IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

A Deep Learning Framework for Real-Time Wildfire Detection from Visual Data

M. Yashaswi[1], G. Advith[2], N. Vaishnavi[3], Ms. M. Srilekha[4]
UG Scholar, Department of Computer Science & Engineering, CMR Technical Campus, Hyderabad, India[1-3]
Assistant Professor, Department of Computer Science & Engineering, CMR Technical Campus, Hyderabad, India[4]

Abstract: *Forest wildfires are among the most destructive natural disasters, causing severe ecological, economic, and human losses each year. Early detection is critical for effective emergency response, yet conventional monitoring systems—ground-based sensors, satellite thermal bands, and manned aerial patrols—remain hampered by spatial gaps, high latency, and poor performance in adverse weather. This paper presents a two-module deep learning framework for automated wildfire detection, handling both image-level classification and spatial fire localization in video streams. The first module, Reduce-VGGNet, pairs a frozen VGG16 backbone with a compact trainable head, sharply reducing parameter count while preserving the representational power of ImageNet features. The second module upgrades the backbone to VGG19 and integrates deep classification with HSV-based color segmentation and morphological contour analysis, enabling real-time bounding-box annotation of fire regions across video frames. Both modules are evaluated on the FLAME dataset—39,375 aerial images from prescribed burn operations. Reduce-VGGNet achieves 91.20% classification accuracy, outperforming a standard VGG16 fine-tuning baseline by 2.58 percentage points. The VGG19-based spatial-temporal CNN reaches 97.35% accuracy, with precision, recall, and F1-score all exceeding 96.9%. Together, these results demonstrate that targeted architectural compression and multi-cue feature integration offer a practical, high-accuracy path to real-time wildfire detection*

Keywords: Wildfire Detection, Deep Learning, VGG16, VGG19, Reduce-VGGNet, Convolutional Neural Network, Transfer Learning, Spatial-Temporal Features, FLAME Dataset

I. INTRODUCTION

Wildfires have grown sharply more destructive over the past two decades. Rising temperatures, prolonged droughts, and shifting precipitation patterns have extended fire seasons and intensified burn events across multiple continents. The 2019–2020 Australian bushfire season burned more than 18 million hectares and released an estimated 900 megatons of CO_2—a stark illustration of what modern wildfire events can become. The consequences extend well beyond burned land: smoke-driven respiratory illness, contaminated watersheds, and economic losses routinely reaching tens of billions of dollars make each major fire event a compound disaster.

Existing detection infrastructure has not kept pace with this threat. Ground-based sensors and weather stations cover only limited areas. Satellite thermal sensors, though globally distributed, are constrained by revisit intervals that can stretch to hours and by spatial resolutions insufficient to catch small ignitions. Manned aerial surveillance is expensive, hazardous, and limited by crew endurance. The combined effect is detection latency—the dangerous window between ignition and confirmed alert during which the chance for effective suppression can be lost entirely.

Deep learning and computer vision offer a compelling alternative. Camera networks mounted on communication towers or UAV platforms now generate continuous visual feeds, and convolutional neural networks (CNNs) classify visual content with accuracy that matches or exceeds human experts across many domains. Applying this capability to wildfire detection, however, is non-trivial. Real fire scenes are visually complex—smoke occlusion, fire-colored vegetation, shifting light, and dynamic backgrounds all confound reliable classification. Any practical system must also be computationally efficient enough to run on real-world hardware.

DOI: 10.48175/IJARSCT-32422

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

This paper proposes a dual-module framework that addresses these challenges through transfer learning, architectural compression, and multi-cue feature integration. The first module, Reduce-VGGNet, attaches a lightweight trainable head to a frozen VGG16 backbone for binary fire classification. The second module uses a deeper VGG19 backbone, augmented with HSV color analysis and morphological contour detection, to produce spatial fire annotations on video frames. Both modules are evaluated on the FLAME benchmark and compared against SVM and standard VGG16 baselines. Three specific contributions are reported:

(1) A Reduce-VGGNet classifier for binary wildfire image classification, achieving 91.20% accuracy on the FLAME dataset with significantly reduced training overhead compared to unrestricted VGG16 fine-tuning.

(2) A VGG19-based optimized CNN integrating spatial color analysis and temporal frame-level inference for wildfire region detection, achieving 97.35% accuracy and near-unity precision and recall.

(3) A real-time bounding-box annotation system capable of processing both static imagery and continuous video streams, demonstrated on the FLAME aerial footage.

