International Journal of
Computational Learning and Intelligence
Vol. 05 | Issue 03 | 2026

RESEARCH ARTICLE OPEN ACCESS

Multi-Platform Forest Fire Detection Using Deep Learning and IoT: A Review

Vinay Kumar . Sourav . Nikhil Kumar . Pradeep Chouksey . Parveen Sadotra . Mayank Chopra

Department of Computer Science and Informatics,
Central University of Himachal Pradesh, Shahpur Parisar, India

DOI: **10.5281/zenodo.20056355**
Received: 29 April 2026 / Revised: 05 May 2026 / Accepted: 06 May 2026

*Corresponding author: *vinaykumarbharwal@gmail.com*

Abstract — Forest fires have severe impacts on ecosystems and property, but current detection methods continue to suffer from detection latency, coverage and environmental limitations. In this paper, we review the state of the art in integrated forest fire detection systems that leverage deep learning, Internet of Things (IoT) sensor networks, unmanned aerial vehicle (UAV) monitoring, and satellite remote sensing. Over two dozen recent works are reviewed to discuss object detection algorithms, hybrid deep learning models, sensor fusion techniques, satellite remote sensing, and land use/land cover (LULC) change analyses for predictive fire risk mapping. Key research challenges are identified in the areas of integration, data, environmental adaptability, efficiency, and the use of contextual information. Popular benchmark datasets, performance metrics and system characteristics are also presented. Based on this review, a visionary research proposal is provided detailing the design and approach for developing a holistic multi-platform detection system to achieve detection within 5 minutes with an accuracy of more than 95% and false alarm rate of less than 5% in different ecosystems.

Keywords—forest fire detection, deep learning, IoT, UAV, sensor fusion, edge computing, satellite imagery, LULC analysis, CNN-ViT hybrid, YOLO.

I. INTRODUCTION

Wildfires are among the world's most devastating natural hazards, impacting more than 350 million hectares of vegetation, emitting two to three petagrams of carbon and incurring more than $100 billion in damages every year [1]. The increasing incidence of wildfires due to climate change, deforestation and human activities calls for advanced detection techniques that can detect fires in their early stages when their spread can be managed. Existing techniques have known pitfalls. Land-based patrols and watchtowers are hampered by topographical obstacles, visibility issues, and cost. Sun-synchronous satellites provide global coverage but revisit rates of 12-24 hours are too slow for fast-moving fires [2]. Such limitations drive the exploration of smart, multi-platform fire detection systems. The advent of deep learning, low-cost IoT sensors, unmanned aerial vehicles (UAVs) and geostationary satellites has revolutionized detection [3]. Current fire detection approaches use hierarchical designs that combine multiple sensor platforms (IoT sensors, UAVs, satellites, fixed cameras) with distributed (edge, fog and cloud) computing to achieve optimal detection speed, accuracy and scalability. While each technology has shown excellent results (86-99% accuracy in detection), their integration into coherent systems for operational deployment remains a significant challenge [4]. In this paper, we survey more than 25 papers on deep learning architectures, multi-sensor systems, IoT sensor networks, satellites, and land use and land cover (LULC)-based risk maps. We highlight five research gaps, provide an overview of data sets and performance measures, compare the performance of different platforms, and propose a research roadmap to direct research towards integrated, reliable wildfire detection systems.

II. DEEP LEARNING APPROACHES FOR FIRE DETECTION

A. YOLO-Based Detection Systems

The YOLO (You Only Look Once) family of realtime fire detectors is most popular because of its accuracy-speed trade-off [5]. YOLO approaches object detection as a single regression task unlike two-stage detectors (R-CNN, Fast RCNN) that separate region proposals from classification and thus achieve realtime performance vital for time-critical fire detection. YOLOv5 has 98% accuracy and 200ms latency when deployed on edge devices (NVIDIA Jetson Nano) without requiring cloud computing [6].

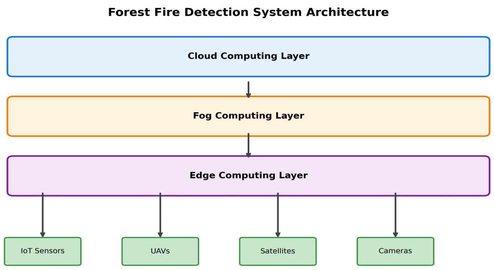

Fig. 1. Typical three-tier fog-edge-cloud architecture integrating IoT sensors, UAVs, and satellites reported in reviewed literature.

