International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I11P119
Volume 72 Issue 11, 175-183, November 2024

Original Article

Forest Fire Detection Using Multi-Modal Learning on RGB/IR Dataset

Tilottama Goswami[1], Apuru Rohan[2], Varkala Sujith Atesh[3], Kovvur Ram Mohan Rao[4]

[1,2,3,4]*Department of Information Technology, Vasavi College of Engineering, Telangana, India.*

[1]*Corresponding Author : tgoswami@staff.vce.ac.in*

Received: 14 June 2024 Revised: 02 October 2024 Accepted: 06 November 2024 Published: 29 November 2024

Abstract - *A forest fire is an uncontrolled fire that spreads rapidly and occurs in forests or other wildlands. Early detection of these kinds of fires is crucial to protecting the environment, vegetation, and wildlife. Several methods have been proposed to detect early forest fires and respond rapidly to emergency teams. The method proposed in the paper improves detection accuracy and identifies the fire's accurate position. This work uses multimodal learning, using two kinds of data from the same scene, IR and RGB, to detect forest fires. It proposes a deep learning-based methodology for detecting fire and smoke pixels at an accuracy much higher than the usual single-channel RGB or IR images. It classifies them into three classes using a deep learning methodology: - Fire without smoke - Fire with smoke - Smoke without fire. The models proposed are trained on the FLAME2 Dataset. The images are classified with a good accuracy of 99\%. In addition, the work uses image processing methods that detect the position of fire in the image and provide some real-time analysis of the feed.*

Keywords - *Deep learning, Ensemble learning, K-fold validation, Segmentation, CNN.*

1. Introduction

There are some natural and man-made causes for the start of a forest fire. Some natural calamities, like lightning strikes and volcanic eruptions, can ignite fire. Additionally, fireworks, improperly discarded cigarettes, and sparks from power lines can spark forest fires at the negligence of human beings. Forest fires adversely affect the environment, causing air pollution and contributing to greenhouse warming and other problems. It also destroys habitats, leading to loss of biodiversity. The remains of the forest fire ashes cause soil erosion. [1] Numerous strategies have been adopted to mitigate forest fires, but preventing damage has proven challenging. The only task to reduce the damage is to detect the wildfires as soon as possible and take necessary action to prevent their spread. Various approaches to detecting wildfires include image sensors, fire sensors, and many more. Different approaches fit for different needs and situations. The work proposes a method to detect wildfires using RGB and IR images of the scene. Infrared cameras detect the heat emitted by objects. Every object with a temperature above absolute zero emits infrared radiation, with hotter objects emitting more. The detected infrared radiation is converted into an electronic signal and then processed to create an image. These images display varying temperatures in varying shades, usually grayscale or false colour. To make the images more interpretable, different colours can represent different temperatures. For example, hotter areas might appear red or white, while cooler areas might appear blue or green. The contrast between the hot fire and the cooler surroundings (including clouds and the forest) makes the fire stand out clearly in infrared images. This allows for precise identification of the fire's location and intensity. Having single channel RGB or IR images for detecting forest fires is inaccurate. That is because each of the methods has its problems. Using only RGB to detect forest fires is not viable because the fire produces smoke that can obstruct the view and hinder finding the spots of fire under the cover of the fog/smoke of the fire. This can be mitigated by using infrared sensors, which can capture the fires even through thick smoke and give us a pretty good contrasting image of the fire with respect to the surroundings, as IR is the reading for the heat produced by a body. Just using the IR is not the answer either [2]. IR sensors detect the light reflected off water bodies or the reflections from the sun at an angle, which might give false readings. This is the reason for using both the IR and RGB image frames 1:1: to have the best of both worlds. It solves the problem of smoke covering the RGB sensor using the IR sensor, and false readings can be avoided using the RGB sensor's counterpart. Ensemble learning is a powerful tool in machine learning that increases prediction accuracy by combining the outputs of different models. This mechanism is applied in the work to combine different models for forest fire detection to detect the presence of fire accurately.[3] Segmentation is a low-level operation of dividing the image into different partitions by determining the similarity and discontinuity or finding the edges or boundaries.[4] There are many deep learning approaches to image segmentation, such as training a convolutional neural network to detect the segments in the images [5]. Predicting fire segments in an image labelled fire is beneficial for gaining various insights, such as the number of areas with fire, calculating the number of resources required to reduce the fire, etc. The work proposes a method using HSV (Hue, Saturation, Value) to determine the areas with fire using the IR image of the scene [6].

Despite the advancements in wildfire detection technologies, current methods still struggle to provide accurate and timely identification of fires under complex environmental conditions. Traditional RGB sensors are limited by visibility issues caused by smoke. In contrast, infrared (IR) sensors often suffer from false detections due to reflections from non-fire heat sources like sunlight or water bodies. Existing solutions focus on single-sensor approaches, which either fail in certain scenarios or generate inaccurate readings. This creates a critical research gap in developing a more robust system that can reliably detect fires in varying conditions by integrating multiple sensor modalities. The proposed study addresses this gap by combining RGB and IR data to overcome the limitations of individual sensor types, providing a more effective and accurate method for early wildfire detection.

