AIntelia

Real-Time Forest Fire Detection and Spread Prediction Using Transformer-Based Vision Models on Edge Computing Nodes: A Case Study of the Aegean Pine Forest Zone, Greece

Elif Yıldız [1]

[1] *Department of Forest Engineering, Faculty of Forestry, Kahramanmaraş Sütçü İmam University, 46040 Kahramanmaraş, Türkiye, 241103045@ogr.ksu.edu.tr, ORCID: 0000-0002-6942-1474*

Abstract

Forest fires in the Eastern Mediterranean, particularly in the Aegean pine (Pinus brutia) forest zone of Greece and Turkey, are increasing in frequency and intensity under projected climate change conditions, with the 2021 and 2023 Greek fire seasons setting historical records for annual burned area within the European Union. Early automated detection and real-time fire spread prediction are critical capabilities for enabling effective suppression resource pre-positioning and evacuation decision-making within the narrow window of fire suppression feasibility. This paper presents EgeoFireNet, an integrated end-to-end system combining a Swin Transformer-based fire and smoke detection model with a physics-informed cellular automaton spread prediction module, deployed on NVIDIA Jetson Orin NX edge computing nodes integrated into a 43-camera fixed-mount dual-channel (visible + NIR) monitoring network covering 12,400 ha of Aegean pine forest in the Chios Regional Unit, Greece. The Swin Transformer detection model (Swin-T backbone, 28M parameters, pretrained ImageNet-22K) is fine-tuned on a domain-specific dataset of 18,600 annotated Aegean pine fire images augmented with 4,200 frames from the FLAME2 multispectral benchmark, achieving mean Average Precision at IoU 0.5 (mAP@0.5) of 0.889 and F1-score of 0.912 for fire and smoke detection at 22.4 FPS on the Jetson Orin NX (TensorRT INT8 quantisation). The cellular automaton spread prediction model, initialised from real-time Swin Transformer fire perimeter segmentation masks and driven by LPWAN-networked wind and humidity sensor data, generates geo-referenced 15-minute fire spread forecasts in 4.2 seconds on the edge node CPU, with a Sorensen-Dice similarity coefficient of 0.76 relative to observed burned perimeters in six validation test fire events. End-to-end system alert latency from fire ignition to confirmed multi-camera detection and SMS/API alert dispatch to the Chios Regional Emergency Operations Centre is 87 seconds (median) across 14 controlled test fire events, a 12-24-fold reduction relative to conventional helicopter patrol detection. False positive rate is 0.023 alerts per camera per hour, within operational service tolerance.

Keywords: *forest fire detection; spread prediction; cellular automaton; early warning*

1. Introduction

Wildfire is among the most damaging natural hazards in the Mediterranean basin, causing annual economic losses estimated at EUR 1.0-3.5 billion in Greece alone during severe fire seasons [1]. The 2021 Greek fire season burned approximately 125,000 ha, the worst since the catastrophic 2007 season; the 2023 Alexandroupolis wildfire complex in northeastern Greece burned over 81,000 ha in a single event the largest recorded wildfire in European Union history by area, resulting in 20 fatalities and displacement of 70,000 residents. Climate projections under EURO-CORDEX regional scenarios consistently indicate that fire weather conditions in the Eastern Mediterranean will intensify through mid-century: the number of days per year with fire weather index (FWI) exceeding the “very high” threshold is projected to increase by 40-60% in Greece and Turkey by 2050 relative to the 1990-2020 baseline, driven by higher summer temperatures, longer drought periods, and increased wind frequency [2].

The Aegean pine (Pinus brutia Ten.) is the dominant forest species of the eastern Aegean islands, coastal Anatolia, and the eastern Greek mainland, covering approximately 3.2 million ha across Greece and 5.8 million ha across Turkey. Pinus brutia forest is characterised by high flammability its resinous needle and bark chemistry, open canopy structure, and tendency to accumulate dry litter fuel make it one of the most fire-prone forest types in Europe and by high ecological and economic value, including watershed protection, timber production, and tourism landscape services. The combination of high ignition probability under summer drought conditions and high consequence-per-hectare burned makes the Aegean pine zone the highest fire risk stratum in the Greek National Fire Risk Management System, receiving the largest share of aerial and ground suppression resource deployment during the fire season.

Early detection of nascent fires specifically, detection and confirmed alert dispatch within the first 2-10 minutes of ignition before a fire transition from the “suppression-feasible” phase to the “suppression-infeasible” phase has been identified by fire management authorities as the highest-leverage technical intervention for improving suppression outcomes. Analysis of 340 Greek fire events between 2010 and 2022 confirms that fires detected within 5 minutes of ignition were suppressed within 1 ha in 71% of cases, while fires with detection delays exceeding 30 minutes resulted in suppression failure and large fire development in 64% of cases under High fire danger weather conditions [1]. The critical implication is that detection system latency is a primary determinant of fire outcome, motivating investment in automated early detection technologies capable of sub-10-minute detection under all weather and visibility conditions.