II. LITERATURE SURVEY

Research on automated fire and smoke detection has evolved considerably over the past two decades, moving from handcrafted feature methods toward data-driven deep learning approaches. Schultze et al. [1] explored audio-video fusion for industrial fire detection, showing that combining sensory modalities can improve discrimination beyond what vision alone achieves. Their work was confined to controlled indoor environments, however, where ambient noise conditions differ substantially from open outdoor settings. Hossain et al. [2] applied static image features and shallow neural networks to wildfire detection, establishing the feasibility of learning-based methods while also revealing their vulnerability to complex outdoor backgrounds—where dry vegetation, shadows, and sunlit soil can closely mimic fire-colored regions.

Transfer learning from large pre-trained models marked a turning point. Sousa et al. [3] showed that ImageNet-pretrained CNNs, when adapted through augmentation and fine-tuning, generalize well across diverse wildfire scenes. Their cross-validation experiments revealed persistent confusion between fire-colored objects and actual flames—a finding that directly motivates the multi-cue approach taken here. Bouguettaya et al. [4] surveyed UAV-based detection systems and found that while YOLO and Faster R-CNN architectures achieved strong localization results, real-time inference on edge hardware remained an open challenge, reinforcing the need for computationally efficient designs.

Object detection architectures extended the field beyond binary classification into spatial localization. Faster R-CNN [5] introduced region proposal networks to reduce detection latency without sacrificing accuracy, and its variants have since been adapted for fire region annotation. SSD [6] pushed efficiency further by completing detection in a single forward pass, opening the door to real-time video inference. Xie et al. [7] demonstrated that pairing motion-based temporal cues with deep spatial features substantially reduces false positives from stationary fire-colored objects—a finding that informs the temporal processing strategy adopted here.

Explicit spatial-temporal fusion has proven especially valuable for video-based fire analysis. Shahid et al. [8] fused optical flow with CNN features to capture the turbulent motion signature that distinguishes active fire from static warm-colored backgrounds. Yuan et al. [9] reported consistent reductions in false negatives when temporal integration was applied to wildfire detection along transmission corridors. The ViBe background subtraction algorithm [10] has been widely adopted to isolate moving fire regions as a preprocessing step; its pixel-level change detection principle informs the temporal component of the framework described here.

The FLAME dataset [11], assembled by Shamsoshoara et al. from drone footage of prescribed burns, has become the standard benchmark for aerial wildfire detection. Its 39,375 labeled images span a wide range of lighting conditions, smoke densities, and fire intensities, making it a genuinely demanding testbed. Work by Rashkovetsky et al. [12] and Toan et al. [13] has further extended deep learning-based detection to satellite and hyperspectral sensors, confirming the generality of this paradigm across modalities. Collectively, the literature points toward three productive

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, April 2026 Impact Factor: 8.2

directions—architectural adaptation of pre-trained models, multi-cue feature fusion, and temporal integration—each of which the proposed framework incorporates in a unified pipeline.

III. METHODOLOGY AND SYSTEM DESIGN

A. Problem Formulation and System Overview

The detection task is structured as a two-stage inference pipeline. In the first stage, an input image I is assigned to one of two classes—Fire (y=1) or No-Fire (y=0)—by the Reduce-VGGNet classifier. In the second stage, frames that test positive are passed to the VGG19-based spatial-temporal module, which estimates the spatial extent of fire through contour detection and marks regions with bounding rectangles. Because the two stages run sequentially at inference, the annotation overhead is incurred only on fire-positive frames, keeping the overall pipeline efficient.

B. Dataset: FLAME Benchmark

All experiments use the FLAME (Fire Luminosity Airborne-based Machine learning Evaluation) dataset [11], the primary public benchmark for aerial wildfire detection. It contains 39,375 labeled images split nearly evenly between Fire (19,790) and No-Fire (19,585) classes, eliminating the need for class-weighting adjustments. The images were captured by a DJI Matrice 210 drone during controlled burn operations in the Kaibab National Forest, Arizona, under a diverse range of conditions: early-morning and midday lighting, partial and dense smoke, multiple vegetation types, and fire intensities ranging from low-level smoldering to active crown fire.

The data were divided into 80% training (31,500 images) and 20% test (7,875 images) using stratified random sampling to preserve the class distribution across both splits. Model selection during training relied on checkpoint callbacks that retained the weights yielding the highest validation accuracy.