YOLO-UFS uses the C3MNV4 module for feature extraction, Bidirectional Feature Pyramid Networks (BiFPN) to fuse features at multiple scales, Adaptive-Focus Intersection over Union (AF-IoU) loss for small object detection, and Normalized Attention Mechanisms (NAM) to refine spatial and channel-wise features, attaining 97.2% accuracy for UAVs [7]. The C3-MNV4 module employs Cross Stage Partial (CSP) connections and MobileNetV4 bottleneck to reduce computational complexity by 40% compared to traditional convolutional layers while maintaining feature representation. An improved YOLOX with CSP-ML (Multi-Level Feature Extraction Structure) and CBAM (Convolutional Block Attention Module) has 96.3% accuracy with 45% fewer parameters for mountain scenarios [8]. The GXLD detector combines dark-channel-prior defogging and YOLOX-L-Light with 92.6% reduction in parameters and 87.47% accuracy in foggy environments [9].

B. Transformer-Based and Hybrid Architectures

The implementation of transformers brought about a revolutionary advancement for fire detection systems. The CNN-ViT model achieves 99% accuracy when tested with IIITDMJSmoke and USTCSmokeRS and Khan datasets that contain smoke covering less than 2% of the image area. The system combines convolutional layers which detect basic features like edges and textures with multi-headed self-attention blocks that build the complete structure and LSTM which tracks the movement of smoke patterns through time and space. The system enables detection of smoke patterns which maintain their geographical distribution throughout wide image sections particularly in satellite images where smoke detection remains difficult because of its continuous spatial patterning. The Adaptive Hierarchical Multi-Headed CNN (AHMHCNN) with modified Convolutional Block Attention Module (CBAM) tackles multi-resolution and multi-scale problems in aerial images [11]. It features Explainable Artificial Intelligence (XAI) using Gradient-weighted Class Activation Mapping (Grad-CAM) to enhance interpretability, crucial for real-world applications where false alarms are costly.

C. Anchor-Free and Encoder-Decoder Architectures

FuF-Det uses an anchor-free encoder-decoder network with dedicated RECAB (Residual Efficient Channel Attention Block), AAFRM (Attention-based Adaptive Fusion Residual Module) and Coordinate Attention blocks to preserve fire information in foggy environments [12]. It surpassed eight baseline deep learning models, such as YOLOv5, Faster RCNN, and RetinaNet, on foggy fire detection, with 94.3% mAP compared to the best baseline of 89.7%. Coordinate Attention uses spatial information for the attention mechanism to accurately locate fires.

Table. 1 Comparative Analysis of Deep Learning Models For Forest Fire Detection

Method	Year	Platform	Accuracy	Key Innovation
YOLOv5 [6]	2022	Edge	98%	Edge deployment, <200ms latency
YOLO-UFS [7]	2025	UAV	97.2%	BiFPN, NAM, AF-IoU loss

YOLOX+CSPML [8]	2024	UAV	96.3%	Small object mountain detection
GXLD [9]	2023	UAV	87.47%	92.6% param reduction, fog
CNN-ViT [10]	2024	Satellite	99%	Ultralightweight, 2% smoke area
AHMHCNN [11]	2024	Multi	—%	Multi-scale + XAI interpretability
FuF-Det [12]	2023	UAV/Gnd	94.3%	Anchor-free, fog conditions

III. MULTI-MODAL SENSOR FUSION

A. Visual and Infrared Image Fusion

RGB cameras can provide accurate spatial resolution together with precise texture detection but their performance decreases when they need to identify smoke under bright light conditions and dim lighting environments. Thermal infrared cameras can detect heat sources regardless of light conditions but their spatial resolution capabilities remain limited and they become easily deceived by other heat sources. Multimodal fusion uses the detailed information which exists in each separate image to create new understanding.

The Fire Fusion Network (FF-Net) unites visual and thermal images derived from UAVs through cross-modal attention modules which analyze the visual and thermal features to create attention weights that identify areas with shared space which probably contain fire [13]. The DJI Matrice 300 RTK drone operated an uncooled vanadium-oxide microbolometer which detected infrared wavelengths between 8-14μm and achieved 50mK sensitivity across a temperature range from -25°C to 450°C together with a 48MP CMOS camera that provided centimeter-precise RTK positioning. The system achieved a reduction in false alarms which decreased from 23.4% (visual) and 18.7% (thermal) to 8.2% (fused) while increasing detection accuracy from 15.6% (visual) to 4.7% (fused) which proved the multi-modal fusion approach was effective based on its complementary nature.

B. Environmental Sensor Fusion

Environmental sensors that measure air composition, temperature, humidity and pressure provide additional fire indicators which detect fires before smoke or flames become visible. Fire produces distinct chemical signals (CO, CO2, volatile organic compounds) and creates microclimate changes which enable fires to be detected earlier than through visual methods. A sensor fusion smoke detector features the Arduino Nicla Sense ME microcontroller with BMP390 barometric pressure, BME688 multi-gas (TVOC, CO2 equivalent, H2, ethanol) and SGP30 TVOC/CO2 sensors [14]. Recursive Feature Elimination Cross-Validation (RFECV) feature selection revealed temperature, TVOC, CO2 equivalent and humidity as the most important features. The deep neural network model developed using the identified feature set and achieved 0.99 accuracy with 0.99 precision, 1.0 recall, and a very short latency of 34µs - allowing real-time inference at the sensor edge.

