



OPEN ACCESS

EDITED BY

Jawad Ahmad,
Swedish College of Engineering and
Technology, Pakistan

REVIEWED BY

R Suresh,
M S Ramaiah University of Applied Sciences,
India
Harish Kumar Muniswamy,
Christ University, India

*CORRESPONDENCE

Satish Kumar,
✉ satishkumar.vc@gmail.com
Ketan Kotecha,
✉ director@sitpune.edu.in

RECEIVED 29 May 2024

ACCEPTED 06 September 2024

PUBLISHED 03 September 2024

CITATION

Joshi DD, Kumar S, Patil S, Kamat P, Kolhar S and
Kotecha K (2024) Deep learning with ensemble
approach for early pile fire detection using
aerial images.

Front. Environ. Sci. 12:1440396.

doi: 10.3389/fenvs.2024.1440396

COPYRIGHT

© 2024 Joshi, Kumar, Patil, Kamat, Kolhar and
Kotecha. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

Deep learning with ensemble approach for early pile fire detection using aerial images

Dhyey Divyeshkumar Joshi^{1,2}, Satish Kumar^{1,3*}, Shruti Patil^{1,3},
Pooja Kamat³, Shrikrishna Kolhar³ and Ketan Kotecha^{1,3*}

¹Symbiosis Centre of Applied Artificial Intelligence, Symbiosis International (Deemed University), Pune, India, ²Devang Patel Institute of Advance Technology and Research, Charusat University, Anand, Gujarat, India, ³Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune, Maharashtra, India

Wildfires rank among the world's most devastating and expensive natural disasters, destroying vast forest resources and endangering lives. Traditional firefighting methods, reliant on ground crew inspections, have notable limitations and pose significant risks to firefighters. Consequently, drone-based aerial imaging technologies have emerged as a highly sought-after solution for combating wildfires. Recently, there has been growing research interest in autonomous wildfire detection using drone-captured images and deep-learning algorithms. This paper introduces a novel deep-learning-based method, distinct in its integration of infrared thermal, white, and night vision imaging to enhance early pile fire detection, thereby addressing the limitations of existing methods. The study evaluates the performance of machine learning algorithms such as random forest (RF) and support vector machines (SVM), alongside pre-trained deep learning models including AlexNet, Inception ResNetV2, InceptionV3, VGG16, and ResNet50V2 on thermal-hot, green-hot, and white-green-hot color images. The proposed approach, particularly the ensemble of ResNet50V2 and InceptionV3 models, achieved over 97% accuracy and over 99% precision in early pile fire detection on the FLAME dataset. Among the tested models, ResNet50V2 excelled with the thermal-fusion palette, InceptionV3 with the white-hot and green-hot fusion palettes, and VGG16 with a voting classifier on the normal spectrum palette dataset. Future work aims to enhance the detection and localization of pile fires to aid firefighters in rescue operations.

KEYWORDS

aerial imaging¹, deep learning², fire piles monitoring³, thermal infrared images⁴, RGB images⁵, wildfire detection system⁶

1 Introduction

Wildfires have recently caused significant damage to forests, wildlife habitats, farms, residential areas, and entire ecosystems. Data from the National Interagency Fire Center (NIFC) reveals that between 2012 and 2021 in the United States, 22,897 fires ravaged more than 1,082,125 acres, resulting in over \$7.4 billion in losses. Fires are rare but can quickly lead to extensive harm and property damage. The National Fire Protection Association (NFPA) predicts that in 2022, the United States Emergency Fire Service will respond to approximately 1,089,844 fires, leading to 1,149 civilian deaths and an estimated \$18 billion in property damage (National Fire News), (Statistics). Early identification of fires without

false alarms was critical to reduce such catastrophes. As a result, various active fire protection methods were created and are widely used in the current world. There were several methods available for detecting and monitoring fires. Satellite imagery, remotely piloted vehicles (RPV), and sensors were examples of this. These systems, however, do not enable real-time fire detection and monitoring. Some disadvantages of these technologies include: i) fires were not noticed in their early stages, ii) satellites have larger time delays, and iii) sensors become an infeasible option due to their limited range. Even if spectral imaging is employed, it may not provide an accurate result at night or in bad weather. To fight against fire at the right moment, individuals were now engaged to watch the forest, risking their lives. These troubling facts inspire academics to develop innovative early pile burn identification and monitoring methods.

In particular, technical innovations in airborne surveillance technologies can provide early rescuers and tactical soldiers with more precise fire behavioral information for improved fire catastrophe prevention and mitigation. Using computer vision and remote tracking imagery, remotely piloted aircraft proved highly effective for real-time fire monitoring and detection. Unmanned aerial vehicles (UAVs), also known as drones, offer a quicker, more agile, and cost-effective method for forest surveillance. Enhancing UAVs with remote sensors can significantly improve current techniques. Additionally, UAVs can operate in hazardous areas that are unsafe for human access. Traditional methods of finding and monitoring flames included stationing personnel in observation stations or sending choppers or corrected airborne vehicles to surveil fires using long-wave thermal vision (Habiboğlu et al., 2012, Rafiee et al., 2011, Jadon et al., 2019). Current research has examined wireless mesh network-based Internet of Things (IoT) advancements, but such infrastructures would require additional funding and verification before providing meaningful information. Satellite data was frequently used for international fire assessment at larger scales, albeit at relatively low resolution and with repeat imagery availability controlled by satellite circular trends (Afghah et al., 2018a).

Given these systems' limits and problems, the use of Unmanned Drones for fire monitoring has grown in recent decades. UAVs provide innovative characteristics and benefits such as quick deployment, mobility, a broader and adjustable field of view, and little operator engagement (Afghah et al., 2018b, Aggarwal et al., 2020, Afghah et al., 2019). UAVs can offer the following advantages: they cover a large area, even in overcast conditions, work at all hours of the day and night, are readily recovered, and are reasonably affordable, electric UAVs are also beneficial to the environment, transport various payloads for various missions and effectively cover target area (Casbeer et al., 2006, Yuan et al., 2015a) and, most critically, missions may be completed automatically without or with minimal human pilot/operator interaction. Drone technology development has increased in recent years. Nowadays, unmanned aerial vehicles (UAVs) are employed in emergency relief scenarios and activities such as forest fires and flooding, especially as a short-term solution whenever domestic connections crash due to destroyed infrastructure, miscommunication, or bandwidth limits (Shamsoshoara et al., 2020a, Shamsoshoara, 2019).

Artificial intelligence (AI) and machine learning advancements in recent years have boosted the success of image categorization, detection, legitimate forecasting, and other applications. Artificial

intelligence (AI) can be utilized in numerous fields. The advancement of nanotechnology in semiconductor materials has led to the creation of advanced tensor and graphical processing units, which offer exceptional processing power for data-driven approaches. Moreover, modern drones and aerial vehicles can be equipped with compact edge technology TPU and GPU units for real-time processing, enabling early detection of fires and preventing major disasters (NVIDIA Jetson Nano Developer Kit), (Edge TPU).

Several guided learning algorithms rely on massive training sets to produce a suitably accurate model (Wu et al., 2020). Using a public fire dataset, their research detected fires using pre-trained convolutional neural networks such as mobile and AlexNet architecture. That dataset, however, was based on images of the fire acquired from the ground. To our knowledge, the fire luminosity airborne-based machine learning evaluation (FLAME) dataset was the only airborne imagery collection for pile-fire research that could be used to build fire modeling or advanced analytics systems for airborne surveillance systems. The authors used their own gathered normal-spectrum fire dataset to detect fire using XceptionNet's pre-trained convolutional neural network (CNN) architectures (Shamsoshoara et al., 2021).

In this article, researchers have utilized the FLAME dataset for pile burns. It is a drone-captured pile of fire clips and photos taken during organized slashed piles in North Arizona. Pile burns were extremely useful for studying spot fires and early-stage fires. Following thinning and restoration initiatives, forest management generally uses pile fires to clear up forest leftovers, such as branches and leaves. Various UAVs took the imagery with distinct viewpoints, magnification levels, and camera kinds, including conventional and infrared cameras. Forestry interventions were an essential strategy for reducing biofuels, and blazing chopped piles were often the most cost-effective and secure technique for re-moving gash. Large burning regions were often hard to manage, resulting in massive deforestation from wildfires. As a result, any controlled burnings were accompanied by comprehensive forest management strategies to avoid fire. Slash or debris burning was common for certain farms and families to burn leaves, crop remnants, and other agricultural trash. This method was both faster and handier than taking them away. However, being near woods increases the danger of wildfires and affects the environment, not to mention that it was outlawed in many nations and states across the United States. Using automated aerial surveillance devices can significantly decrease the effort of forest managers. Google's Federated Transfer Learning (FTL) is highlighted as a promising decentralized learning methodology that can be integrated with UAV-based fire detection systems, enhancing the efficiency and accuracy of models operating with limited resources. (Bonawitz et al., 2019). Portable technologies, such as unmanned aerial vehicles (drones), employ localized information to train a model with raw data the drone collects. They then communicate anticipated data with the FTL. Later, the FTL gets all model parameters via various drones and starts combining them before releasing an optimal model to all drones so they can concur rapidly. FTL techniques provide data secrecy and confidentiality; these techniques are well suited to narrow bandwidth and portable devices with limited charging capacity. Furthermore, it exchanges model parameters

with the FTL rather than all images and videos, reducing latency output.

The current FLAME dataset was gathered utilizing several drones (Shamsoshoara et al., 2021, Shamsoshoara et al., 2020b). As a result, the FTL solutions in this data collection are vital for industrial difficulties and open issues (Brik et al., 2020). The FLAME collection contains aerial imagery and clips of two sorts of palettes: normal range spectrum palettes and thermal fusion palettes, taken by advanced high-resolution cameras attached with drones which give four output visuals such as normal-spectrum for RGB color-space, Fusion heat map, white-fusion, and lastly, green-fusion having thermal palettes. In this work, the authors focused on individual binary classification tasks to assess the dataset's utility for extreme fire control problems. This paper presents a comprehensive comparative study of the performance of machine learning and deep learning models on the FLAME dataset. Gabor filter and Sobel filters extract features, and RF and SVM classifiers are used to detect early pile fire detection. Transfer learning is used to design and implement deep learning models. An ensemble voting classifier is used to further improve the deep learning models' performance. The proposed deep learning technique demonstrated robust performance on the FLAME dataset, which consists of aerial imagery for pile burn surveillance, including both normal spectrum and thermal heat map frames.

2 Related work

Recent deep learning applications include detecting and classifying visual objects, audio recognition, and natural language processing. To increase performance, researchers have done several experiments on fire detection based on deep learning. Deep learning varies significantly from conventional computer vision-based fire detection. As a result, the effort required to identify the appropriate handcrafted qualities has been redirected toward establishing a strong network and preparing training data. Another distinction was that the detector/classifier may be obtained by simultaneously training with the features in the same neural network.

Frizzi et al. (2016) proposed a fire detection network based on CNN, where features were simultaneously learned by training a multilayer perceptron (MLP) neural network classifier. Zhang et al. (2016) also presented a CNN-based fire detection system that works in a chain. If a fire is discovered, a fine-grained patch classifier is employed to pinpoint the fire patches precisely. Muhammad et al., 2019 suggested a fire detection system based on a finely tuned CNN fire detector. Squeeze Net-inspired (Ye et al., 2013) optimum neural network architecture was proposed for fire identification, positioning, and cognitive comprehension of the fire situation. Saydirasulovich et al. (2023) used the YOLOv8 model to detect forest fire smoke using UAV images. Abdusalomov et al. (2023) used the Detectron2 platform based on deep-learning methods for fast and accurate forest fire detection.

A unit in the deep layer of CNN has a huge receptive field; therefore, its activation may be viewed as a feature containing a big region of context information. This was another advantage of using CNN to train features for fire detection. Shamsoshoara et al. (2021) produced the FLAME dataset for pile burns in the forest of Northern Arizona, one of the key reference papers researchers have

investigated. The FLAME collection contains aerial imagery and clips of two sorts, normal range spectrum and thermal fusion spectrum, taken by advanced high-resolution cameras that provide four output visuals: normal-spectrum view, Fusion heat map view, white-hot view, and lastly, green-hot view. The video frames were employed for two purposes: fire presence or absence classification and fire detection. The authors used an artificial neural network for fire absence or presence classification and U-Net architecture for fire detection. Xuan Truong and Kim (2012) used an adaptive Gaussian mixture model to detect moving regions, a fuzzy c-means algorithm to segment fire regions, and SVM applied on parameters extracted to detect fire and non-fire regions. Rinsurongkawong et al. (2012) proposed a method for early fire detection based on the Lucas-Kanade optical flow algorithm.