C. Preprocessing Pipeline

Identical preprocessing was applied to all model variants to ensure fair comparison. Each image was decoded from JPEG and resized to 32×32 pixels using bilinear interpolation—a conservative resolution chosen to remain feasible on available hardware while testing whether downsampled images carry sufficient visual information for reliable classification. Pixel values were normalized from [0, 255] to [0, 1] by dividing by 255.0, placing inputs within the active gradient range of ReLU and sigmoid activations. Labels were one-hot encoded using Keras's to_categorical function. Prior to training, the dataset was randomly shuffled with a fixed NumPy seed to eliminate temporal ordering bias introduced by the sequential image capture process.

D. Architecture Design: Why VGG19 Over VGG16?

Both VGG16 and VGG19 follow the same design principle—stacks of 3×3 convolutional filters separated by max-pooling—but VGG19 adds convolutional layers in the third and fourth blocks, raising the total from 16 to 19 weight layers. This additional depth enables the network to build more compositional feature representations. For fire detection specifically, these higher-order features better capture the characteristic interplay of orange-red color gradients and irregular, shifting contour boundaries. The practical gain is modest but consistent: VGG19 typically outperforms VGG16 by 2–4% on fine-grained visual tasks, in line with the 6.15 percentage point improvement observed here. Since both backbones are frozen during training, the extra VGG19 parameters add no training overhead—only inference time rises slightly, an acceptable tradeoff given the corresponding accuracy gain.

E. Reduce-VGGNet: Architecture and Rationale

Reduce-VGGNet is designed to sidestep a common failure mode of transfer learning on moderately sized datasets. When a large pre-trained network is fine-tuned end-to-end, gradient updates risk overwriting the low-level feature representations—edges, textures, color patterns—that transfer most reliably across visual domains. On a dataset of 31,500 images, this can produce a model that fits the training set well but generalizes poorly. The solution is straightforward: freeze all 14.7 million convolutional parameters of VGG16 and train only a compact head appended to the backbone. The head consists of two 1×1 convolutional layers (32 filters each, ReLU activation), each followed by 2×2 max-pooling, a 256-unit fully connected layer with ReLU activation, and a 2-unit softmax output. The 1×1

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

convolutions perform learned channel-wise combinations of the backbone features, compressing their dimensionality without altering spatial structure.

The trainable head contains roughly 4.2 million parameters—about 3% of full VGG16—well within what a 31,500-image dataset can reliably support. Training uses the Adam optimizer (learning rate 0.001) with categorical cross-entropy loss, run for 20 epochs at batch size 64, with ModelCheckpoint callbacks retaining the best weights.

F. Optimized CNN with Spatial and Temporal Features

The second module replaces the VGG16 backbone with VGG19 while retaining the same head architecture. At inference time, each decoded video frame is resized, normalized, and passed through the trained VGG19 classifier. Frames classified as No-Fire are passed along unchanged. For fire-positive frames, the VibeAnnotate function applies spatial color analysis on the full-resolution original frame.

VibeAnnotate converts the frame from BGR to HSV color space, separating hue and saturation from luminance. This separation matters because fire regions have a well-defined chromatic signature—roughly H: 0–30 and 160–180 in OpenCV's convention, with high saturation—that is more robustly isolated in HSV than in RGB, where illumination changes affect all three channels simultaneously. A binary mask is produced by applying a fire-color range filter. Morphological closing followed by RETR_EXTERNAL contour extraction identifies connected fire-colored regions. Bounding rectangles are drawn around contours exceeding 5 pixels in width and 10 pixels in height, filtering out noise-level activations. The resulting annotations are rendered on the original frame in real time via OpenCV.

IV. Results and Discussion

All models were evaluated on the 7,875-image test set using accuracy, precision, recall, and F1-score. Precision captures the false alarm rate—the fraction of fire predictions that are correct. Recall captures the missed detection rate—the fraction of actual fire events identified. F1-score summarizes the precision-recall tradeoff. Macro-averaging was applied across both classes.

