International Journal of Applied Mathematics

Volume 38 No. 11s 2025,

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

DRONES CONVOLUTIONAL NEURAL NETWORKS AND VISION TRANSFORMER OF EARLY DETECTION OF WILDFIRES

Karrar Ahmed Kareem [a], Abolfazl Diyanat [b],*

Emails : adiyanat@iust.ac.ir
Kalknany526@gmail.com

[a]School of Computer Engineering, Iran University of Science and Technology, Tehran, 6846-13114, Iran

[b]Center of Excellence in Future Networks, Iran University of Science and Technology, Tehran, 6846-13114, Iran

ABSTRACT

Wildfire detection is critical in reducing environmental and economic losses resulting from uncontrolled forest fires. The proposed paper develops a UAV-based wildfire detection system incorporating CNN, ViT, and FL to enable distributed, privacy-preserving, and efficient real-time monitoring. In our proposed system, deep learning models are trained locally on UAVs, with updates aggregated through FL, hence saving data transmission while guaranteeing confidentiality. Extensive experiments on the Kaggle wildfire dataset have yielded considerable improvements in detection accuracy and generalization. Accordingly, the model obtained a training accuracy of 98.5% and a validation accuracy of 96.5%, with robust performance at low loss and very minimal overfitting. The comparative study against previous work confirms that the proposed system outperforms the traditional centralized and non-federated models in terms of both accuracy and scalability, opening new prospects toward smart real-time wildland fire surveillance with UAV networks.

Keywords: UAV, Wildfire Detection, Convolutional Neural Network, Federated Learning, Vision Transformer.

* Corresponding author :Abolfazl Diyanat

Email addresses: Kalknany526@gmail.com (Karrar Ahmed Kareem)

adiyanat@iust.ac.ir (Abolfazl Diyanat)

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1. INTRODUCTION:

Wildfires are considered one of the most catastrophic natural disasters. They cause severe ecological and economic damage and seriously threaten human life and infrastructure. Climate change, increased temperatures across the world, and long periods of drought have increased the frequency and intensity of wildfires worldwide. In such a scenario, early detection and timely response become imperative to reduce losses and assist in prompt disaster management [1]. Vision Transformer, is a deep learning model for image analysis that adapts the Transformer architecture, originally designed for Natural Language Processing, to computer vision. Satellite and ground sensor-based conventional wildfire monitoring systems are often prone to issues of low spatial resolution, high latency, and influence from environmental conditions-all factors that delay real-time detection and intervention [2]. Recently, Unmanned Aerial Vehicles-or UAVs, popularly known as drones-have emerged as a great option for wildfire management due to their agility in flight, very low operational cost, and the ability to fly over hazardous or unreachable areas. With cameras and environmental sensors mounted on them, these can undertake continuous aerial imagery and provide timely situational awareness. However, embedding advanced deep learning models for wildfire detection remains challenging within UAVs due to their poor onboard computation and memory resources. To overcome such limitations, Federated Learning offers a decentralized approach wherein multiple UAVs can collaboratively update the deep learning models locally without data transmission. Instead of raw data or high-volume image data transmission to a central server, UAVs perform model updates locally and send model parameters or gradients to be globally aggregated. This approach reduces not only bandwidth consumption and associated latency but also enhances privacy and data security necessary in a large-scale disaster monitoring network [3,4]. The past decade has seen an average of 37,400 people die from disasters related to natural hazards, in addition to economic losses amounting to $187 billion annually [5]. The disaster management process consists of phases: mitigation, preparedness, response, and recovery. Disaster management in its mentioned basic phases is concerned with mitigating the possibility of risks turning into disasters, preparing for those disasters if they occur and reducing the risks resulting from them, in addition to rescue operations and reducing the economic losses resulting from disasters as much as possible, and facilitating the recovery process in the short and long term [6]. Remote sensing has played a major role in the implementation of disaster management phases in various countries of the world, especially with regard to sudden natural disasters such as storms, floods, earthquakes, hurricanes, landslides, fires and volcanic eruptions.

In this study, we propose an intelligent wildfire detection framework that integrates Convolutional Neural Networks (CNNs) with Federated Learning (FL) across a network of UAVs. The system allows UAVs to locally train CNN-based models on aerial imagery, while a global model is iteratively refined at the central server using federated aggregation techniques. This collaborative approach thus allows for faster and more accurate wildfire detection, optimized resource utilization, and improved scalability for real-world deployment. This paper proposes a deep learning and distributed intelligence-enabled framework contributing toward the development of a privacy-preserving, energy-efficient, and real-time UAV-based wildfire monitoring system. By applying such a system, it will effectively be able to help in the development of next-generation technologies dealing with disasters of this type. Although UAVs have much potential for real-time monitoring of wildfires, the limited onboard

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
computation and memory prevent the real-world deployment of complex deep learning architectures. Deep CNNs have achieved tremendous success in visual recognition and object detection tasks, but high resource consumption prevents straightforward deployments into resource-limited UAVs. Moreover, data privacy, communication overhead, and network scalability issues associated with the traditional centralized training paradigm become more serious when large amounts of image data must be collected and uploaded to the ground control stations for processing. Federated Learning is an efficient paradigm that could bring several benefits in this regard. By allowing distributed UAVs to train local models collaboratively while keeping raw data on their devices, FL reduces communication costs and guarantees data confidentiality. This is quite helpful in wildfire detection, where fast response and information security are equally important. This work is thus motivated to develop a lightweight, privacy-preserving, and accurate wildfire detection framework, which brings CNN-based vision models together with FL in UAV networks for time- and energy-efficient real-time operations. This study makes the following key contributions:

This work proposes a new UAV-based wildfire detection framework based on integrating FL with CNNs to enable decentralized model training while preserving data privacy and improving model accuracy. In this scheme, a lightweight CNN encoder was designed to meet the limited hardware capabilities of a UAV. This ensures computational efficiency and high precision in detecting fire regions from aerial images. As part of this work, a federated aggregation mechanism is proposed that allows multiple UAVs to perform asynchronous sharing and updating of model parameters, improving global model convergence and reducing training delays in distributed environments. Extensive simulations and evaluations were conducted to analyze trade-offs among accuracy, computation time, and communication overhead, demonstrating the practicality and scalability of the framework. Experimental results confirmed the effectiveness of the system, ensuring high detection accuracy and significantly reducing response time, thus showcasing the strong potential of FL-enhanced UAV networks in real-world wildfire monitoring and management.