Table. 2: Comparative Analysis of Multi-Modal Sensor Fusion and IoT Systems

Study	Sensors	Perf.	Key Feature Limitation	
FF-Net [13]	RGB+ Thermal	FA: 8.2%	Cross-modal attention	UAV-only, custom dataset
Env. Fusion [14]	Gas+ Pres- sure	Accuracy: 0.99	RFECV, 34μs	Indoor testing only
IoT-Rasp. [15]	Fisheye Cam	1.5km range	YOLOv4, DSM map	No real fire tests
EEWBP [16]	Temp/Gas/ Hum	92%+ PDR	Energy efficient Simulation cluster only	
IoRT [17]	Multi- sensor	—	UAV+UGV Simulation suppress. only	

IV. IOT AND WIRELESS SENSOR NETWORKS

A. Distributed IoT Architectures

IoT-based systems with Raspberry Pi 4, 220-degree fisheye cameras and YOLOv4 vision-based models can detect fires over a range of 1.5km with 94% accuracy on Mt. Buk Jeong [15]. The YOLOv4 model was quantised using TensorFlow Lite to reduce model size from 244MB to 61MB, and inference time from 1.2 seconds to 0.3 seconds per frame, allowing 3-5 frames per second processing. Fire detections raise alarms using 4G LTE or Wi-Fi, and Digital Surface Model (DSM) data is used to triangulate fire locations with ±15 meter accuracy.

B. Energy-Efficient Clustering Protocols

The Energy-Efficient Weighted-Based Protocol (EEWBP) prolongs WSN lifetime by 34% relative to LEACH by using energy-weighted cluster head election, based on energy level, node degree, trust value and sensing range [16]. The protocol has setup and steady-state rounds with collision-free TDMAbased transmission, packet delivery ratio of over 92%, and up to 18-23% lower end-to-end delay compared to other protocols.

C. Cyber-Physical Integration and Fog/Edge Computing

The Internet of Robotic Things (IoRT) approach combines IoT sensors, UAV swarms, and unmanned ground vehicles (UGVs) to detect and extinguish fires [17]. UAVs with Intel RealSense D435 depth sensors autonomously follow a search pattern, detect fires with onboard YOLOv5 algorithm, and have fireball extinguishers. UGVs use LiDAR-based mapping for navigation and water cannons or foam sprays. The system runs on Robot Operating System (ROS) middleware using MQTT, but is primarily tested in simulation. The fog-edge-cloud trilayer architecture optimally offloads computation: the edge layer uses lightweight models (8-15MB) on Raspberry Pi or Jetson Nano hardware with 150-250ms inference time [6], [10]; the fog layer collects data from 20-50 edge nodes, performs ensemble validation, coordinates UAV missions, and reduces cloud bandwidth by 50-70% [14], [16]; the cloud layer integrates satellite data, classifies LULC, archives historical data, retrains models, and displays user dashboards [18], [19]. Quantization with TensorFlow Lite INT8 achieves 70-80% model size reduction with 3-5% loss in accuracy.

V. SATELLITE-BASED DETECTION SYSTEMS

A. Geostationary Satellite Innovations

Geostationary satellites offer persistent monitoring of large areas every 10-15 minutes. The Multi-Scale Spatial-Temporal Features (MSSTF) model with Himawari8/9 Advanced Himawari Imager data produces 93.7% accuracy and 4.2% false alarms by implementing multi-kernel attention-based convolution (MKAC) and long short-term tracking (LSTT) [18]. MKAC employs multiple kernel convolutional layers (3×3, 5×5, 7×7) to capture fire patterns of different scales, and LSTT uses LSTM recurrent neural networks to learn temporal fire dynamics to remove false alarms from persistent hot objects Stacking ConvLSTM prediction of brightness temperatures allows 99.5% accuracy, 7-minute fire detection time and 71second processing per disk image [19]. This pre-emptive approach to fire detection compares predicted temperatures versus observed temperatures to detect fires before they are detected by threshold-based approaches. Bivariate analysis using the high 3.9µm/11µm brightness temperature ratio of fire also enhances performance.

B. Multispectral Analysis for Burn Mapping

Multispectral techniques using Normalized Burn Ratio (NBR) and GLCM texture features, and employing Random Forest, SVM, and CNN, reach 98% accuracy on Landsat-8 burn scar detection [20]. Analysis of feature importance showed 40% of the classification power came from GLCM texture features, 35% from NIR reflectance, and 25% from NBR, demonstrating the importance of engineered spatial features augmenting the raw spectral reflectance. Sentinel-2-based systems with GLCM and Random Forest achieve 86% operational accuracy for active fire detection [21]. The new GeoXO (Geostationary Extended Observations) satellite system in the 2030s will enhance the key 3.9µm fire-detection band from 2km (GOESR) to 1km resolution, allowing detection of fires as small as 50100m2 (compared to 200-400m2 today).