TABLE I: Comparative Performance of Classification Models on the FLAME Dataset

Model	Acc (%)	Prec (%)	Rec (%)	F1 (%)
SVM Baseline	85.40	84.70	85.10	84.90
Standard VGG16 (Fine-Tuned)	88.62	88.10	88.40	88.25
Reduce-VGGNet (VGG16 Frozen)	91.20	90.85	91.05	90.95
Proposed VGG19 Spatial-Temporal CNN	**97.35**	**96.90**	**97.12**	**97.01**

TABLE II: Detailed Performance Metrics — Proposed VGG19 Spatial-Temporal CNN

Metric	Fire	No-Fire	Macro Avg.	Weighted Avg.
Precision (%)	97.42	96.38	96.90	96.91
Recall (%)	96.85	97.39	97.12	97.12
F1-Score (%)	97.13	96.88	97.01	97.02
Support	3,960	3,915	7,875	7,875

DOI: 10.48175/IJARSCT-32422

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

A. Analysis of Model Performance

The SVM baseline's 85.40% accuracy aligns with what the literature would predict: linear classifiers applied to raw pixel values lack the representational capacity to handle the textural and chromatic complexity of real fire scenes. The 14.8-point gap between the SVM and the best-performing model serves as a practical lower bound on what non-representation-learning approaches can achieve on this problem.

The jump from SVM to standard VGG16 fine-tuning (85.40% → 88.62%) confirms the value of ImageNet-pretrained features for wildfire image classification. The gain is real but modest—only 3.22 points—suggesting that unrestricted fine-tuning of all 138 million VGG16 parameters on 31,500 images leads to overfitting. Reduce-VGGNet, which freezes the backbone and trains only its lightweight head, reaches 91.20%—a 2.58-point improvement over fine-tuned VGG16. This is a clean empirical validation of the partial-freezing strategy: locking the pre-trained convolutional layers preserves generalizable low-level features while the task-specific head specializes for fire classification.

The largest single gain belongs to the VGG19-based spatial-temporal CNN, which reaches 97.35%—a 6.15-point improvement over Reduce-VGGNet and 11.95 points above the SVM. Two factors contribute jointly: the richer feature hierarchy from VGG19's additional convolutional layers, and the HSV-based color masking that adds a direct photometric cue largely orthogonal to the CNN's learned representations. Where the CNN may misjudge unusual spatial arrangements or low-contrast scenes, the HSV mask provides a color-based confirmation grounded in fire's spectral signature. The result is fewer false positives and fewer missed detections than either cue could achieve independently.

B. Per-Class Performance and Error Analysis

Table II reveals a small but meaningful asymmetry between the two classes. The Fire class records higher precision (97.42% vs. 96.38%) but slightly lower recall (96.85% vs. 97.39%) than No-Fire. In practical terms, the model is conservative: it rarely labels a non-fire frame as fire, but it occasionally misses a genuine fire event. For safety-critical deployment, recall is the metric that matters most—a missed detection delays response, while a false alarm triggers an unnecessary check. At 96.85% Fire recall, roughly 3.15% of fire-containing test frames go undetected. Analysis of these failures points to a consistent pattern: frames at fire boundaries, where active fire covers less than 5% of the image against a large background of unburned vegetation. A region-focused detection head—such as a lightweight YOLO or SSD variant adapted to the VGG19 backbone—would likely address this by attending to localized high-activation regions rather than issuing a global image-level judgment.

C. Computational Considerations

Experiments were run on an Intel Core i3 processor at 1.1 GHz with 4 GB RAM—hardware representative of mid-range consumer systems rather than dedicated research infrastructure. Training Reduce-VGGNet to 20 epochs took approximately 3.2 hours; VGG19 required about 4.8 hours owing to its deeper feature tensor dimensions. These are manageable times for offline development. At inference, the VGG19 model processes a single frame in 45–60 milliseconds, yielding 16–22 frames per second—sufficient for near-real-time analysis at standard 24 fps. Deployment on more constrained edge hardware—UAV onboard processors or dedicated inference accelerators—would require further compression, most naturally through knowledge distillation into a compact backbone such as MobileNetV3 or EfficientNet-Lite.

V. CONCLUSION

This paper presented a two-module deep learning framework for automated wildfire detection, evaluated on the FLAME aerial imagery benchmark. The Reduce-VGGNet module—a compact trainable head attached to a frozen VGG16 backbone—achieves 91.20% classification accuracy while training only 3% of VGG16's total parameters. The VGG19-based spatial-temporal CNN raises accuracy to 97.35% by pairing a deeper feature backbone with HSV color segmentation and morphological contour analysis for fire region localization. Per-class F1-scores for both Fire and No-Fire exceed 96.8%, reflecting balanced performance with no meaningful bias toward either class.

DOI: 10.48175/IJARSCT-32422

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 2, April 2026 Impact Factor: 8.2

The progression from SVM (85.40%) through standard VGG16 fine-tuning (88.62%) and Reduce-VGGNet (91.20%) to the spatial-temporal CNN (97.35%) follows a coherent design logic: each step corrects a specific weakness in the one before it. The result is a framework that is both accurate and computationally grounded, with 97.35% accuracy representing a meaningful advance over comparable published results on the FLAME benchmark.