3 Proposed methodology

Image identification is among the most challenging tasks in computer vision. Image identification in the past algorithms used RGB or Red and Yellow or only green channel comparison to distinguish distinct items in frames or films, such as fire.

RGB luminosity comparing methods, for example, which normally employ a specified threshold to locate the fire, may identify sunset and sunrise as false positives. A guided machine learning technique and certain advanced CNN principles are used in this work to detect pile fires in a region using drone-camera-captured footage, as shown in Figure 1. When pile-fire and no-pile-fire components coexist in a mixed image, the frame is designated "pile-fire," when there is no pile burn, the frame is labeled "no-pile-fire." The Forward-Looking-Infrared (FLIR)-Vue-Pro-R-Camera, Zen-muse X4S, and DJI Phantom 3 cameras are used to identify four different types of recorded clips, such as white-hot, green-hot, thermal heat mapping, and normal spectrum of imagery. The binary classification models employed in this study are the Resnet50V2 network, the AlexNet network, the InceptionV3 network, and a VGG-16 network integrated with an ensemble voting classifier. The selection of ResNet50V2, AlexNet, InceptionV3, and VGG-16 models was based on their proven efficacy in handling complex image classification tasks and their ability to generalize well across varied datasets, making them suitable for the challenging task of early fire detection in aerial imagery. Since pile-fire detection is a binary classification problem with two possible outcomes—pile-fire and no-pile-fire—the output layer consisted of two neurons and employed a sigmoid activation function. The sigmoid activation function is represented by the following Equation 1.

$$p(\text{target.label} = \text{pile-fire}) = \sigma(\text{target.label} = \text{pile-fire} | \zeta(\theta)) = 1 / (1 + e^{-\zeta(\theta)}) \quad (1)$$

where $\zeta(\theta)$ was the output layer value obtained from the input neuron, each neuron is nothing more than a pixel value of each pixel, then summed it up with the hidden network weights. The output value of the Sigmoid activation function was the probability of pile-fire detection based on the loaded frames into the network. While training the proposed pre-trained

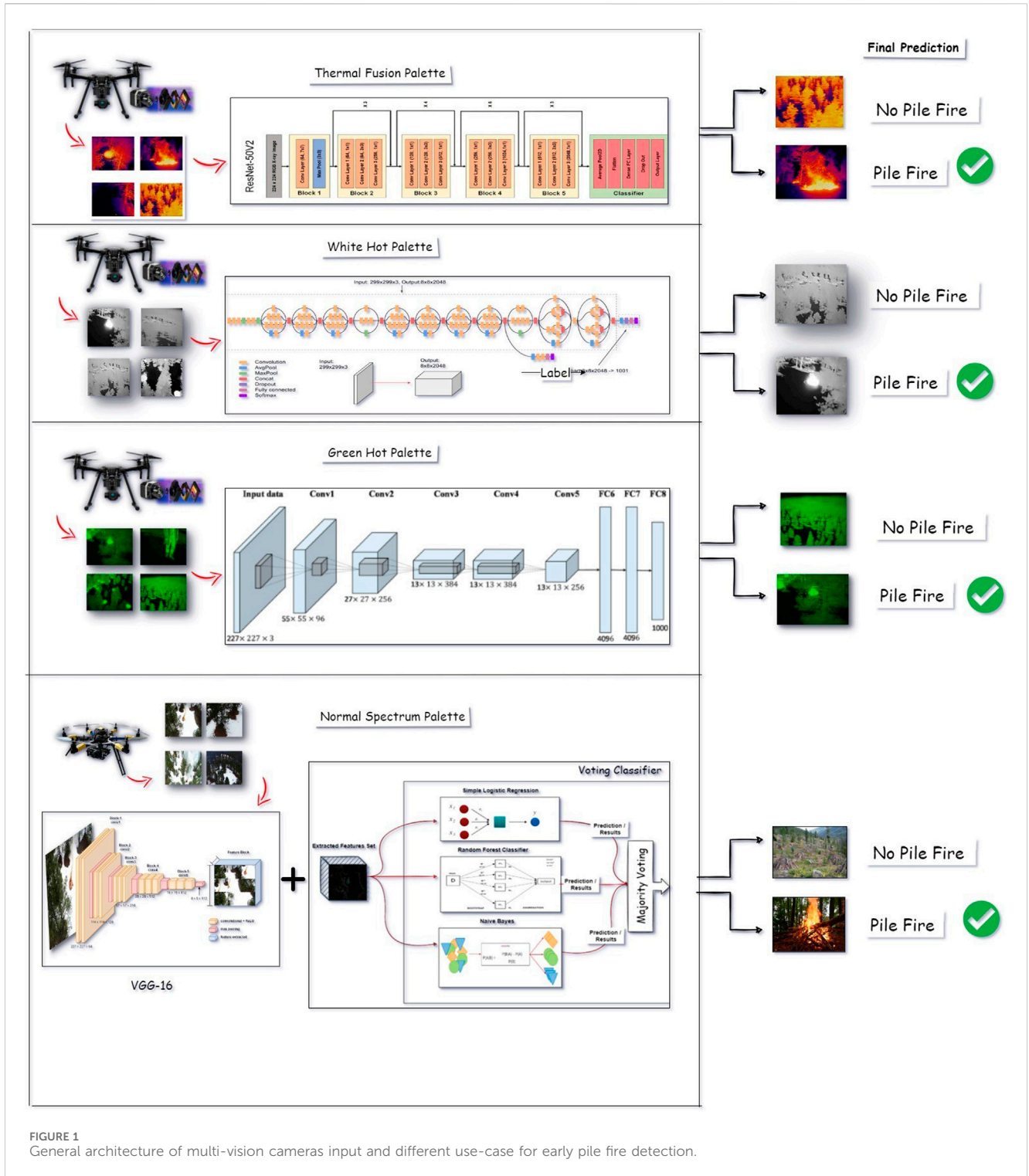


FIGURE 1 General architecture of multi-vision cameras input and different use-case for early pile fire detection.

networks, the network weights are calculated using weight initialization techniques; also, a loss function is employed using the tensor-flow library to improve network accuracy and determine the optimal weight values. Although the challenge in this part involves binary-classification, the error function studied is a binary-cross-entropy binary cross entropy which is calculated as shown in Equation 2 below:

$$LBCE(z, \hat{z}) = -\frac{1}{N} \sum z_i * \log(p(\hat{z}_i)) + (1 - z_i) * \log(1 - p(\hat{z}_i)) \quad (2)$$

where N, at each epoch is the sample population in each batch utilized to adjust the gradient descent, and z is the actual true label for the frames training dataset. Classified as pile-fire (z = 1)

and $z = 0$, if classified as no-pile-fire. The projected probability of a frame belonging to the fire class was given by (y) . During the learning phase, the Adamax optimizer optimizer regulates the error function and the optimal weights on each hidden layer.

3.1 Residual neural network–Resnet-50: Version 2.0

ResNet50V2 is a typical pre-trained convolutional neural network capable of image recognition, auto-encoding, and classification. ResNet50V2 is a variant of the ResNet model with forty-eight convolutions, one max-pooling layer, and one average-pooling layer (Xuan Truong and Kim, 2012), (Rinsurongkawong et al., 2012). ResNet50V2 is an upgraded version of the ResNet50 architecture that outperformed ResNet50 and ResNet101 on the ImageNet repository. ResNet50V2 modified the transmission formulation of the linkages between blocks. ResNet50V2 creates a 10×10 2048 feature map from the input picture on its final feature ex-tractor layer. The residual network is generated from a 50-layer residual network architecture and includes 177 layers in total. The weights of the pre-trained network's previous layers (1–174) are frozen. The weights of several early layers can be frozen to accelerate network training and considerably prevent over-fitting to new data.

3.2 AlexNet: Version 2010

The AlexNet model has enhanced network depth compared to LeNet-5. AlexNet encompasses eight tiers of trainable parameters, as discussed in (Hu et al., 2018), (Muhammad et al., 2019). Its architectural design consists of five layers, comprising a max-pooling layer, three fully connected layers, and a final layer with an output layer employing the ReLU activation function to produce outputs. Dropout layers are also used to avoid overfitting in their model. This model takes images with dimensions $227 \times 227 \times 3$ as input. The resultant feature vector has the dimensions $(13 \times 13 \times 256)$. The number of filters rises as input images move further into the levels.

3.3 Inception-Net: Version 3.0

The InceptionV3 is a pre-trained convolution neural network used for image classification. The InceptionV3 is a modified version of the base models like the InceptionV1, launched by Google in 2014. The InceptionV3 model is a slightly advanced and optimized version of the inception-v1 model. Many techniques were used to optimize and stabilize the model (Shafiq and Gu, 2022), (Zhou et al., 2016). As a result, they created deeper networks than the Inception-v1 and v2 models, though it needed to be more effective, it took less computational power. The Inception-v3 model has forty-two layers, which is slightly higher than the previous base models. Several improvements have been made to Inception v3, such as factorization being fast as smaller convolutions, spatial factorization being converted to asymmetric convolutions, and the efficient grid size being reduced. These were the primary changes made to the InceptionV3 model.

