components of InceptionV3 is an improved inception module. This module enables parallel multi-scale convolutions, allowing the model to correctly capture diverse and complex patterns in wildfire images. This architectural advancement might be a key factor in InceptionV3's superior performance across all evaluated metrics, as it effectively handles the complexity and diversity of wildfire imagery [9, 39].

Finally, we compared the performance of our proposed ensemble deep learning models with recently reported deep learning-based wildfire detection systems in the literature. Additionally, we evaluated our models against state-of-the-art algorithms trained on the same benchmark dataset (DeepFire dataset) and reported in the literature. In every case, our proposed models outperformed the other models across all the performance metrics used.

Table 3. Performance comparison of our proposed approaches with other reported models

Model	Accuracy	Precision	Recall	Error Rate
Weighted Ensemble Deep Learning Model	0.997	0.995	1.000	0.003
Ensemble Deep Learning Model	0.997	0.995	1.000	0.003
VGG16 [40]	0.950	0.960	0.940	---
ResNet50 [41]	0.880	---	---	---
YOLO [42]	---	0.74	---	---
FFireNet [23]	0.980	0.974	0.995	0.016
VGG19 [21]	0.950	0.957	0.942	0.005

By leveraging the complementary strengths of ResNet50, InceptionV3 and VGG16 models, our ensemble deep learning approach not only achieved higher accuracy but also exhibited greater robustness and reliability, making it one of the most effective solutions for wildfire detection reported in the literature using the DeepFire dataset as the benchmark.

5. Conclusions

In this study, we reviewed recent literature to understand the progress researchers have made in enhancing the performance and reliability of automated wildfire detection systems. Despite these advancements, our thorough literature review revealed that none of the existing studies considered assigning weight to the learning capabilities of individual models when developing an ensemble deep learning model. To address this gap, we trained and tested two ensemble deep learning models using the DeepFire dataset. The experimental results demonstrated that our proposed ensemble models outperformed the separate ResNet50, InceptionV3, and VGG16 models across all performance metrics used in the study. However, there is still room for improvement in future work to enhance the robustness of computer vision-based wildfire detection systems. Future research should consider implementing our proposed models on larger datasets, including the DeepFire, satellite imagery, and other benchmark datasets captured under diverse environmental scenarios. Additionally, future research should focus on reducing the size of our proposed models to improve their efficiency further.

References

[1] M. Castelli, L. Vanneschi, and A. Popovič, “Predicting burned areas of forest fires: An artificial intelligence approach,” *Fire Ecol.*, vol. 11, no. 1, pp. 106–118, 2015, doi: 10.4996/fireecology.1101106.

[2] K. Bot and J. G. Borges, “A Systematic Review of Applications of Machine Learning Techniques for Wildfire Management Decision Support,” *Inventions*, vol. 7, no. 1, 2022, doi: 10.3390/inventions7010015.

[3] A. S. Mahdi and S. A. Mahmood, “Analysis of Deep Learning Methods for Early Wildfire Detection Systems: Review,” *IICETA 2022 - 5th Int. Conf. Eng. Technol. its Appl.*, no. October, pp. 271–276, 2022, doi: 10.1109/IICETA54559.2022.9888515.

[4] K. Dimitropoulos, O. Gunay, K. Kose, F. Erden, and F. Chaabene, “Flame Detection for Video-Based Early Fire Warning,” *Prog. Cult. Herit. Preserv. 4th Int. Conf. EuroMed 2012, Limassol, Cyprus, Oct. 29 – Novemb. 3, 2012*, pp. 378–387, 2012, [Online]. Available: http://link.springer.com/book/10.1007/978-3-642-34234-9/page/5.

[5] J. San-Miguel-Ayanz, N. Ravail, V. Kelha, and A. Ollero, “Active fire detection for fire emergency management: Potential and limitations for the operational use of remote sensing,” *Nat. Hazards*, vol. 35, no. 3, pp. 361–376, 2005, doi: 10.1007/s11069-004-1797-2.