This paper first reviews existing literature and research in the field. Following this outline is the methodology employed in the study. Subsequently, it presents the results obtained from the experiments and concludes by summarising the notable findings, discussing their implications, and outlining the future scope of the research.

2. Literature Review

In the past decade, significant attention has been directed towards forest fire detection, driven by the escalating reports of forest fire incidents globally and their consequential impact on society and the environment. Various methodologies have been proposed to address this challenge, including camera-based systems, Wireless Sensor Network (WSN)-based solutions, and machine learning applications. These approaches exhibit a spectrum of advantages and drawbacks, each with its performance metrics in detection accuracy. Leveraging multiple sensor sources and strategically deploying sensor nodes in satellite-inaccessible areas, wireless sensor networks emerge with a more optimistic outlook for accurate and timely detection. Consequently, they have gained prominence as a preferred technology across multiple domains. Melanophila is an insect, commonly known as fire beetle, with extraordinary sensitivity to infrared radiation using a specialised sensor organ. The same kind of sensor was developed in [7] and was used to predict the fire. The photomechanical IR receptors in Pyrophilous Melanophila beetles were made and used to detect forest fires. There are many other approaches to detecting forest fires using cameras and sensors. Sensor-based technologies to detect forest fires involve temperature and carbon dioxide sensors [8]. This approach compares the temperature and carbon dioxide readings at two different places to detect the fire. This approach accounts for 93.3% of accuracy. Forest fires are also predicted using noise [9]. A work uses spectral analysis of noise to detect forest fires. The noise sensors are placed in some parts of the forest, and the WSN is used to communicate among all the sensors and predict the forest fires early. The sensors are not recommended because they are placed in the forest, where they will be exposed to various physical conditions, and they may fail in some critical conditions.

Besides using sensors, another approach to detecting forest fires is to use video feed from a camera sensor or an Unmanned Aerial Vehicle (UAV). The data collected from the UAV has many advantages with respect to resolution and clarity [10]. The video feed is analysed by using technology to predict the fire. Many image processing techniques are proposed to detect forest fires from video. The real-time video feed can be fed from drones, and this method also has some disadvantages, such as when the forest is covered by smoke, the video will also not be clear enough to detect the forest fire. Many approaches have been proposed to overcome this by a bit.

The image data taken using IR cameras are converted into frequency using cosine transforms [11]. The smoke and fire areas are detected using the properties of smoke and fire colour, respectively. Morphological operations are applied to remove the noise, and the regions are connected using the connected component labelling technique. Many deep learning methodologies have been applied to the data to detect forest fires. The data type may vary in different cases, such as using only IR or RGB data. There are various datasets available for applying deep learning frameworks. Datasets may contain only IR imagery, RGB imagery or both RGB and IR imagery of the scene [12]. There can be others.

The referenced study in [13] utilizes infrared (IR) data for detecting forest fires by employing RGB and YCbCr color models to differentiate fire pixels from background elements. By separating luminance and chrominance from the original image, the method applies MATLAB for processing. The technique is validated on a specific image captured by a camera containing visible fire. Additionally, a supplementary method calculates and analyzes the fire image to distinguish genuine fire occurrences from false positives. Another approach processes the fire image, displaying results through terminal nodes and graphs via Wavelet Analyzer 5.0.

This methodology successfully identifies fire presence and provides related data, confirming the effectiveness of infrared and other channel camera sensors in enhancing fire detection due to the thermal emissions of fires. An additional algorithm is introduced for smoke detection, utilizing video sequences from Internet Protocol (IP) cameras and leveraging smoke characteristics, such as color, movement, and growth patterns. This scheme identifies potential smoke regions based on motion and color properties, reduces noise through morphological operations, and evaluates the growth patterns of these candidate smoke regions over time using connected component labeling. The evaluation indicates the smoke detection method's reliability, yielding false negative and false positive error rates of approximately 4% and 2%, respectively.

A deep learning model [14] has been trained on Landsat-8 imagery, in addition to RGB and IR data, specifically for detecting active fires and burning biomass. This method integrates both optical (Red, Green, and Blue) and thermal image modalities to enhance the model's representational power. Residual convolution and separable convolution blocks are employed to capture more detailed features from coarse datasets. Experimental results indicate an impressive overall accuracy of 97.35%, showcasing the model's strong detection ability, even for small active fires. The study includes imagery from various fire-prone regions such as Australian and North American forests, the Amazon rainforest, Central Africa, and Chernobyl (Ukraine).

Some methods like [15] are used to keep complexity low and detect the fire in real-time at a lower cost. These systems are embedded into edge devices and detect forest fires in real-time. Other than using UAVs for capturing videos of forests, a method uses satellites to detect forest fires [16]. This method offers satellite imagery promptly, capturing active fire spots and associated data to pinpoint wildfires' location, extent, and intensity. FIRMS tools and applications supply geospatial data and services to support fire management efforts and inform the public. While global data are accessible within three hours of satellite observation, real-time active fire detections are provided for the U.S. and Canada.