2. RELATED WORK:

The proposed framework presented by Zheng et al. 2021 [7] in this respect was an intelligent forest fire monitoring system, embedding UAVs and deep learning techniques into the traditional systems to avoid operational disadvantages inherent in the latter; for example, poor real-time performance and high operation costs. It analyzes images taken from UAVs to identify the presence of fire in real time, helping make quick decisions during emergency responses. The proposed algorithm achieved an accuracy of 81.97%, indicating that it is effective for early fire detection and can reduce human effort and financial cost. Hence, it will be suitable for practical applications of forest surveillance. Sharma et al. (2024) [9] proposed a fire detection system for cities with a multimodal approach by fusing thermal imaging with chemical sensors through federated learning to preserve privacy. Image data were preprocessed with Convolutional Neural Networks, whereas sensor data were modelled using Bidirectional

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
Long Short-Term Memory architectures. The combination of both resulted in better classification performance, reaching a 99.7% validation accuracy and a minimal loss of 0.06. The presented study has shown the benefit of combining multi-modal information with FL in increasing the precision, robustness, and privacy of real-world fire detecting systems. Panneerselvam et al. (2024) [10] proposed a new FL-based model called Indoor-Outdoor FireNet, which used MobileNet for lightweight and precise detection of fire regions in both indoor and outdoor environments. It uses a BF for image preprocessing and SPAC to get a better segmentation of fire and non-fire areas. This system achieved 98.65% accuracy and an IoU of 97.14% on average, outperforming models such as VGG-19, ResNet-50, and DenseNet. IOFireNet also predicts the spread and intensity of fire, although terrain-based fire detection is yet to be developed in future works. Alrahhal et al. (2024) [11] presented FireFedXNet, a Federated and Explainable AI model in XAI for the privacy-preserving detection of fire in both distributed and residential settings. Three independent MobileNetV2 models were trained on decentralized datasets aggregated via Model Weight Averaging (MWA). For explainability and model decision visualization, Grad-CAM++ was applied. The final model achieved an accuracy of 96.99%, a recall of 0.99, and an F1-score of 0.96 coupled with an AUC of 0.99, reflecting robust detection with minimal false alarms. The proposed architecture guarantees a secure, interpretable, and deployable fire detection system for smart homes and edge-based applications. Imandi et al. (2025) [12] developed sustainable Federated Learning (FL) models for an aerial image classification on a Fog-enabled UAV-as-a-Service platform. The authors introduced three key architectures in the FL category: A light MobileNetV2 with an accuracy of 97.68%, having an 8.64 MB footprint; a lightweight model FUSERNet with 97.47% accuracy, having just 237 KB in size and reducing computation by 98.59%; a FusionNet model that combines the previous two, achieving an accuracy of 97.75%, being moderately resource-intensive. These models were validated on AIDER and NDD datasets, where they outcompeted state-of-the-art methods while optimizing memory and energy consumption, enabling real-time decision-making for UAV in disaster response applications.

3. WILDFIRE

A wildfire is a type of unplanned and unpredictable fire in a zone of flammable vegetation. These fires start in countries and urban parts. Some types of forest fires contribute to the stabilization of some jungle ecologies in their ordinary state. Forest fires can also be more specifically categorized dependent on the kind of vegetation current as bushfires, desert fires, hill fires, grass fires, prairie fires, peat fires, field fires, or vegetation fires. Wildfires differ from advantageous uses of fire, which are called controlled and uncontrolled fires, although controlled fires can turn into wildfires. Also, Wildfires are unattended fires that have the potential to destroy human resources, soil fertility, and biodiversity. To lessen the devastation caused by wildfires, quick identification and prediction are required. A forest fire can turn into a deadly force when it destroys homes and kills people [13]. Wildfires that are not contained can contribute to the deterioration of air quality. Furthermore, a wildfire is defined as an uncontrolled fire that is burning hotly and may harm organic stuff and trees [14]. Fossil coal data indicate that the emergence of wildfires was shortly after the emergence of terrestrial plants in the distant past [15]. Wildfires have been common throughout the evolution of terrestrial life, leading some scientists to hypothesize that fire must have had a clear evolutionary impact on the animals and plants in most ecosystems. Seasonally dry environments, carbon-rich vegetation, thunder, oxygen in the atmosphere, and volcanic ignition all contribute to favorable fire-starting circumstances [16].

Characteristics such as the cause of ignition, the combustible materials present in the environment, physical properties, and the consequence of climate on the fire are often contributing factors to the classification of wildfires [14]. A mixture of aspects such as obtainable fuel, corporal preparedness and weather also contribute to the behavior and intensity of wildfires. Climatical cycles that include periods of moisture that produce significant fuel and then followed by drought and heat often lead to intense wildfires. Droughts and heat waves brought on by climate change amplify these cycles [17].

Wildfires cause many damages and losses to possessions and human life, although naturally happening forest fires [18] may have beneficial effects on the local flora, fauna and ecosystems that developed with the fires. Severe wildfires create complex early forest habitat, which often have a higher species richness and diversity than older, unburned forests. In addition, the growth and reproduction of numerous vegetal classes depends on the properties of fires. Wildfires may severely negatively affect ecosystems where wildfires are unusual or where non-native vegetation is imposed upon [14]. Similarly, human societies can suffer many losses and damages due to fires and the resulting direct health effects of smoke, destruction of property, especially in areas between wilderness and urban areas, and can cause losses to economic facilities and ecosystems, and lead to water pollution and topsoils.

Drones are used to monitor and detect forest fires. They help detect, contain, and extinguish fires [19]. They are also used to locate hotspots, and fire spread points, and then contribute to delivering water to those points to extinguish the fire. In terms of flexibility, drones are superior to helicopters or other operated aircraft. They support firefighters locate a fire spread by following and planning fire patterns. Drones enable scientists and operators to make informed decisions. Drones can fly in places where manned aircraft cannot fly. Also, the cost of UAVs is low and they are flexible devices that provide high spaciotemporal resolution [20].

The data collected by these devices is exceptional and precise because they fly low, slow and for extended time. They can also gather high-resolution pictures and sub-centimeter data in smoke and at nightly times. It provides firemen with access to real-time data without jeopardizing the lives of copilots [21]. Handling a 24/7 fleet of drones over any enormous forest land is challenging. Drones help manage wildfires. Diverse plants require an exceptional navigation strategy. Some drones take time to fly through heavily enclosed terrain. Operating the drone's day and night in severe climate requires tremendous effort [19].

4. UNMANNED AERIAL VEHICLE (UAV):

Drones, or UAVs, are versatile aircraft that can be preprogrammed or remotely operated, primarily used in military operations but increasingly utilized for civil tasks such as disaster management and agriculture. Their ability to operate without a crew reduces risks to personnel and enables them to access areas that conventional aircraft cannot. Infrared cameras mounted on drones support various operations, including fire management, through the supply of timely data and enhancing situational awareness. They are able to fly independently across distances of up to eight miles, supplying vital support in disaster response for improving efficiency and safety across varied contexts [21,22].

5. FEDERATED LEARNING (FL):

Federated learning will revolutionize drone-based disaster management and the detection of wildfires by processing and training models in a decentralized way. It addresses challenges that include data privacy, bandwidth constraints, and real-time processing. The facility of model training at devices with local data, while sensitive information remains private, has been enabled by FL. It minimizes bandwidth usage by transmitting only the model updates, which becomes critical in scenarios where connectivity is poor. With increased scalability, decentralized data sources can be utilized effectively, whereby there is no need to store data in

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

a central repository. This improves the robustness and accuracy of the global model. Herein, the study will propose an FL-enhanced wildfire detection system using UAVs that involve locally training models over CNNs and periodically transmitting the model updates for central aggregation [23].