VI. LAND USE AND LAND COVER ANALYSIS FOR FIRE RISK

Land use/land cover (LULC) change analysis is critical to fire risk analysis. A multitemporal analysis of LULC change in semi-arid Algeria between the 1990s and 2020 using Random Forest, SVM, CART and Naive Bayes classifiers on Landsat and Sentinel-2 satellite images reveals considerable forest loss from recurrent fires, crop variations and urban expansion [22]. This retrospective analysis locates high risk areas with vegetation loss and human-forest interface. Linking spatial analysis in GIS and machine learning-based LULC classification allows the development of dynamic risk maps that can effectively inform deployment of IoT sensor networks, UAV surveillance routes and satellite monitoring priorities. Yet, existing research considers LULC analysis separately from detection platforms. Integration of historical LULC change with active detection systems for risk-based predictive monitoring is underexplored - and remains a key research gap. Table III lists satellite systems.

Table. 3: Comparative Analysis of Satellite-Based Detection Systems

System	Satellite Accuracy FA	(%)	Key Innovation
MSSTF [18]	Himawari8/9 93.7%	4.2	MKAC + LSTT
ConvLSTM [19]	Himawari8/9 99.5%	—	BT prediction, 7min
RF+SVM+ CNN [20]	Landsat-8 98%	—	NBR + GLCM texture
RF+GLCM [21]	Sentinel-2 86%	—	40% texture features

VII. RESEARCH GAPS

Analysis of more than 25 studies identifies five key issues with the state of the art.

- **Integration Deficiencies:** Platforms are largely standalone, with real-time integration of satellite, UAV and ground sensor data being mostly conceptual. There is no common communication and data-sharing protocol such as MQTT or CoAP, limiting cross-platform integration [17].
- **Dataset Fragmentation:** There's no standardized benchmark offering temporally aligned multi-modal (RGB, thermal, environmental, satellite) data across multiple ecosystems and environments. Variability in annotations and geographic extent limits reproducibility and comparisons across studies [4].
- **Environmental Robustness:** Robustness of models decreases in fog, clouds, dust, and at night. Cross-ecosystem testing (semi-arid, temperate and tropical forests; different vegetation densities) is limited, affecting applicability of results [7], [10].
- **Computational Constraints:** Edge-based deployment demands an ultra-lightweight model (<10MB, <1GFLOP, <5W) with accuracy over 95% but current solutions still face the trade-off between accuracy and computational cost [8], [10]. Many approaches are only tested in simulations, not real fires.
- **Underutilization of Historical Context:** Land use/land cover data assist in fire risk assessments, but dynamic incorporation with real-time fire detection systems for predictive risk-aware alerting and pre-positioning of resources has yet to be operationalized [22].

VIII. METHODOLOGY: PROPOSED INTEGRATED SYSTEM

Based on the gaps identified, we outline the design of a unified multi-platform forest fire detection system addressing all five deficiencies.

A. Three-Tier Hierarchical Architecture

The new system design uses a three-layered edge-fog-cloud approach. The edge layer hosts ultra-lightweight (<10MB) CNN-ViT hybrid models on IoT sensors, Raspberry Pi and UAVs with <200ms inference time for real-time detection of fire candidates. TensorFlow Lite INT8 quantization reduces the model size by 75% without compromising >5% accuracy. The fog computing layer creates mid-tier processing nodes, to integrate data from several edge nodes, ensemble validation, UAV swarm management and RFECV sensor parameter selection. The cloud computing layer offers historical data storage, model retraining, multi-temporal LULC classification over Landsat-8 and Sentinel-2 data, satellite data fusion, and stakeholder customizable visualization platforms.

B. Deep Learning Model Architecture

The detection model is based on CNN-ViT architecture with: (i) convolutional modules for local feature extraction of smoke edges, textures, and gradients; (ii) multiheaded self-attention layers for global modeling of spatial context allowing detection of spatially connected smoke plumes; (iii) temporal LSTM layers tracking fire dynamics and propagation across images; and (iv) a multi-task learning head for simultaneous detection of smoke and flames, fire intensity, and fire spread trajectories. Training uses a combination of IIITDMJSmoke (12,000 images), USTCSmokeRS (8,500), Khan (3,200), and He Smoke (2,800) datasets complemented with custom field data for target ecosystems (80/10/10 train-validation-test split).