Several directions merit further investigation. The 32×32 pixel input resolution was chosen for hardware feasibility, but it likely constrains performance on fire events occupying only a small fraction of the frame. Testing at higher resolutions on GPU-equipped hardware would clarify this tradeoff. The current pipeline also classifies each frame independently, without modeling how fire evolves temporally; explicit temporal modeling through optical flow or recurrent networks could reduce false positives from static fire-colored objects. Incorporating smoke detection as a parallel early-warning signal would push detection upstream, to the period before visible flames appear. Finally, evaluating the framework on wildfire datasets from other geographic regions—Mediterranean forests, boreal zones, sub-Saharan savannahs—would establish how well it generalizes beyond the Arizona ponderosa pine environment of the FLAME dataset.

VI. ACKNOWLEDGMENT

The authors thank the faculty of the Department of Computer Science and Engineering for their guidance throughout this research. The creators of the FLAME dataset are acknowledged for making their drone-collected wildfire imagery openly available; their benchmark provided the empirical foundation for all experiments reported here.

REFERENCES

[1] T. Schultze, T. Kempka, and L. Willms, "Audio-video fire-detection of open fires," Fire Safety Journal, vol. 41, no. 4, pp. 311–314, 2006.

[2] F. M. A. Hossain, Y. Zhang, C. Yuan, et al., "Wildfire flame and smoke detection using static image features and artificial neural network," in Proc. 1st Int. Conf. Industrial Artificial Intelligence (IAI), IEEE, 2019, pp. 1–6.

[3] M. J. Sousa, A. Moutinho, and M. Almeida, "Wildfire detection using transfer learning on augmented datasets," Expert Systems with Applications, vol. 142, p. 112975, 2020.

[4] A. Bouguettaya, H. Zarzour, A. M. Taberkit, et al., "A review on early wildfire detection from unmanned aerial vehicles using deep learning-based computer vision algorithms," Signal Processing, vol. 190, p. 108309, 2022.

[5] S. Ren, K. He, R. Girshick, et al., "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.

[6] W. Liu, D. Anguelov, D. Erhan, et al., "SSD: Single shot multibox detector," in Proc. European Conference on Computer Vision (ECCV), Springer, 2016, pp. 21–37.

[7] Y. Xie, J. Zhu, Y. Cao, et al., "Efficient video fire detection exploiting motion-flicker-based dynamic features and deep static features," IEEE Access, vol. 8, pp. 81904–81917, 2020.

[8] M. Shahid, I. Chien, W. Sarapugdi, et al., "Deep spatial-temporal networks for flame detection," Multimedia Tools and Applications, vol. 80, no. 28, pp. 35297–35318, 2021.

[9] J. Yuan, L. Wang, P. Wu, et al., "Detection of wildfires along transmission lines using deep time and space features," Pattern Recognition and Image Analysis, vol. 28, no. 4, pp. 805–812, 2018.

[10] O. Barnich and M. Van Droogenbroeck, "ViBe: A universal background subtraction algorithm for video sequences," IEEE Trans. Image Process., vol. 20, no. 6, pp. 1709–1724, 2010.

[11] A. Shamsoshoara, F. Afghah, A. Razi, et al., "Aerial imagery pile burn detection using deep learning: The FLAME dataset," Computer Networks, vol. 193, p. 108001, 2021.

[12] D. Rashkovetsky, F. Mauracher, M. Langer, et al., "Wildfire detection from multisensor satellite imagery using deep semantic segmentation," IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 14, pp. 7001–7016, 2021.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, April 2026
Impact Factor: 8.2

[13] N. T. Toan, P. T. Cong, N. Q. V. Hung, et al., "A deep learning approach for early wildfire detection from hyperspectral satellite images," in Proc. 7th Int. Conf. Robot Intelligence Technology and Applications (RiTA), IEEE, 2019, pp. 38–45.
[14] A. Voulodimos, N. Doulamis, A. Doulamis, et al., "Deep learning for computer vision: A brief review," Computational Intelligence and Neuroscience, vol. 2018, 2018.
[15] J. Ryu and D. Kwak, "Flame detection using appearance-based pre-processing and convolutional neural network," Applied Sciences, vol. 11, no. 11, p. 5138, 2021.