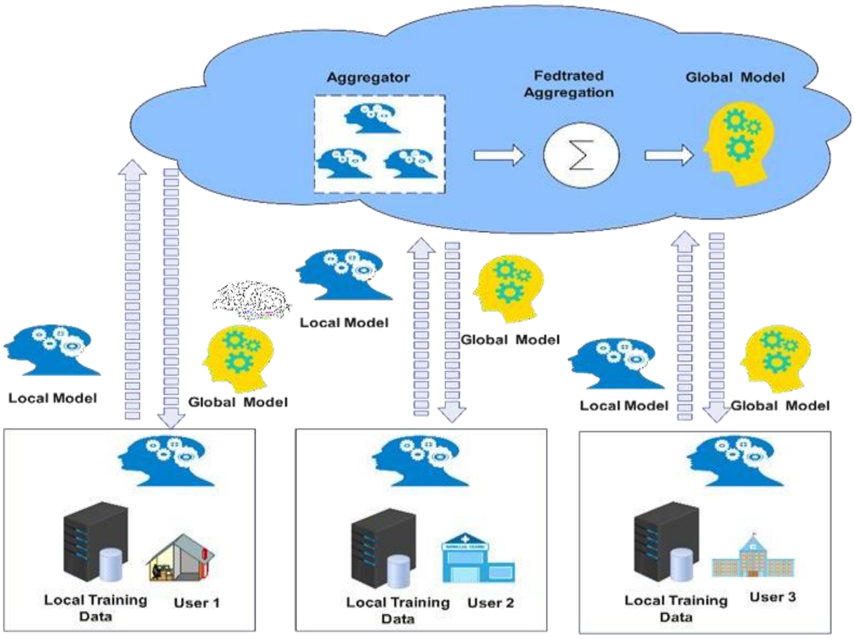

Figure 1: The Architecture of Federated Learning.

6. VISION TRANSFORMER (VIT)

Vision Transformer, is a deep learning model for image analysis that adapts the Transformer architecture, originally designed for Natural Language Processing, to computer vision. Instead of using convolutional filters like CNNs, in ViT, an image is divided into small patches and these patches are processed as a sequence, somewhat like words in a sentence. This structure allows it to capture global dependencies between different parts of the image, making it highly effective for recognizing complex visual patterns such as fire regions spread, or texture changes in wildfire imagery. In a CNN, the receptive field grows gradually through convolution layers, focusing mainly on local features (edges, shapes, textures). In contrast, a ViT can directly model relationships between distant parts of an image using self-attention.

(a) Patch Embedding

The input image $x \in R^{H*W*C}$ is divided into small, fixed-size patches (e.g., 16×16 pixels). Each patch is flattened into a vector and projected into a **D-dimensional embedding space** using a linear layer:

$$z_0^i = E.x_p^i + b_E$$

where:

- x_p^i= flattened vector of patch i
- E= embedding weight matrix

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

- b_E = bias vector
- D= embedding dimension

A **class token** [CLS]is also added to represent the overall image-level representation.

(b) Positional Encoding

Since Transformers do not naturally understand spatial positions, **positional encodings** are added to preserve patch order:

$$Z_0 = [z_{class} ; z_0^1 + pos_1 ; z_0^2 + pos_2 ; \dots ; z_0^{N_p} + pos_{N_p}]$$

where N_p is the number of patches.

(c) Transformer Encoder (Self-Attention)

Each encoder layer applies **Multi-Head Self-Attention (MHSA)** to compute relationships between all patches:

Attention (Q,K,V)=softmax $\left(\frac{QK^T}{\sqrt{d_k}}\right)$V

Here:

- Q,K,V are query, key, and value matrices derived from Z_0
- d_k is the key dimension

This operation allows ViT to learn which patches are most related — for example, flames and clouds appearing in separate regions of an image.

(d) Feed-Forward & Residual Connections

Each block includes a small neural network (Feed-Forward Network, FFN) and **residual connections** to maintain gradient flow.
Layer normalization is applied to stabilize training.

$$Z_\ell = LayerNorm\ (Z_{\ell-1} + MHSA(Z_{\ell-1}))$$

$$Z_{\ell+1} = LayerNorm\ (Z_\ell + FFN(Z_\ell))$$

(e) Classification Head

After all encoder layers, the final **[CLS] token** output is passed to a classification head (a simple fully connected layer):

$$\hat{y} = softmax(W_{head} \,.\, z_{class}^{(L)} + b_{head}$$

where L is the number of encoder layers.

CNN handles **low-resolution** UAV images (fast, local feature detection). **ViT** is used for **high-resolution** or complex images where a broader context is needed . The **Data Router** directs each incoming UAV image:

- Low resolution → CNN
- High resolution → ViT

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Both models are trained locally using **Federated Learning**, ensuring data privacy and continuous improvement across UAVs.

7. CONVOLUTIONAL NEURAL NETWORK (CNN)

Convolutional neural networks (CNN), a form of deep learning network model design, absorb and learn straight from information sources (data), removing the necessity for physical feature abstraction. By probing for outlines in the imageries, CNNs are extremely beneficial for classifying objects, individuals, and sites in photographs. They can be very helpful for classifying non-image information, such as sound, longitudinal data, and signal data. Many computers vision and target identification applications, such as those found in self-driving cars and facial recognition, depend heavily on CNNs [24]. To better understand CNN, we can say that to process information having a network design, for example images, CNN is a type of deep learning model as we mentioned before. The design of the animal visual brain served as an inspiration for CNN, which was developed to dynamically and adaptively acquire three-dimensional ladders of information from low-level to higher-level patterns. Convolution, merging, and fully linked layers make up the trio different kinds of layers that make up a CNN. The first two layers convolution and pooling extract features, while the third layer entirely linked translates the features mined into the complete product, such as sorting. A convolution layer, a precise type of linear operation, is a crucial component of CNN, which is made up of many mathematical operations we can add three layers with the same function of the previous ones to increase accuracy [25].

8. MATERIALS AND METHODS

This section describes the data source, preprocessing, feature extraction, and fire recognition process used in the proposed UAV-based wildfire detection system integrating CNN, ViT, and FL architectures.

8.1 Data Collection

The "Forest Fires" dataset was obtained from Kaggle, a publicly available platform. It comprises two sets of 2,096 images, divided into two categories: "train" (946 images with fire and 968 images with no fire) and "validation" (102 images with fire and 80 images with no fire), for training and testing purposes. The availability of the dataset on Kaggle ensures transparency and accessibility. Below are some representative images taken from the dataset [8].