3.4 Visual geometry group networks–VGG-Net: Version 16

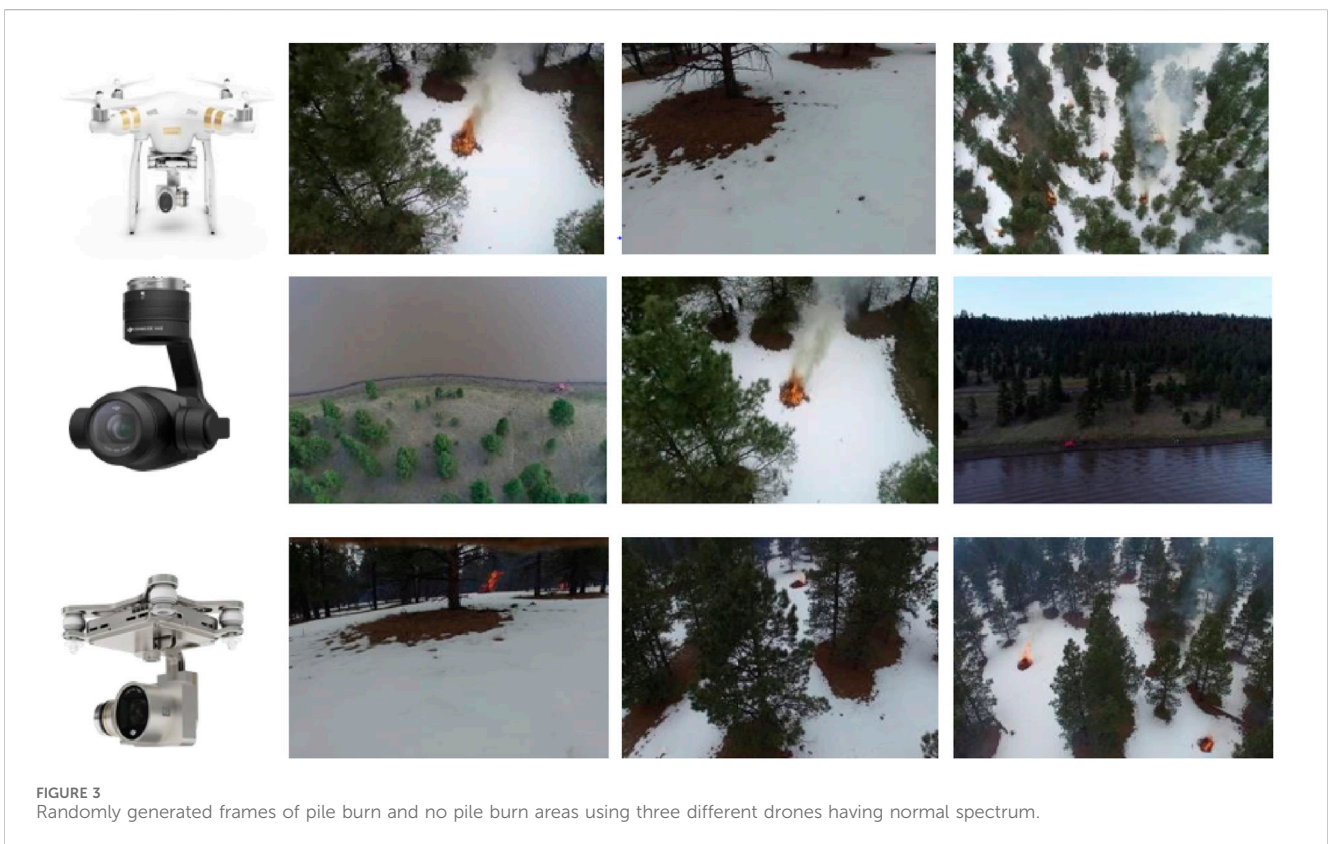
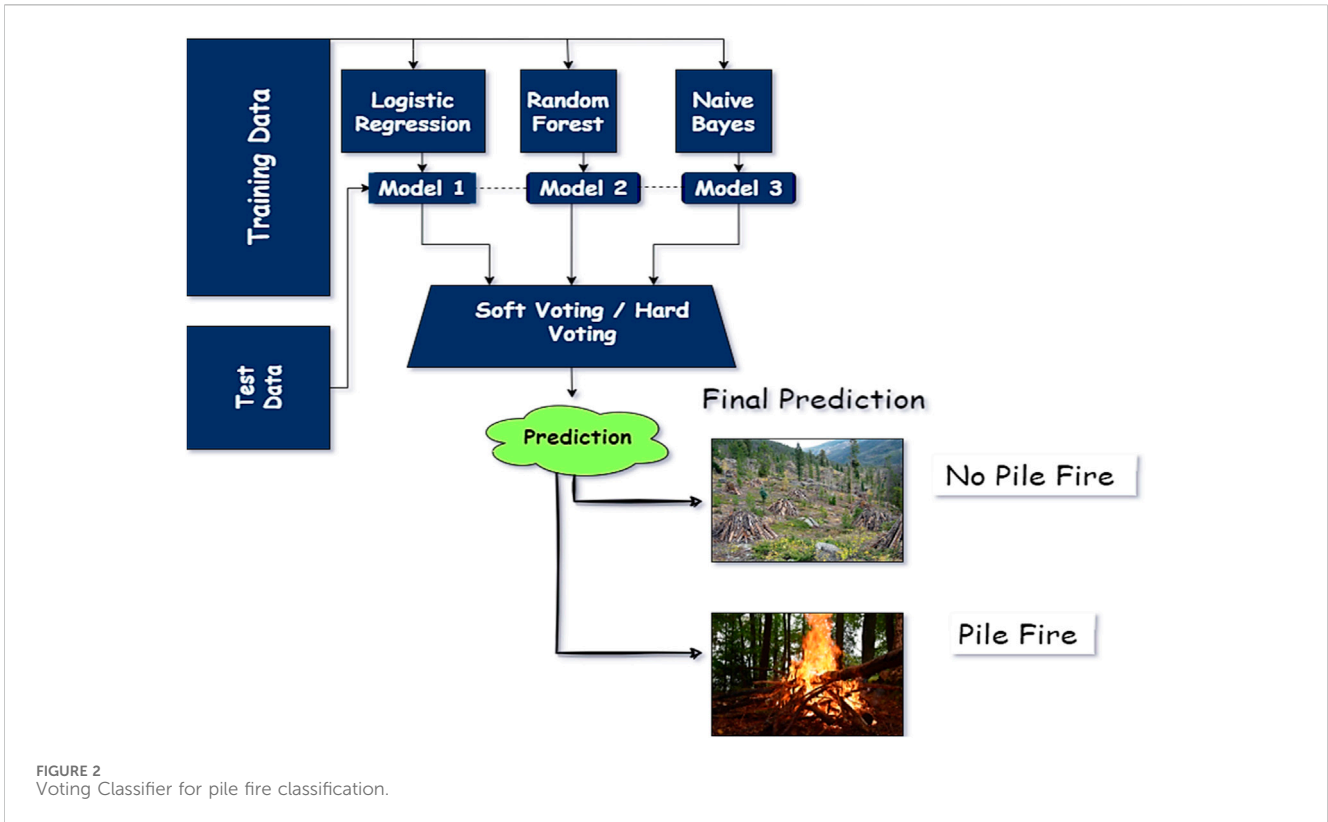
The VGG-16 is also a modified version of the VGG model with sixteen convolution layers. As per the observation of the VGG-16 architecture, it contains sixteen convolutional layers. This model also has the same 3×3 size filters on each layer (Kundu et al., 2018), (Szegedy et al., 2017), and (Wang et al., 2017). This study used the VGG-16 model as a pre-trained model, which was trained on an ImageNet dataset. The parameters can be used as hyper-parameters to improve the accuracy. This network setup accepts a fixed size of 227×227 image pixels with three channels - Red, Green, and Blue. The only image preprocessing performed on the image is normalizing each pixel's RGB values. Then, the image is passed through the ReLU activation function with a stack of two convolution layers having a filter size of 3×3 . These 2 convolution layers include 64 filters, each with stride and padding by 1. This setup is also helpful in preserving spatial resolution. In the last layer, the dimension of the output layer is kept the same as the dimensions of the input image. A window size of 2×2 and a stride of two pixels were applied to the images.

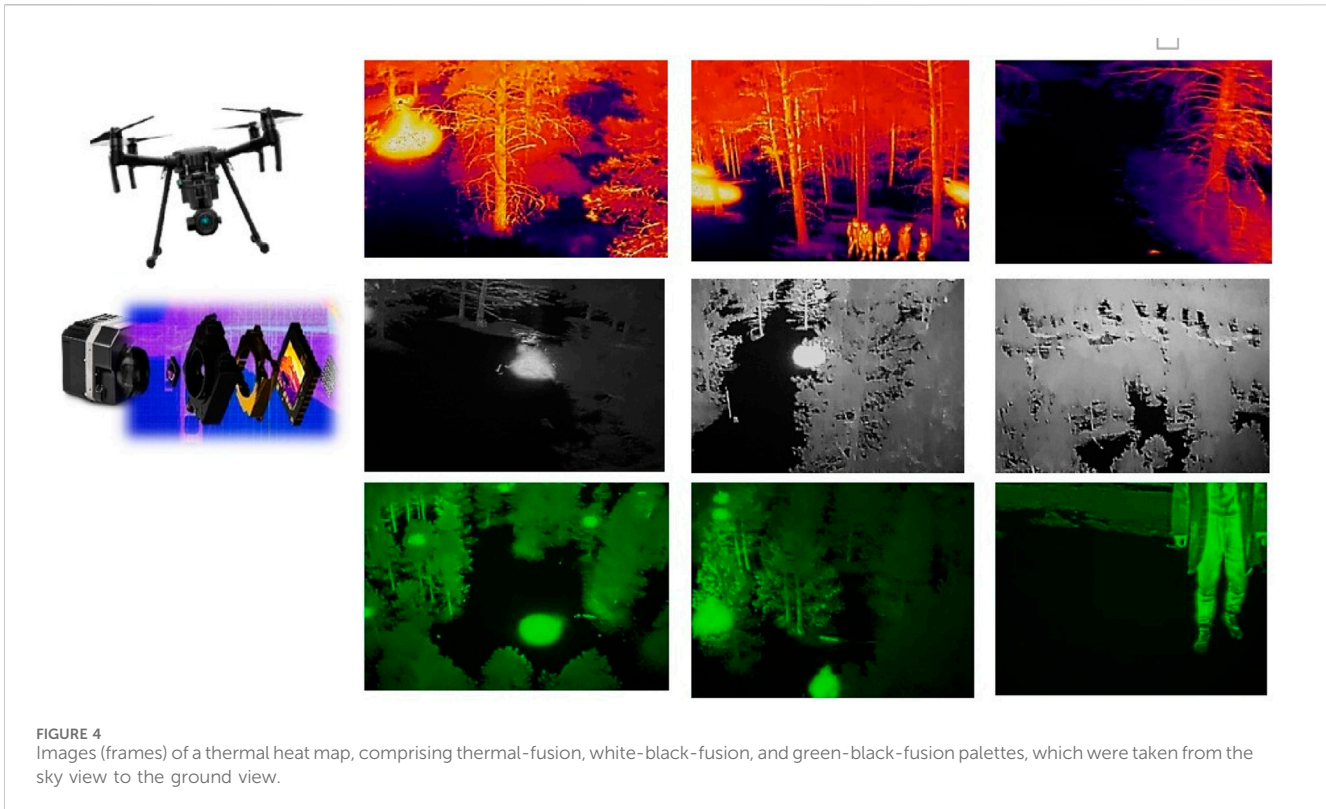
3.5 Ensemble vote classifier–Voting classifier

A voting classifier is an ensemble learning technique. Voting classifiers can aggregate multiple base model outputs and give the best result for a specific problem. The voting aggregation method considers each base model probability and combines them to produce an average probability of a model (Mueller et al., 2013), (Borges and Izquierdo, 2010), (Qureshi et al., 2016). There are two methods of voting criteria. Hard voting gives an integer number as a probabilistic output class, while soft voting means a floating-point-based probabilistic output class. Thus, a voting classifier can be used as a hard or soft vote for optimal results. The proposed model is trained using three basic estimators for pile fire classification: Logistic Regression, RF, and Gaussian naive Bayes as shown in Figure 2. Developers have the flexibility to select between hard and soft vote aggregators using the parameters “voting = ‘soft’” or “voting = ‘hard’”. In the proposed model, a soft voting classifier is utilized with a weight distribution of 1:2:1, where the RF model is assigned twice the weight compared to others.

4 Dataset

This section contains information about the captured images and recorded clips captured by drones. These video clips were played at different frames-per-second speeds. To create an image (frame) dataset from a video clip, it is edited and cut into frames, where each frame may contain a black-out, pile-fired, or no-pile-fired type of background. The FLAME collection, hosted on the IEEE POST website, is freely accessible. Users can download the full high-resolution video clips and corresponding images, which are organized into four categories based on the type of drone camera and palette used. The dataset is





structured to facilitate easy extraction of frames for analysis. The FLAME collection has a total of four different recordings, which were captured using different drone cameras, such as normal-range-spectrum with RGB color-space shown in Figure 3 and thermal-heat-map, white-hot-fusion and green-hot-fusion palettes shown in Figure 4.

4.1 Normal-spectrum palette

The Zenmuse-X4S and the Phantom-3 Camera were used to capture the Normal-Spectrum Palette. Many videos and photos with pile-fire and no-pile-fire imagery were accessible in the FLAME repository. Two clips were recorded using the Zenmuse-X4S Camera. One was a 16-min video with a resolution of $1,280 \times 720$ and a frame rate of 29 frames-per-second (fps). At $1,280 \times 720$ quality and 29 frames per second, a further 6 min of video was available for pile-fire from the start of the burning. The size of the videos was 1.14 GB and 479 MB, respectively. The video captured from the Phantom-3 Camera was of length 17 min with a frame resolution of $3,840 \times 2,160$. Figure 3 shows some images (frames) of pile-fire and no-pile-fire clips captured by the drone. These clips can be found in the suggested flame dataset. All videos were shot using different palettes—Normal-Spectrum, Thermal-Fusion, White-Hot, and Green-Hot—each captured with specialized drone cameras. All videos were shot at 640×512 resolution and 30 frames per second. There were several videos with varying lengths of fire and no-fire types. Table 1 (a) shows the sample set distribution for normal spectrum palette images.

4.2 Thermal-fusion palette

The Thermal-Fusion Palette was captured using a Forward-Looking-Infrared (FLIR) Vue Pro R camera. A solitary video clip was accessible in the FLAME repository, showcasing frames depicting both scenes with pile fires and scenes without pile fires. The video of length 24 min and has 640×512 resolution with 30 frames-per-second (fps). The size of the video clip was 2.83 GB. Table 1 (b) shows the sample set distribution for thermal fusion palette images.

4.3 White-hot palette

The White-Hot palette was recorded using a Forward-Looking-Infrared (FLIR)-Vue-Pro-R-camera. Only one video clip was available on the FLAME repository, which has frames containing both pile-fire and no-pile-fire imagery. The video of length 2 min and has 640×512 resolution with 30 frames-per-second (fps). The size of the video was 43.3 MB. Table 1 (c) shows the sample set distribution for white-hot palette images.