C. Multi-Modal Sensor Fusion

Hardware Platform: Visual sensors use 48MP CMOS cameras (1/2-inch sensors). Thermal sensors are uncooled vanadium-oxide microbolometers (8-14μm, 50mK sensitivity, 25°C to 450°C). Environmental sensors comprise barometric BMP390, multi-gas (TVOC, CO2 equivalent, H2, ethanol) BME688 and TVOC/CO2 SGP30 sensors. Fusion Strategy: Intermodal attention blocks calculate similarity between thermal and visual features, and compute weights for areas with correlated signatures. Time consistency removes sporadic noise like dust or car exhaust. Land use and land cover (LULC) context layers provide reference to vegetation type maps to help eliminate false alarms from nonfire sources. *D. Satellite Integration*

The approach uses geostationary (Himawari-8/9, GOESR) and polar satellites (MODIS, VIIRS, Sentinel-2). The MSSTF model provides continuous large-area monitoring while Stacking ConvLSTM enables prediction through brightness temperature forecasting. Fire detection and burn scar assessment benefit from multispectral features which include Normalized Burn Ratio (NBR) and Grey-Level Co-occurrence Matrix (GLCM) texture attributes and Burned Area Index (BAI).

E. LULC Risk Modeling

The research uses three machine learning techniques which include Random Forest SVM and CNN to create land use land cover maps from Landsat 8 and Sentinel 2 data for the period from 2010 to 2025. The system calculates a dynamic Forest Fire Index which uses sensor data to determine fire risk based on the distance to previous fire locations. Self-Organizing Maps (SOM) create clusters of geographical risk areas which need concentrated monitoring. The process of monitoring land use land cover changes creates feedback loops between predictive risk assessment and detection which depend on sensor network configuration and UAV patrol routes and adaptive alert threshold settings.

F. WSN Protocol and Communication Framework

The EEWBP protocol controls sensor grouping through its function which selects cluster heads based on their energy level and node degree and trustworthiness and sensing capacity. TDMA scheduling enables successful delivery of more than 92 percent of packets while extending network operation time by 34 percent when compared to the LEACH protocol. The study evaluates LoRaWAN and NB-IoT as power-saving alternatives to 4G LTE for backhaul purposes in remote installations. The implementation of variable sleep/wake schedules which depend on fire risk index (FFI) measurements leads to extended battery life during low-risk periods while maintaining operational readiness for high-risk situations. Communication standardization employs MQTT for lightweight IoT messaging, CoAP for constrained device communication, and REST APIs for cloud dashboard integration. Data formats follow standardized GeoJSON for fire location reporting and NetCDF for satellite data exchange, enabling device-agnostic composition across heterogeneous sensing platforms.

G. Validation Methodology

There are three phases of system validation. Phase 1 involves controlled burn experiments in three environments (semi-arid, temperate and tropical) accompanied by ground truth data collection of GPS locations of fire ignition, ground truth measurements from thermal cameras and manual smoke observations. Phase 2 is monthly testing over 100km2 over different weather conditions (fog, rain, high winds and vegetation seasons). Phase 3 includes benchmarking across different ecosystems, and with state-of-the-art systems, such as MODIS fire products, and fire agency notifications. Evaluation metrics include accuracy, F1-score, detection time (from fire start to alert), false alarm rate, localization accuracy (GPS error), efficiency (FLOPs, inference time, power consumption) and communication efficiency (packet loss rate, end-to-end delay). Uniform reporting in the three phases facilitates benchmarking with previous studies and the proposed benchmark dataset can be used as a resource. Table IV summarizes the key LULC study reviewed.

Table. 4: Land Use/Land Cover Change Analysis For Fire Risk Assessment

Study	Region	Period	Methods Key Finding
[22]	Algeria (semiarid)	1990–2020	RF, SVM,Forest decline CART, NB from fires; + GIS urban growth; no real-time link

Table V summarizes primary datasets used across reviewed studies. Typical studies combine 2–4 datasets totaling 15,000– 35,000 images with 80/10/10 train-validation-test splits.

Table. 5: Summary of Benchmark Datasets Used in Reviewed Studies

Dataset	Images	Type	Focus
IIITDMJSmoke [10]	12,000	RGB	Smoke, varied lighting
USTCSmokeRS [10]	8,500	RGB	Smoke, surveillance
Khan [10]	3,200	RGB	Smoke and flame
He Smoke [10]	2,800	RGB	Early-stage smoke
Himawari-8/9 [18], [19]	Time series	Satellite	Brightness temperature
MODIS MOD14 [23]	Annual	Satellite	Active fire product
Sentinel-2 [20], [21]	Multiyear Multi-spectralBurn scar mapping		

IX. BENCHMARK DATASETS AND EVALUATION METRICS

The studies report on five metrics: (i) *detection performance* - accuracy, precision, recall, F1score and specificity (state-of-the-art has 90-99% accuracy across different platforms); (ii) *temporal metrics*- detection time (2-15 minutes from fire start to alert), processing latency (50-500ms); (iii) *reliability* - false alarm rate (1-15%), localization accuracy (50-200m, depending on the platform); (iv) *computational efficiency* - model size (5-50MB), FLOPs (1-20G), inference time (100800ms, measured on edge device), and power consumption (215W); and (v) *network performance* - packet delivery rate (8595%), end-to-end latency (200-800ms), and network lifetime (6-18 months). Table VI compares representative systems across key performance dimensions.

