This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3712753

HQDL: A Hybrid Quantum–Deep Learning Framework for Real-Time Fire Detection Using UAV Remote Sensing

Sufyan Danish, Waseem Ullah, *Member, IEEE,* Muhammad Fayaz, L.Minh Dang, Hyoung-Kyu Song, Hyeonjoon Moon

***Abstract*—Fire detection in remote-sensing imagery plays an important role in reducing the impact of various types of fire, including forest fires and other fire-related hazards. Traditional CCTV-based systems have limited viewpoints, restricted coverage, and poor ability to monitor large or remote areas. UAV-based remote sensing offers better mobility and wider coverage, but onboard storage and computing limitations make real-time fire detection challenging. Although CNN and Vision Transformer-based methods achieve high accuracy, their computational complexity often restricts deployment on resource-constrained UAV edge devices. To address these limitations, this paper presents a hybrid-quantum deep learning (HQDL) fire detection framework that integrates the MobileViT lightweight backbone with feature-extraction capabilities enabled by quantum-enhanced processing. Previous studies have integrated classical CNN backbones with additional processing layers for fire detection. In contrast, our proposed HQDL framework incorporates quantum convolution (QConv) and quantum dense (QDense) layers using parameterized quantum circuits. This enables the model to capture complex nonlinear dependencies and higher-order feature interactions that classical hybrid layers cannot represent. The framework was evaluated on FD, FLAME, and ADSF datasets, achieving accuracies of 98.42%, 87.08%, and 96.83%, respectively. HQDL maintains a compact model size of 19.3 MB with 5.03 million trainable parameters and achieves 28.42 FPS on CPU, 86.48 FPS on GPU, and 10.49 FPS on Raspberry Pi. Visual analysis using class activation maps shows that the model focuses on fire-relevant regions while suppressing background distractions. Overall, HQDL provides a scalable and computationally efficient framework for real-time UAV-based fire detection.**

***Index Terms*—Fire Detection, Quantum Computing, Hybrid Model, Quantum Deep Learning, UAV, Drone**

I. Introduction

DISASTER management has emerged as a critical interdisciplinary research domain covering computer engineering, health sciences, and environmental engineering, underscoring the importance of integrated approaches to mitigate complex risks. Disasters are generally classified into natural, such as floods, earthquakes, and fires, and human-made incidents, including nuclear plant accidents, acts of terrorism, and hazardous material releases [1], [2]. Fires are among the most devastating disasters, causing widespread destruction to human communities, wildlife environments, and ecosystems [3]. According to data from the Korean National Fire Agency, there were 40,030 fires in South Korea in 2019, causing 284 casualties and 2,219 injuries. On average, 110 fire incidents and approximately 0.8 fire-related deaths were recorded per day, leading to property losses estimated at KRW 2.2 billion [4]. In 2025, more than 102 million hectares of land had already burned by mid-year, with Africa accounting for nearly half of the total, highlighting the ongoing global importance of the wildfire crisis [5]. Similarly, according to "Our World in Data", based on the Global Wildfire Information System, the annual area burnt by wildfires globally has ranged between 300 to 400 million hectares per year since 2012 [6]. Timely fire detection and classification are important for minimizing damage and enabling a rapid response to fire emergencies. Traditional fire detection systems based on CCTV cameras have limited coverage areas and are constrained by their static positioning. This fails to detect fires outside their blind spots, thereby reducing their reliability. To overcome this limitation, UAVs have become a popular alternative for monitoring larger, more remote areas in real time and, therefore, provide greater flexibility by covering large areas and inspecting potential fire sources at closer range [7]. However, the primary constraints for UAVs are their limited onboard storage and computing resources. Therefore, it is impractical to implement a resource-intensive deep learning model on board UAVs to enable real-time fire detection.

In recent years, several methods for fire detection have been developed. Initially, fire detection methods employed scalar sensor technologies, i.e., smoke detectors, flame detectors, and temperature detectors [8]. Scalar sensors are relatively inexpensive and easy to install. However, they are generally used to monitor indoor areas. Their ability to detect fires across large outdoor areas is limited in accuracy, and they typically require human intervention [9]. In response to the limitations of scalar sensor technology, researchers have increasingly turned to vision-based fire detection methods. Vision-based fire detection methods employ visual data collected from cameras and sensors to provide a wider field of view, provide real-time monitoring capabilities, and provide critical fire-related information, such as the size of the affected area, fire progressions, and environmental conditions (e.g., smoke), all of which reduce the amount of human intervention required compared to previous methods. Nevertheless, there are still many challenges associated with vision-based fire detection methods, including the diversity of fire appearance

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2024-00461244) and by National Research Foundation of Korea (NRF) grant funded by the Korea government, Ministry of Science and ICT (MSIT) (RS-2026-25473871).

S. Danish, M.Fayaz and H. Moon are with the Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea (e-mail: sufyandanish@sju.ac.kr; muhammadfayaz@sju.ac.kr; hmoon@sejong.ac.kr). L. M. Dang and H.K. Song are with the Department of Information and Communication Engineering and Convergence Engineering for Intelligent Drone, Sejong University, Seoul 05006, Republic of Korea (e-mail: minhdl@sejong.ac.kr; songhk@sejong.ac.kr) W. Ullah with the Department of Machine Learning, Mohamed Bin Zayed University of Artificial Intelligence, United Arab Emirates.

characteristics (e.g., shapes, sizes, colors) and the presence of smoke and other occlusions that impede accurate classification and localization. To improve the accuracy of vision-based fire classification methods, researchers have developed two approaches, including traditional machine learning (TML) and deep learning (DL).