4.4 Green-hot palette

The Green-Hot Palette was recorded using Forward-Looking-Infrared (FLIR)-Vue-Pro-R-camera. Only one video clip was present on the FLAME repository with frames containing both pile-fire and no-pile-fire imagery. The video of length 5 min and has 640×512 resolution with 30 frames-per-second (fps). The size of the video

TABLE 1 Training samples and Testing samples of (a) normal spectrum palette, (b) thermal fusion palette, (c) white-hot palette and (d) green-hot palette.

| (a) Normal spectrum palette | | |
|-----------------------------|-----------------|----------------|
| Class Label | Training Sample | Testing Sample |
| Pile Fire | 25,018 | 6,242 |
| No-Pile Fire | 14,357 | 6,107 |
| Total | 43,375 | 12,353 |
| (b) Thermal Fusion Palette | | |
| Class Label | Training Sample | Testing Sample |
| Pile Fire | 686 | 100 |
| No-Pile Fire | 518 | 118 |
| Total | 1,204 | 218 |
| (c) Whit-hot palette | | |
| Class Label | Training Sample | Testing Sample |
| Pile Fire | 914 | 100 |
| No-Pile Fire | 392 | 114 |
| Total | 1,306 | 214 |
| (d) Green-hot palette | | |
| Class Label | Training Sample | Testing Sample |
| Pile Fire | 722 | 118 |
| No-Pile Fire | 522 | 100 |
| Total | 1,244 | 218 |

clip was 146 MB. Table 1 (d) shows the sample set distribution for green-hot palette images.

5 Image pre-processing and augmentation

5.1 Image pre-processing

Image processing of flame is an important part of fire safety engineering since pixel-based fire detection systems are increasing rapidly. The border extraction method was one of the most critical phases in fire image processing, serving as a precursor and laying the groundwork for subsequent processing. There were several motives because flame margin identification was crucial. For starters, the flame borders provide a framework for measuring many aspects of flame dynamics, such as structure, volume, orientation, and steadiness. Second, establishing flame borders can reduce the file size processed and screen out unwanted data within the frame, such as noise in the background (Celik, 2010), (Habiboğlu et al., 2011). Border identification, in other words, can preserve significant structural characteristics of the flames while shortening time consumption. Finally, border identification can be used to split a

group of flames. This was useful in smothering many flames in boiler tubes with a multi-burner system. Moreover, prompt identification of flame borders can trigger a fire alert and offer firefighters information on the type of fire, combustible substances, the outside of the flame, and so on. For example, the mobility of a monitored flame border can be used to distinguish between actual and spurious fire alarms (Çelik and Demirel, 2009), (Umar et al., 2017), (Binti Zaidi et al., 2015). In this study, the authors have extracted features from an image/frame using pixel value, Gabor, and Sobel filters. Figure 5 depicts a sample image of a thermal fusion palette and flame edge detection phases.

5.2 Image augmentation

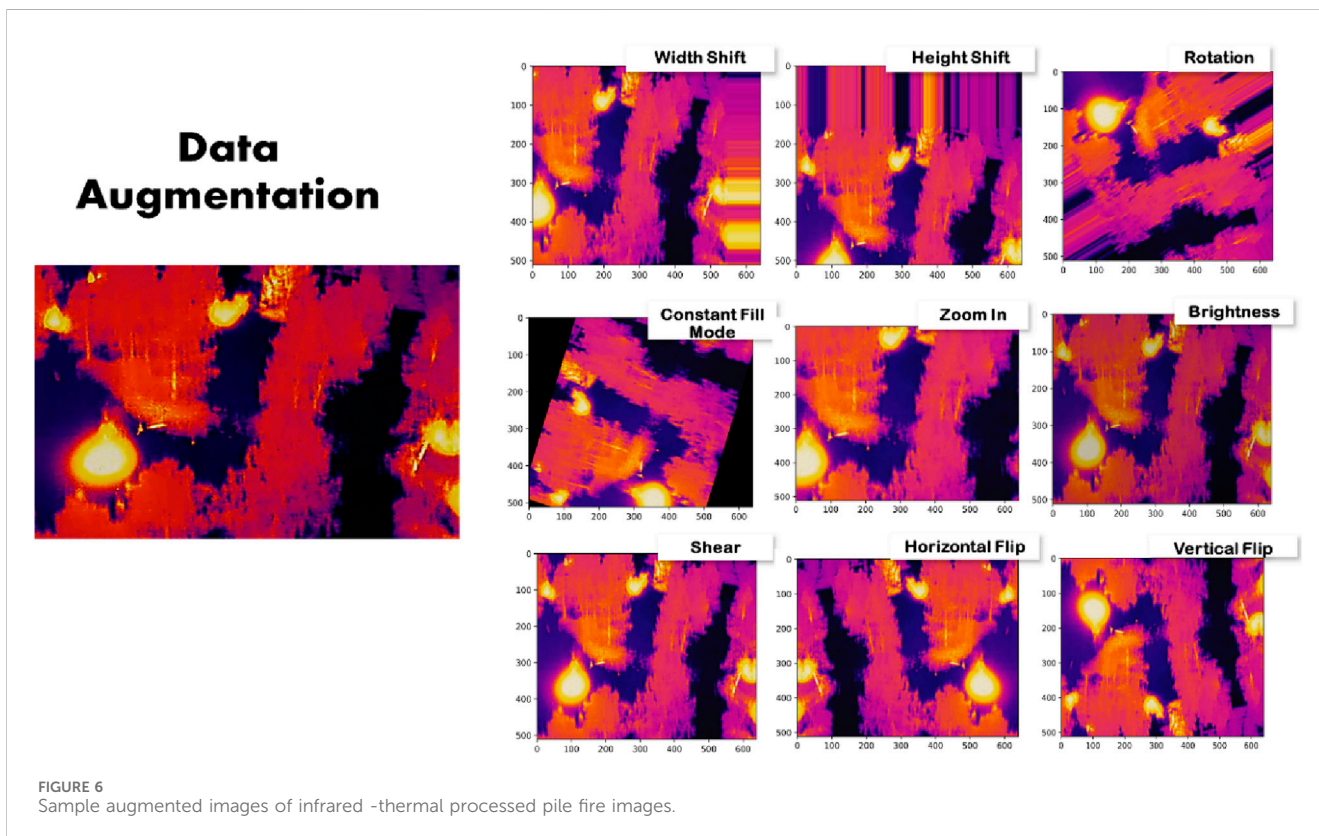
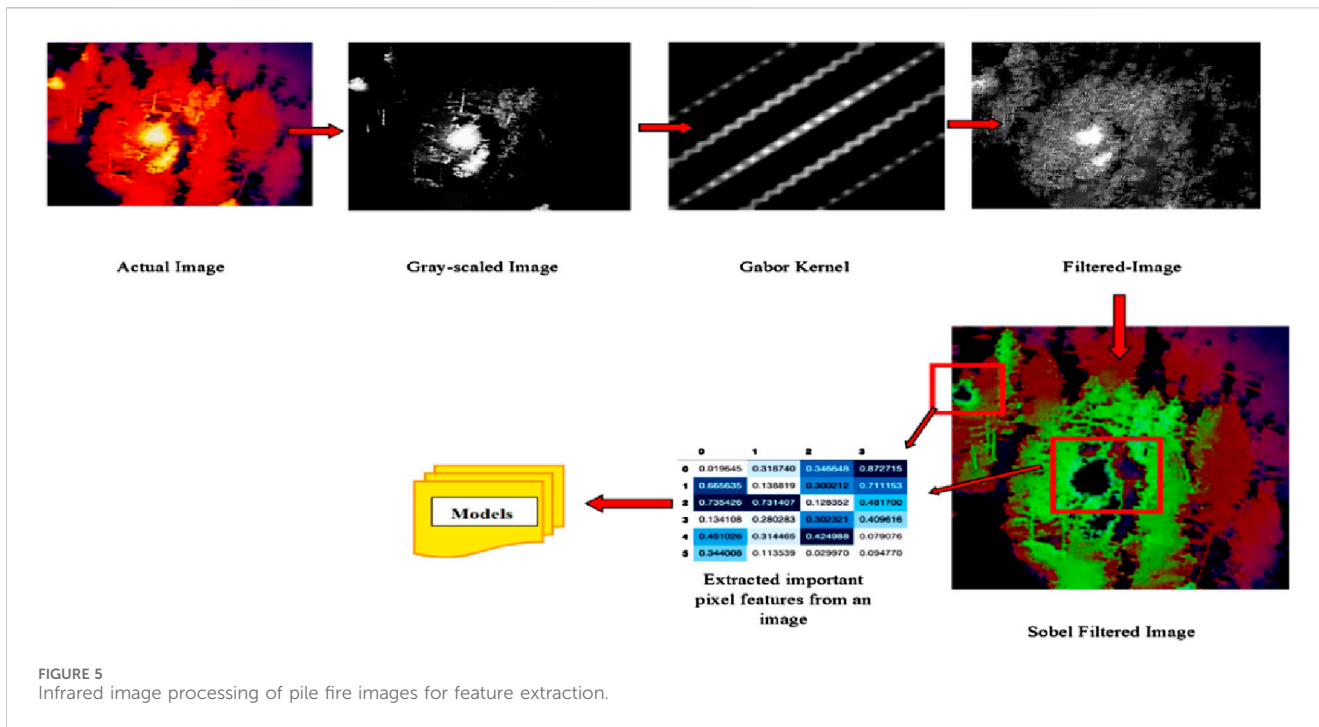
In this study, we utilized image augmentation techniques to diversify the training dataset, thereby enhancing the model's exposure to a wider range of image variations. By altering factors such as orientation, brightness, and size randomly, the model is compelled to recognize image subjects across different contexts. Unlike image preprocessing, which is applied to both training and test sets, image augmentation is exclusively implemented on the training data. Various methods such as translation, rotation, lighting adjustments, and color manipulation can be employed for data augmentation, which can be categorized into offline and online stages. Offline augmentation involves expanding the training dataset before training begins, while online augmentation increases the diversity of images seen during training. In this research, offline augmentation was employed to create a dataset for plant seedling categorization. This involved resizing, cropping, and horizontally flipping training photos into three sizes. Additionally, data augmentation was illustrated in Figure 6, depicting augmentation and transformation on thermal fusion frames across different spectrums. Such augmentation is vital for enhancing the system's robustness and accuracy in estimating thermal heat within a pile.

6 Experimental results and analysis

6.1 Traditional machine learning techniques and its comparative analysis

6.1.1 Thermal-fusion palette

The total number of frames in the training portion was 1,204, out of which 686 were of class "pile-fire" and 518 frames of class "pile-no-fire." All the frames were scrambled before dumping into the model. The dataset contained many raw frames, blackout frames, and so on. As a result, image processing was required to extract the features from the frame. A total of three filters were applied to each frame to extract features from it. So, the first feature extracted was the pixel value. Each image has a pixel value representing how bright the image is in a specific region of the image or how much intensity of the color is present in that region. The image is distributed into three channels: red, blue, and green. Thus, each channel is represented by 8 bits, and a total of three channels have range from 0 to 255). The second feature extractor was the Gabor filter; now, a total of four Gabor filters were applied on each frame with different parameters. The parameters were theta value, which ranges



from 0.0 to 2.0; sigma value, which ranges between 1 and 3; lambda value of 0.78; and gamma value of 0.5. The third feature extractor was the Sobel filter or the Sobel edge detector, a gradient-based method that looks for significant changes in an image’s first derivative. To evaluate the accuracy of the test dataset,

218 frames, including 100 pile-fire-labeled frames and 118 pile-no-fire-labeled frames, were fed into traditional machine learning algorithms such as RF and SVM. Table 2 shows the training parameters used for both algorithms, and Table 3 shows the classification report, volume of the model, inference CPU time,

TABLE 2 Training parameters of (A) RF classifier and (B) SVM classifier.

| (A) | | (B) | |
|-------------------|--------|-------------------------|--------|
| Parameter | Method | Parameter | Method |
| N_estimators | 100 | C | 2 |
| Criterion | Gini | Kernel | 'rbf' |
| Max_depth | 12 | decision_function_shape | 'ovo' |
| Min_samples_split | 2 | gamma | 'auto' |
| Random-state | 42 | max_iter | 200 |

and flops. Figures 7A, B show the confusion matrix and ROC-AUC curve for RF algorithm applied on thermal heat map frames to detect fire piles.

6.1.2 White-hot palette

The total number of frames in the training portion was 1,306 frames, out of which 914 frames were of class “pile-fire” and 392 frames of class “pile-no-fire.” In image processing for this frame, doing the same thing here by applying three filters on each frame to extract features from the frame. So, the first feature extracted was the pixel value. Each image has a pixel value representing how bright the image is in a specific region of the

image or how much intensity of the color is present in that region. The image was distributed into one channel only: Black and White. Thus, each channel is represented by 8 bits, and a total of one channel has 8 bits (ranging from 0 (darker) to 255 (brighter)). The second feature extractor was the Gabor filter; now, four Gabor filters were applied on each frame with different parameters. The parameters were theta value, which ranges from 0.0 to 1.0; sigma value, which ranges between 1 and 4; lambda value of 0.85; and gamma value of 0.7. The third feature extractor was the Sobel filter or the Sobel edge detector, a gradient-based method that looks for significant changes in an image’s first derivative. 100 Pile-Fire-labeled frames and 114 Pile-No-fire-labeled frames were used in traditional machine learning methods, such as RF and SVM, to evaluate the accuracy of the test dataset. Table 4 shows the report of training accuracy on the training set and testing accuracy on the testing set. Figures 8A, B show the confusion matrix and ROC-AUC curve for RF algorithm applied on white hot palette frames to detect fire piles.