[6] S. Dogan *et al.*, “Automated accurate fire detection system using ensemble pretrained residual network,” *Expert Syst. Appl.*, vol. 203, no. January, p. 117407, 2022, doi: 10.1016/j.eswa.2022.117407.

[7] P. Barmpoutis, P. Papaioannou, K. Dimitropoulos, and N. Grammalidis, “A review on early forest fire detection systems using optical remote sensing,” *Sensors (Switzerland)*, vol. 20, no. 22, pp. 1–26, 2020, doi: 10.3390/s20226442.

[8] M. J. Sousa, A. Moutinho, and M. Almeida, “Wildfire detection using transfer learning on augmented datasets,” *Expert Syst. Appl.*, vol. 142, p. 112975, 2020, doi: https://doi.org/10.1016/j.eswa.2019.112975.

[9] V. E. Sathishkumar, J. Cho, M. Subramanian, and O. S. Naren, “Forest fire and smoke detection using deep learning-based learning without forgetting,” *Fire Ecol.*, vol. 19, no. 1, 2023, doi: 10.1186/s42408-022-00165-0.

[10] M. Iman, H. R. Arabnia, and K. Rasheed, “A Review of Deep Transfer Learning and Recent Advancements,” *Technologies*, vol. 11, no. 2, pp. 1–14, 2023, doi: 10.3390/technologies11020040.

[11] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, “An Adaptive Ensemble Machine Learning Model for Intrusion Detection,” *IEEE Access*, vol. 7, pp. 82512–82521, 2019, doi: 10.1109/ACCESS.2019.2923640.

[12] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, “A forest fire detection system based on ensemble learning,” *Forests*, vol. 12, no. 2, pp. 1–17, 2021, doi: 10.3390/f12020217.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017, doi: 10.1145/3065386.

[14] J. Schmidhuber, “Deep learning in neural networks: An overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015, doi: https://doi.org/10.1016/j.neunet.2014.09.003.

[15] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, and S. W. Baik, “Efficient Deep CNN-Based Fire Detection and Localization in Video Surveillance Applications,” *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 49, no. 7, pp. 1419–1434, 2019, doi: 10.1109/TSMC.2018.2830099.

[16] B. Kim and J. Lee, “A video-based fire detection using deep learning models,” *Appl. Sci.*, vol. 9, no. 14, 2019, doi: 10.3390/app9142862.

[17] C. Bahhar *et al.*, “Wildfire and Smoke Detection Using Staged YOLO Model and Ensemble CNN,” *Electron.*, vol. 12, no. 1, pp. 1–15, 2023, doi: 10.3390/electronics12010228.

[18] kmalbek B. AbdusalomAov, B. M. S. Islam, R. Nasimov, M. Mukhiddinov, and T. K. Whangbo, “An Improved Forest Fire Detection Method Based on the Detectron2 Model and a Deep Learning Approach,” *Sensors*, vol. 23, no. 3, 2023, doi: 10.3390/s23031512.

[19] A. S. Almasoud, “Intelligent Deep Learning Enabled Wild Forest Fire Detection System,” *Comput. Syst. Sci. Eng.*, vol. 44, no. 2, pp. 1485–1498, 2023, doi: 10.32604/csse.2023.025190.

[20] S. T. Seydi, V. Saeidi, B. Kalantar, N. Ueda, and A. A. Halin, “Fire-Net: A Deep Learning Framework for Active Forest Fire Detection,” *J. Sensors*, vol. 2022, 2022, doi: 10.1155/2022/8044390.

[21] A. Khan, B. Hassan, S. Khan, R. Ahmed, and A. Abuassba, “DeepFire: A Novel Dataset and Deep Transfer Learning Benchmark for Forest Fire Detection,” *Mob. Inf. Syst.*, vol. 2022, 2022, doi: 10.1155/2022/5358359.

[22] A. V. Jonnalagadda and H. A. Hashim, “SegNet: A segmented deep learning based Convolutional Neural Network approach for drones wildfire detection,” *Remote Sens. Appl. Soc. Environ.*, vol. 34, no. February, 2024, doi: 10.1016/j.rsase.2024.101181.