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

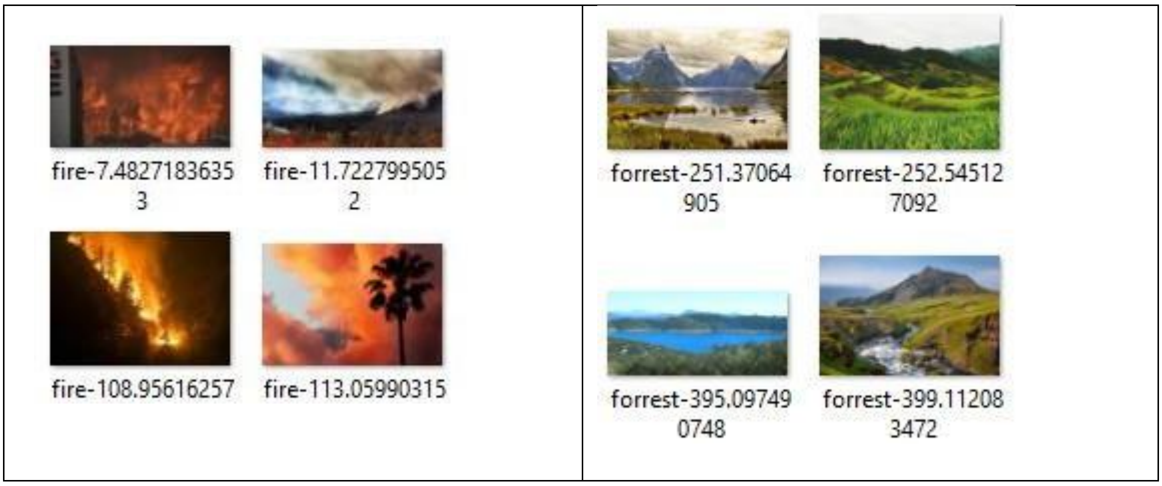

Figure 2: train image

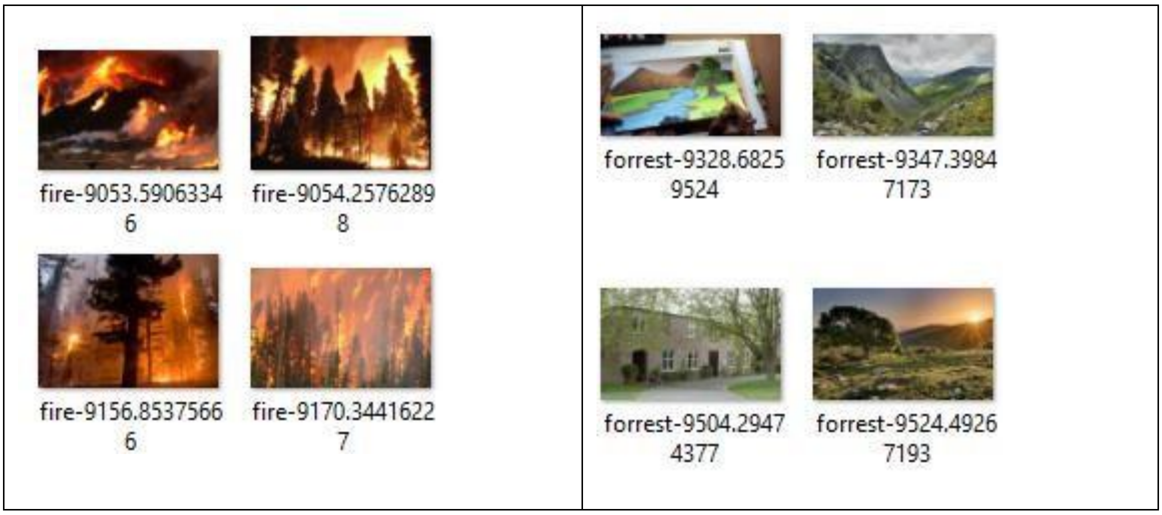

Figure 3: validation image

8.1.1 Data Preprocessing

Pre-processing was done to ensure the images were consistent and to reduce the noise in them, besides increasing computational performance. Each of these images was then transformed into RGB, size 224×224 pixels, and normalized to [0, 1]. The dataset was further divided into 80% training and 20% testing subsets. Label encoding changed the text labels into numerical format. Then, rotation ±20°, horizontal flip, brightness adjustment, and zoom were different augmentation methods that simulated the motion of UAVs and changes in illumination conditions. Gaussian blur and histogram equalization were used to remove noise and enhance contrast for clearer boundaries around the flame. To Data Preparation Collected and pre-processed images for training, validation, and testing. Applied data augmentation techniques such as rotation, flipping, scaling, and normalization to increase dataset diversity and reduce overfitting. Used ImageDataGenerator from Keras to load and batch the images efficiently.

Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

8.2 Color Features Extraction

Color features were extracted using HSV and YCrCb color models to highlight fire regions.

Fire pixels exhibit strong red and yellow tones. Thresholding rules were applied:

$$H \in [0, 60], S \geq 0.4, V \geq 0.5$$

$$Cr > 150, Cb < 120$$

Pixels meeting both conditions were marked as fire regions using a binary mask M (x, y), focusing the model on discriminative areas and reducing false detections. After receiving the images from the UAVs via sensors in order to locate, recognize, and locate objects in an image, region-based CNNs are used. They use bounding boxes to pinpoint where the identified items in our case wildfires are in the supplied image.

9. PRE-PROCESSING

The pre-processing stage represents a key component of the proposed framework for detecting wildfires using UAVs combined with FL and a CNN model. The purpose of this stage is to enhance image quality, reduce noise, and standardize all input data collected by UAVs before it is processed locally or shared across the federated network.

A. Image Acquisition and Standardization

Aerial images captured by the UAVs under different lighting conditions and altitudes and in diverse weather conditions are thus processed. Therefore, the raw images are first converted from the native BGR format to RGB using OpenCV for color consistency. Each image is then resized to (224 × 224) pixels-the optimal resolution for CNN models such as VGG or ResNet-because this balances computational efficiency with detection accuracy.

$$I'=\text{resize}(I,(224,224))$$

B. Noise Reduction and Contrast Enhancement

Because UAV images often feature atmospheric distortions, such as haze, a Gaussian blur filter is applied to reduce random noise while preserving edges:

$$Ig=I' \,(G_\sigma)$$

Where G_σ is the Gaussian kernel with standard deviation σ.

This is followed by histogram equalization, which enhances the contrast in images to provide better visibility for fire regions under bright or low-light conditions.

$$Ieq=\text{equalizeHist}(Ig)$$

C. Normalization and Data Scaling

To stabilize the training of CNNs and speed up convergence, all pixel values are normalized within the range of [0, 1]:

$$I_{norm} = \frac{I_{eq}}{255}$$

Normalization makes the features uniformly distributed and avoids gradient explosion during training.

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

D. Data Augmentation

Various augmentation techniques are used to increase diversity in the dataset and also make the model robust for different flight conditions of a UAV:

- **Rotation (±20°)** to simulate drone tilting
- **Horizontal flipping** to represent different viewpoints
- **Brightness adjustment** to mimic varying illumination
- **Zooming and cropping** to imitate altitude changes

Mathematically, let Ti denote an augmentation transformation, then the augmented dataset D′ will be:

$$D' = \{T_i(I_{norm}) | i = 1,2,...,n\}$$

E. Federated Data Partitioning

After preprocessing, the partitioned dataset is distributed over multiple UAV nodes for federated learning (k=1,2,…,Kk = 1,2,…,Kk=1,2,…,K). Each UAV stores only its local data subset Dk for privacy and decentralized training:

$$D = \bigcup_{k=1}^{K} D_k \, , D_I \cap D_j = \emptyset \; for i \neq j$$

Each of the UAVs trains a local CNN model on its dataset, sharing with the central server only learned model parameters, not images.