6.1.3 Green-hot palette

The total number of frames in the training portion was 1,244, of which 722 were of class “pile-fire” and 522 frames of class “pile-no-fire.” In image processing for this frame, doing the same thing here by applying three filters on each frame to extract features from the frame. So, the first feature extracted was the pixel value. Each image has a pixel value representing how bright the image is in a specific

TABLE 3 Accuracy for evaluation of the pile fire for thermal fusion classification (224 × 224).

| Model name | Size of model (MB) | Dataset type | Training accuracy | Testing accuracy | Precision | Recall | F1-score | AUC |
|------------------|--------------------|--------------------------|-------------------|------------------|-----------|--------|----------|-------|
| RF | 49.67 | Thermal- Fusion | 0.997 | 0.995 | 0.99 | 0.99 | 0.99 | 0.987 |
| | | Palettes | | | | | | |
| SVM (RBF Kernel) | 200 | Thermal- Fusion Palettes | 0.994 | 0.981 | 0.98 | 0.97 | 0.98 | 0.995 |

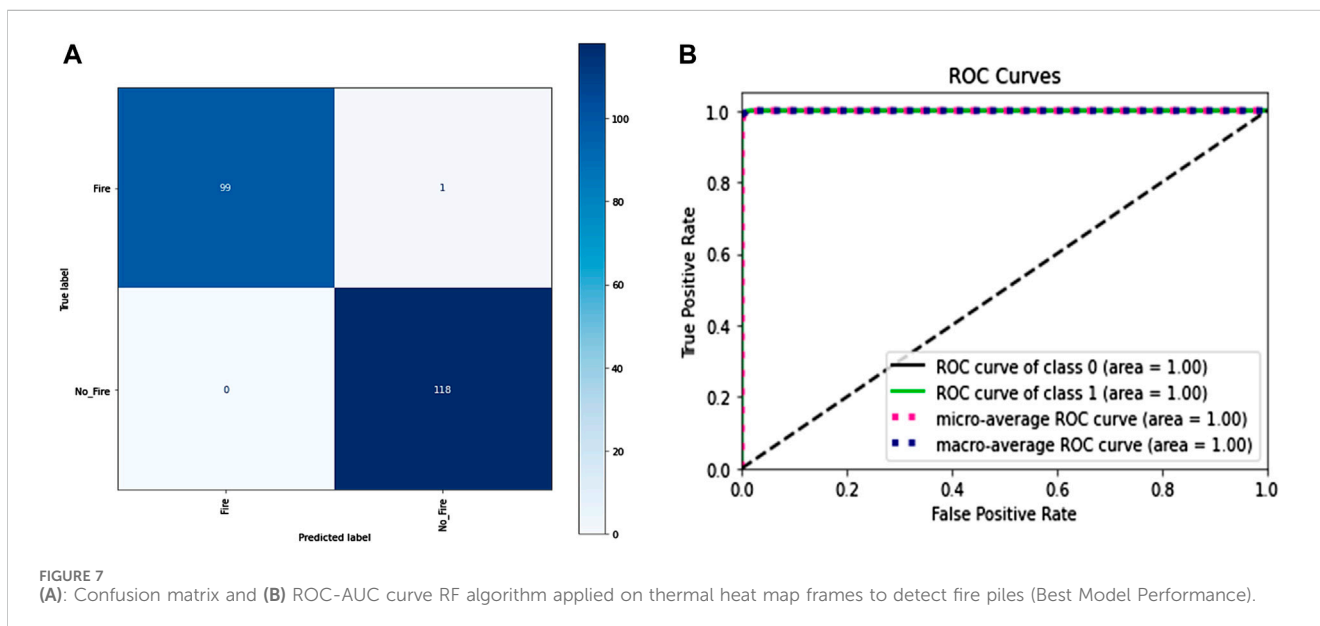
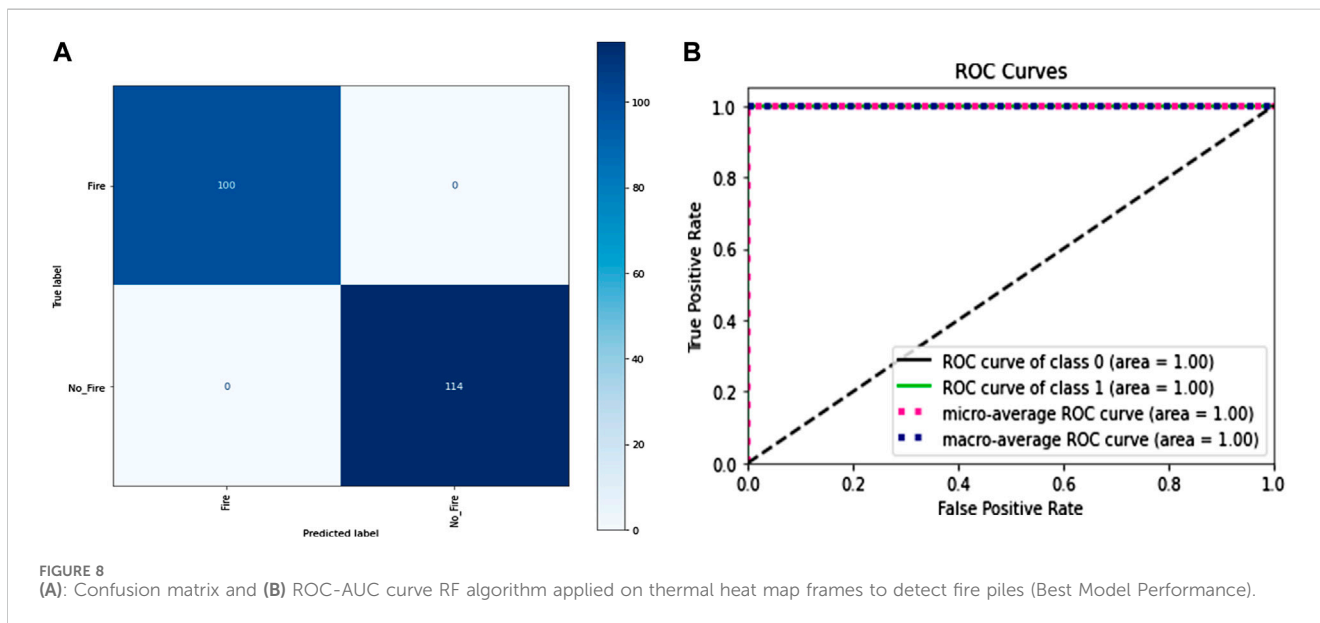


FIGURE 7 (A): Confusion matrix and (B) ROC-AUC curve RF algorithm applied on thermal heat map frames to detect fire piles (Best Model Performance).

TABLE 4 Accuracy for evaluation of the pile fire for white-hot palette classification (224 × 224).

| Model name | Size of model (MB) | Dataset type | Training accuracy | Testing accuracy | Precision | Recall | F1-score | AUC |
|------------------|--------------------|--------------------|-------------------|------------------|-----------|--------|----------|-------|
| RF | 49.67 | White-Hot Palettes | 0.998 | 0.996 | 0.99 | 0.99 | 0.99 | 0.996 |
| SVM – RBF Kernel | 200 | White-Hot Palettes | 0.99 | 0.983 | 0.99 | 0.98 | 0.985 | 0.997 |



region of the image or how much intensity of the color is present in that region. The image was distributed into one channel only: Black and Green. But here, the green channel often contains larger bits. This is due to our eyes’ increased sensitivity to green color and decreased sensitivity to blue or other colors (ranging from 0 (darker) to 255 (brighter)). The second feature extractor was the Gabor filter. A Gabor filter with different parameters was applied to each frame. The parameters were theta value, which ranges from 0.0 to 2.0; sigma value, which ranges between one and 2; lambda value of 0.78; and gamma value of 0.65. The third feature extractor was the Sobel filter or edge detector, a gradient-based method that looked for significant changes in an image’s first derivative. To evaluate the accuracy of the test dataset, 218 frames, including 118 Pile-Fire-labelled frames and 100 Pile-No-fire-labelled frames, were fed into traditional machine learning algorithms such as RF and SVM. Table 5 shows the report of training accuracy on the training set and testing accuracy on the testing set. Figures 9A, B show the confusion matrix and ROC-AUC curve for RF algorithm applied on white hot palette frames to detect fire piles.

6.2 Knowledge transferring technique: Transfer learning (TL)

There are ample advantages to using the RF algorithm. Still, the main drawback of using RF is that it has more decision trees, making

the algorithm computationally expensive and complex. It is also not preferable for live streams or real-time analysis. Furthermore, manual feature extraction is difficult to eliminate because it requires business and domain logic to build a robust model that can replicate and capture trends and patterns from data (Çelik et al., 2007), (Khalil et al., 2021), (Chino et al., 2015), (Verstockt et al., 2013). Transfer learning was, therefore, increasingly employed to save cost and time by removing the need to train several machine learning models from the start to execute identical tasks. This paper uses a transfer learning approach to reduce computational costs. The initial layers, except the last dense layer, were frozen, and pre-trained models with trainable parameters contributing to the last layer were used, which reduced the time required for the training and compared to training the entire model from scratch. Pre-trained models were used to compensate for the lack of labeled training data held by an organization. Table 6 shows the training parameters used to implement transfer learning approaches in this study.

6.2.1 Thermal-fusion palettes

The total number of frames in the training portion was 1,204, of which 686 were of class “pile-fire” and 518 frames of class “pile-no-fire.” All the frames were scrambled before being sent to the network. Furthermore, data augmentation was performed using different techniques before feeding the training samples into the model. The initial learning rate was set to 0.0001, and the Adamax optimizer was used to regulate weight and minimize the loss

TABLE 5 Accuracy for evaluation of the pile fire for green hot palettes classification (224 × 22).

| Model name | Size of model (MB) | Dataset type | Training accuracy | Testing accuracy | Precision | Recall | F1-score | AUC |
|------------------|--------------------|--------------------|-------------------|------------------|-----------|--------|----------|-------|
| RF | 49.67 | Green-Hot Palettes | 0.998 | 0.958 | 0.96 | 0.96 | 0.96 | 0.997 |
| SVM – RBF Kernel | 200 | Green-Hot Palettes | 0.994 | 0.912 | 0.91 | 0.92 | 0.91 | 0.992 |

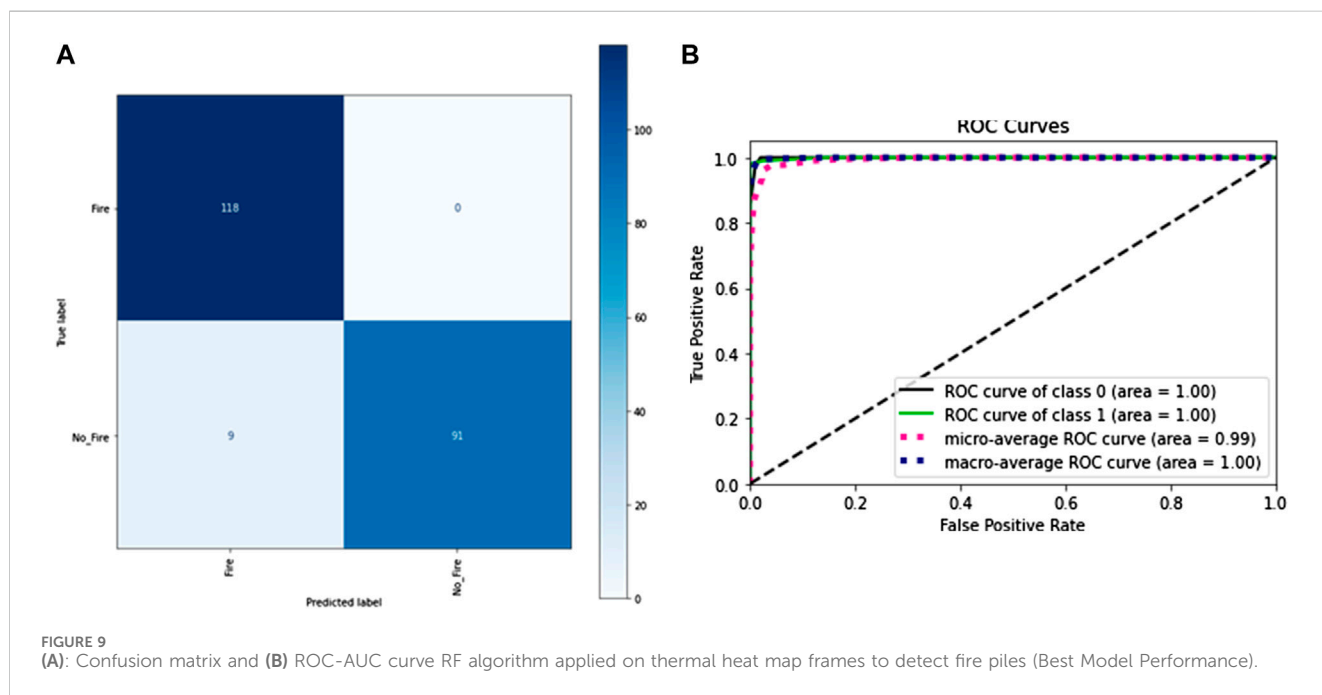


TABLE 6 Training parameters of transfer learning neural networks.