However, the operational costs are high, and real-time availability is often limited. Additionally, though satellite imagery offers sufficient spatial resolution for fire detection, its infrequent updates pose a challenge. More frequent updates sacrifice image granularity, rendering satellite imagery less suitable for forest fire detection.

Convolutional neural networks are powerful deep learning tools that can process grid structured data like images and identify the fire from the image. The methods [17] and [18] propose CNN architectures that are specially customised for forest fire detection and can accurately detect forest fires. Some object detection algorithms are faster than the standard CNN. The work [19] uses elastic YOLO v3 to detect the fire. This work is specially designed to work at nighttime, and [20] uses a single shot multibox detector, an object detection algorithm that performs forest fire detection in a single shot, enabling faster processing speed.

Recently, numerous approaches have been developed to detect forest fires, including camera-based systems, Wireless Sensor Networks (WSN), and machine learning techniques. Each method offers unique advantages, such as high detection accuracy or low cost, but also presents limitations. Sensor-based technologies like infrared (IR) and RGB cameras have been widely adopted but face challenges in environments with smoke or reflective surfaces, leading to false positives or reduced accuracy.

Deep learning models, including Convolutional Neural Networks (CNNs) and object detection algorithms, have further improved detection accuracy, with some systems tailored for specific conditions like nighttime or real-time detection using UAVs or satellite imagery. Despite the advancements, many of these methods struggle to balance detection accuracy with real-time processing and adaptability to various environmental conditions. This study addresses these gaps by proposing a novel solution that combines RGB and IR images, along with ensemble learning and deep learning-based segmentation techniques, to improve both the precision and speed of wildfire detection, even under challenging conditions.

3. Methodology

This work can be divided into two sections briefly. These two sections are depicted in the block diagram (Figure 1) as the last two stages. The first section presents a comparative study of different models used to classify an image into three sections, namely: - 1) Fire without smoke, 2) Fire with smoke, 3) Smoke without fire, and the next section describes a method to identify the segments in the image containing fire. The two sections form a chain, and an image predicted with fire is only advanced to the next section of detecting fire segments.

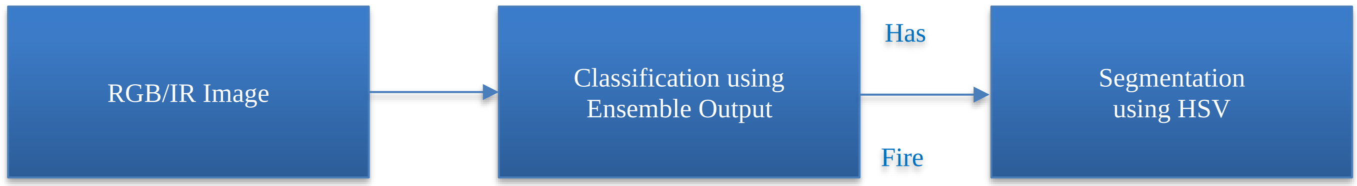

Fig. 1 Block diagram of the system

3.1. Dataset

The dataset that was used is the Flame2 dataset [12]. This contains two main sections: The main section contains raw and manipulated aerial imagery collected during a prescribed fire in an open canopy pine forest in Northern Arizona in November 2021. The primary dataset contains seven raw, unlabelled RGB and IR video pairs, labelled original resolution RGB/IR frame pairs, and 254p x 254p RGB/IR frame pairs. Both sets of frame pairs are derived from the 7 raw video pairs.

There are 7 video pairs which are present along with the 254p videos, each with resolution:

1. {3840p x 2160p RGB, 640p x 512p IR} length of video: 291 seconds
2. {3840p x 2160p RGB, 640p x 512p IR} length of video: 404 seconds
3. {1920 x 1080 RGB, 640p x 512p IR} length of video: 301 seconds
4. {1920 x 1080 RGB, 640p x 512p IR} length of video: 267 seconds
5. {1920 x 1080 RGB, 640p x 512p IR} length of video: 312 seconds
6. {1280 x 720 RGB, 640p x 512p IR} length of video: 234 seconds
7. {1280 x 720 RGB, 640p x 512p IR} length of video: 198 seconds

There are 53,541 pairs of RGB and IR images. The frame pairs are labelled with a “Fire/NoFire” label and a “Smoke/NoSmoke” label. Two experts have labelled both frame pair sets on a frame-by-frame basis. The “Fire/NoFire” label indicates whether the experts observe fire in each pair's RGB or IR image. The “Smoke/NoSmoke” label indicates whether the experts visually estimate that smoke fills at least 50% of the 254p x 254p RGB frame.

3.2. Classification

This section describes various models that detect fire in the image [21] and a technique to increase classification accuracy. The models used in [21] are –

1. Logistic Regression.
2. LeNet5 one stream.
3. LeNet5 (Early & Late Fusion).
4. Flame one stream.
5. Flame (Early & Late Fusion).
6. MobilenetV2.