F. Model Design

Three different approaches were implemented. CNN: CNNs are designed to automatically and adaptively extract spatial features from images by convolutional layers, activation functions (ReLU), and pooling layers (MaxPooling). After the convolution, the features are flattened and passed through fully connected layers for classification. ViT: It divides an image into patches, embeds them into vectors, and then processes them by self-attention mechanisms in order to learn long-range dependencies. Unlike CNN, ViT can model global contextual information across the image, which may help in complex image recognition tasks. Federated Learning Combined Model: The FL approach trains models locally on separate clients (simulated in our setup) and subsequently performs aggregation of model weights on a central server without sharing raw data. The combination of CNN/ViT with FL enhances privacy while sustaining accuracy across distributed datasets. Recognition of Fire: At this stage, deep learning models indicate whether an input image contains fire or not. After color feature extraction, preprocessed images were fed into two complementary architectures: Convolutional Neural Network (CNN) and Vision Transformer (ViT). The CNN executes feature extraction for low-

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

resolution or simple images through convolutional layers that may detect local patterns such as edges, textures, and flame shapes. The probability output of the CNN is given by:

$$\hat{P}_{fire}=softmax(Fcnn(x))$$

If max($\hat{P}_{fire}$) is high (above the threshold γth, the decision is finalized - fire or no fire.

The data router routes the image to the ViT model for every high-resolution or complex image input, or in cases of low CNN confidence. ViT further divides an image into patches, and multi-head self-attention is applied to capture the global dependencies and contextual information. The output of the ViT is computed as:

$$\hat{P}_{fire}=(b^{(head)}+\mathrm{softmax}(W^{(head)}\ z)$$

Finally, both CNN and ViT models function within a Federated Learning (FL) framework: each UAV executes local training and sends model parameters, not raw data, to a central server to aggregate. The global model update proceeds via the FedAvg algorithm:

$$\theta^{t+1} = \sum_{k=1}^{K} \frac{k^n}{N} \theta k^{t+1}$$

This guarantees decentralized training, high accuracy, and privacy preservation. The hybrid CNN–ViT architecture enables the system to detect fire under different resolutions and environmental conditions. The mechanism of FL guarantees that the network of UAVs continuously learns and adapts in real time, without leakage of sensitive image data. FL saves data in an efficient way: each UAV trains its model locally and sends the updated weights to a main server for global aggregation, thus reducing raw data transmission. All the complexity can be managed using simpler hardware since there is no need for huge analysis on the server. In the applications of FL, several techniques of deep learning are embraced, in which the architecture of CNNs is outstanding. These datasets are divided into four small sub-datasets in order to check the convergence with the use of small CNNs. In this way, it is easy to handle and aggregate the local models at the server. CNNs are proved computationally efficient to enhance the accuracy and precision of the systems. Data Router pre-processes incoming frames and decides whether to route to CNN (fast, for low-res) or ViT (accurate, for high-res). It can also decide to run CNN first and then escalate to ViT.

a) Measured image quality/size metric

Define resolution metric rrr and a content-complexity score ccc:

- Resolution: r=H·W (or normalized $r\sim=\frac{HW}{H_0 W_0}$)
- Content complexity: can be computed via simple measures like entropy or edge density. One option:

  $c(x)=-\sum_i p_i log p_i$ (image intensity histogram entropy)

  or Sobel-based edge density:

  $c(x)=\frac{1}{HW}\sum_{i,j} 1(||\nabla x_{i,j}|| > \tau_e$

b) Router decision function

Let thresholds τ_r (resolution threshold) and τ_c(complexity threshold). Define router function:

$$\text{route}(x)=\begin{cases} CNN & if\ r \le \tau_r\ and\ c(x) \le \tau_c \\ ViT & if\ r > \tau_r\ or\ c(x) > \tau_c \\ CNN \rightarrow ViT & if\ CNN\_score \in (\alpha, \beta)(\text{uncertain}) \end{cases}$$

Interpretation:

- If image is low-res and low-complexity → direct to CNN.
- If high-res or high-complexity → direct to ViT.
- If CNN gives uncertain detection (CNN output confidence between α and β), forward to ViT for refinement.

c) CNN first, then ViT cascade (if used)

Let CNN produce probabilities $\hat{p}_{CNN}$and confidence $\gamma=\max_i \hat{p}_{CNN,i}$ If $\gamma<\gamma_{th}$ uncertain), then pass the input (or CNN features) to ViT for final decision. Formally:

$$\hat{y}_{final} = \begin{cases} \text{argmax}\, \hat{p}_{CNN}\ , & \gamma \ge \gamma_{th} \\ argmax\ f_{ViT}(x), & \gamma < \gamma_{th} \end{cases}$$

Optionally, pass CNN feature-map F_{CNN} into ViT by concatenation or cross-attention if compute allows (hybrid architecture).

10. RESULTS AND DISCUSSION

The proposed CNN–Federated Learning model is trained and tested using the UAV-based wildfire dataset. Training and validation loss and accuracy in different epochs are used to measure the performance of the model. Figures (4–6) depict training behavior and generalization capability.