Early TML fire detection methods employed hand-crafted feature extraction techniques that primarily relied on color analysis and image processing to detect fire and smoke [10], [11]. Despite this, initial TML approaches suffered from high false alarm rates due to the diversity and variability of fire characteristics. The early TML techniques were subsequently improved by incorporating motion features (optical flow and spatiotemporal analysis) to capture the dynamic nature of fires better [12]. Xu et al. [13] introduced a Modified Pixel Swapping Algorithm, combined with mixed-pixel separation and threshold-weighted fusion, to improve forest fire detection. Their algorithm accomplished higher accuracy and lower false alarm rates than existing TML methods. However, TML-based methods will continue to have limitations in dynamic environments, such as camera movement, varying light conditions, or when moving objects mimic the behavior of a fire. Such environmental inconsistency compromises the reliability of TML methods, and therefore, they are generally unsuitable for real-time applications in outdoor settings. Thus, recent research has focused on developing DL models that are more robust and adaptable to complex fire detection tasks. Advances in computer vision using DL methods have significantly improved the performance of fire detection in remote sensing [14]. Convolutional Neural Networks (CNNs) have become popular for fire detection due to their ability to effectively extract local features, leading to increased accuracy [15], [3], [16]. Although lightweight CNN architectures have been proposed to overcome the computational constraints of CNNs [17], [18], they lack sufficient representation capacity to detect small-scale or distant fires under adverse environmental conditions accurately. Some researchers have used CNNs with attention mechanisms to enhance performance further [19], [20], [21], [22], [23], [24]. However, hybrid attention-based CNN models are computationally expensive during both training and inference, limiting their use in real-time applications on resource-constrained platforms such as UAVs and IoT devices, which lack the processing power needed for high-resolution, low-latency analysis. Vision Transformers (ViTs) address some of these limitations by modeling long-range dependencies and capturing global scene information through self-attention mechanisms applied across image patches. Chen et al. [25] Combining CNN and transformers for classifying fire and smoke scenes in remote sensing images leverages the strengths of both architectures to enhance classification accuracy and efficiency. Yar et al. [26] introduced a Vision Transformer (ViT)-inspired model incorporating shifted patch tokenization for spatial detail, locality-based self-attention to optimize the softmax temperature, and dense layers to reduce model complexity. Other researchers have also used transformer-based fire detection and classification models, such as [27], [28], [29]. However, ViTs also suffer from high computational overhead due to the quadratic complexity of the self-attention mechanism. The limitations of these approaches prevent their deployment in real time on resource-constrained UAV platforms and in remote sensing fire-detection applications, where efficiency is crucial for quick response times.

To address these challenges, quantum machine learning (QML) has emerged as a promising direction for enhancing classical deep learning models through quantum-inspired feature transformations. In this study, we propose a hybrid quantum deep learning (HQDL) framework for UAV-based fire detection in remote sensing imagery. The proposed framework is not designed as a simple extension of MobileViT with additional layers. Rather, it introduces a dual-stage quantum-enhanced feature-learning strategy. First, MobileViT is used to extract compact local–global spatial representations from UAV imagery. Second, a Quantum Convolution (QConv) layer processes the spatial feature maps to enhance local nonlinear fire-related patterns. Third, a Quantum Dense (QDense) layer operates on the compact global representation to model higher-order feature dependencies before classification. This design differentiates HQDL from existing CNN-attention, CNN-transformer, and lightweight hybrid fire detection models, which mainly rely on classical attention, feature fusion, or convolutional enhancement modules. The main contributions of this paper include:

- We Integrate Quantum Convolution (QConv) and Quantum Dense (QDense) layers with MobileViT in a comprehensive dual-stage framework, significantly enhancing both local and global feature representations. This extensive integration enables the model to capture complex nonlinear and higher-order relationships in UAV-based fire imagery, providing a substantial improvement in modeling capability and accuracy over existing hybrid deep learning methods.
- We differentiate the proposed HQDL framework from conventional hybrid CNN-transformer, CNN-attention, and lightweight fire detection models by applying quantum-enhanced processing at both spatial and global representation levels.
- We conduct an expanded ablation study, including w/o MobileViT, w/o QConv, w/o QDense, w/o both quantum layers, single-QConv, single-QDense, and a same-depth classical counterpart, to verify the individual and combined contributions of each component.
- Extensive experimental validation of the proposed HQDL model on multiple fire detection datasets demonstrates that the framework achieves superior classification accuracy and robustness compared with state-of-the-art methods, highlighting the effectiveness of the dual-stage quantum-enhanced feature learning in capturing complex fire patterns.
- Grad-CAM visualizations indicate that the model consistently highlights high-intensity flame regions, providing interpretable and explainable fire localization.

The rest of the paper is organized as follows. Section II presents the design and components of the proposed HQDL model. Section IV details the quantum circuit design, and Section V describes the training process, including the model's configuration parameters, the characteristics of the datasets used, and the results of the experiments. Finally, Section VI concludes the paper, discusses the main limitations of the study, and outlines possible avenues for further research.

II. METHODOLOGY

The challenges of fire detection in remote sensing imagery require models that are compact, computationally efficient, and robust to different environmental conditions. To achieve this, we propose a new HQDL architecture that integrates a pre-trained MobileViT backbone with dual-stage quantum-enhanced processing. The MobileViT backbone provides both fine-grained local image features and long-range contextual information from UAV imagery. These features are then passed through a Quantum Convolution (QConv) layer, which uses parameterized quantum circuits (PQCs) to process the spatial patches of the input image, thereby incorporating non-linear relationships into the feature representation. Following this QConv layer, a GAP (Global Average Pooling) layer is applied to reduce the feature representation's dimensionality. This GAP layer will be used to prepare the feature representation for a classical fully connected layer. Finally, this classical fully connected layer will be used to prepare the feature representations for use in the next stage of the proposed architecture, i.e., a Quantum Dense Layer (QDense). This Quantum Dense Layer will apply another PQCs to the feature representations to capture higher-order interdependencies, and dropout will be added to the layer to ensure that the model remains robust. The final classical output layer produces logit values, which are transformed into probability values via a sigmoid activation and thus provide fire/no-fire classification results. The proposed architecture is therefore differentiated at two levels. At the spatial level, QConv receives Mobile-ViT feature maps and performs quantum-enhanced nonlinear transformation of local feature responses. At the global level, QDense receives the aggregated representation and models higher-order dependencies before classification. This two-stage design differs from simply increasing the depth of MobileViT or adding a conventional fully connected head. It also differs from existing hybrid CNN-transformer or attention-based fire detection models because the proposed model incorporates parameterized quantum circuits into both spatial and global representation learning. HQDL leverages quantum-inspired feature transformations at both local and global levels, enhancing expressiveness without increasing computational cost. The resulting model is lightweight, UAV-deployable, and capable of detecting small fire signatures under complex remote-sensing conditions. An illustration of the HQDL pipeline is shown in Figure 1.