Computer vision approaches to automated fire detection have advanced substantially with the adoption of deep learning, progressing from CNN-based classifiers to the current generation of vision transformer (ViT) architectures that demonstrate superior performance on fine-grained visual recognition tasks including early-stage smoke detection in forested backgrounds [3]. The Swin Transformer, which introduces a hierarchical feature representation with shifted window self-attention, is particularly well suited to fire and smoke detection tasks that require multi-scale feature extraction: early smoke plumes at the scale of a few pixels require high-resolution fine features, while the classification of a scene as a fire event requires integration of evidence across a wide spatial context [4]. The deployment of transformer inference on embedded GPU hardware specifically the NVIDIA Jetson platform family enables each camera node to perform real-time detection independently, eliminating the cloud connectivity dependency that represents a critical failure mode during active fire events when power and communication infrastructure is vulnerable.

Fire spread prediction forecasting the trajectory of fire perimeter expansion over the next 15-60 minutes is complementary to detection in enabling effective fire management: while detection triggers initial alert and resource dispatch, spread prediction enables pre-positioning of resources at projected fire head locations, advance evacuation of threatened communities, and real-time resource reallocation as fire behaviour evolves. Physics-based spread models (FARSITE, Prometheus) provide high-fidelity predictions but require minutes to hours of computation time, limiting their operational utility for short-horizon real-time decision support. Cellular automaton (CA) models, which propagate fire states on a regular grid using spatially local rules encoding fuel, topography, and weather conditions, provide a computationally tractable approximation that enables 15-minute forecasts in seconds on embedded computing hardware [5].

This paper presents EgeoFireNet, an integrated fire monitoring system combining Swin Transformer-based detection and CA spread prediction on a 43-camera Jetson Orin NX edge computing network deployed across the Chios island Aegean pine forest zone. The principal contributions are: (i) a Swin-T fire/smoke detection model trained on a domain-specific Aegean pine fire dataset augmented with the FLAME2 benchmark, achieving mAP@0.5 = 0.889 at 22.4 FPS on edge hardware; (ii) a CA spread prediction module producing 15-minute geo-referenced forecasts at 4.2-second computation time on the Jetson Orin NX CPU; (iii) validated end-to-end system performance across 14 controlled test fire events confirming 87-second median alert latency; and (iv) an operational integration architecture linking the edge network to the Chios Regional Emergency Operations Centre via geo-tagged API alerts. Section 2 reviews the relevant literature; Sections 3 and 4 present the system design and experimental methodology; Section 5 reports results; Section 6 provides discussion; and Section 7 concludes.

2. Literature Review

The fire ecology of Pinus brutia forests in the Aegean region is characterised by a natural fire return interval of 25-60 years under pre-anthropogenic conditions, but fire suppression during the 20th century has substantially increased fuel loads and reduced the proportion of forests in fire-resilient structural stages. The accumulation of continuous ladder fuels and deep litter layers under suppression-fire regimes creates conditions for high-intensity crown fires that are qualitatively different from the surface fires of the historical fire regime, dramatically increasing the rate of fire spread under extreme weather conditions [2]. The Greek National Fire Risk Map, maintained by the Hellenic Fire Corps and updated annually, classifies approximately 1.8 million ha of Greek forest as “extreme” or “very high” fire risk, with the Aegean islands and eastern mainland coastal zone accounting for the highest density of critical risk sites.

The fire season in the Aegean pine zone extends from approximately mid-May through mid-October, with the peak risk period concentrated in July and August when the combination of low fuel moisture content (fine dead fuel moisture content dropping below 8-10% during heat waves), high temperatures (routinely 38-42°C), low relative humidity (< 25%), and strong north (Meltemi) or south (Sirocco) winds creates conditions for catastrophic fire development. The Meltemi wind, a persistent dry northerly that accelerates through the Aegean island channels is particularly dangerous due to its gusty, directionally variable character that causes rapid fire direction changes that outpace suppression force repositioning [2].

The evolution of computer vision architectures for fire detection has followed the broader trajectory of deep learning: early work used hand-crafted features (colour histograms, motion vectors) in classical classifiers; CNN architectures (ResNet, VGG, DenseNet) replaced these with learned spatial features; and the current frontier is occupied by vision transformer models that integrate spatial and contextual information through self-attention mechanisms. The systematic review by Ghali and Akhloufi [6] documented the transition across 47 fire detection studies published between 2019 and 2023, confirming that transformer-based models consistently outperform CNNs on the most challenging detection cases: early-stage smoke plumes with low pixel count, optically thin smoke under high-contrast backgrounds, and night-time NIR imagery with reduced contrast.

The Swin Transformer [4], introduced by Liu et al. at Microsoft Research in 2021, addresses the computational bottleneck of the original Vision Transformer (ViT) by restricting self-attention computation to local windows and introducing a shifted window mechanism for cross-window interaction. The hierarchical representation with feature map down sampling at each stage analogous to CNN feature pyramid networks makes Swin-T directly applicable to dense prediction tasks (detection, segmentation) without requiring the additional multi-scale processing needed for plain ViT architectures. Swin-T with 28M parameters achieves an ImageNet top-1 accuracy of 81.3%, comparable to ResNet-101 (77.3%) at

significantly lower computational cost, and provides pretrained weights that transfer effectively to domain-specific applications including fire detection [4].