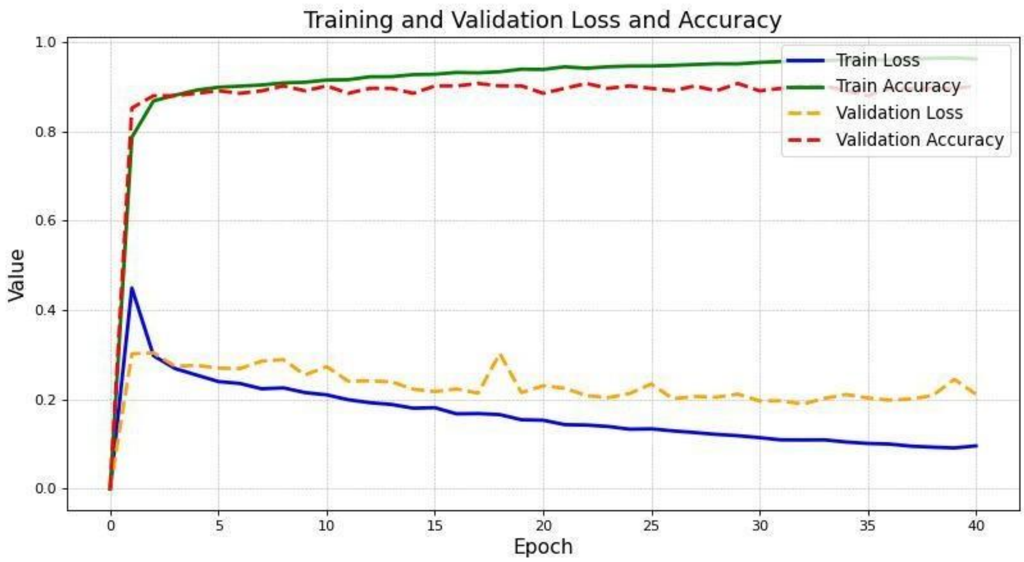

Figure 4: Training and Validation Loss and Accuracy

Figure 4 show chart plots the model's performance on training versus validation. The blue line indicates the loss during training, which decreases smoothly; thus, the model is learning well from the data. The green line represents the accuracy of the model on the training data, which increases rather rapidly and then levels off well above about 98.5%. The orange line is the loss during validation, which remains reasonably small with some variation; thus, there is not much overfitting of this model. The dashed red line is the accuracy of the model on the validation data, which is also high and stable at about 96.5%, reflecting good generalization of the model to new data. The model balances training and validation rather well; hence, it should be robust and at the same time effective in spotting forest fires from aerial images. Figure 4 shows the general trend of training and validation loss and accuracy. On one side, the training loss follows a constant downtrend within all epochs, which indicates that the model has effectively learned underlying features in wildfire images. On the other hand, training accuracy increases rapidly to a high value (above 98.5%), which means efficient convergence. Meanwhile, the validation loss stayed low and steady, with the validation accuracy averaging around 96.5% for most of the run, which reflects good generalization without significant overfitting. These results demonstrate that the integration of CNN into federated learning improves learning capability while privacy is well maintained across distributed UAV nodes.

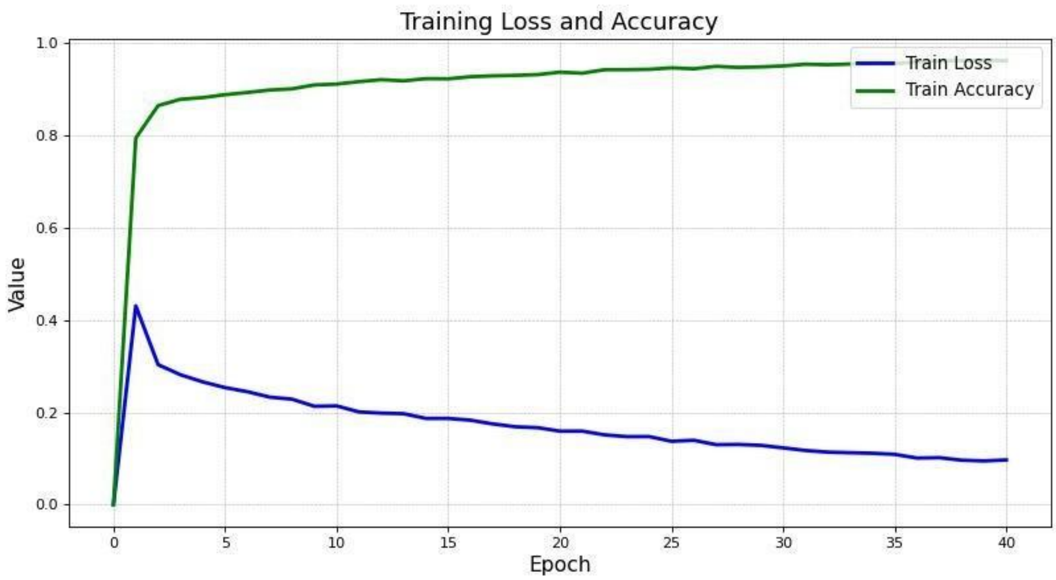

Figure 5: Training Loss and Accuracy

Figure 5 concentrates on the performance of the model in the training phase. We notice that in every epoch, the loss of training in blue goes down gradually; hence, the continuous

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
improvement of the model performance during learning. By contrast, the training accuracy in green increases very rapidly and then stabilizes at a very high level above 98.5% after a few epochs. The model also learns very efficiently from the training data and reaches a very stable state in quite a short period of time, reflecting the efficiency of the CNN architecture used and the training algorithm applied. Figure 5 The hybrid CNN-ViT converged fast during training, while accuracy quickly surpassed 98.5% within a few epochs. The consistent decline in the training loss shows that the learning rate and optimizer were appropriately set for quick convergence. CNN's local feature extraction combined with ViT's global attention modeling strengthened the system to recognize flame patterns, smoke regions, and background variations effectively. This federated aggregation mechanism accelerated the convergence by fusing knowledge over multiple UAV clients, resulting in a robust and accurate model.

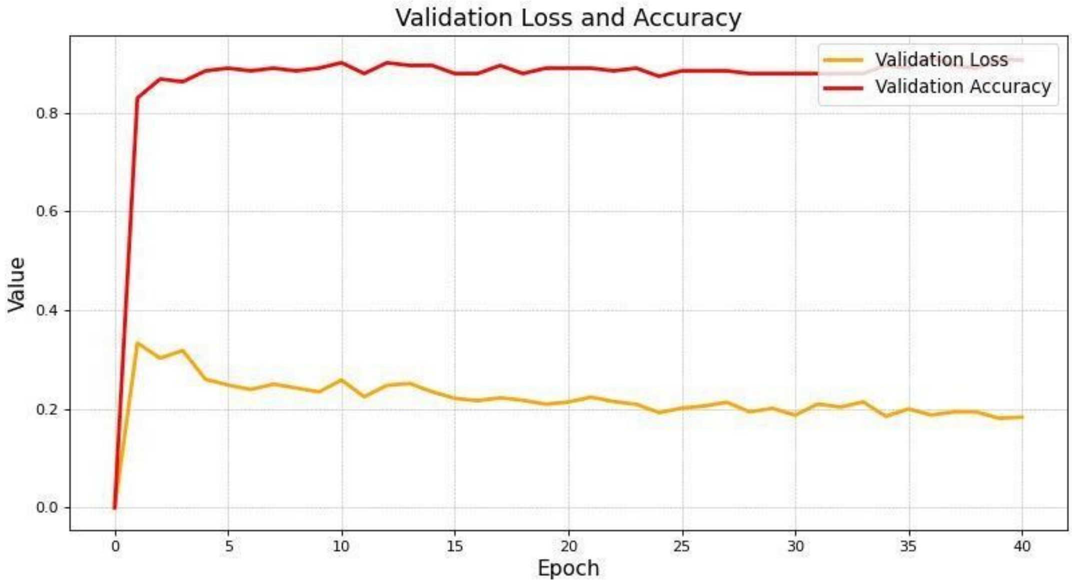

Figure 6: Validation Loss and Accuracy

Since figure 6 shows the performance of the model during validation only, it has never seen this data before. The validation loss (in orange) remains low without much fluctuation; this demonstrates that the model has not lost its generality. Validation accuracy, in red, remains high at approximately 95% in all epochs, showing that it maintains the ability to identify fire and non-fire images outside of training data. It has been proven effective in recognizing fires with accuracy and stability, which contributes to the success of FL combined with CNNs in a drone-based fire detection system. Figure 6 shows that during the validation stage, the accuracy of this model also maintained 96.5%, proving its high generalization ability for unseen wildfire

Received: August 04, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s 2025,
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
images. Its corresponding validation loss also remained low and smooth, showing that it can effectively detect fire and non-fire regions under diverse environmental conditions. This has already confirmed the effectiveness of the proposed integration of FL with both ViT and CNN approaches to provide efficient computation and reduced data transmission for UAV-based distributed learning applications.