| Parameters | Method |
|---------------------|-------------------------------------|
| Epoch | 5–50 |
| Iteration per epoch | 25 |
| Steps per epoch | 15 |
| Weight decay | 0.00004 |
| Learning rate | 0.001 > 0.0005 > 0.00025 > 0.000125 |
| Learning rate decay | 0.00001 (Iteration equal step size) |
| Loss Function | binary_crossentropy |
| Activation Function | Sigmoid |
| Optimizer | Adamax |
| Batch size | 50 |

function. In addition, a batch size of 50 was employed while training the model. To test the performance and error loss of the testing samples, 218 frames were fed into the pre-trained networks, in which 100 frames were labeled as “pile-fire” and 118 frames as “pile-no-fire.” Five models were fine-tuned and

trained on the Thermal-Fusion Palette dataset. Among them, Resnet50V2 performs better on test and train datasets of the thermal-fusion palette. Table 7 shows the error-loss and accuracies of multiple models. The classification “pile-fire versus no-pile-fire” obtained 99.08% testing accuracy. Figures 10A, B depict training versus validation loss and accuracy curves for Resnet50V2 model.

6.2.2 White-hot palettes

The total number of frames in the training portion was 1,306 frames, out of which 914 frames were of class “pile-fire” and 392 frames of class “pile-no-fire.” To test the performance and loss of the testing samples, 214 frames were fed into the pre-trained networks, in which 100 frames were labeled as “pile-fire” and 114 frames as “pile-no-fire.” Five models were fine-tuned and trained on the White-Hot Palette dataset. Among them, InceptionV3 performs well on test and train datasets of the white-hot palette. Table 8 shows the training and testing accuracy, model weight, and model inference time on CPU taken while training the model with Bflops and AUC score. In inceptionv3, the researcher got a testing accuracy of around 99.86% for pile fire classification. Figures 11A, B depict training versus validation loss and accuracy curves for the InceptionV3 model.

TABLE 7 Transfer learning enhances the performance of pile-fire classification for thermal fusion. (227 × 227).

| Model name | Size of model (MB) | Inference time (ms) | Training accuracy | Testing accuracy | Precision | Recall | F1-score | AUC | BFlops |
|--------------------|--------------------|---------------------|-------------------|------------------|-----------|--------|----------|--------|--------|
| AlexNet | 233 | 539 | 0.8512 | 0.8303 | 0.9175 | 0.7542 | 0.8278 | 0.8086 | 0.0122 |
| Inception ResnetV2 | 92 | 42.2 | 0.9538 | 0.9700 | 0.9478 | 1.00 | 0.9732 | 0.9727 | 0.1322 |
| InceptionV3 | 215 | 130.2 | 0.9728 | 0.9817 | 0.9914 | 0.9746 | 0.9829 | 0.9991 | 0.0567 |
| VGG16 | 528 | 69.5 | 0.8945 | 0.9037 | 0.9145 | 0.9068 | 0.9106 | 0.9350 | 0.3069 |
| Resnet50V2 | 98 | 45.6 | 0.9845 | 0.9908 | 0.9915 | 0.9915 | 0.9915 | 0.9946 | 0.0696 |

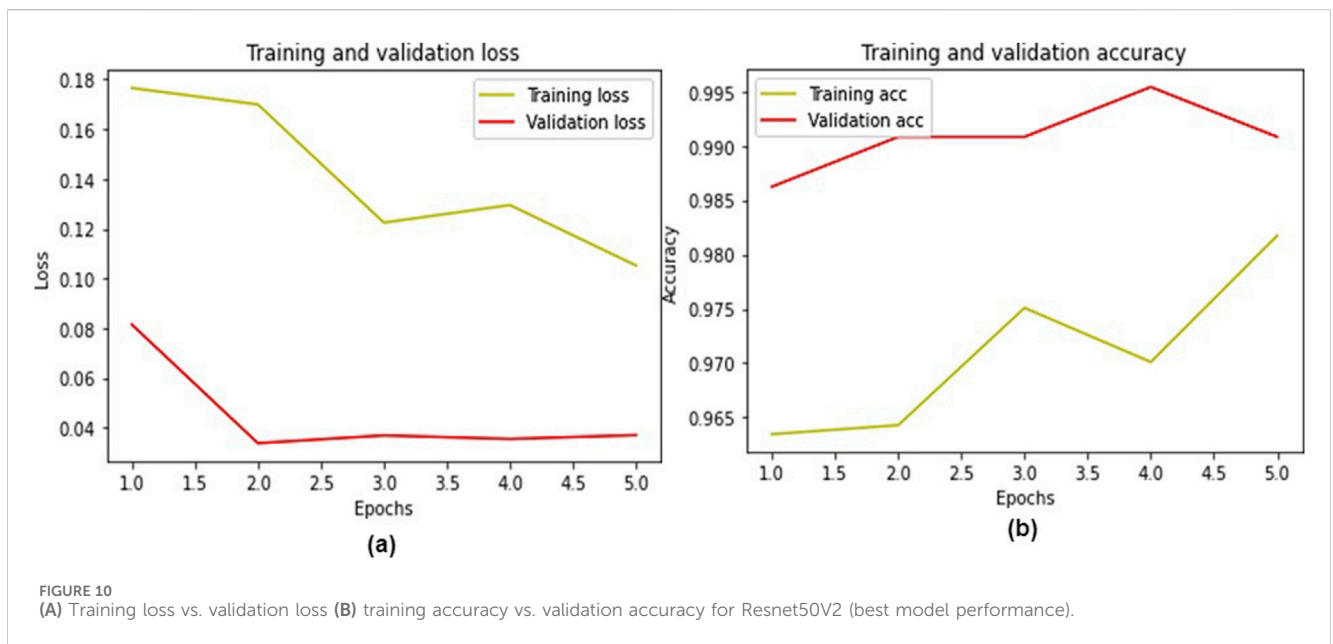


FIGURE 10 (A) Training loss vs. validation loss (B) training accuracy vs. validation accuracy for Resnet50V2 (best model performance).

TABLE 8 Transfer learning accuracy for evaluating pile fire classification for white hot palette (227 × 227).

| Model name | Size of model (MB) | Inference time (ms) | Training accuracy | Testing accuracy | Precision | Recall | F1-score | Auc-curve | BFlops |
|--------------------|--------------------|---------------------|-------------------|------------------|-----------|--------|----------|-----------|--------|
| AlexNet | 233 | 539 | 0.9478 | 0.9019 | 1.00 | 0.8158 | 0.8985 | 0.9578 | 0.0122 |
| Inception ResnetV2 | 92 | 42.2 | 0.9850 | 0.9950 | 1.00 | 0.9903 | 0.9945 | 1.00 | 0.1322 |
| InceptionV3 | 215 | 130.2 | 0.9990 | 0.9986 | 1.00 | 1.00 | 1.00 | 1.00 | 0.0567 |

6.2.3 Green-hot palettes

The total number of frames in the training portion was 1,244, of which 722 were of class “pile-fire” and 522 frames of class “pile-no-fire.” To test the performance and loss of the testing samples, 218 frames were fed into the pre-trained networks, in which 118 frames were labeled as “pile-fire” and 100 frames as “pile-no-fire.” Five models were fine-tuned and trained on the Green-Hot Palette dataset. Among them, InceptionV3 performed better on a test set of the green-hot palettes. Table 9 shows the accuracy, inference CPU time, and flops while training and testing the model. In the classification of “pile-fire versus no-pile-fire,” the researcher obtained a testing accuracy of 99.89%. Figures 12A, B depict training versus validation loss and accuracy curves for the InceptionV3 model.

6.2.4 Normal-spectrum palettes

The total number of frames in the training portion was 43,375, of which 25,018 were of class “pile-fire” and 14,357 were class “pile-no-fire.” To test the performance and loss of the testing samples, 12,343 frames were fed into the pre-trained networks, comprising 6,242 frames labeled as “pile-fire” and other 6,107 frames labeled as “pile-no-fire.”

Thus, three pre-trained models were fine-tuned and trained on a normal-spectrum palette dataset. Among them, VGG16 performs well on test and train datasets of a normal-spectrum palette. The frames (data) collected from the Matrice 200 drone and Zenmuse X4S camera were used to train the model, while the frames collected from the Phantom drone

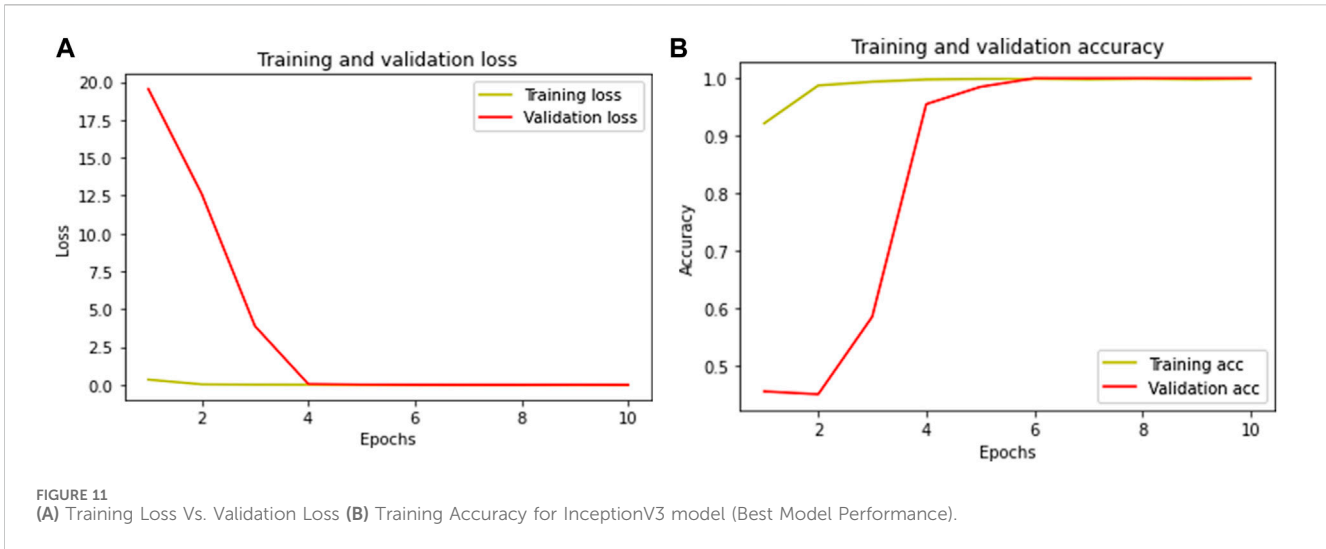
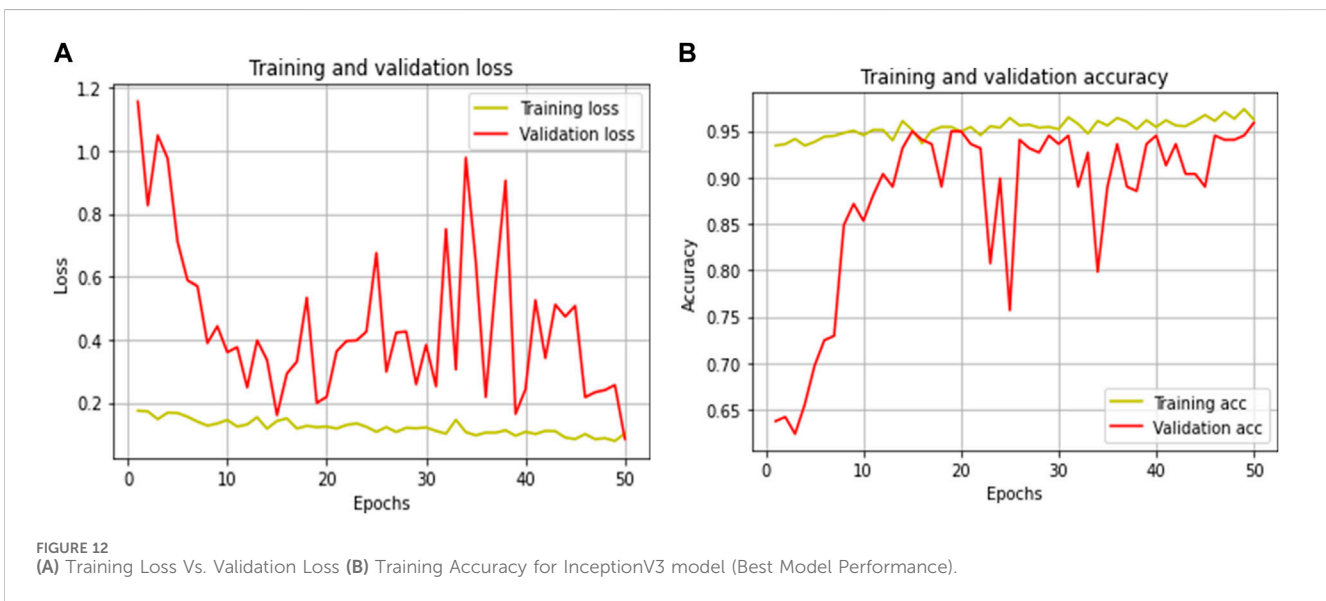