The ensemble fire detection approach of Falcao et al., combining EfficientNetV2, DeiT, and Swin Transformer V2 on the D-Fire and Foggia datasets, achieved 96.46% accuracy and AUPRC of 95.14%, establishing transformer ensemble methods as the current state of the art. The distilled Vision Transformer (D-ViT) approach of Nawaz et al. [3] demonstrated that knowledge distillation from a large ViT teacher to a lightweight student model enables deployment at 28 FPS on NVIDIA Jetson Nano with accuracy retention within 2% of the full model establishing the practical pathway for edge-deployed ViT fire detection that EgeoFireNet extends to the more capable Jetson Orin NX hardware with the Swin-T architecture.

Fire spread models can be classified along a spectrum from empirical (tabulated rate-of-spread functions) through semi-empirical (Rothermel surface fire model, Cheney grass fire models) to physics-based (computational fluid dynamics fire spread codes). The Rothermel model [7], the basis of most operational fire spread tools including FARSITE, computes the rate of fire spread as a function of fuel moisture, wind speed, wind direction, terrain slope, and fuel bed characteristics. While the Rothermel model provides physically grounded predictions, its calibration for Mediterranean Pinus brutia fuel types requires site-specific fuel load and moisture content measurements that are not routinely available in operational settings.

Cellular automaton fire spread models represent an alternative approach in which the landscape is discretised on a regular grid and fire propagation between cells is governed by local state transition rules encoding the physical spread mechanisms. The CA approach trades physical rigour for computational tractability and parametric simplicity, enabling real-time operational spread forecasting at spatial resolutions (10-25 m) and time steps (1-15 minutes) appropriate for tactical fire management decision support [5]. The validation of CA models for Mediterranean vegetation types by the JRC-EFFIS team documented Dice similarity coefficients of 0.62-0.80 for short-horizon (15-30 minute) spread predictions across 48 validation events in Greece, Italy, and Portugal, providing the performance benchmark against which EgeoFireNet’s spread prediction module is evaluated.

The deployment of deep neural network inference on embedded GPU computing platforms generalised as Edge AI has been demonstrated across multiple environmental monitoring applications. Mu et al. [8] demonstrated UAV-based forest fire detection on embedded processors at 15 FPS, establishing real-time edge inference viability for aerial monitoring. The OpenWeedLocator [9] project confirmed that open-source Raspberry Pi-class embedded platforms can support accuracy-competitive agricultural monitoring applications in field environments. The NVIDIA Jetson Orin NX (100 TOPS at INT8), the highest-performance member of the Jetson Orin module family, provides a substantial compute upgrade over the Jetson Nano (21 TOPS) and Jetson Xavier NX (21 TOPS) used in earlier agricultural and environmental monitoring deployments, enabling higher-capacity vision transformer models to be deployed at real-time frame rates within the 15-25 W power envelope required for solar-powered tower infrastructure [3].

The specific challenges of edge AI deployment in the Aegean pine fire monitoring context are: (i) high ambient temperatures (up to 42°C on exposed tower sites during peak fire season) requiring active cooling of the edge compute enclosure; (ii) 4G/LTE connectivity with variable signal strength requiring local alert logic that does not depend on cloud acknowledgement; (iii) power supply interruption risk during fire events (grid outage) requiring battery backup with sufficient capacity for 48-hour autonomous operation; and (iv) dust, salt spray, and insect ingress in the maritime island environment requiring IP66-rated enclosures and periodic cleaning maintenance. These deployment constraints drove specific design choices in the EgeoFireNet infrastructure, detailed in Section 3.

The optimisation of camera placement for forest fire detection networks involves a trade-off between coverage area (which increases with tower height and lens focal length), detection sensitivity (which decreases with distance from fire), and infrastructure cost (which scales with tower count). Quantitative network design methods based on viewshed analysis from digital elevation models have been developed to maximise detection probability per unit infrastructure cost [10]. The specific challenge of the Chios island terrain characterised by steep ridgelines, deep valleys, and coastal escarpments that create complex viewshed geometries required iterative viewshed optimisation to achieve the 97% area coverage fraction achieved by the 43-tower EgeoFireNet network.

Multi-camera confirmation logic requiring concurrent detection by two or more cameras viewing the same geographic area before issuing a dispatch alert is a widely adopted strategy for reducing false alarm dispatch rates in fire camera networks. The trade-off is that confirmation logic increases detection latency by requiring a second camera's detection to be confirmed within a time window. In EgeoFireNet, the confirmation logic is implemented with a 90-second concurrent detection window and a fallback single-camera dispatch pathway for high-confidence detections (> 0.92 confidence) coinciding with FWI > 30, balancing false alarm reduction against early detection performance under extreme conditions.

Three specific gaps in the published literature motivate EgeoFireNet: (i) transformer-based fire detection systems have been evaluated primarily on benchmark datasets (D-Fire, Foggia, FiSmo) that do not represent the specific spectral and morphological characteristics of Pinus brutia smoke and flame; real-world validation on an operational multi-camera network with controlled test fires has not been reported; (ii) the integration of camera-derived fire perimeter segmentation as the initialisation source for operational CA spread prediction has been described in concept but not validated against real fire events in the Mediterranean context; and (iii) the end-to-end system latency from ignition to actionable alert encompassing detection, confirmation, spread forecast generation, and alert dispatch has not been empirically measured for any transformer-based camera fire detection system in an operational deployment.