The models mentioned above are trained typically on the FLAME2 Dataset [12]. This dataset contains both IR and RGB imagery taken from an Unmanned Aerial Vehicle (UAV). The images are human labelled and classified into three classes, as mentioned earlier. The RGB and IR videos collected by UAVs are pre-processed. The procedure of preprocessing used by [1] is mentioned below. There is no change in the preprocessing procedure in our work. The UAV video feed undergoes preprocessing to ensure that the resolutions of the RGB and IR images are precisely aligned. This involves cropping and scaling the pre-processed IR and RGB frames to achieve 254p resolution, ensuring they have similar fields of view and zoom levels. Each RGB frame correlates directly with an IR frame extracted from the 254p video using Adobe Media Encoder.

While RGB videos have a 1.78:1 aspect ratio, IR videos have a 1.25:1 aspect ratio, necessitating cropping for a 1:1 aspect ratio. Following conversion into frames, the IR video typically contains slightly more frames than the RGB video, with approximately 20 additional IR frames for every 10,000 frames of IR/RGB video. Frameshift arises from slight asynchrony in IR/RGB camera recording times and differences in frame rates. To rectify this, a corresponding number of IR frames are removed from the start and end of each video pair's IR frames. The frame drift varies non-linearly, ranging from 5 to 18 frames. Perfect matching is achieved after aligning IR/RGB frames and removing disparities. The architecture of the models mentioned above was provided by [21] and trained using pre-processed data. Accuracy, F1-score, precision and recall of the model are recorded after training the models for fire detection. The training time was long, and the computing power was huge. The training process was sped up using a small portion of repetitive data.

The dataset is composed of 3 classes – 1) Fire without smoke, 2) Fire with smoke, 3) Smoke without fire. The training dataset was reduced to 3% of each output class, so every class has equal importance in the training procedure. The training was done using 5-fold validation to avoid overfitting the models. The train data was divided into five equal random sets or folds. In each iteration, 4 folds were used for training, and the remaining 1-fold was used for testing. So, in the first round, fold 1 is used as the testing set and folds 2-5 as the training set. In the second round, fold 2 was used as the testing set, and folds 1, 3, 4, and 5 were used as the training set. And so on, until it has used each fold as the testing set once. This technique is illustrated in Figure 2. After the model is trained, the models are tested on the remaining dataset, and the results are recorded. The results depicted a slight increase in the performance compared to the ones given in [21], but more importantly, this k-fold method improved the quality of the trained models drastically. The models trained with a regular 80-20 train test split produced very inconsistent models; some were 50% accurate, others were 90% and so on. So, the practical test train split is 80-20, but k-fold cycles are used throughout the whole dataset. Using the k-fold method, the models were very consistent, with a slight difference in accuracy between multiple training sessions. The models proposed in [21] are trained using k-fold validation, and the outputs of these models were combined. The above procedure of combining the outputs of different models is called ensemble learning, which improves the accuracy of fire detection. Ensemble learning combines the output of all models by a voting mechanism. The class predicted by most of the models is considered the output class. This mechanism is depicted in Figure 3. If any model predicts the wrong output class, the rest of the models may predict the correct output class, improving the overall accuracy of detection. The classification output is influenced by the different models plugged for prediction. The ensemble output is created only on these models.