[23] S. Khan and A. Khan, “FFireNet: Deep Learning Based Forest Fire Classification and Detection in Smart Cities,” *Symmetry (Basel).*, vol. 14, no. 10, 2022, doi: 10.3390/sym14102155.

[24] J. Brownlee, “Develop Deep Learning Models On Theano And TensorFlow Using Keras,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2019.

[25] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance,” *Appl. Soft Comput.*, vol. 97, p. 105524, 2020, doi: 10.1016/j.asoc.2019.105524.

[26] V. N. G. Raju, K. P. Lakshmi, V. M. Jain, A. Kalidindi, and V. Padma, “Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification,” *Proc. 3rd Int. Conf. Smart Syst. Inven. Technol. ICSSIT 2020*, no. Icssit, pp. 729–735, 2020, doi: 10.1109/ICSSIT48917.2020.9214160.

[27] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[28] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[29] N. S. Shadin, S. Sanjana, and N. J. Lisa, “COVID-19 Diagnosis from Chest X-ray Images Using Convolutional Neural Network(CNN) and InceptionV3,” *2021 Int. Conf. Inf. Technol. ICIT 2021 - Proc.*, vol. 3, no. September 2012, pp. 799–804, 2021, doi: 10.1109/ICIT52682.2021.9491752.

[30] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.

[31] C. Lin, L. Li, W. Luo, K. C. P. Wang, and J. Guo, “Transfer learning based traffic sign recognition using inception-v3 model,” *Period. Polytech. Transp. Eng.*, vol. 47, no. 3, pp. 242–250, 2019, doi: 10.3311/PPtr.11480.

[32] S. Tammina, “Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images,” *Int. J. Sci. Res. Publ.*, vol. 9, no. 10, p. p9420, 2019, doi: 10.29322/ijsrp.9.10.2019.p9420.

[33] Z. Zahisham, C. P. Lee, and K. M. Lim, “Food Recognition with ResNet-50,” *IEEE Int. Conf. Artif. Intell. Eng. Technol. IICAIET 2020*, pp. 0–4, 2020, doi: 10.1109/IICAIET49801.2020.9257825.

[34] J. Cao, M. Yan, Y. Jia, X. Tian, and Z. Zhang, “Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals,” *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, 2021, doi: 10.1186/s13634-021-00740-8.

[35] A. F. Agarap, “Deep Learning using Rectified Linear Units (ReLU),” no. 1, pp. 2–8, 2018, [Online]. Available: http://arxiv.org/abs/1803.08375.

[36] X. Wang, Y. Zhong, L. Jin, and Y. Xiao, “Scale Adaptive Graph Convolutional Network for Skeleton-Based Action Recognition,” *Tianjin Daxue Xuebao (Ziran Kexue yu Gongcheng Jishu Ban)/Journal Tianjin Univ. Sci. Technol.*, vol. 55, no. 3, pp. 306–312, 2022, doi: 10.11784/tdxbz202012073.

[37] D. Chicco and G. Jurman, “The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification,” *BioData Min.*, vol. 16, no. 1, pp. 1–23, 2023, doi: 10.1186/s13040-023-00322-4.

[38] H. El Massari, Z. Sabouri, S. Mhammedi, and N. Gherabi, “Diabetes Prediction Using Machine Learning Algorithms and Ontology,” *J. ICT Stand.*, vol. 10, no. 2, pp. 319–338, 2022, doi: 10.13052/jicts2245-800X.10212.

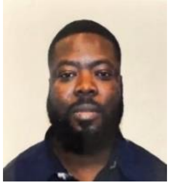
[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[40] L. Zhang, M. Wang, Y. Fu, and Y. Ding, “A Forest Fire Recognition Method Using UAV Images Based on Transfer Learning,” *Forests*, vol. 13, no. 7, 2022, doi: 10.3390/f13070975.

[41] S. Treneska and B. R. Stojkoska, “Wildfire detection from UAV collected images using transfer learning,” no. August, 2021, [Online]. Available: https://www.researchgate.net/publication/353732371.