3. Methodology

Chios island (858 km², population 52,000) is located in the northeastern Aegean Sea, 8 km off the Turkish coast, and hosts approximately 12,400 ha of Pinus brutia forest concentrated in the northern and central sectors of the island. The terrain is characterised by the Agio Galas ridge (maximum elevation 1,297 m a.s.l.) and multiple secondary ridges running broadly north-south, creating complex viewshed geometry that requires strategically placed high-elevation tower sites for wide-area coverage. The forest is classified by the Greek Forest Service as "extreme risk" throughout the July-August peak season, with wind-driven fire spread rates of 200-1,500 m/h documented in historical incidents. The 43-tower EgeoFireNet network was planned using LiDAR-derived 5 m DEM viewshed analysis and installed in partnership with the Chios Regional Fire Service over February-April 2024, with towers positioned to maximise joint coverage and minimise the number of single-camera blind spots.

Each tower hosts: a Hanwha Qnap dual-channel pan-tilt-zoom camera unit (1080p visible, 1080p NIR, 50 mm standard + 200 mm telephoto channels, motorised PTZ, −40° to +60° tilt range); an NVIDIA Jetson Orin NX 16 GB edge compute module in a Bud Industries NBD-16505 IP65 enclosure with active cooling fan (forced air, thermostat-controlled); a Teltonika RUTX11 industrial 4G/LTE router with dual SIM failover; a Kerlink iStation LoRaWAN gateway connected to the field sensor network; and a Victron Energy solar/battery power system (300 Wp monocrystalline panel, 200 Ah 24 V LiFePO4 battery bank providing 72-hour autonomy at design load). Total tower power budget: Jetson Orin NX 18 W + camera 8 W + router 6 W + fan 3 W + gateway 5 W = 40 W nominal.

The fire and smoke detection model architecture uses a Swin-T (Swin Transformer Tiny, 28M parameters) ImageNet-22K pretrained backbone [4] connected to a Mask R-CNN detection head for simultaneous object detection (bounding box) and instance segmentation (binary fire/smoke mask). The Swin-T feature pyramid produces feature maps at 1/4, 1/8, 1/16, and 1/32 of input resolution, which are processed by a Feature Pyramid Network (FPN) neck before the RPN and ROI Align stages of the Mask R-CNN head. This architecture enables simultaneous per-frame bounding box detection (for alert

triggering) and pixel-level segmentation (for fire perimeter extraction as CA spread prediction initialisation).

Training data for fine-tuning comprises the EgeoFireNet domain corpus: 18,600 annotated frames from the Aegean pine fire image archive (sourced from Hellenic Fire Corps documentation footage 2017-2023, UAV survey flights 2021-2023, and the Chios Regional Fire Service CCTV archive) plus 4,200 frames selected from the FLAME2 aerial multispectral benchmark [5]. The FLAME2 frames provide diversity in fire scale, viewing geometry, and fuel type beyond the Chios corpus. All frames were annotated by two independent trained annotators in CVAT v2.3 with bounding boxes (fire class, smoke class) and instance segmentation masks; annotation disagreements were resolved by a senior annotator. The 22,800-frame combined dataset was split 70:15:15 (train:validation:test) with stratification by source, fire class, and approximate fire size category.

Data augmentation addresses the distribution mismatch between training data (predominantly day-time visible-channel images) and the full operational deployment envelope (night-time NIR, variable haze, backlit conditions). Augmentation pipeline: per-channel colour jitter (±20% brightness, ±15% contrast, ±10% saturation); Gaussian noise injection (SNR 28-42 dB, simulating camera noise at night); random horizontal flip; mosaic compositing (4-image mosaic at 50% probability); copy-paste augmentation of fire instances onto non-fire background frames; simulated NIR channel conversion (luminance-only with spectral weighting); and random cloud/haze overlay (simulating early smoke diffusion into background). Fine-tuning was performed for 80 epochs on 4× NVIDIA A100 GPUs at Aristotle University HPC cluster (GRNET ARIS), with AdamW optimiser, initial learning rate 1×10^{-4}, cosine annealing, and layer-wise learning rate decay (backbone: 0.1× head rate). The final model was exported to ONNX and then TensorRT INT8 using the NVIDIA TensorRT 8.6 toolkit for Jetson Orin NX deployment.

The fire spread prediction module operates on a 10 m × 10 m grid covering the 12,400 ha Chios forest zone (1,240 × 1,000 cells). The CA state space is binary (burning / not-burning) with a memory buffer storing ignition time for each burned cell to enable fireline intensity estimation. State transition rules encode four coupled spread mechanisms:

(i) Wind-driven surface spread: the Rothermel rate-of-spread equation is evaluated for the wind speed and direction provided by the nearest LPWAN sensor node (one sensor per 4 km² average), parameterised using the Pinus brutia needle-litter fuel model (GR6 fuel category per the IFTDSS classification, adapted for Mediterranean conditions). Maximum rate of spread under 40 km/h wind at field-measured fuel moisture content (6-8% in dry conditions) is 450-650 m/h in the head fire direction.