III. MODEL ARCHITECTURE

A. MobileViT Backbone

The first stage of the proposed framework is a Mobile-ViT backbone that serves as a feature-extraction module. The mobile-vit architecture is a hybrid architecture that combines the advantages of CNNs and transformers. Convolutional layers are used early to capture fine-grained local spatial features such as edges, textures, and small fire signatures. These local features are important in remote sensing imagery where fires may appear as small or fragmented regions. Following the convolutional blocks, the Mobile-ViT architecture incorporates transformer layers that model global dependencies across the entire image. In contrast with convolutions, which operate on receptive fields of limited size, transformers use self-attention mechanisms to relate distant pixels and regions. This allows the backbone to capture contextual information, such as smoke plumes or surrounding vegetation patterns, which are important for distinguishing fire events from other bright objects.

Formally, given an input image $X \in \mathbb{R}^{B\times 3\times 256\times 256}$, the MobileViT (MVT) backbone outputs feature maps:

$$F = \mathrm{MVT}(X), \quad F \in \mathbb{R}^{B\times 256\times 7\times 7} \tag{1}$$

where B is the batch size, 256 is the number of output channels, and the spatial resolution is reduced to 7×7. This compact representation balances computational efficiency with rich feature extraction.

B. Quantum Convolution Layer

Let $X \in \mathbb{R}^{H\times W\times C}$ be the input feature map, where H and W are spatial dimensions and C is the number of channels. In classical convolution, a kernel $K \in \mathbb{R}^{k\times k\times C}$ is applied to each patch $X_p \in \mathbb{R}^{k\times k\times C}$ as:

$$y = f\left(\sum_{i,j,c} K_{i,j,c} \cdot X_p(i,j,c) + b\right), \tag{2}$$

where f is a nonlinear activation function and b is a bias term. Each patch is transformed by the kernel into a scalar or vector, producing the feature map.

In HQDL, the Quantum Convolution (QConv) layer replaces classical convolution with a quantum-inspired transformation. Each spatial patch X_p of the MobileViT feature map F is encoded into a quantum circuit. This encoding uses angle embedding, where classical feature values are mapped to rotation angles of qubits:

$$|\psi(X_p)\rangle = \bigotimes_{i=1}^{q} R_x(\phi_i)|0\rangle, \tag{3}$$

Here, $q = 5$ is the number of qubits, and ϕ_i are the classical feature values from the patch. Once encoded, the qubits are entangled using BasicEntanglerLayers, which apply parameterized rotations followed by CNOT gates. This entanglement enables the circuit to capture correlations between features that classical convolutions may miss. The output of the circuit is obtained by measuring the expectation values of Pauli-Z operators:

$$z_i(j,k) = \langle \psi(X_p(j,k)) | Z_i | \psi(X_p(j,k)) \rangle, \quad i = 1, \ldots, q \tag{4}$$

The outputs from all patches are aggregated to produce the quantum-enhanced feature map:

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3712753

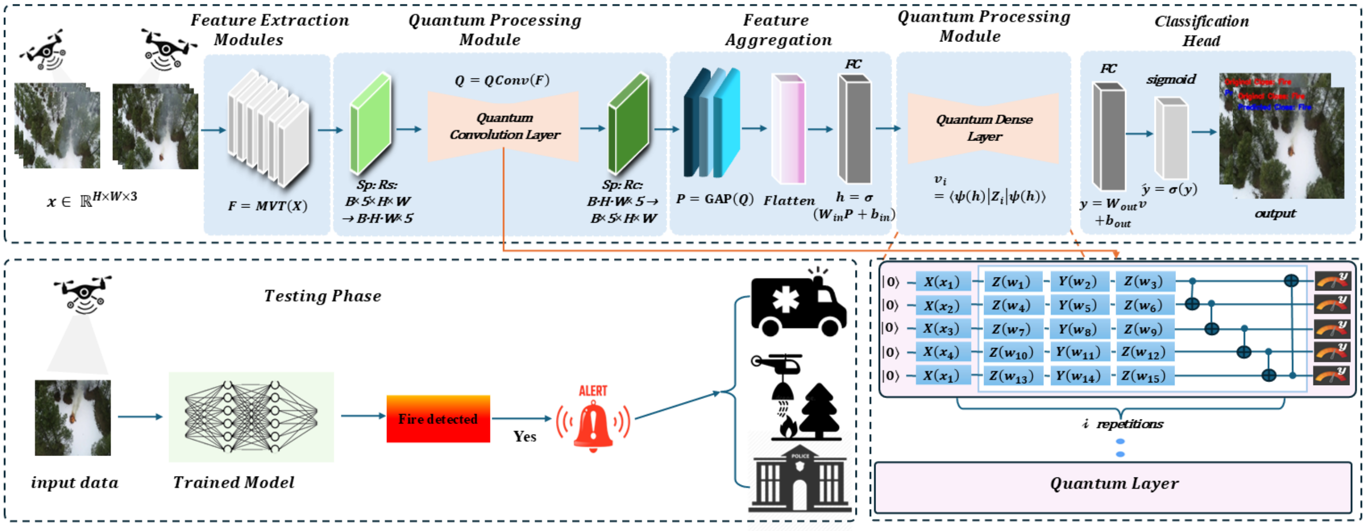

Fig. 1. Proposed hybrid framework for UAV-based fire classification using MobileViT with quantum-enhanced learning. MobileViT extracts lightweight local–global representations, which are processed by a quantum convolution (quanvolution) layer to generate quantum feature maps from spatial patches. Aggregated features are then passed to a quantum dense (variational quantum) layer to model global feature interactions via entanglement and produce classification logits for final classification.

$$Q = \mathrm{QConv}(F), \quad Q \in \mathbb{R}^{B\times q\times 7\times 7}. \tag{5}$$

Mathematically, QConv parallels classical convolution by processing spatial patches, but differs by applying a trainable quantum circuit that acts as a flexible kernel to extract richer, nonlinear feature representations. The combination of angle embedding, entanglement, and measurement enables QConv to capture local nonlinear patterns, directly replacing classical convolution while enhancing MobileViT's feature maps with quantum-inspired features. This formulation explicitly demonstrates How QConv replaces classical convolution and provides the additional representational power necessary to model complex fire patterns in UAV imagery.