[42] M. Mukhiddinov, A. B. Abdusalomov, and J. Cho, “A Wildfire Smoke Detection System Using Unmanned Aerial Vehicle Images Based on the Optimized YOLOv5,” *Sensors*, vol. 22, no. 23, 2022, doi: 10.3390/s22239384.

Authors’ Profiles



Lawrence Owusu is a PhD Computational Data Science and Engineering student at North Carolina Agricultural and Technical State University. He holds BSc. Natural Resources Management from Kwame Nkrumah University of Science and Technology, Ghana, M.S. Agricultural and Environmental Systems and M.S. Data Science and Engineering from North Carolina A and T State University, Greensboro, USA. His research interests deep learning, data security, wildfire detection, and quantum computing.

Robert B. Eshun is a PhD Computational Science and Engineering candidate at North Carolina A & T State University specializing in ML/AI enabled methods for cancer diagnosis and gene discovery. He obtained his BSc. Computer Science and Physics and MPhil. Computational Nuclear Science and Engineering at the University of Ghana. Robert has interests at the intersection of Bioinformatics and Machine Learning particularly in the areas of deep learning and graph-based models applied to histopathology mages, neuro-images and genomic profiles for feature extraction and discovery, predictive modeling of health-oriented tasks and integrative analysis of the data modalities for informative decision support.

Dr. Leila Hashemi-Beni is an Associate Professor at North Carolina Agricultural and Technical State University. She holds BSc. Civil Engineering from the University of Isfahan, M.S. in Civil Engineering from University of Tehran and PhD in Geomatics Science from Laval University, Iran. Her research interests include 3D modeling, Sensor Integration, multi-source, multi-temporal, geospatial data modeling, UAV and satellite remote sensing, spatial-temporal data analysis, remote sensing image classification, Transportation and Urban Planning

Dr. Ali AlQahtani is an Assistant Professor at North Carolina Agricultural and Technical State University, where he founded and directs the Context-based Authentication Laboratory (CAL). He holds B.S. in Computer Science from Grambling State University, M.S. Computer and Information Science from Arkansas University, USA, M.S.Eng. Electrical Engineering and Ph.D. Cyberspace Engineering from Louisiana Tech University, USA. His research interests are cybersecurity, context-based authentication, user authentication, and the design of Internet of Things (IoT)

Dr. Masud R. Rashel holds PhD. in Earth and Space Sciences from the University of Évora since 2018. He completed his MSc and BSc in computer science and engineering from United International University, Bangladesh in 2014 and 2008, respectively. Previously, he served as an Assistant Professor at Daffodil International University in Bangladesh. Currently, he works as a Postdoctoral Researcher at the University of Évora, Portugal, specializing in renewable energy, machine learning, and the Internet of Things (IoT) within engineering and natural sciences

Dr. AKM K. Islam is an Assistant Professor and Graduate Coordinator for the Department of Computational Data Science and Engineering at North Carolina Agricultural and Technical State University, USA. He holds B.S. in Computer Science and Engineering from Jahangirnagar University, Bangladesh, MPH in Computing and Informatics from Bournemouth University, M.S. and PhD in Computer Science from Georgia State University, USA. He is an investigational and experimental computer scientist in the field of data mining, machine learning, and computational biology, who likes to contrive novel real-world problems with practical solutions through interdisciplinary collaborations that can affect the existing world and paves the way for future. His research interest focuses on designing and implementing machine learning and deep learning algorithms for biomedical image processing and bioinformatics gene expression analysis.

How to cite this paper: Lawrence Owusu, Robert B Eshun, Leila Hashemi-Beni, Ali AlQahtani, Masud R Rashel, AKM K. Islam, "Reduction of Misclassifications in Wildfire Detection: A Weighted Ensemble Deep Learning Approach", International Journal of Intelligent Systems and Applications(IJISA), Vol.16, No.5, pp.53-67, 2024. DOI:10.5815/ijisa.2024.05.05