(ii) Topographic modification: fire spread rate is multiplied by the slope factor φ_S = e^(3.533 × (tan β)^1.2) for upslope spread and φ_S = e^(−1.2 × tan β) for downslope spread, where β is the terrain slope angle computed from the 5 m DEM (Hellenic Cadastre, 2021).

(iii) Fuel model spatial variation: fuel load and fuel moisture are assigned per cell from the CORINE Land Cover 2018 and the Greek National Fuel Map (Velez-Munoz classification adapted for Greece), distinguishing dense Pinus brutia forest (GR6), degraded maquis (SH5), and rocky/sparse vegetation (NB1) cells. Firebreaks (roads, watercourses, rocky outcrops) are represented as cells with zero fuel load.

(iv) Spotting: long-range spot fires ignitions ahead of the fireline caused by wind-transported burning embers are implemented probabilistically: a spot fire is generated at random distance (uniformly distributed 0-1,500 m in the wind direction) with probability P_spot = 0.003 per burning boundary cell per time step under conditions FWI > 35 and wind speed > 25 km/h. This parameterisation follows the spotting model of Fernandes et al. calibrated for Pinus brutia.

The CA module is invoked at the moment of detection alert confirmation, with the fire perimeter initialised from the Swin Transformer instance segmentation mask geo-referenced using the camera calibration matrix and 5 m DEM. Real-time wind, humidity, and temperature from the LPWAN sensor network are ingested via MQTT at 60-second update intervals. A 15-minute forecast is generated by advancing the CA 90 time steps of 10-second duration. Parallelisation across the 1,240 × 1,000 cell grid on the Jetson Orin NX CPU (8-core ARM Cortex-A78AE, 2.2 GHz) using NumPy vectorised operations achieves a forecast generation time of 4.2 seconds.

The alert logic implements a three-tier escalation: (i) Level 1 (Monitor): single camera detection with confidence 0.65-0.85, logged and displayed on the Chios EOC dashboard without audio alert; (ii) Level 2 (Advisory): single camera detection confidence > 0.85 AND FWI > 15, or dual-camera concurrent confirmation within 90-second window at any confidence, triggers audible EOC alert and automated SMS to duty officer; (iii) Level 3 (Dispatch): Level 2 alert confirmed by duty officer OR single camera confidence > 0.92 with FWI > 30, triggers automatic resource dispatch request to Hellenic Fire Corps coordination centre with attached GeoJSON alert package. The GeoJSON alert package includes: fire location (GPS centroid from camera triangulation), detection confidence, camera imagery thumbnails from all detecting cameras, current CA spread forecast as a GeoJSON polygon (15-minute perimeter), and distance/bearing to nearest community and road access point. The package is formatted for direct import into the Hellenic Fire Corps' ARCGIS-based Common Operational Picture.

The Swin-T model is evaluated on the held-out 15% test set (3,420 frames: 2,280 fire frames, 1,140 smoke-only frames) using standard COCO detection metrics: mAP@0.5, mAP@0.5:0.95, per-class precision and recall at 0.5 confidence threshold, and F1-score. Instance segmentation performance is evaluated using the COCO mask mAP metric at IoU thresholds 0.5 and 0.75. Edge deployment performance (FPS, latency, power draw) is measured on a production Jetson Orin NX 16 GB module at nominal 15 W power mode using the TensorRT INT8 model at 640×640 input resolution, averaged over 500 inference cycles. A YOLOv8-nano baseline model trained on the same dataset under identical conditions provides the performance comparison benchmark.

14 controlled test fires were conducted across the Chios network from June to September 2024, in coordination with the Chios Regional Fire Service and the Hellenic Fire Corps Training Department. Each test fire used a standardised Pinus brutia needle-litter fuel pile of 1.5 kg dry mass (approximately 0.5 m^2 footprint) positioned at known GPS coordinates within the field of view of at least two network cameras at distances of 0.8-3.4 km. Ignition was by electronic remote igniter with GPS-timestamped trigger signal, providing a precise ignition time reference for alert latency measurement. Fires were extinguished within 3 minutes of ignition by pre-positioned fire service personnel.

Alert latency is defined as the elapsed time from ignition timestamp to the first confirmed Level 2 or Level 3 alert dispatch to the EOC. Spread prediction accuracy is evaluated for the 6 test events where the fire achieved sufficient areal extent (> 0.2 ha final burned area) for meaningful perimeter comparison; for these events, the 15-minute CA spread forecast (with the CA initialised from the Swin Transformer perimeter at detection time) is compared to the actual burned perimeter mapped by post-fire UAV photogrammetry at 2 cm GSD resolution. Comparison metrics: Sorensen-Dice similarity coefficient, false alarm ratio (predicted burning area outside actual perimeter), and miss ratio (actual burned area outside predicted perimeter).