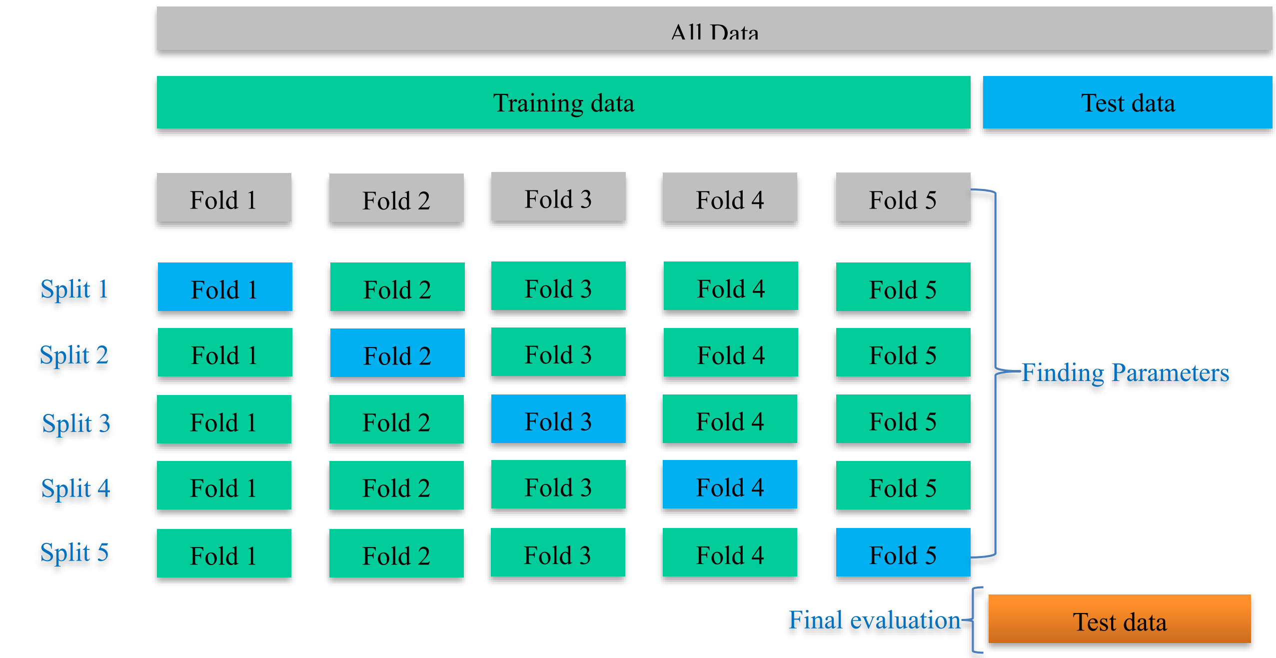

Fig. 2 Illustration of 5-fold cross validation

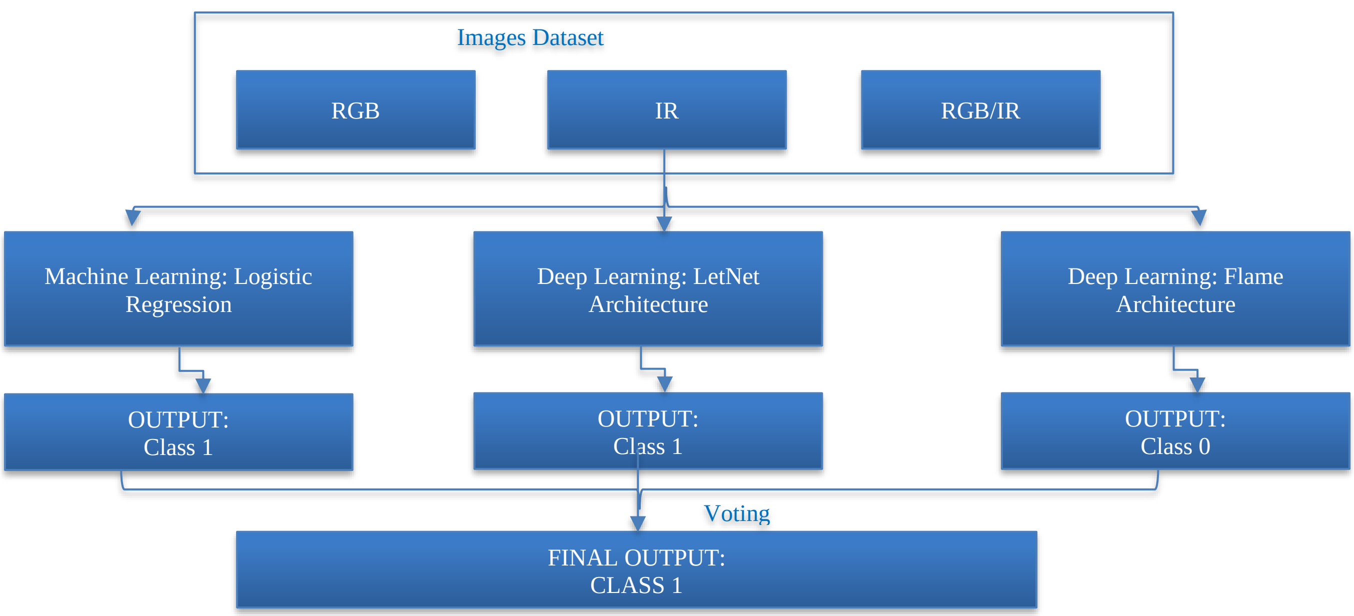

Fig. 3 Working of ensemble

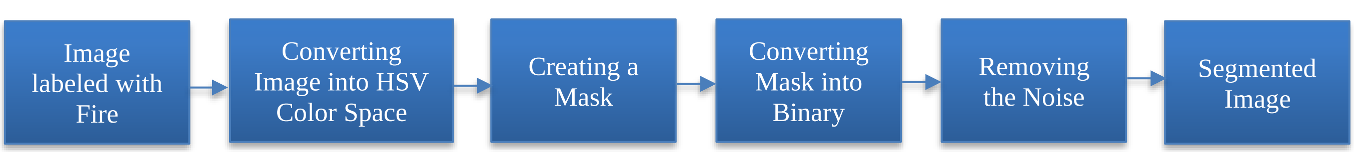

Fig. 4 Block diagram for segmentation

Table 1. Metrics of evaluation

Model Name	Mode	Accuracy	Precision	F1 Score	Recall
Flame	RGB	97%	0.98	0.98	0.98
Flame	IR	86%	0.86	0.86	0.86
Flame	Early Fusion	97.66%	0.98	0.98	0.98
Flame	Late Fusion	98%	0.98	0.98	0.98
LeNet5	RGB	98.90%	0.95	0.95	0.95
LeNet5	IR	97.70%	0.95	0.95	0.95
LeNet5	Early Fusion	98.96%	0.99	0.99	0.99
LeNet5	Late Fusion	98.30%	0.95	0.95	0.95
Logistic	RGB	93.05%	0.99	0.99	0.99
Logistic	IR	95.23%	0.99	0.99	0.99
Logistic	Early Fusion	94.86%	0.99	0.99	0.99
MobilenetV2	RGB	84.09%	0.85	0.85	0.85
MobilenetV2	IR	82%	0.83	0.84	0.83