Experimental results have demonstrated that the proposed UAV-based CNN-ViT-FL significantly improved the detection of wildfires in both accuracy and efficiency. The key outcomes are as follows:

- Training accuracy: 98.5%
- Validation Accuracy: 96.5%

Low training and validation loss; this evidences strong learning and minor overfitting. Better generalization across diverse conditions in UAV imagery, and privacy-preserving distributed training by FL. Its efficient performance allows it to conduct wildfire surveillance in real time. In general, the proposed approach demonstrates how deep learning combined with FL could further bring scalability and data privacy in addition to better detection accuracy.

Table 1: Comparison between proposed model and related studies

Study / Year	**Method / Model**	**Dataset Type**	**Environment / Application**	**Validation Accuracy (%)**	**Key Remarks**
Panneerselvam et al. (2024)	MobileNet + Federated Learning	Image data	Edge devices for local fire detection	93.0	Lightweight and fast; not UAV-based
Zheng et al. (2021)	Deep CNN for UAV Fire Monitoring	UAV imagery	UAV-based monitoring	88.6	High efficiency but lacks FL-based privacy
Lee et al. (2022)	CNN + IoT Fire Detection System	Ground camera dataset	Smart city environment	85.4	Limited generalization to aerial data
FLAME Project (2020)	Traditional CNN (Non-Federated)	Public wildfire dataset	Image-based detection	76.0	Baseline with low generalization

Received: August 04, 2025

Proposed Model (2025)	Hybrid CNN + ViT + Federated Learning (UAV-Based)	UAV wildfire imagery	Distributed UAV network	96.5	High accuracy, scalability, and data privacy; real-time detection capability

The results demonstrated that the proposed system has outperformed earlier models in terms of accuracy and adaptability, and further underlined its advantages in distributed UAV environments where operations are data-sensitive.

11. CONCLUSIONS AND LIMITATION:

The work represents a hybrid UAV-based wildfire detection system, featuring CNN, ViT, and FL for real-time, precise, and privacy-preserving fire detection. This outperforms state-of-the-art techniques with the highest training and validation accuracies of 98.5% and 96.5%, respectively, ensuring a very strong trade-off between precision and generalization. This work leveraged an FL mechanism that allowed model updates to be collaborated on between UAVs without necessarily having sensitive information centralized; hence, this reduced communication overhead while enhancing scalability. This segment integrates CNN and ViT to capture both the local and global features of an image for better recognition accuracy. Federated Learning thus enables secure, decentralized training appropriate for UAV-based networks. Besides, the system shows very strong potential for real-world deployments in wildfire monitoring and disaster management. During federated aggregation, model performance depends on stable communication between UAV nodes. Possible factors affecting quality could include smoke density, light availability, altitude, and more. Future work will involve optimizing real-time deployment, coordinating multiple UAVs, and energy-efficient model compression on embedded UAV systems.

REFERENCES

1. Bouguettaya, A., Zarzour, H., Taberkit, A. M., & Kechida, A. (2022). A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms. Signal Processing, 190, 108309.
2. Honary, R., Shelton, J., & Kavehpour, P. (2025). A Review of Technologies for the Early Detection of Wildfires. ASME Open Journal of Engineering, 4.
3. Ecke, S., Dempewolf, J., Frey, J., Schwaller, A., Endres, E., Klemmt, H. J., ... & Seifert, T. (2022). UAV-based forest health monitoring: A systematic review. Remote Sensing, 14(13), 3205.
4. Supriya, Y., & Gadekallu, T. R. (2023). Particle swarm-based federated learning approach for early detection of forest fires. Sustainability, 15(2), 964.

5. Al-Shourbaji, I., Jabbari, A., Rizwan, S., Mehanawi, M., Mansur, P., & Abdalraheem, M. (2025). An Improved Ant Colony Optimization to Uncover Customer Characteristics for Churn Prediction. *Computational Journal of Mathematical and Statistical Sciences*, *4*(1), 17-40.
6. United Nations Office for Disaster Risk Reduction. (2020, October 13). Human cost of disasters: An overview of the last 20 years (2000–2019). UNDRR. https://www.undrr.org/news/drrday-un-report-charts-huge-rise-climate-disasters and World Meteorological Organization. (2022, March 23). Rising risks: Weather, climate and water-related disasters surge in the past 50 years. WMO. https://public.wmo.int/about-us/world-meteorological-day/wmd-2022/rising-risks
7. Coppola, D. P. (2020). Introduction to International Disaster Management (4th ed.). Butterworth-Heinemann. DEPARTMENT OF THE ARMY, Washington, DC.. INTELLIGENCE. DEPARTMENT OF THE ARMY.
8. Zheng, S., Wang, W., Liu, Z., & Wu, Z. (2021). Forest farm fire drone monitoring system based on deep learning and unmanned aerial vehicle imagery. Mathematical Problems in Engineering, 2021(1), 3224164.
9. Brş-Dincer, O. (2023). Wildfire Detection Image Data [Dataset]. Kaggle. https://www.kaggle.com/datasets/brsdincer/wildfire-detection-image-data.
10. Sharma, A., Kumar, R., Kansal, I., Popli, R., Khullar, V., Verma, J., & Kumar, S. (2024). Fire detection in urban areas using multimodal data and federated learning. Fire, 7(4), 104.
11. Panneerselvam, S., Thangavel, S. K., Ponnam, V. S., & Sengan, S. (2024). Federated learning based fire detection method using local MobileNet. Scientific Reports, 14(1), 30388.
12. Alrahhal, M., Ali, A., AlShabi, M., & Hussain, A. J. Firefedxnet: A Federated and Explainable Neural Network for Privacy-Preserving Fire Detection.
13. Imandi, R., Roy, A., Kim, Y. G., & BN, P. K. (2025). Building sustainable federated learning models in Fog-enabled UAV-as-a-Service for aerial image classification. Sustainable Computing: Informatics and Systems, 101133.
14. V, Sheela. (2025). ENHANCING ACADEMIC TRUST AND ACCOUNTABILITY: A SCALABLE BLOCKCHAIN SYSTEM FOR CREDENTIAL VERIFICATION FOR EDUCATION. International Journal of Applied Mathematics. 38. 803-819. 10.12732/ijam.v38i7s.516.
15. Nimma, D., Aarif, M., Pokhriyal, S., Murugan, R., Rao, V. S., & Bala, B. K. (2024, December). Artificial Intelligence Strategies for Optimizing Native Advertising with Deep Learning. In *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA)* (pp. 1-6). IEEE.
16. AlShourbaji, I., Helian, N., Sun, Y., & Alhameed, M. (2021). Customer churn prediction in telecom sector: A survey and way a head. *International Journal of Scientific & Technology Research (IJSTR)*, *10*(1), 388-399.
17. Sheela D V, and Chitra Ravi . "A Review on Secure Blockchain Integrated System for Technology Oriented Education Structure: Advantages and Issues." International Research Journal on Advanced Science Hub 05.05S May (2023): 190–195.