C. Global Pooling and Flattening

To reduce the dimensionality of the quantum feature maps, a global average pooling (GAP) operation is applied. GAP aggregates spatial information across the 7×7 grid, producing a compact vector for each channel:

$$P = \mathrm{GAP}(Q), \quad P \in \mathbb{R}^{B\times q} \tag{6}$$

This operation ensures that each channel contributes equally to the representation, while discarding spatial redundancy. The pooled features are then flattened into a 2D tensor, preparing them for fully connected layers.

D. Classical Fully Connected Layer

The flattened vector P is projected into a higher-dimensional space through a classical fully connected (FC) layer. This step bridges the quantum convolutional output with the quantum dense layer:

$$h = \sigma(W_{in}P + b_{in}), \quad h \in \mathbb{R}^{B\times n} \tag{7}$$

where W_{in} and b_{in} are trainable parameters, and σ denotes the ReLU activation function. This transformation expands the feature space, enabling richer representations before quantum processing.

E. Quantum Dense Layer

The Quantum Dense Layer (QDense) applies a parameterized quantum circuit to the vector h. Each input vector is encoded into qubit states, followed by trainable rotations and entangling gates:

$$|\psi(h)\rangle = U(\theta)R_x(h)|0\rangle^{\otimes q} \tag{8}$$

where $U(\theta)$ is the variational unitary transformation with trainable parameters θ.

The output is obtained by measuring expectation values:

$$v_i = \langle\psi(h)|Z_i|\psi(h)\rangle, \quad v \in \mathbb{R}^{B\times q} \tag{9}$$

This layer introduces nonlinear quantum transformations that enhance the model's expressive power, enabling it to capture subtle patterns in fire imagery.

F. Classification layer

To prevent overfitting, dropout regularization with probability $p = 0.5$ is applied to the quantum dense output v. The final fully connected layer maps these features to logits:

$$y = W_{out}v + b_{out}, \quad y \in \mathbb{R}^{B\times 2} \tag{10}$$

A sigmoid activation converts logits into probabilities:

$$\hat{y} = \sigma(y), \quad \hat{y} \in [0, 1] \tag{11}$$

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3712753

where $\hat{y}$ represents the probability of fire presence. This binary output is suitable for fire/no-fire classification tasks in remote sensing.

IV. Quantum Circuit Design

The HQDL framework integrates Quantum Convolution (QConv) and Quantum Dense (QDense) layers implemented via parameterized quantum circuits (PQCs). These layers serve as nonlinear feature transformers within the hybrid architecture, complementing the MobileViT backbone for UAV-based fire detection. Their design explicitly accounts for qubit count, circuit depth, and noise to balance expressiveness and classical simulation feasibility.

Each quantum layer consists of three stages, including embedding, variational transformation, and measurement. This structure maps classical features into quantum states, applies trainable unitary operations, and collapses them back into classical outputs for downstream processing.

A. Embedding

Classical features are encoded into quantum states using angle embedding, where each value rotates a qubit around the Bloch sphere. For a feature vector $x = (\phi_1, \phi_2, \ldots, \phi_q)$:

$$|\psi(x)\rangle = \bigotimes_{i=1}^{q} R_x(\phi_i)|0\rangle \tag{12}$$

Here, $R_x(\phi_i)$ is a rotation around the X-axis by angle ϕ_i, and $|0\rangle$ is the ground state, while the $q = 5$ denotes the number of qubits per feature vector. This mapping ensures that continuous classical values are smoothly represented in quantum states, enabling nonlinear transformations in the variational stage.

B. Variational Gates

The embedded quantum states are processed by parameterized rotational and entangling CNOT gates:

$$U(\theta) = \prod_{l=1}^{L} \left(R_z(\theta_l)R_y(\theta_l)R_x(\theta_l) \cdot \text{CNOT}\right) \tag{13}$$

where $L = 2$ represents the circuit depth, and θ are trainable parameters. Rotations explore each qubit's Hilbert space, while entangling CNOTs capture correlations between features.

C. Measurement

Finally, the quantum state is collapsed into classical outputs by computing expectation values of Pauli-Z operators:

$$v_i = \langle\psi(x)|Z_i|\psi(x)\rangle, \quad i = 1, \ldots, q \tag{14}$$

The outputs from PQCs are concatenated to form v, which is passed to the classical dense layer for final classification. This design ensures that quantum circuits act as nonlinear feature transformers, bridging quantum and classical representations.

D. Motivation of Design and Noise Considerations

These quantum circuit designs are used as non-linear feature transformers in the hybrid architecture of the network. In the embedding stage, classical features are converted into quantum states in the variational gate stage; these quantum states are enriched with trainable correlations, and in the measurement stage, classical output is generated that reflects the quantum transformation. These three processing stages allow the QConv and QDense layers to identify complex and subtle spatial patterns (e.g., faint hotspots, diffuse smoke, or irregular boundaries) in fire imagery. Thus, the quantum-enhanced processing complements the MobileViT backbone of this model, enabling it to be both compact and perform well across a variety of difficult remote-sensing conditions.

All experiments for the quantum layers were conducted on ideal classical simulators, providing a controlled environment to evaluate the representational advantages of the quantum circuits. While actual quantum hardware may introduce decoherence and gate errors, classical simulations enable a rigorous assessment of QConv and QDense's ability to extract nonlinear local and global features from UAV fire imagery. The chosen number of qubits (q = 5) and circuit depth (L = 2) were selected to balance expressiveness and computational feasibility. Furthermore, comprehensive ablation studies demonstrate that both QConv and QDense layers significantly improve HQDL's for fire detection. For real-time deployment, the model was also tested on a Raspberry Pi, demonstrating overall performance and validating the design's effectiveness for complex fire detection tasks.