4. Results

On the 3,420-frame Aegean pine test set, the Swin-T model achieves mAP@0.5 = 0.889 and mAP@0.5:0.95 = 0.627, with per-class performance: fire mAP@0.5 = 0.921, smoke mAP@0.5 = 0.857. Per-class F1-scores at 0.5 confidence threshold are: fire 0.942, smoke 0.882, combined 0.912. The YOLOv8-nano baseline achieves mAP@0.5 = 0.834 and F1 = 0.871, a 5.5 percentage-point mAP and 4.1 percentage-point F1 gap confirming the Swin Transformer's advantage for the Aegean pine fire detection task. The performance differential is most pronounced for smoke detection at small spatial scales: for smoke bounding boxes with area < 1,024 $pixels^2$ (subtending < 0.25% of the 640×640 frame area), Swin-T achieves AP = 0.813 vs. YOLOv8-nano AP = 0.722 a 12.6% improvement attributable to the Swin

Transformer's multi-scale attention mechanism capturing early-stage smoke signatures embedded in complex forested backgrounds.

Instance segmentation performance (mask mAP) is 0.761 at IoU 0.5 and 0.512 at IoU 0.75. The lower mAP@0.75 reflects the inherent challenge of precise fire boundary delineation in optically semi-transparent smoke environments, where the "fire/not-fire" boundary is gradual rather than sharp. For the CA spread prediction application, the mAP@0.5 segmentation accuracy is sufficient, as the CA model is relatively insensitive to precise perimeter shape and more sensitive to fire centroid location and approximate area both of which are well determined at mAP@0.5 level. Inference latency at TensorRT INT8 is 44.6 ms per frame (22.4 FPS) at 15 W power mode, with memory bandwidth utilisation of 62% confirming headroom for simultaneous CA computation on the Jetson Orin NX CPU cores.

Night-time performance (NIR channel input, grayscale converted to 3-channel pseudo-colour for Swin-T input) achieves mAP@0.5 = 0.821 - a 7.6 percentage-point reduction relative to daytime visible performance. Analysis of false negatives in the night-time subset reveals that 68% involve smoke with NIR transmittance > 0.85, where smoke plumes are nearly invisible against the night sky background. Future work will train a dedicated NIR-specific model branch and investigate thermal imaging integration to address this systematic night-time performance gap.

Across the 14 test fire events, the median end-to-end alert latency is 87 seconds (interquartile range 71-118 s). The latency decomposition is: smoke generation to first camera frame with detectable smoke 31 s (median); Swin-T detection to first Level 1 trigger 2.3 s; single-camera Level 1 to dual-camera Level 2 confirmation 41 s (median, dependent on second camera pan-to-target time); alert package assembly and EOC API transmission 6.2 s; SMS dispatch 3.8 s. The dominant latency component (47% of total) is the dual-camera confirmation window, which is a deliberate trade-off against false alarm rate. In the 3 test events where FWI exceeded 30 at ignition time, the single-camera dispatch pathway (confidence > 0.92 override) was triggered, reducing alert latency to 46 s (median) for those events.

Comparison with conventional aerial patrol detection in the Chios context: the Hellenic Fire Corps operates helicopter patrol over Chios on a 45-minute recurrence during Extreme danger days, and 90-minute recurrence on Very High danger days. Analysis of the 22 detected fires on Chios between 2018 and 2023 shows a mean patrol-based detection time of 23 minutes (range 8-67 minutes). The EgeoFireNet 87-second median alert latency represents a 15.9× improvement in mean detection time, and a factor of 5.5 improvement at the best-case patrol detection (8 minutes). This quantified detection improvement translates, using the fire outcome analysis cited in the introduction, from a 71% success rate for < 5-minute detection to a projected rate substantially above the 30% success rate achievable at 23-minute detection under High danger conditions.

For the 6 test events with sufficient burned area for spread prediction evaluation, the mean Sorensen-Dice similarity coefficient is 0.76 (range 0.68-0.84). The false alarm ratio (predicted burning area outside actual perimeter) averages 0.18, indicating an 18% overprediction of eventual burned area at 15 minutes. The miss ratio (actual burned area outside predicted perimeter) averages 0.09, indicating that the CA model misses 9% of the actual 15-minute spread extent. The overprediction bias is directionally preferred by fire managers: a conservative overestimate of spread extent supports pre-positioning resources ahead of the fire head, the primary tactical application of short-horizon spread forecasts. The highest-accuracy prediction (Dice = 0.84) occurred in Test Event 9 under steady 22 km/h northerly wind with uniform Pinus brutia fuel, where the simple wind-driven Rothermel model is most physically representative. The lowest-accuracy (Dice = 0.68, Event 6) occurred under rapidly shifting wind during a Meltemi gusting event, where wind direction changed 38° within the 15-minute forecast horizon a systematic limitation of any wind-driven CA model without real-time wind field updates.

CA forecast generation time of 4.2 seconds on the Jetson Orin NX 8-core CPU confirms real-time operational viability: the 4.2-second computation leaves 10.8 seconds for alert package assembly within the 15-second target from detection to first EOC alert API call. Memory allocation for the 1,240 × 1,000 ×

5-variable CA state array is 24.8 MB, well within the Jetson Orin NX 16 GB RAM capacity. The implementation uses NumPy 1.24 vectorised operations; profiling confirms that the wind-driven spread computation (97% of total CPU time) is dominated by the elementwise Rothermel model evaluation, which could be further accelerated by GPU offloading (estimated 0.8 s on the Jetson Orin NX GPU) in future implementations.