Table. 6: Comparative Performance of Representative Detection Systems

System	Accuracy (%)	False Alarm (%)	Detection (min)
MSSTF Satellite [18]	93.7	4.2	7.0
FF-Net UAV [13]	91.8	8.2	5.5
IoT Ground [15]	94.0	12.8	6.2
CNN-ViT [10]	99.0	13.5	3.8
YOLO-UFS [7]	97.2	9.3	4.5

X. COMPARATIVE ANALYSIS AND DISCUSSION

The CNN-ViT model has the best accuracy (99%) but worst false alarm rate (13.5%) as it is sensitive to visually similar nonfire. Satellite systems have the lowest false alarms (4.2%) through multi-spectral verification, but are slowest (7.0 minutes) due to low orbit. UAV fusion is the most well-rounded (91.8% accuracy, 8.2% false alarms, 5.5 minutes). Ground IoT has low-latency continuous monitoring (94% accuracy, 6.2 minutes) but the highest false alarms (12.8%) due to environmental noise. Multi-modal fusion outperforms single-modality for 30-65% fewer false alarms (visual-thermal) with environmental sensors providing robustness to fog and poor night-time visibility [13], [14]. Inter-ecosystem studies have 3-8% accuracy variations, highlighting the importance of diverse training data [4], [10]. In terms of efficiency, lightweight YOLO models are edge-deployable, with transformers offering higher accuracy. CNN-ViT hybrids are most accurate with moderate complexity, and thus in the "sweet spot" for edge deployment. Early fusion results in an accuracy improvement of 2-5% over late fusion [13].

A. Technological Maturity and Integration Gaps

The literature surveyed demonstrates high TRL of single detection technologies. Convolutional neural networks (CNNs) and variants of YOLO and CNN combined with Vision Transformer (CNN-ViT) are at >95% accuracy in benign conditions. Satellites have high detection (99.5%, 7 minutes) and multi-modal systems have significantly lower false alarms than single-modality systems.

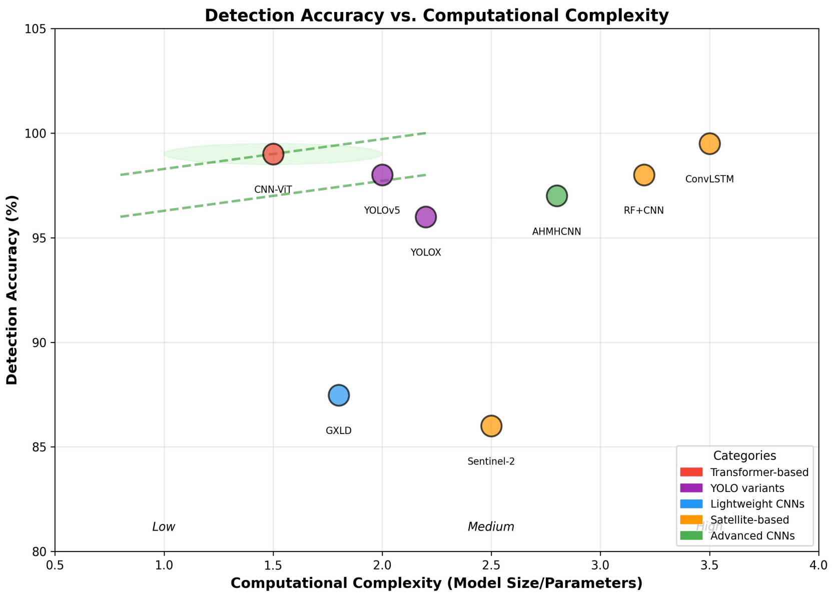

Fig. 2. Accuracy vs. computational complexity trade-off across deep learning architectures reported in surveyed literature. The ideal zone (dashed ellipse) represents high accuracy with low computational requirements, suitable for edge deployment.

But there are key integration needs most research develops and tests system independently without data sharing between platforms. Satellite, UAV and ground sensor communications is still largely theoretical and only applied in a few cases [17]. Lack of communication, data and interoperability standards limits collaboration between platforms. The real world brings a range of challenges from variable environmental impacts (extreme weather, interference from wildlife) to sensor drift (due to prolonged exposure to the environment) to communication in fire smoke and other atmospheric conditions and power management for remote battery operated systems.

B. Environmental Robustness

Adverse environmental conditions reduce performance. Visual-only systems are less accurate in fog, clouds, dust or night-time [9], [12]. Multisensory systems, which are based on visual, thermal and context information, have been more robust [13], [14], but have not been evaluated under multiple adverse conditions. Generally, the studies report accuracy of the system in a lab environment or a specific location, without cross-ecosystem validation in semi-arid, temperate and tropic forests with different vegetation density, terrain and weather, making it hard to conclusively prove the system can be deployed across different ecosystems.