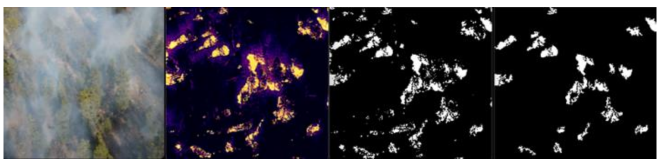

Fig. 5 Images in different stages of segmentation

3.3 Segmentation

The IR images predicted to have fire are only advanced to the next section of IR segmentation. IR segmentation is done using HSV color segmentation. HSV indicates hue, saturation, and value. This segmentation method in image processing partitions the image into regions based on the HSV color space. This approach separates pixels based on hue, saturation, and value components, allowing thresholds and ranges for each component to be set. This approach was considered a good fit for this task compared to other segmentation techniques like Otsu and adaptive thresholding.

By observing the RGB and IR images of the same scene, the HSV colour space for fire was developed. The colour of fire in the IR image appears predominantly as orange red. The work proposes HSV colour space for fire detection, which was tuned with multiple hit and trial experiments.

The block diagram for segmentation is depicted in Figure 4, which gives a glimpse of the HSV segmentation procedure. The images in the dataset are in RGB format and consist of three channels. The IR images predicted to have fire are taken and converted into a single channel. The image consisting of 3 channels is converted into a single channel now. Then, the colour space mask is applied to the image, resulting in a binary image. This binary image has a pixel value of 1, where the IR image pixel matches the colour space of the mask, and 0 in other cases. IR images contain some noise, which causes the output binary image also to contain noise. This noise is eliminated by knowing the area of each segment. Each segment whose area is less than the threshold is considered as noise and removed. A visual depiction of the image is shown in Figure 5.

4. Experimental Analysis

This section provides detailed information about the datasets used for the experiments, the performance metrics used and the quantitative analysis and interpretation.

4.1. Datasets

The dataset [21] utilised in this study was collected through an Unmanned Aerial Vehicle (UAV) equipped with two sensors: one capturing RGB video using the M2EA RGB Cam and the other capturing IR video with the M2EA IR Cam. The dataset comprises 7 pairs of synchronised RGB and IR video sequences, with both videos operating at a frame rate of 30 Frames Per Second (FPS).

The RGB videos exhibit varying resolutions, including 3840x2160 and 1920x1080, while the IR videos maintain a standardised resolution of 640x512. This dual-modal dataset offers a valuable resource for exploring the synergy between RGB and IR imagery, with applications ranging from fire detection to dynamic environmental analysis.

4.2. Performance Analysis

The experimental results of the models are kept in Table 1. The metrics used are accuracy, precision, F1-score, and recall. The work experiments on using the k-fold technique to train the model mentioned in [21] and using an ensemble to classify the images. Additionally, the work uses segmentation techniques to detect the position of the fire.

Firstly, the usage of the K-fold technique in the training of the machine learning models results in a more consistent trained model and also gives us a viable solution for situations where the dataset is minimal, as these small datasets can be used for training the machine learning models whereas previously would need a more extensive dataset.