V. RESULTS AND DISCUSSIONS

A. Experimental Setup and Evaluation Metrics

This section describes the experimental setup, the fire detection benchmarks, and the metrics used to evaluate the performance of the proposed approach and compare it to state-of-the-art methods. An ablation experiment has been performed to assess the contributions of each component in the proposed architecture. All experiments were run on a Windows 11 workstation equipped with an NVIDIA processor, 64GB of system RAM, and an NVIDIA GeForce RTX 3090 Ti GPU. The HQDL model was implemented using PyTorch and PennyLane. Training was performed for 50 epochs at a batch size of 16 across all datasets. The Adam optimizer with a learning rate of 0.0001 was used for optimization. Standard classification metrics have been used to evaluate the proposed model's performance, including accuracy, precision, recall, and F1 score [15]. These metrics are defined as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{21}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{22}$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{23}$$

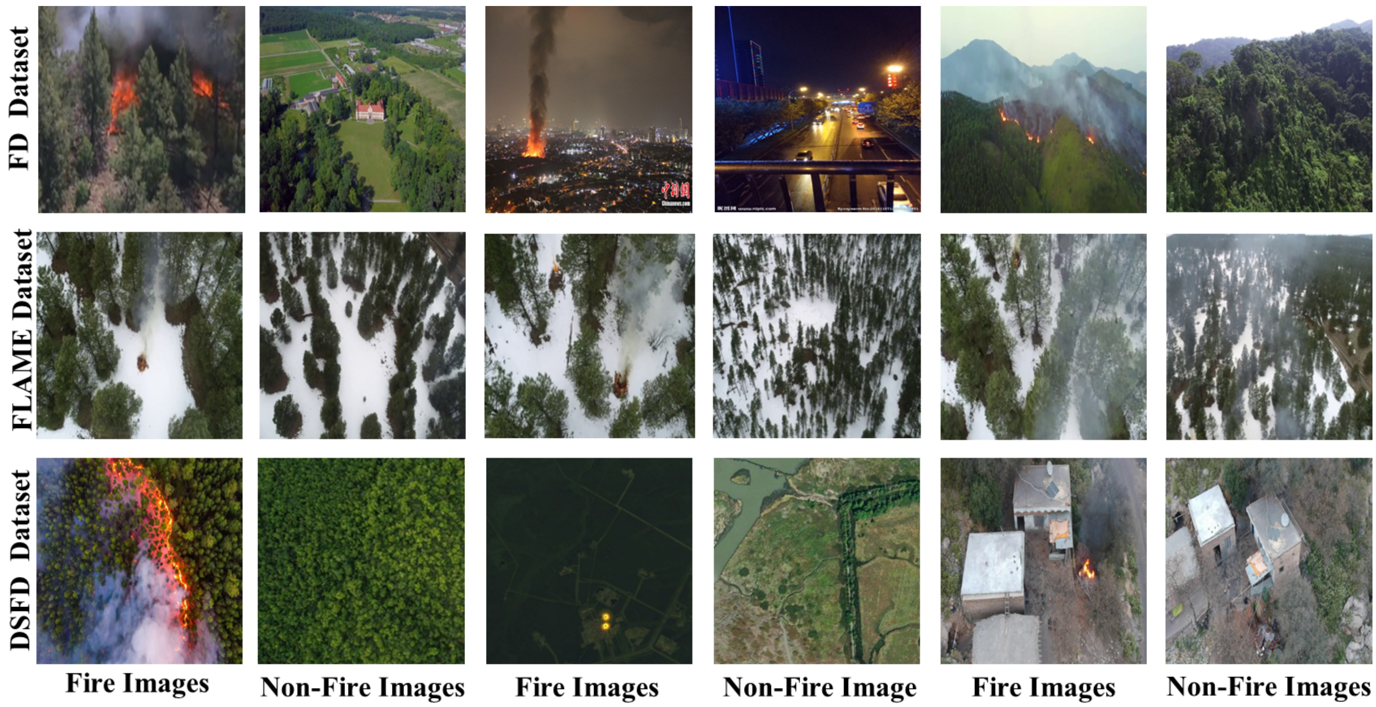

Fig. 2. Representative fire and non-fire image samples from the FD, FLAME, and ADSF datasets, illustrating diverse remote-sensing scenarios with variations in scene complexity, background clutter, illumination conditions, and fire scale, as encountered in UAV-based fire monitoring applications.

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively.

B. Datasets

To evaluate the proposed HQDL framework for drone-based fire detection, three data sets were used. Each dataset was chosen to encompass different aspects of fire imagery, ranging from UAV data acquisition to a variety of satellite image scenes, to ensure the model would be both robust and generalized.

1) FLAME Dataset:: The FLAME dataset was developed to address the limitations of CCTV-based fire datasets, which often suffer from limited coverage, limited mobility, and occlusion. The Flame dataset has been collected using multiple drones with various types of heterogeneous sensors, including FLIR Vue Pro R, Phantom 3, and Zenmuse X4S cameras. Only the RGB camera data were considered in this paper, which contains 39,375 individual RGB image frames, each with a resolution of 254 x 254 pixels at 96 dpi. The entire dataset was split into two parts, 80% for training and 20% for validation. Shamshora et al. [30] created a completely independent test dataset containing 8,617 RGB images. The use of these images on FLAME provides a basis for testing real-world drone-based conditions for fire detection.

2) Aerial Drone and Satellite Fire (ADSF):: The ADSF dataset [20] integrates drone and satellite imagery to produce a comprehensive benchmark for fire scene classification. There are two classes, fire and non-fire, each with 3000 images. Drone images were obtained with a DJI Spark, recording from multiple angles and at altitudes ranging from 10 m to 70 m. Furthermore, to improve the dataset's diversity, a frame-skipping mechanism was implemented, eliminating redundant frames, resulting in 1000 unique drone images. The remaining samples were obtained from satellite imagery that captured several environmental conditions (e.g., daytime, nighttime, fog, cloud cover), drone images that did not contain any fire regions, and other images from online repositories. The dataset was randomly split into 70% for training, 20% for validation, and 10% for testing. The combination of drone and satellite images in the ADSF dataset creates a significant resource for testing fire detection models in both aerial and orbiting contexts.

3) FD Dataset:: The FD dataset [19] is a large-scale benchmark dataset created by integrating the two previously-established datasets of BoW-Fire and Foggia. Additional images were systematically collected from multiple online sources to expand the dataset's size and diversity. The final version of the dataset consists of two classes, fire and non-fire, each with 25,000 samples. The dataset has heterogeneous resolutions and dpi values. Due to its size and balanced composition, the FD dataset is among the largest and most inclusive available for research on fire scene classification. Examples of the dataset are shown in Fig. 2, where the first three columns illustrate fire scenes and the last three illustrate non-fire scenes. dataset's

C. Ablation Study: Componentwise Contributions

An ablation study was conducted on the FLAME dataset to evaluate the contribution of the main architectural components of the proposed HQDL framework. In particular, we examined the effects of the MobileViT backbone, the quantum convolutional layer, the quantum dense layer, and a classical matched-depth counterpart. The results are reported in Table I.