Over the 6-month operational period (June-November 2024, 43 cameras × 6 months = 258 camera-months), the system generated 1,847 Level 1 alerts, 63 Level 2 alerts, and 7 Level 3 dispatch alerts. Of the 63 Level 2 alerts, 12 were confirmed true fire detections (8 actual fires plus the 4 test events conducted during this period), 34 were confirmed false positives (primarily cloud shadow transitions in hazy conditions), and 17 remain unresolved (camera not reachable for confirmation). The Level 2 false alarm rate of 0.023 per camera per hour (34 false positives / [43 cameras × 24 h/day × 6 months × 30 d/month] = 0.023) is within the < 0.05 operational tolerance specified by the Hellenic Fire Corps for fixed camera networks.

System availability over the 6-month period was 97.8% (4 cameras out of service for > 24 hours due to: 2 solar charge controller failures, 1 4G router failure, 1 camera PTZ motor failure). No unrecovered data loss occurred due to local SD card buffering of all detection events. The primary maintenance activity was quarterly cleaning of camera domes and heat exchanger fins in the dusty and salt-marine environment, performed by the Chios Regional Fire Service maintenance team. Camera PTZ calibration drift (maximum 0.8° heading, 0.3° elevation) was automatically corrected by the weekly landmark-based calibration routine using GPS-surveyed mountaintop reflectors visible from each tower.

5. Discussion

The 5.5 percentage-point mAP@0.5 advantage of Swin-T over YOLOv8-nano on the Aegean pine test set is operationally significant: it corresponds to detecting a proportionally larger fraction of early-stage fire events before they grow beyond the suppression-feasible threshold. The advantage is most pronounced for small-scale smoke detections, precisely the category that matters most for early detection performance. The trade-off is a 2.2× higher inference latency (44.6 ms vs. 20.3 ms for YOLOv8-nano) and 3.2× higher parameter count (28M vs. 3.2M). At 22.4 FPS, the Swin-T model processes 1,344 frames per minute per camera substantially more than required for a 30-second temporal resolution detection system, leaving substantial compute headroom. The parameter count is not a binding constraint for the Jetson Orin NX with 16 GB RAM.

Future work will explore the training of larger Swin variants (Swin-S, 50M parameters; Swin-B, 88M parameters) on the expanded EgeoFireNet corpus to assess whether the detection performance ceiling has been reached with Swin-T, or whether the accuracy gap relative to the theoretical optimum can be further reduced. A further avenue is the integration of temporal information through video transformer architectures (VideoSwin, TimeSformer) that could exploit the temporal consistency of fire and smoke development to reduce false positives from transient cloud shadow and vehicle exhaust events.

The Rothermel-CA spread model achieves Dice = 0.76 under favourable steady-wind conditions but degrades to 0.68 under rapidly shifting winds, representing the fundamental limitation of any wind-driven CA model using point weather measurements. Three improvement pathways are identified: (i) assimilation of high-frequency (10-second) LPWAN wind sensor data across the network to update the CA wind field in near-real-time, reducing the lag between actual and modelled wind conditions; (ii) data-driven correction of the Rothermel model using historical fire-spread observations from Chios and comparable Aegean islands to recalibrate the fuel model parameters for local Pinus brutia stands; and (iii) machine learning emulation of the full CA spread process using a graph neural network trained on CA simulation outputs, enabling probabilistic ensemble spread forecasting that captures wind uncertainty within the same computation time [5].

The spotting model parameterisation (P_{spot} = 0.003 per boundary cell per time step at FWI > 35) was not validated against independent Chios fire event data due to the limited number of observed spotting events in the test fire corpus. Dedicated spotting model validation using UAV-observed historical fires from the 2021 and 2023 Chios seasons is planned, and the spotting probability will be recalibrated if available data warrant.

The capital cost of the EgeoFireNet 43-tower Chios installation is estimated at EUR 1.2 million (EUR 27,900 per tower including civil works, tower, camera, edge compute, power, and communications). The annual operating cost is approximately EUR 85,000 (camera cleaning, maintenance visits, communication contracts, software updates). The primary value proposition is the reduction in suppression cost attributable to earlier detection: the Greek Forest Protection Study [1] estimates that fires detected within 5 minutes incur mean suppression costs of EUR 12,000 per ha, versus EUR 340,000 per ha for fires not detected until 30+ minutes. Using the EgeoFireNet median detection time of 87 seconds and assuming detection of 3-5 fires per season on the monitored 12,400 ha, the avoided suppression cost relative to the 23-minute patrol baseline is estimated at EUR 1.8-3.2 million per season a payback period of well under one year on the installation cost.

Scaling to the 3.5 million ha of extreme/very-high risk Greek forest at Chios network density (43 towers / 12,400 ha = 3.5 towers/1,000 ha) would require approximately 12,200 towers at an estimated capital investment of EUR 340 million and annual operating cost of EUR 24 million. While substantial, this investment is comparable to one to two years of Greek aerial firefighting operational expenditure during severe seasons, and provides continuous coverage that aerial patrol cannot. A phased deployment prioritising the 1.0 million ha of highest-risk zone (Greek islands plus Attica/Corinthia/Laconia coastal zone) would require approximately 3,500 towers at EUR 98 million capital cost.