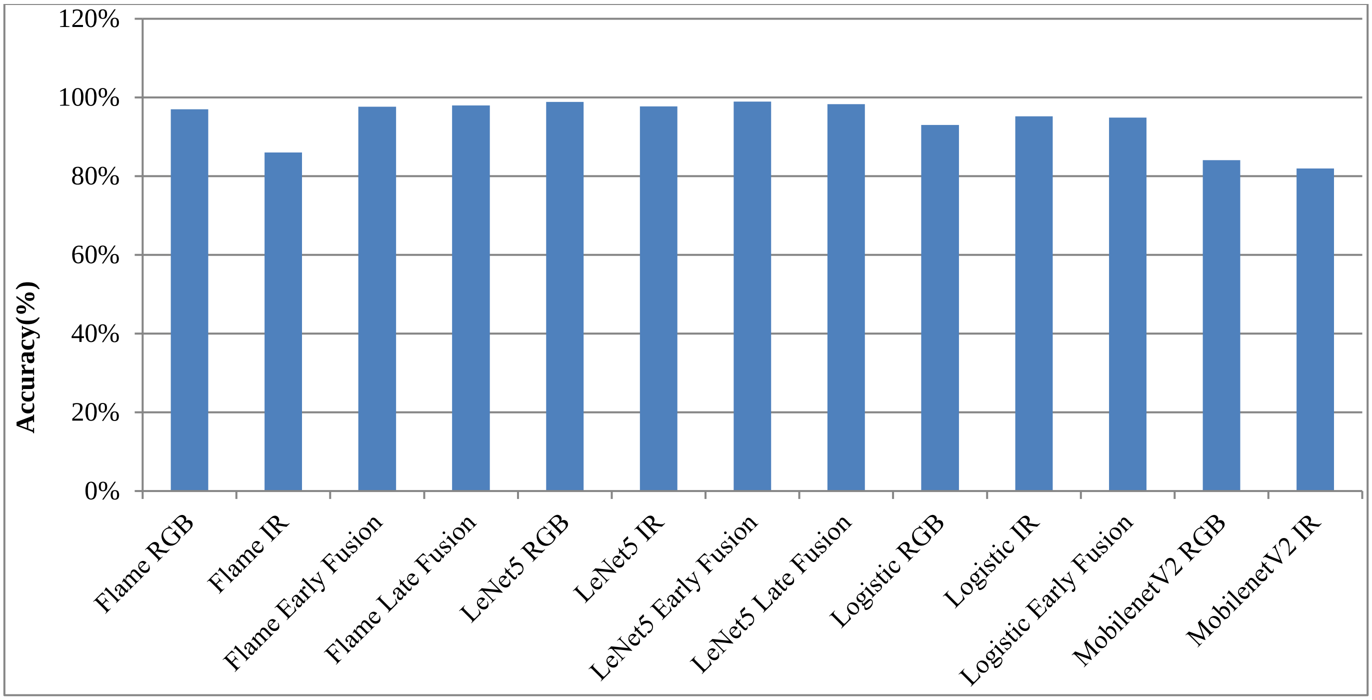

Fig. 6 Model accuracy comparison

The K-fold technique provides a notable enhancement in performance metrics like F1 score, precision, recall, and a slight increase in accuracy, indicating the models' improved capability to predict outcomes accurately. Secondly, ensemble learning ensures robustness even in scenarios where some of the models fail; the majority will almost always give the correct result due to the high accuracies of the models. Overall, the combination of ensemble learning and K-fold cross-validation proves instrumental in bolstering the reliability and effectiveness of machine learning models. Thirdly, segmentation using the HSV technique correctly segments the image into fire. This approach worked correctly when the image contained fire. Therefore, all the images predicted to have been fired by the ensemble are only advanced to the segmentation stage. The HSV segmentation worked better than the other segmentation techniques, like Otsu thresholding and adaptive thresholding methods. Finally, the approach of ensemble learning and the use of K-fold cross-validation to train the models for fire detection has improved fire detection. The accuracy increased by almost 5% overall, and the ensemble improved it further. The results are compared with [21], and there is a significant increase in all the evaluation metrics. LeNet5 has the highest accuracy among all other models. Figure 6 indicates a comparison of the accuracies of different models. The graph says that all the models have outstanding accuracy.

5. Real World Application and Deployment

There might be some challenges in deploying the proposed method of using UAVs/ drones for monitoring and detecting forest fires as most of the forests come under government jurisdiction; it tends to have very strict aviation laws, which might hinder the deployment of such a process.
On the other hand, having a small fleet of monitoring drones across vulnerable forest areas would dramatically increase the chances of putting out fire, as the drones serve another excellent purpose: geolocation. A modern drone is equipped with many sensors apart from cameras, such as temperature sensors, accelerometers, and gyroscopes, and it has built-in connectivity to the internet.

Having this much real-time information enables having a robust system in which there can be a routine in which the fleet of drones can take turns monitoring, alerting and notifying a central system that processes the feed of video data and identifies the coordinates of the forest fire. A very interesting problem arises in the allocation/management of the drones, which must be charged/maintained regularly. This compels a job sequencing and maintenance solution that could also be automated with clever algorithms. The data being recorded could also act as a data reserve that stores historical data on the forest, which can be used for other purposes and as data for other systems. The areas of interest the models provide are very informative maps of the forest, which enable targeted strategies for subduing forest fires.

6. Conclusion and Future Scope

The study presents an innovative method to detect forest fires at their inception, harnessing the combined potential of RGB and IR imagery. The deep learning LeNet5 has the highest accuracy in classifying the image into 3 classes, with

98.96%. Our pioneering approach revolves around the meticulous refinement of existing state-of-the-art models, achieved through the rigorous implementation of a k-fold cross-validation training regimen. This meticulous process culminates in the remarkable attainment of cent percent accuracy, facilitated by the strategic utilisation of ensemble learning techniques. Furthermore, it delineates a specialised methodology to identify fire-affected regions within IR images, particularly in optimising accuracy within predicted fire zones. The effectiveness of ensemble learning is underscored by its ability to consolidate outputs from a diverse array of models, thus effectively mitigating the inherent risks associated with individual model failures and substantially bolstering overall reliability. Moreover, our research opens avenues for real-time analytics, enabling the dynamic evaluation of fire characteristics such as flame contours. This comprehensive analysis offers invaluable insights into the magnitude of devastation and the intricate dynamics of fire propagation. Additionally, the study proposes the seamless integration of our detection system with an outbound messaging platform, which expedites emergency response efforts by promptly alerting response teams and providing comprehensive situational assessments.