The baseline Hybrid Quantum Neural Network (HQCNN), which uses classical convolutional and dense neural networks, achieved an accuracy of 83.32%, precision of 77.68%, recall of 80.60%, and F1-score of 79.11%. This demonstrates that classical CNN-based architectures can learn meaningful fire-related features but still have limitations in capturing complex patterns in UAV imagery. When MobileViT was integrated with a quantum-dense layer (w/o QCon), accuracy increased to 85.87%, while recall improved substantially to 96.78%. Compared with HQCNN, this configuration improved accuracy by 2.55 percentage points and recall by 16.18 percentage points. This indicates that the MobileViT backbone provides stronger local–global feature representations, while the QDense layer enhances the model's ability to detect fire instances, particularly in terms of sensitivity. To evaluate the role of the QDense layer, a variant without QDense was tested, which consisted of MobileViT, QConv, and a classical classifier. This model achieved 84.31% accuracy and 80.61% F1-score. Compared with this variant, the Proposed HQDL model improved accuracy by 2.77 percentage points, recall by 18.22 percentage points, and F1-score by 5.11 percentage points. This result suggests that QDense significantly improves fire-class sensitivity after quantum convolutional feature processing. The importance of the MobileViT backbone is also evident from the variant in which MobileViT was replaced with ResNet-18 while keeping QConv and QDense.This variant achieved significantly lower performance, with 58.74% accuracy and 54.62% F1-score. The proposed HQDL model outperformed this configuration by 28.34 percentage points in accuracy and 31.10 percentage points in F1-score. This result highlights that MobileViT's ability to capture both local details and global context is essential for UAV-based fire detection, where fire regions are often small, irregular, and affected by complex backgrounds. To assess whether the performance gains were due solely to increased model complexity, two classical variants were evaluated. The MobileViT with classical layers and no quantum modules achieved 78.80% accuracy and 69.94% F1-score, whereas the classical counterpart of the same depth achieved 73.17% accuracy and 55.25% F1-score. Compared with these two non-quantum variants, the proposed HQDL improved accuracy by 8.28 and 13.91 percentage points, respectively, and improved F1-score by 15.78 and 30.47 percentage points, respectively. These comparisons show that the performance gain is primarily due to the integration of quantum-inspired feature transformations, resulting in a 13-31% improvement in F1-score over classical, matched-depth variants. The proposed HQDL model achieved the best overall performance, with 87.08% accuracy, 98.97% recall, and 85.72% F1-score. Compared with HQCNN, the proposed model improved accuracy by 3.76 percentage points, recall by 18.37 percentage points, and F1-score by 6.61 percentage points. Compared with the HQDL single-Q-layer configuration, the proposed HQDL model improved accuracy by 1.21 percentage points, recall by 2.19 percentage points, and F1-score by 1.42 percentage points. These improvements demonstrate that integrating multiple quantum layers with transform-based feature extraction methods significantly enhances deep neural networks' ability to generalize and be robust to variability in drone-based fire detection.

TABLE I
ABLATION STUDY OF THE PROPOSED HQDL FRAMEWORK ON THE FLAME DATASET

Variant	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
HQCNN	83.32	77.68	80.60	79.11
w/o QConv layer	85.87	76.66	96.78	84.30
w/o QDense layer	84.31	80.47	80.75	80.61
w/o MobileViT	58.74	49.14	61.47	54.62
w/o both quantum layers	78.80	81.83	61.06	69.94
Classical counterpart	73.17	84.64	41.01	55.25
Proposed HQDL	**87.08**	**75.61**	**98.97**	**85.72**

To summarize, the ablation studies demonstrated that quantum layers are not merely auxiliary enhancements, but rather necessary elements to achieving state-of-the-art performance in drone-based fire detection. Furthermore, the progression from the CNN-based HQNN to the classical counterpart and ultimately to the proposed HQDL framework demonstrates How quantum entanglement, variational depth, and transformer-based feature extraction can improve neural network representation and generalization. This research provides a new benchmark for hybrid quantum-classical deep learning, demonstrating that carefully designed quantum modules can outperform purely classical models on many complex tasks in remote sensing and geoscience.

D. Comparative Evaluation With Existing Methods

This section presents a comparative evaluation of the proposed HQDL model against SOTA approaches across three benchmark datasets, including FLAME, FD, and ADSF. The metrics considered include accuracy, precision, recall, and F1-score. Bold values in the tables indicate the best performance, while underscored values represent the second-best results.

1) Performance on the FLAME Dataset: Table II summarizes the results on the FLAME dataset. Classical CNN-based models such as Xception [31], EfficientNet variants [31], and MobileNetV3 [31] exhibit moderate performance, with accuracies ranging from 51.64% to 80%. Additional comparative methods, including UNIMODAL [32], Xception [33], further broaden the evaluation and report accuracies between 77.01% and 83.43%. Ensemble learning approaches improve performance, achieving 85.12% accuracy and an F1-score of 84.77%. More advanced fire-specific architectures, including DFAN [20], OFAN [24], and MViTs [26], achieve balanced results with F1-scores around 84.3 to 84.7%.Recent methods such as the 2-D CNN fixed-point implementation [34], LLM-based model [35], and EFNet-CSM [15] provide additional strong baselines, with EFNet-CSM achieving 85.44% accuracy and 85.15% F1-score. The proposed HQDL model surpassed all these approaches, achieving 87.08% accuracy, 98.97% recall, and an F1-score of 85.72%. Compared with existing methods, HQDL improves accuracy by 1.64 to 35.44 percentage points and F1-score by 0.57 to 40.75 percentage points. These gains highlight the effectiveness of the hybrid quantum–classical layers in capturing complex nonlinear dependencies in drone imagery, particularly under challenging conditions such as occlusion and variable zoom.