C. Simulation-to-Deployment Gap

Another issue in many of the studies is the simulation deployment workflow. The EEWBP WSN system is only simulated in ns-3 but not deployed in the forest [16]. The IoRT cyber-physical system for fire detection and suppression is only simulated in Gazebo with no wildfire deployment [17]. The IoT-fog-cloud FFI framework predicted fire with 87% accuracy [24]. We believe this is the final and most important step to gain acceptance of the system in the field, via long-term (more than 12 months) deployment in different ecosystems, and in partnership with fire managers.

XI. FUTURE DIRECTIONS AND EXPECTED OUTCOMES

A. Proposed Research Directions

Integrated Systems: Future research should develop integrated systems to synergise the edge, fog and satellite layers with the open protocol (MQTT, CoAP) and test them in experiments (not only simulations). Our system is aimed at detecting wildfires in less than 5 minutes with 95% accuracy and 5% false alarm in various environments. Distributed computing using smart load balancing of computing layers (based on urgency and availability) is vital for resiliency.

Smart Sensing: LiDAR for fuel and terrain assessment, hyperspectral (400-2500nm) for fire-vegetation discrimination and acoustic for flames, in addition to RGB-thermal, to address visually-occluded (occluded by smoke) wildfires. This can be seamlessly introduced into the existing UAV and ground sensor networks as well as the current RGB-thermal-environmental fusion.

Predictive Modeling: Long-horizon, spatiotemporal models combining physics-based fire behaviour models and deep learning can allow short-term (1-6 hours) fire spread prediction, to help move from detection towards risk assessment and management. This will involve using spatiotemporal features (wind speed and direction, fuel moisture, slope and vegetation density) to the model as time varying inputs, along with the sensor streams.

Robust Learning: Domain adaptation, uncertainty estimation, continual learning and self-supervised learning techniques are needed to ensure good performance across ecological regions and seasons without retraining. Current modelling work for North American and Mediterranean European fires do not apply to tropical fires and boreal fires [23].

Efficiency and Sustainability: The opportunity for energy harvesting from solar and wind for remote nodes needs to be explored, as well as sleep/awake modes driven by fire risk. We envisage low-power wide-area networks (LPWANs) such as LoRaWAN and NB-IoT to extend the range of sensors and reduce power consumption for remote monitoring of forests.

Benchmarking and Long-term Deployments: Community data sets with co-timed RGB, thermal, environmental and satellite data have to be made available for a number of ecosystems, seasons and fires to allow benchmarking. Benchmarking tools will need to report the detection time, false alarms, detection accuracy, computational efficiency and costs as well as accuracy. 12+ month deployments with fire agencies are thought to be the key to operationalising fire detection.

B. System Integration Roadmap

Planning for operational deployment has 4 phases. The first stage (months 1-6) is component validation of the CNN-ViT model, multi-modal fusion, EEWBP WSN and LULC on benchmark datasets and in laboratory environments. In the subsystem integration phase (months 7-12), communication protocols are investigated and standardised and the UAV satellite-ground sensor data streams are validated in laboratory conditions. In the controlled field testing (month 13-24) controlled burns in three ecosystems are used to calculate detection accuracy with ground truth and the first operational benchmark data set is developed. During the operational field stage (months 25-36), the entire system will be deployed in 100km2 areas in collaboration with forestry and fire management agency with system performance monitoring and continuous system learning for model adaptation.

C. Expected Outcomes

The integrated research program will develop (1) a fully operational multi-platform system with <5-minute detection time, >95% accuracy, and <5% false alarm in various environments; (2) open-source CNN-ViT hybrid model architecture for on-edge deployment with <10MB model size and <200ms inference time; (3) interoperability protocols to allow integration of different sensing platforms; (4) a community multi-modal benchmark dataset with aligned RGB, thermal, environmental and satellite images with rich annotations in three different ecosystems; (5) a dynamic LULC integrated risk model that uses prediction for resource allocation and context-aware alerting; and (6) a prototype system to forestry and fire management agencies, including deployment guidelines and training manuals.

XII. CONCLUSION

This paper surveys integrated multi-platform forest fire detection systems in more than 25 recent studies, and suggests future research to overcome the challenges. The techniques are effective: deep learning (86-99% accuracy, 99% with hybrid transformers) on benchmarks; multi-modality data fusion from 12-23% (single modsility) to 4-8% (fused) false alarms; satellite based monitoring with 7-minute detection time and broad coverage; and low power WSN protocols for 34% longer lifetime. Five issues need to be resolved: integration issues to realise platform inter-operability; data inconsistency to enable repeatability; environmental constraints to fog, cloud, dust and darkness to enhance environmental resilience; computational limitations to ensure accuracy and efficiency on the edge; and the underutilisation of historical LULC knowledge for risk assessment. Our integrated approach overcomes these five challenges in a single edge-fog-cloud system: lightweight CNN-ViT deep learning, multi-modal sensor fusion, satellite and LULC-based risk modelling. To reach the goal of less than 5-minute detection, more than 95% accuracy, and less than 5% false alarms in all environmental conditions, the design of multi-platform fire detection systems needs long-term interdisciplinary research from computer vision, remote sensing, sensor networks, ecology and fire management.