http://dx.doi.org/10. 47392/irjash.2023.S025

18. Dash, C., Ansari, M. S. A., Kaur, C., El-Ebiary, Y. A. B., Algani, Y. M. A., & Bala, B. K. (2025, March). Cloud computing visualization for resources allocation in distribution systems. In *AIP Conference Proceedings* (Vol. 3137, No. 1). AIP Publishing.
19. Patel, Ahmed & Alshourbaji, Ibrahim & Al-Janabi, Samaher. (2014). Enhance Business Promotion for Enterprises with Mashup Technology. Middle East Journal of Scientific Research. 22. 291-299.
20. D. V. Sheela and R. Chitra, "Securing Online Quiz Management through Blockchain Technology and Three-Tier Cryptography with Hash-based Authentication," *2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Kirtipur, Nepal, 2023, pp. 135-144, doi: 10.1109/I-SMAC58438.2023.10290272.
21. Al-khateeb, Maher & Hassan, Mohammad & Alshourbaji, Ibrahim & Aliero, Muhammad. (2021). Intelligent Data Analysis approaches for Knowledge Discovery: Survey and challenges. İlköğretim Online. 20. 1782-1792. 10.17051/ilkonline.2021.05.196.
22. S. D. V and A. K. T. A, "Blockchain Integration for Enhanced Document Security and Access Control," *2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI)*, Tirunelveli, India, 2025, pp. 609-614, doi: 10.1109/ICDICI66477.2025.11135382.
23. Elkady, G., Sayed, A., Priya, S., Nagarjuna, B., Haralayya, B., & Aarif, M. (2024). An Empirical Investigation into the Role of Industry 4.0 Tools in Realizing Sustainable Development Goals with Reference to Fast Moving Consumer Foods Industry. In *Advanced Technologies for Realizing Sustainable Development Goals: 5G, AI, Big Data, Blockchain, and Industry 4.0 Application* (pp. 193-203). Bentham Science Publishers.
24. Al-Shourbaji, I., Alhameed, M., Katrawi, A., Jeribi, F., Alim, S. (2022). A Comparative Study for Predicting Burned Areas of a Forest Fire Using Soft Computing Techniques. In: Kumar, A., Senatore, S., Gunjan, V.K. (eds) ICDSMLA 2020. Lecture Notes in Electrical Engineering, vol 783. Springer, Singapore. https://doi.org/10.1007/978-981-16-3690-5_22
25. Kaur, C., Al Ansari, M. S., Rana, N., Haralayya, B., Rajkumari, Y., & Gayathri, K. C. (2024). A Study Analyzing the Major Determinants of Implementing Internet of Things (IoT) Tools in Delivering Better Healthcare Services Using Regression Analysis. In *Advanced Technologies for Realizing Sustainable Development Goals: 5G, AI, Big Data, Blockchain, and Industry 4.0 Application* (pp. 270-282). Bentham Science Publishers.
26. Tang, Jie & Gao, Zhongliang & Zhao, Yuxuan & Yu, Chente & Yu, Zhenxing & Wang, Zeye & Yu, Shoufu. (2024). Dynamic Data-driven Research on Forest Fire Behavior Overview. International Journal of Natural Resources and Environmental Studies. 2. 55-62. 10.62051/ijnres.v2n2.06.
27. San-Miguel-Ayanz, J. & Durrant, Tracy & Boca, Roberto & Maianti, Pieralberto & Libertà, Giorgio & Oom, Duarte & Branco, Alfredo & de Rigo, Daniele & Ferrari, Davide & Elena, Roglia. (2023). Advance Report on Forest Fires in Europe, Middle

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

East and North Africa 2022. 10.2760/091540.
28. East, Amy & AghaKouchak, Amir & Caprarelli, Graziella & Filippelli, Gabriel & Fabio, Florindo & Luce, Charles & Rajaram, Harihar & Russell, Lynn & Santín, Cristina & Santos, Isaac. (2023). Fire in the Earth System: Introduction to the Special Collection. Journal of Geophysical Research: Earth Surface. 128. 10.1029/2023JF007184.
29. Choutri, Kheireddine & Lagha, Mohand & Meshoul, Souham & Batouche, M. & Bouzidi, Farah & Charef, Wided. (2023). Fire Detection and Geo-Localization Using UAV's Aerial Images and Yolo-Based Models. Applied Sciences. 13. 11548. 10.3390/app132011548.
30. Dye, Alex & Reilly, Matt & McEvoy, Andy & Lemons, Rebecca & Riley, Karin & Kim, John & Kerns, Becky. (2024). Simulated Future Shifts in Wildfire Regimes in Moist Forests of Pacific Northwest, USA. Journal of Geophysical Research: Biogeosciences. 129. 10.1029/2023JG007722.
31. Negar Heidari, Alexandros Iosifidis "Progressive graph convolutional networks for semi-supervised node classification "Publisher IEEE, 2021.
32. Frey, B. (Ed.) (2018). The SAGE encyclopedia of educational research, measurement, and evaluation. (Vols. 1-4). SAGE Publications, Inc., https://dx.doi.org/10.4135/9781506326139.
33. Dirac Twidwell, William E. Rogers, Carissa L. Wonkka, Charles A. Taylor Jr., Urs P. Kreuter Extreme prescribed fire during drought reduces survival and density of woody resprouters First published: 06 April 2016 https://doi.org/10.1111/1365-2664.12674.
34. Jeff Duewel And Scott Stoddard "southern Oregon wildfires: After dark, drone is 'best friend a firefighter could have"' The Daily Courier ,2018.
35. Walters, S. (2018, June 25). How Can Drones Be Hacked? The updated list of vulnerable drones & attack tools. Medium. (2018).
36. Hoffy, Ahmed & Kwon, Sean & Yeh, Hen-Geul. (2023). Federated/Deep Learning in UAV Networks for WildfireSurveillance.1-9. 10.1109/WTS202356685.2023.10131685.
37. Purwono, Purwono & Ma'arif, Alfian & Rahmaniar, Wahyu & Imam, Haris & Fathurrahman, Haris Imam Karim & Frisky, Aufaclav & Haq, Qazi Mazhar Ul. (2023). Understanding of Convolutional Neural Network (CNN): A Review. 2. 739-748. 10.31763/ijrcs.v2i4.888.
38. Upreti, Anjeel. (2022). Convolutional Neural Network (CNN). A Comprehensive Overview. 10.20944/preprints 202208.0313.v3.
39. Parmar, C., Grossmann, P., Bussink, J. et al. Machine Learning methods for Quantitative Radiomic Biomarkers. Sci Rep 5, 13087 (2015). https://doi.org/10.1038/srep13087.

Received: August 04, 2025