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3712753

TABLE II
PERFORMANCE COMPARISON ON THE FLAME DATASET

Approach	Acc:	Pre:	Recall	F1S
Xception [30]	78.41	–	–	78.12
Small-Xception [30]	76.23	–	–	–
US-DSAN+ResNet50 [36]	63.90	78.20	60.70	66.20
FT-ResNet50 [37]	79.48	–	–	–
EfficientNet-B5 [31]	75.82	–	–	73.90
EfficientNet-B4 [31]	69.93	–	–	65.51
EfficientNet-B3 [31]	65.81	–	–	64.02
EfficientNet-B2 [31]	66.04	–	–	60.71
InceptionV3 [31]	80.88	–	–	79.53
DenseNet169 [31]	80.62	–	–	79.40
MobileNetV3-Small [31]	51.64	–	–	44.97
MobileNetV3-Large [31]	65.10	–	–	61.91
UNIMODAL [32]	77.01	76.26	71.84	77.34
Xception [33]	81.3			
Ensemble Model [31]	85.12	–	–	84.77
DFAN [20]	83.79	83.00	84.55	83.76
OFAN [24]	84.41	84.12	84.49	84.30
MViTs [26]	84.90	84.58	84.93	84.75
2-D CNN (Fixed-point) [34]	81.49	–	–	–
LLM model [35]	71.2	81.3	46.8	59.4
EFNet-CSM [15]	85.44	84.88	85.97	85.15
HQDL(Proposed model)	**87.08**	**75.61**	**98.97**	**85.72**

TABLE III
PERFORMANCE COMPARISON ON THE FD DATASET

Approach	Acc:	Pre:	Recall	F1S
ResNet50 [38]	90.07	84.80	97.60	90.00
ANetFire [39]	87.20	83.30	93.20	87.90
GNetFire [48]	92.30	88.00	98.00	92.80
CNNFire [40]	87.30	84.60	91.30	87.90
EMNFire [41]	92.80	88.30	98.70	93.20
ViT-B/32 [29]	92.18	92.36	92.19	92.18
EFDNet [19]	95.30	93.50	97.40	95.40
DFANcomp [20]	95.70	95.50	96.30	95.90
US-DSAN+ResNet50 [36]	95.80	98.00	93.60	95.70
MIAPC [49]	96.00	94.90	97.30	96.10
DFAN [20]	96.17	96.00	97.00	96.00
OFAN [24]	96.54	97.00	96.00	97.00
IEFDNet [44]	95.18	92.93	97.80	95.30
FDSFEWF [45]	96.67	96.19	96.69	97.20
FlareNet [42]	97.06	98.00	95.00	96.00
Ms-Net [43]	97.38	97.80	96.82	97.35
MViTs [26]	97.70	97.00	98.00	98.00
ADFireNet [47]	98.12	98.25	98.05	98.11
DirFireNet [46]	96.50	96.20	97.10	96.65
Proposed model	**98.42**	**98.36**	**98.48**	**98.42**

2) Performance on the FD Dataset:: Table III reports results on the FD dataset [19]. Earlier CNN-based methods such as ResNet50 [38], ANetFire [39], and CNNFire [40] achieve accuracies between 87 to 92%. Attention-based and multi-scale CNN frameworks such as EMNFire [41] improve recall to 98.70%, demonstrating strong sensitivity to fire events. Transformer-based ViT-B/32 [29] achieves balanced performance with accuracy and F1-score around 92%. Recent architectures such as DFAN [20], OFAN [24], FlareNet [42], and MS-Net [43] achieve accuracies above 96%, with OFAN [24] reporting an F1-score of 97.0%. Recent fire-specific architectures, including IEFDNet [44], FDSFEWF [45], Ms-Net [43], MViTs [26], and DirFireNet [46], further demonstrate the rapid progress of lightweight and attention-based fire detection models, with reported accuracies ranging from 95.18% to 97.70%. Furthermore, the hybrid, transformer-inspired approach, MViT [26], achieves 97.70% accuracy and 98.0% F1-score. ADFireNet [47] sets a strong baseline with 98.12% accuracy and 98.11% F1-score. The proposed model achieves the best overall performance, with 98.42% accuracy, 98.36% precision, 98.48% recall, and an F1-score of 98.42%. HQDL improves accuracy by 0.30 to 11.22 percentage points and F1-score by 0.31 to 10.52 percentage points while maintaining a more balanced precision–recall trade-off. These results confirm the robustness of the hybrid architecture in handling diverse fire scenarios, including small-scale fires and visually complex backgrounds.

3) Performance on the ADSF Dataset:: Table IV presents results on the ADSF dataset, which combines drone and satellite imagery. Classical models such as EFDNet [19] and DFAN [20] achieve accuracies of 88 to 89%, while ADFireNet [47] improves performance to 90.86%. MobileNetV3 + MSAM [22] achieves 93.50% accuracy and balanced precision and recall values. The more recent variant MSoftFireNet [9] achieves 95.50% across all metrics, establishing a stronger baseline. The proposed model achieves the highest performance, with 96.83% accuracy, 97.0% precision, 96.68% recall, and 96.84% F1-score. This represents a +1.3% improvement over the strongest baseline and a margin of nearly +9% compared to earlier CNN-based approaches. These gains confirm the HQDL's superior generalization capability, particularly across diverse fire scenarios in the ADSF dataset.

TABLE IV
PERFORMANCE COMPARISON ON THE ADSF DATASET

Model	Acc(%)	Pre(%)	Recall(%)	F1S(%)
EFDNet [19]	88.00	87.50	88.00	87.50
DFAN [20]	89.36	86.01	94.00	89.84
ADFireNet [47]	90.86	90.90	90.86	89.84
MobileNetV3 + MSAM [22]	93.50	93.57	93.51	93.51
MSoftFireNet [9]	95.50	95.50	95.50	95.50
Proposed model	**96.83**	**97.0**	**96.68**	**96.84**

Across all four data sets, the proposed HQDL model consistently outperforms SOTA models. On the FLAME dataset, it improves upon EFNet-CSM [15] by approximately +1.0% on the same four evaluation metrics. In addition, on the FD dataset, the proposed model achieved the best performance in terms of accuracy and F1-scores. On the ADSF dataset, the proposed model outperforms the latest MSoftFireNet [9] by 1.5%, providing evidence that it generalizes across different types of fire scenes. Overall, these results provide evidence that the proposed HQDL model is effective. ZX-calculus analysis ensured there was no redundancy in the quantum layers, while Fourier analysis demonstrated the model's ability to capture a wide spectrum of frequency information in fire patterns. These theoretical insights explain the empirical improvements observed across datasets. The proposed model establishes a new benchmark for drone-based fire detection, demonstrating both accuracy and generalizability under remote sensing conditions.