REFERENCES

1. Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. *Proceedings of the National Academy of Sciences, 113*, 11770–11775.
2. Finney, M. A., et al. (2011). A simulation of probabilistic wildfire risk components for the continental United States. *Stochastic Environmental Research and Risk Assessment, 25*, 973–1000.
3. Sudhakar, S., et al. (2020). Unmanned aerial vehicle (UAV) based forest fire detection and monitoring for reducing false alarms in forest-fires. *Computer Communications, 149*, 1–16.
4. Saleh, A., et al. (2024). Forest fire surveillance systems: A review of deep learning methods. *Heliyon, 10*.
5. Redmon, J., et al. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 779–788).
6. Mahdi, A. S., & Mahmood, S. A. (2022). An edge computing environment for early wildfire detection. *Annals of Emerging Technologies in Computing, 6*, 56–68.
7. Luo, Z., et al. (2025). YOLO-UFS: A novel detection model for UAVs to detect early forest fires. *Forests, 16*(5), 743.
8. Luan, T., et al. (2024). Enhanced lightweight YOLOX for small object wildfire detection in UAV imagery. *Sensors, 24*, 2710.
9. Huang, J., et al. (2023). Real-time forest fire detection by ensemble lightweight YOLOX-L and defogging method. *Sensors, 23*, 1894.
10. Chaturvedi, S., et al. (2024). Ultra-lightweight convolution-transformer network for early fire smoke detection. *Fire Ecology, 20*, 83.
11. Mowla, M. N., et al. (2024). Adaptive hierarchical multi-headed convolutional neural network with modified convolutional block attention for aerial forest fire detection. *IEEE Access*.
12. Pang, Y., Wu, Y., & Yuan, Y. (2023). FuF-Det: An early forest fire detection method under fog. *Remote Sensing, 15*, 5435.
13. Liu, Y., et al. (2023). Forest fire monitoring method based on UAV visual and infrared image fusion. *Remote Sensing, 15*, 3173.
14. Julian, J., et al. (2024). Design of smoke detection system using deep learning and sensor fusion with recursive feature elimination cross-validation. *IAES International Journal of Artificial Intelligence, 13*, 1658–1667.

15. Lee, C.-H., et al. (2023). Development of IoT-based real-time fire detection system using Raspberry Pi and fisheye camera. *Applied Sciences, 13*, 8568.
16. Kaur, P., et al. (2023). Early forest fire detection using a protocol for energy-efficient clustering with weighted-based optimization in wireless sensor networks. *Applied Sciences, 13*, 3048.
17. Battistoni, P., et al. (2023). A cyber-physical system for wildfire detection and firefighting. *Future Internet, 15*, 237.
18. Zhang, L., et al. (2025). Near-real-time wildfire detection approach with Himawari-8/9 geostationary satellite data integrating multi-scale spatial-temporal feature. *International Journal of Applied Earth Observation and Geoinformation, 137*, 104416.
19. Xu, Z., et al. (2025). Adaptive early wildfire monitoring based on spatiotemporal prediction and Himawari 8/9. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.
20. Lakshmanaswamy, P., et al. (2024). Prioritizing the right to environment: Enhancing forest fire detection and prevention through satellite data and machine learning algorithms. *Remote Sensing in Earth Systems Sciences, 7*, 472–485.
21. Zhou, B., et al. (2024). Enhancing active fire detection in Sentinel-2 imagery using GLCM texture features in random forest models. *Scientific Reports, 14*, 31076.
22. Mohammed, G., et al. (2024). Assessing the impact of anthropogenic activities on land use and land cover changes in the semi-arid and arid regions of Algeria. *Environmental Monitoring and Assessment, 196*, 383.
23. Grari, M., et al. (2022). Early wildfire detection using machine learning model deployed in the fog/edge layers of IoT. *Indonesian Journal of Electrical Engineering and Computer Science, 27*, 1062–1073.
24. Aljumah, A. (2022). IoT-inspired framework for real-time prediction of forest fire. *International Journal of Computers, Communications & Control, 17*.
25. Ahmed, S. T., Kumar, V. V., Singh, K. K., Singh, A., Muthukumaran, V., & Gupta, D. (2022). 6G enabled federated learning for secure IoMT resource recommendation and propagation analysis. *Computers and Electrical Engineering*, *102*, 108210.
26. Pasha, A., Ahmed, S. T., Painam, R. K., Mathivanan, S. K., Mallik, S., & Qin, H. (2024). Leveraging ANFIS with Adam and PSO optimizers for Parkinson's disease. *Heliyon*, *10*(9).
27. Periasamy, K., Periasamy, S., Velayutham, S., Zhang, Z., Ahmed, S. T., & Jayapalan, A. (2022). A proactive model to predict osteoporosis: An artificial immune system approach. *Expert Systems*, *39*(4), e12708.