This integration is a strategic asset, enhancing informed decision-making in critical situations. While competitive performance was observed, SSRN exhibited comparatively slower processing. ResNet and DPyResNet demonstrated distinct advantages, particularly in interpretability and the effective use of contextual information. These findings highlight the strengths and limitations of each approach, offering guidance for choosing the most appropriate method for various hyperspectral image classification tasks. Future efforts will focus on exploring ensemble learning strategies to further advance the classification performance of hyperspectral image analysis techniques.

Acknowledgments

The authors would like to acknowledge Vasavi College of Engineering, Hyderabad, for supporting the work.

References

[1] Richard Skinner et al., “A Literature Review on The Impact of Wildfires on Emergency Departments: Enhancing Disaster Preparedness,” *Prehospital and Disaster Medicine*, vol. 37, no. 5, pp. 657-664, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Yongsong Huang et al., “Infrared Image Super-Resolution: Systematic Review, and Future Trends,” *arXiv*, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ammar Mohammed, and Rania Kora, “A Comprehensive Review on Ensemble Deep Learning: Opportunities and Challenges,” *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 2, pp. 757-774, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Nida M. Zaitoun, and Musbah J. Aqel, “Survey on Image Segmentation Techniques,” *Procedia Computer Science*, vol. 65, pp. 797-806, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[5] S. Prince Mary et al., “A Survey on Image Segmentation Using Deep Learning,” *Journal of Physics: Conference Series*, vol. 1712. no. 1, pp. 1-10, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[6] D. Hema, and S. Kannan, “Interactive Color Image Segmentation Using HSV Color Space,” *Science and Technology Journal*, vol. 7, no. 1, pp. 37-41, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] Herbert Bousack et al., “Towards Improved Airborne Fire Detection Systems Using Beetle Inspired Infrared Detection and Fire Searching Strategies,” *Micromachines*, vol. 6, no. 6, pp. 718-746, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[8] P. Raghavendra Reddy, and P. Kalyanasundaram, “Novel Detection of Forest Fire Using Temperature and Carbon Dioxide Sensors with Improved Accuracy in Comparison Between Two Different Zones,” *Proceedings 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, London, United Kingdom, pp. 524-527, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[9] Alexander A. Khamukhin, and Silvano Bertoldo, “Spectral Analysis of Forest Fire Noise for Early Detection Using Wireless Sensor Networks,” *Proceedings 2016 International Siberian Conference on Control and Communications (SIBCON)*, Moscow, Russia, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[10] Abdelmalek Bouguettaya et al., “A Review on Early Wildfire Detection from Unmanned Aerial Vehicles Using Deep Learning-Based Computer Vision Algorithms,” *Signal Processing*, vol. 190, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[11] Leonardo Millan-Garcia et al., “An Early Fire Detection Algorithm Using IP Cameras,” *Sensors*, vol. 12, no. 5, pp. 5670-5686, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[12] Bryce Hopkins et al., “FLAME 2: Fire Detection and Modeling: Aerial Multi-Spectral Image Dataset,” *IEEE DataPort*, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[13] Norsuzila Ya'acob et al., “Image Processing based Forest Fire Detection Using Infrared Camera,” *Proceedings 2nd International Conference on Space Weather and Satellite Application (ICeSSAT)*, Selangor, Malaysia, vol. 1768, no. 1, pp. 1-13, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[14] Seyd Teymoor Seydi et al., “Fire-Net: A Deep Learning Framework for Active Forest Fire Detection,” *Journal of Sensors*, vol. 2022, no. 1, pp. 1-14, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[15] Arpit Jadon, Akshay Varshney, and Mohammad Samar Ansari, "Low-Complexity High-Performance Deep Learning Model for Real-Time Low-Cost Embedded Fire Detection Systems," *Procedia Computer Science*, vol. 171, pp. 418-426, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[16] S. NASA, "Fire Information for Resource Management System," *FIRMS*, 2020. [Google Scholar] [Publisher Link]

[17] Vladimir Sergeevich Bochkov, and Liliya Yurievna Kataeva, "wUUNET: Advanced Fully Convolutional Neural Network for Multiclass Fire Segmentation," *Symmetry*, vol. 13, no. 1, pp. 1-18, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[18] Pu Li, and Wangda Zhao, "Image Fire Detection Algorithms Based on Convolutional Neural Networks," *Case Studies in Thermal Engineering*, vol. 19, pp. 1-11, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[19] MinJi Park, and Byoung Chul Ko, "Two-Step Real-Time Night-Time Fire Detection in an Urban Environment Using Static ELASTIC-YOLOv3 and Temporal Fire-Tube," *Sensors*, vol. 20, no. 8, pp. 1-17, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[20] Wei Liu et al., "SSD: Single Shot Multibox Detector," *Computer Vision–ECCV 2016: 14th European Conference*, Amsterdam, The Netherlands, vol. 9905, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[21] Xiwen Chen et al., "Wildland Fire Detection and Monitoring Using a Drone-Collected RGB/IR Image Dataset," *IEEE Access*, vol. 10, pp. 121301-121317, 2022. [CrossRef] [Google Scholar] [Publisher Link]