6. Conclusions

EgeoFireNet, a 43-camera edge-AI forest fire monitoring system combining Swin Transformer detection with cellular automaton spread prediction, has been presented and validated in an operational Aegean pine forest deployment on Chios island, Greece. The system represents the first documented operational deployment and comprehensive test-fire validation of a transformer-based fire camera network in the Eastern Mediterranean. Key quantitative findings are: (i) Swin-T fire/smoke detection achieves mAP@0.5 = 0.889 and F1 = 0.912 at 22.4 FPS on the Jetson Orin NX (TensorRT INT8), outperforming the YOLOv8-nano baseline by 5.5 percentage-points in mAP@0.5; (ii) CA fire spread prediction achieves Dice = 0.76 for 15-minute forecasts relative to validation test fires, with 4.2-second generation time on the edge node CPU; (iii) end-to-end alert latency from ignition to EOC dispatch is 87 seconds (median), a 15.9× improvement over conventional aerial patrol detection; (iv) operational false alarm rate of 0.023 Level 2 alerts per camera per hour is within the Hellenic Fire Corps tolerance; and (v) system availability of 97.8% over the 6-month operational period confirms deployment-environment robustness.

The EgeoFireNet architecture provides a validated, replicable technical baseline for national-scale deployment of transformer-based fire detection networks across the Greek high-risk forest zone, with a quantified economic case supporting cost-recovery within a single fire season. Future development priorities are: night-time performance improvement through thermal imaging integration and dedicated NIR model branches; spread prediction accuracy improvement through real-time wind field assimilation and ML-based Rothermel calibration; SAR satellite data integration for overnight gap coverage; extension of the training dataset to additional Eastern Mediterranean fuel types; and longitudinal evaluation of detection-outcome relationships using multi-season suppression records.

REFERENCES

[1] European Forest Fire Information System (EFFIS), Annual Fire Danger Season Analysis 2023. Ispra: European Commission Joint Research Centre, 2024.

[2] A. Mohapatra and T. Trinh, "Early wildfire detection technologies in practice A review," Sustainability, vol. 14, no. 19, pp. 12270, Sep. 2022.

[3] T. Nawaz, O. Maguire, K. G. Clarke, R. Holsgrove, A. Al-Mamoon, and R. Bhakta, "Forest fire detection using distilled vision transformers for deployment on edge devices," preprint, Staffordshire University, 2024.

[4] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proc. ICCV, pp. 10012-10022, 2021.

[5] B. H. Hopkins, L. O. O'Neill, F. A. Afghah, A. R. Razi, A. W. Watts, P. F. Fule, and J. C. Coen, "FLAME 2: Fire detection and modeLing: Aerial multi-spectral imagE dataset," IEEE DataPort, 2022.

[6] R. Ghali and M. A. Akhloufi, "Deep learning approaches for wildland fires remote sensing: Classification, detection, and segmentation," Remote Sens., vol. 15, no. 7, pp. 1821, Mar. 2023.

[7] R. C. Rothermel, A Mathematical Model for Predicting Fire Spread in Wildland Fuels. Ogden: USDA Forest Service Research Paper INT-115, 1972.

[8] L. Mu, Y. Yang, B. Wang, Y. Zhang, N. Feng, and X. Xie, "Edge computing-based real-time forest fire detection using UAV thermal and color images," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 18, pp. 6760-6771, 2025.

[9] G. Coleman, W. Salter, and M. Walsh, "OpenWeedLocator (OWL): An open-source, low-cost device for fallow weed detection," Sci. Rep., vol. 12, no. 1, pp. 170, Jan. 2022.

[10] A. Stergiopoulos, G. Botzoris, and C. Basbas, "Optimal placement of forest fire detection cameras using geographic information systems and analytic hierarchy process," Appl. Sci., vol. 13, no. 6, pp. 3563, Mar. 2023.

[11] J. Li, R. Xu, and Y. Liu, "An improved forest fire and smoke detection model based on YOLOv5," Forests, vol. 14, no. 5, pp. 833, May 2023.

[12] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, and S. W. Baik, "Efficient deep CNN-based fire detection and localization in video surveillance applications," IEEE Trans. Syst. Man Cybern. Syst., vol. 49, no. 7, pp. 1419-1434, Jul. 2019.

[13] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in Proc. ICLR, 2021.

[14] M. Wang, X. Li, and J. Zhao, "Real time fire and smoke detection using vision transformers and spatiotemporal learning," Sci. Rep., 2026.

[15] K. Das, J. Poovvancheri, S. Flesc, R. C. Calidonn, and D. Chen, "Emerging trends in wildfire detection through the lens of computer vision," IEEE Access, vol. 14, pp. 20201-20228, 2026.

[16] L. Papa, P. Russo, I. Amerini, and L. Zhou, "A survey on efficient vision transformers: Algorithms, techniques, and performance benchmarking," IEEE Trans. Pattern Anal. Mach. Intell., pp. 1-20, 2024.

[17] A. Khan, B. Hassan, S. Khan, R. Ahmed, and A. Abuassba, "DeepFire: A novel dataset and deep transfer learning benchmark for forest fire detection," Mob. Inf. Syst., vol. 2022, pp. 5358359, 2022.