TABLE 9 Transfer learning accuracy for evaluating pile fire classification for a green hot palette (227 × 227).

| Model name | Size of model (MB) | Inference time (ms) | Training accuracy | Testing accuracy | Precision | Recall | F1-score | Auc-curve | BFlops |
|--------------------|--------------------|---------------------|-------------------|------------------|-----------|--------|----------|-----------|--------|
| AlexNet | 233 | 539 | 0.9622 | 0.9587 | 0.9473 | 0.9636 | 0.9553 | 0.9933 | 0.0122 |
| Inception ResnetV2 | 92 | 42.2 | 0.9688 | 0.9400 | 0.9645 | 0.9622 | 0.9633 | 0.9826 | 0.1322 |
| InceptionV3 | 215 | 130.2 | 0.9930 | 0.9989 | 1.00 | 1.00 | 1.00 | 1.00 | 0.0567 |
| VGG16 | 528 | 69.5 | 0.9487 | 0.9850 | 0.9787 | 0.9891 | 0.9838 | 0.9983 | 0.3069 |
| Resnet50 V2 | 98 | 45.6 | 0.9396 | 0.9550 | 0.900 | 0.900 | 0.9473 | 1.00 | 0.0696 |

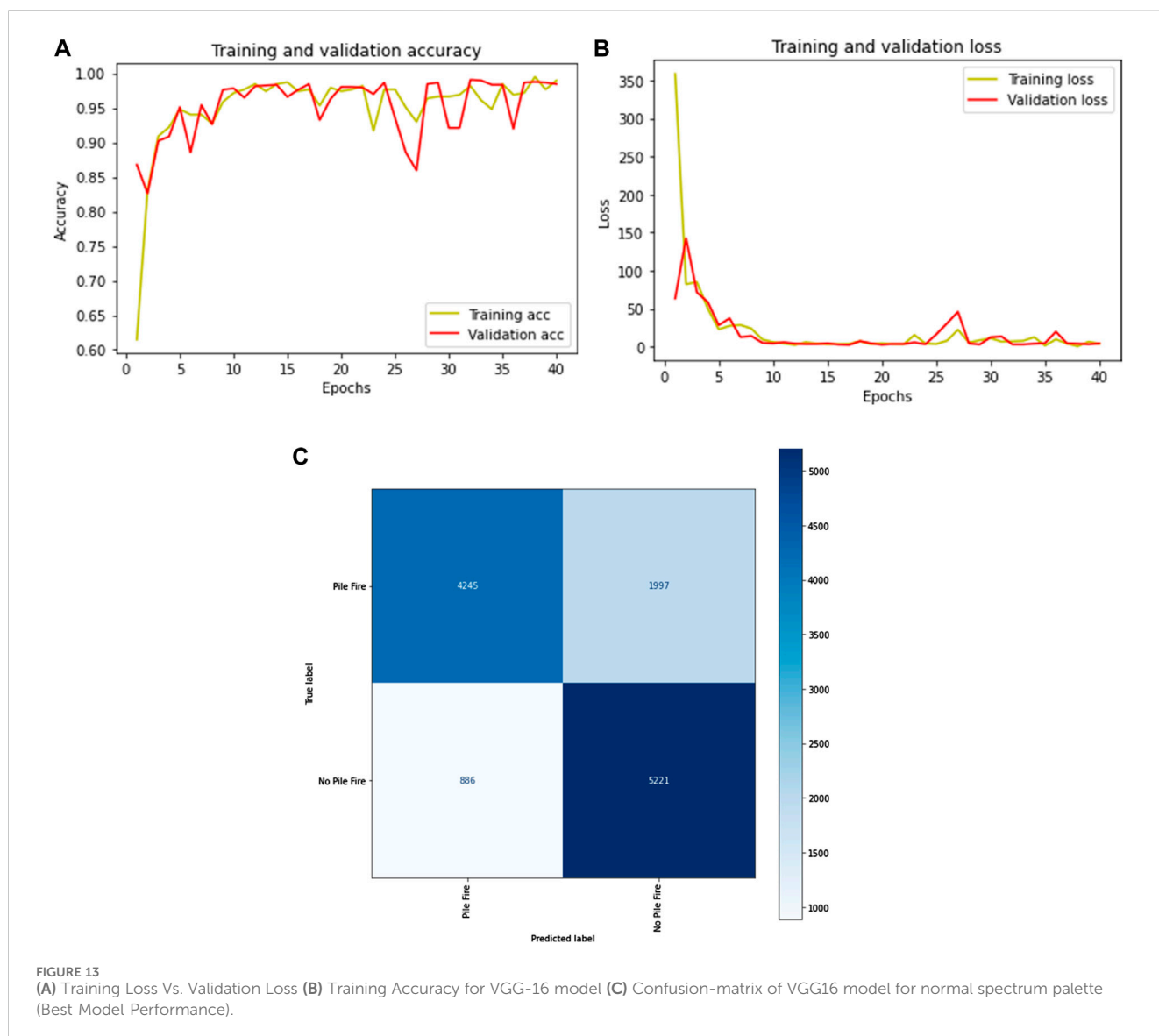


were used to test the model. As a result, there is no overlap between the training and validation samples. Table 10 shows the accuracy, inference CPU time, and flops while training and testing the model. The “pile-fire vs. no-pile-fire” classification attained a

testing accuracy of 77%. Figures 13A, B depict the error loss and accuracy curve graphs generated during training and testing for VGG-16. Figure 13C shows the confusion matrix of the VGG-16 model for detecting fire piles.

TABLE 10 Transfer learning accuracy for evaluating pile fire classification for normal spectrum palette (227 × 227).

| Model name | Size of model (MB) | Inference time (ms) | Training accuracy | Testing accuracy | Precision | Recall | F1-score | AUC | BFlops |
|--------------|--------------------|---------------------|-------------------|------------------|-----------|--------|----------|--------|--------|
| Inception V3 | 215 | 130.2 | 0.9792 | 0.7525 | 0.8117 | 0.5040 | 0.6218 | 0.7152 | 0.0567 |
| VGG16 | 528 | 69.5 | 0.9364 | 0.7700 | 0.7186 | 0.7075 | 0.7130 | 0.7641 | 0.3069 |
| Resnet50 V2 | 98 | 45.6 | 0.9312 | 0.7276 | 0.7811 | 0.4523 | 1.00 | 0.7320 | 0.0696 |



6.2.5 Optimization of normal–Spectrum model: Transfer learning–Using pre-trained knowledge of VGG- 16 for feature extraction and amalgamation with ensemble learning algorithms

VGG-16 has a deep convolutional layered architecture and uses a smaller kernel size, which reduces the trainable and non-trainable parameters. The input layer for this model was ninety single-dimensional feature vectors. In this study, the authors used a VGG-16 with a single-dimensional kernel space in every

convolution layer, followed by a max-pooling layer to extract boundary points of flame features from them. The ninety input vectors as a one-batch and the single-dimensional VGG-16 can grab boundary pixel values as feature vectors with the same weights in each network. Table 11 shows the performance accuracies obtained by amalgamating VGG-16 (feature-extractor) with ensemble learning models. In this study, authors have used many ensemble learning models for the amalgamation process with VGG16 and XGBoost because they selected the optimum tree model. Proposed

TABLE 11 Table of accuracy for evaluation of the pile fire classification of normal spectrum frame using VGG-16 (feature-extractor) + ensemble.

| Model name | Size of model (MB) | Training accuracy | Testing accuracy |
|---|--------------------|-----------------------|-----------------------|
| (Feature-Extractor) VGG16 + XgBoost | 528 MB + 490 MB | 0.9992 | 0.75 |
| (Feature-Extractor) VGG16 + Random Forest Classifier | 528 MB + 49.67 MB | 0.9997 | 0.82 |
| (Feature-Extractor) VGG16 + AdaBoost (Base Estimator – Logistic Regression) | 528 MB + 333 MB | 0.9943 | 0.77 |
| (Feature-Extractor) VGG16 + Gradient Boost | 528 MB + 85.8 MB | 0.9999 | 0.77 |
| (Feature-Extractor) VGG16 + Voting Classifiers (Logistic Regression, RF, Naïve Bayes) | 600+ MB | 0.90 (Hard)/(Soft) | 0.91 (Hard)/(Soft) |

classifiers used in voting were Logistic Regression, RF, and Naïve Bayes. These types of classification algorithms can be beneficial for balancing the unique shortcomings of a group of equally highly performing models.

6.3 Discussion

This study aimed to assess the effectiveness of fine-tuning five pre-trained convolutional architectures for classifying frames into two classes: “Pile Fire” and “No-Pile-Fire”. The primary contribution of this work lies in the systematic evaluation and identification of the most suitable deep transfer learning models for each type of thermal imagery, leading to highly accurate fire detection across diverse environmental conditions. Following are the key contributions.

- **Model-specific optimization:** The study does a deep analysis of the performance of different CNN models applied to various types of thermal imagery. Of the rest, ResNet50V2 reached 99.08% on thermal fusion frames, AlexNet performed at 95.87% on green-hot frames, InceptionV3 reached 99.87% on white-hot frames, while VGG16 did best with 77.37% on the normal spectrum frame type. These results emphasize the importance of appropriate architecture selection based on the characteristics of the data.
- **VGG16 as Feature Extractor:** One of the novelties of this study is to use VGG16 with one-dimensional kernel space in each convolution layer, which could obtain boundary pixel values of flame features effectively. This work shows that the ongoing approach blended with max-pooling can be more informative for fine-grained features required for accurate fire detection.
- **Ensemble Learning Integration:** It also tends to investigate the combination of VGG16 with ensemble learning algorithms such as XGBoost, Logistic Regression, Random Forest, and Naïve Bayes. Such combinations tend to improve predictive performance by cancelling out specific drawbacks of single classifiers and thereby yielding a fire-detection system that is robust in performance and based on the strengths of various algorithms.

7 Conclusion

This study underscores the importance of model selection and optimization in developing reliable fire detection systems, providing

a foundation for future research in enhancing disaster management and environmental monitoring technologies. The FLAME dataset includes thermal imagery processed by various filters to extract information from an infrared image. In this study, two different approaches were used for fire classifications using different image modalities. In the first task, a standard machine learning approach, such as binary pile fire classification, was used to label data and grasp data insights. In the second method, researchers presented a transfer learning strategy to execute the sharing of knowledge concepts, which can result in a far more precise and efficient model. In this study, for thermal fusion frames, ResNet50V2 exhibited the highest accuracy of 99.08%; for green hot frames, InceptionV3 achieved the highest accuracy of more than 99%; for white-hot frames, InceptionV3 exhibited the highest accuracy of 99.87%; and lastly, for normal spectrum frames, VGG16 achieved the highest accuracy of 77.37%. Finally, to improve the performance of the normal-spectrum vision model, we combined transfer learning with ensemble learning to produce the most efficient model with an accuracy of 91%. The scientific community can enhance the given findings by building more advanced systems. Another potential application for this collection is the development of fire detection and segmentation algorithms, such as semantic segmentation or instance segmentation, based on a collective study of several imaging modalities, including conventional and thermal images provided by the researcher. Instance segmentation can be used to segregate pile burns, thermal heat, and forest, animals, and human beings so that we can save them before any disaster occurs in that location. Researchers may also use fire segmentation approaches to identify associated networking and monitoring challenges, such as optimum job scheduling for a fleet of drones to cover pile burns in a certain location in the least amount of time.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://iee-dataport.org/open-access/flame-dataset-aerial-imagery-pile-burn-detection-using-drones-uavs>.

Author contributions

DJ: Writing–original draft, Data curation, Methodology, Software, Validation. SaK: Conceptualization, Supervision,

Writing–review and editing. SP: Conceptualization, Supervision, Writing–review and editing. PK: Conceptualization, Supervision, Writing–review and editing. SHK: Supervision, Writing–review and editing. KK: Conceptualization, Supervision, Writing–review and editing, Funding acquisition.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Research Support Fund (RSF) of Symbiosis International (Deemed University), Pune, India.

References

- Abdusalomov, A. B., Islam, B. M. S., Nasimov, R., Mukhiddinov, M., and Whangbo, T. K. (2023). An improved forest fire detection method based on the Detectron2 model and a deep learning approach. *Sensors* 23 (3), 1512. doi:10.3390/s23031512
- Afghah, F., Razi, A., Chakareski, J., and Ashdown, J. (2019). “Wildfire monitoring in remote areas using autonomous unmanned aerial vehicles,” in *Infocom 2019 - IEEE conference on computer communications workshops, INFOCOM WKSHPs 2019*.
- Afghah, F., Shamsoshoara, A., Njilla, L., and Kamhoua, C. (2018a). “A reputation-based stackelberg game model to enhance secrecy rate in spectrum leasing to selfish IoT devices,” in *Infocom 2018 - IEEE conference on computer communications workshops*.
- Afghah, F., Zaeri-Amirani, M., Razi, A., Chakareski, J., and Bentley, E. (2018b). “A coalition formation approach to coordinated task allocation in heterogeneous UAV networks,” in *Proceedings of the American control conference*.
- Aggarwal, R., Soderlund, A., Kumar, M., and Grymin, D. (2020). “Risk aware SUAS path planning in an unstructured wildfire environment,” in *Proceedings of the American control conference*.
- Binti Zaidi, N. I., Binti Lokman, N. A. A., Bin Daud, M. R., Achmad, H., and Chia, K. A. (2015). Fire recognition using RGB and YCbCr color space. *ARPN J. Eng. Appl. Sci.* 10 (21), 9786–9790. Available at: https://www.arpnjournals.org/jeas/research_papers/rp_2015/jeas_1115_2983.pdf
- Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., et al. Towards federated learning at scale: system design. 2019.
- Borges, P. V. K., and Izquierdo, E. (2010). A probabilistic approach for vision-based fire detection in videos. *IEEE Trans. Circuits Syst. Video Technol.* 20 (5), 721–731. doi:10.1109/tcsvt.2010.2045813
- Brik, B., Ksentini, A., and Bouazza, M. (2020). Federated learning for UAVs-enabled wireless networks: use cases, challenges, and open problems. *IEEE Access* 8, 53841–53849. doi:10.1109/access.2020.2981430
- Casbeer, D. W., Kingston, D. B., Beard, R. W., and McLain, T. W. (2006). Cooperative forest fire surveillance using a team of small unmanned air vehicles. *Int. J. Syst. Sci.* 37 (6), 351–360. doi:10.1080/00207720500438480
- Celik, T. (2010). Fast and efficient method for fire detection using image processing. *ETRI J.* 32 (6), 881–890. doi:10.4218/etrij.10.0109.0695
- Çelik, T., and Demirel, H. (2009). Fire detection in video sequences using a generic color model. *Fire Saf. J.* 44 (2), 147–158. doi:10.1016/j.firesaf.2008.05.005
- Çelik, T., Özkaramanli, H., and Demirel, H. (2007). “Fire and smoke detection without sensors: image processing based approach,” in *European signal processing conference*.
- Chino, D. Y. T., Avalhais, L. P. S., Rodrigues, J. F., and Traina, A. J. M. (2015). “BoWFire: detection of fire in still images by integrating pixel color and texture analysis,” in *Brazilian symposium of computer graphic and image processing*.
- Edge TPU - run inference at the edge | Google cloud. Available at: <https://cloud.google.com/edge-tpu>
- Frizzi, S., Kaabi, R., Bouchouicha, M., Ginoux, J. M., Moreau, E., and Fnaiech, F. (2016). Convolutional neural network for video fire and smoke detection. *IECON 2016 - 42nd Annu. Conf. IEEE Industrial Electron. Soc.*, 877–882. Available at: <https://api.semanticscholar.org/CorpusID:3541656>. doi:10.1109/IECON.2016.7793196
- Habiboğlu, Y. H., Günay, O., and Çetin, A. E. (2011). “Flame detection method in video using covariance descriptors,” in *ICASSP, IEEE international conference on acoustics, speech and signal processing - proceedings*.
- Habiboğlu, Y. H., Günay, O., and Çetin, A. E. (2012). Covariance matrix-based fire and flame detection method in video. *Mach. Vis. Appl.* 23 (6), 1103–1113. doi:10.1007/s00138-011-0369-1
- Hu, C., Tang, P., Jin, W., He, Z., and Li, W. (2018). “Real-time fire detection based on deep convolutional long-recurrent networks and optical flow method,” in *Chinese control conference, CCC*.
- Jadon, A., Omama, M., Varshney, A., Ansari, M. S., and Sharma, R. (2019). FireNet: a specialized lightweight fire and smoke detection model for real-time IoT applications.
- Khalil, A., Rahman, S. U., Alam, F., Ahmad, I., and Khalil, I. (2021). Fire detection using multi color space and background modeling. *Fire Technol.* 57 (3), 1221–1239. doi:10.1007/s10694-020-01030-9
- Kundu, S., Mahor, V., and Gupta, R. (2018). “A highly accurate fire detection method using discriminate method,” in *2018 international conference on advances in computing, communications and informatics, ICACCI 2018*.
- Mueller, M., Karasev, P., Kolesov, I., and Tannenbaum, A. (2013). Optical flow estimation for flame detection in videos. *IEEE Trans. Image Process.* 22 (7), 2786–2797. doi:10.1109/tip.2013.2258353
- Muhammad, K., Ahmad, J., Lv, Z., Bellavista, P., Yang, P., and Baik, S. W. (2019). Efficient deep CNN-based fire detection and localization in video surveillance applications. *IEEE Trans. Syst. Man. Cybern. Syst.* 49 (7), 1419–1434. doi:10.1109/tsmc.2018.2830099
- National Fire News National interagency fire center. Available at: <https://www.nifc.gov/fire-information/nfn>.
- NVIDIA Jetson Nano developer Kit | NVIDIA developer Available at: <https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/jetson-nano/product-development/>
- Qi, X., and Ebert, J. (2009). A computer vision-based method for fire detection in color videos. *Int. J. Imaging* 2 (9 S). Available at: <https://www.usu.edu/cs/people/XiaojunQi/Promotion/IJL.FireDetection.09.pdf>
- Qureshi, W. S., Ekpanyapong, M., Dailey, M. N., Rinsurongkawong, S., Malenichev, A., and Krasotkina, O. (2016). QuickBlaze: early fire detection using a combined video processing approach. *Fire Technol.* 52 (5), 1293–1317. doi:10.1007/s10694-015-0489-7
- Rafiee, A., Dianat, R., Jamshidi, M., Tavakoli, R., and Abbaspour, S. (2011). “Fire and smoke detection using wavelet analysis and disorder characteristics,” in *ICCRD2011 - 2011 3rd international conference on computer research and development*.
- Rinsurongkawong, S., Ekpanyapong, M., and Dailey, M. N. (2012). “Fire detection for early fire alarm based on optical flow video processing,” in *2012 9th international conference on electrical engineering/electronics, computer, telecommunications and information technology, ECTI-CON 2012*.
- Saydirasulovich, S. N., Mukhiddinov, M., Djuraev, O., Abdusalomov, A., and Cho, Y. I. (2023). An improved wildfire smoke detection based on YOLOv8 and UAV images. *Sensors (Basel)* 23 (20), 8374. doi:10.3390/s23208374
- Shafiq, M., and Gu, Z. (2022). Deep residual learning for image recognition: a survey. *Appl. Sci. Switz.* 12, 8972. doi:10.3390/app12188972
- Shamsoshoara, A. Ring oscillator and its application as physical unclonable function (PUF) for password management. 2019.
- Shamsoshoara, A., and GitHub repository, GitHub. 2020b. Fire-detection-uav-aerial-image-classification-segmentationunmannedaerialvehicle.
- Shamsoshoara, A., Afghah, F., Razi, A., Mousavi, S., Ashdown, J., and Turk, K. (2020a). An autonomous spectrum management scheme for unmanned aerial vehicle networks in disaster relief operations. *IEEE Access* 8, 58064–58079. doi:10.1109/access.2020.2982932

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Shamsoshoara, A., Afghah, F., Razi, A., Zheng, L., Fulé, P. Z., and Blasch, E. (2021). Aerial imagery pile burn detection using deep learning: the FLAME dataset. *Comput. Netw.* 193, 108001. doi:10.1016/j.comnet.2021.108001
- Statistics National interagency fire center. Available at: <https://www.nifc.gov/fire-information/statistics>.
- Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. A. (2017). "Inception-v4, inception-ResNet and the impact of residual connections on learning," in *31st AAAI conference on artificial intelligence, AAAI 2017*.
- Umar, M. M., De Silva, L. C., Bakar, M. S. A., and Petra, M. I. (2017). State of the art of smoke and fire detection using image processing. *Int. J. Signal Imaging Syst. Eng.* 10 (1–2), 22. doi:10.1504/ijsise.2017.084566
- Verstockt, S., Beji, T., De Potter, P., Van Hoecke, S., Sette, B., Merci, B., et al. (2013). Video driven fire spread forecasting (f) using multi-modal LWIR and visual flame and smoke data. *Pattern Recognit. Lett.* 34 (1), 62–69. doi:10.1016/j.patrec.2012.07.018
- Wang, T., Shi, L., Yuan, P., Bu, L., Hou, X., and Jiang, C. (2017). HIF1 α -Induced glycolysis metabolism is essential to the activation of inflammatory macrophages. *Mediat. Inflamm. Proc. - 2017 Chin. Autom. Congr.*, 2017, 9029327, doi:10.1155/2017/9029327
- Wu, H., Li, H., Shamsoshoara, A., Razi, A., and Afghah, F. (2020). "Transfer learning for wildfire identification in UAV imagery," in *2020 54th annual conference on information sciences and systems, CISS 2020*.
- Xuan Truong, T., and Kim, J. M. (2012). Fire flame detection in video sequences using multi-stage pattern recognition techniques. *Eng. Appl. Artif. Intell.* 25 (7), 1365–1372. doi:10.1016/j.engappai.2012.05.007
- Ye, D. H., Zikic, D., Glocker, B., Criminisi, A., and Konukoglu, E. (2013). Modality propagation: coherent synthesis of subject-specific scans with data-driven regularization. *5MB MODEL SIZE. ICLR17*. 16 (Pt 1), 606–613. doi:10.1007/978-3-642-40811-3_76
- Yuan, C., Liu, Z., and Zhang, Y. (2015b). "UAV-based forest fire detection and tracking using image processing techniques," in *2015 international conference on unmanned aircraft systems, ICUAS 2015*.
- Yuan, C., Liu, Z., and Zhang, Y. (2017). Aerial images-based forest fire detection for firefighting using optical remote sensing techniques and unmanned aerial vehicles. *J. Intelligent Robotic Syst. Theory Appl.* 88 (2–4), 635–654. doi:10.1007/s10846-016-0464-7
- Yuan, C., Zhang, Y., and Liu, Z. (2015a). A survey on technologies for automatic forest fire monitoring, detection, and fighting using unmanned aerial vehicles and remote sensing techniques. *Can. J. For. Res.* 45 (7), 783–792. doi:10.1139/cjfr-2014-0347
- Zhang, Q., Xu, J., Xu, L., and Guo, H. Deep convolutional neural networks for forest fire detection. In 2016.
- Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., and Torralba, A. (2016). "Learning deep features for discriminative localization," in *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*.