E. Visual Results Analysis

Fig. 3 shows the qualitative visualization results using heat map activations and Grad-CAM for three test data sets (FD,

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3712753

FLAME, and ADSF). The visualizations were conducted to assess the interpretability, localization ability, and robustness of the proposed HQDL architecture across different remote-sensing scenarios. On the FD data set (with heavy smoke and complex flame shapes), the activation maps have shown that the core fire area was consistently highlighted, while the surrounding smoke and background clutter were suppressed. In addition, the Grad-CAM overlays showed that the model is focusing on high-intensity flame areas rather than low-contrast diffuse smoke patterns. Thus, the model reliably distinguishes between flames and smoke even in low-quality images. The proposed model has demonstrated the ability to precisely localize ignitions in small or partially occluded fires in aerial images from the FLAME data set, even when scale variability and background similarity with terrain and vegetation are present. The proposed model demonstrates that poor lighting and those of poor lighting. Overall, these visualizations demonstrate the robustness of the proposed model to various backgrounds found in real-world geospatial imagery. All three data sets also show similar heat maps with compact, organized activation patterns, indicating that the proposed model provides a stable, discriminant feature representation through the use of both MobileViT and the quantum layer. Therefore, the proposed model has produced activation maps that are much more focused and semantically meaningful compared to traditional deep learning models. As a result, the proposed model can accurately localize fires, suppress background noise, and maintain consistent attention behaviors across a range of spatial resolutions and environmental conditions. Thus, providing a suitable solution for the development of real-time, interpretable, and robust fire detection systems based on UAVs for remote sensing.

F. Complexity Analysis

The deployment of fire detection models on UAVs, surveillance systems, and IoT devices requires architectures that provide low memory usage, fewer trainable parameters, and fast inference speed on both desktop and embedded edge hardware. Therefore, the proposed HQDL framework was evaluated in terms of model size, parameter count, CPU/GPU inference speed, and Raspberry Pi edge-device inference speed, as shown in Table V.

Compared to models based on classical convolutional neural networks, such as ANetFire [31] and ResNetFire [36], with sizes of 233 MB and 98 MB, and a parameter count of more than 25 million, the proposed HQDL framework has a much smaller model size of 19.3 MB and fewer parameters at 5.03 million. This corresponds to an approximately 91.7% reduction in model size and 91.6% reduction in parameters compared with ANetFire [31], and an approximately 80.3% reduction in model size and 80.4% reduction in parameters compared with ResNetFire [36]. Similarly, lightweight CNN-based models, such as EMNFire [30], have reduced complexity (with a model size of 13.2 MB and a parameter count of 3.5 million) and Transformer-based models, such as ViT-B-32 [29], are also highly computationally intensive and have been found to be unsuitable for UAV-based real-time detection (requiring 89.8 MB and 88.2 million parameters). In contrast, HQDL provides stronger inference efficiency across multiple hardware settings. In particular, HQDL achieves 28.42 FPS on CPU and 86.48 FPS on GPU. More recent specialized architectures, including DFAN [20], FlareNet [31], and AD-FireNet [37], MAFire-Net [34], and EFNet-CSM [15], achieve high inference efficiency and reduced memory usage compared with earlier CNNs models. However, they still require more memory or parameters than HQDL. For example, DFAN [20] uses 83.6 MB and 23.9M parameters, MAFire-Net [34] uses 74.4 MB and 22.6M parameters, and EFNet-CSM [15] uses 38 MB and 6.4M parameters. In comparison, HQDL provides the same operational capability while reducing model size by 49.2% and parameters by 21.4% compared with EFNet-CSM [15], while also improving inference speed from 25.0 to 28.42 FPS on CPU and from 82.53 to 86.48 FPS on GPU, making it highly suitable for resource-constrained UAV platforms.

To further address real-time edge deployment, HQDL was also benchmarked on a Raspberry Pi platform. As reported in Table V, HQDL achieves 10.49 FPS on Raspberry Pi, outperforming other edge-tested models, including EMNFire [30], GNetFire [31], DFAN [20], SE-EFFNet [31], ADFireNet [37], MAFire-Net [34], and EFNet-CSM [15]. Compared with EFNet-CSM [15], which reports 8.0 FPS on Raspberry Pi, HQDL improves embedded inference speed by approximately 31.1%. Compared with ADFireNet [37], HQDL improves Raspberry Pi inference speed from 7.2 FPS to 10.49 FPS, corresponding to an approximately 45.7% improvement. These results demonstrate that the proposed HQDL framework is not only compact in terms of storage and parameters but also efficient for practical embedded inference on low-power edge devices.

In addition to FPS, inference latency and resource behavior were measured to validate real-time feasibility further. HQDL achieves a mean latency of 35.19 ms on CPU and 11.56 ms on GPU, with 95th percentile latency values of 35.84 ms and 12.25 ms, respectively. On the Raspberry Pi edge platform, HQDL achieved an average latency of 90.53 ms per image, with minimum and maximum latencies of 81.54 ms and 96.91 ms, respectively. During Raspberry Pi inference, the device temperature increased only slightly from 48.8°C to 50.5°C, indicating stable thermal behavior. The model also maintained low memory overhead, with RAM usage remaining within the available 7.87 GB system memory. An approximate energy analysis was also conducted using the Raspberry Pi power specification, and HQDL achieved an estimated average energy consumption of approximately 0.476 J per image. The combination of compact size, reduced parameters, high FPS, low latency, stable thermal behavior, and estimated low energy consumption demonstrates that HQDL is well-suited for UAV-oriented edge fire detection and remote-sensing deployment scenarios.

VI. CONCLUSION AND FUTURE DIRECTION

In this paper, we developed a hybrid deep learning quantum model to detect fires from data captured by UAVs in remote sensing applications. Our proposed hybrid model uses

This article has been accepted for publication in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. This is the author's version which has not been fully e
content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2026.3712753

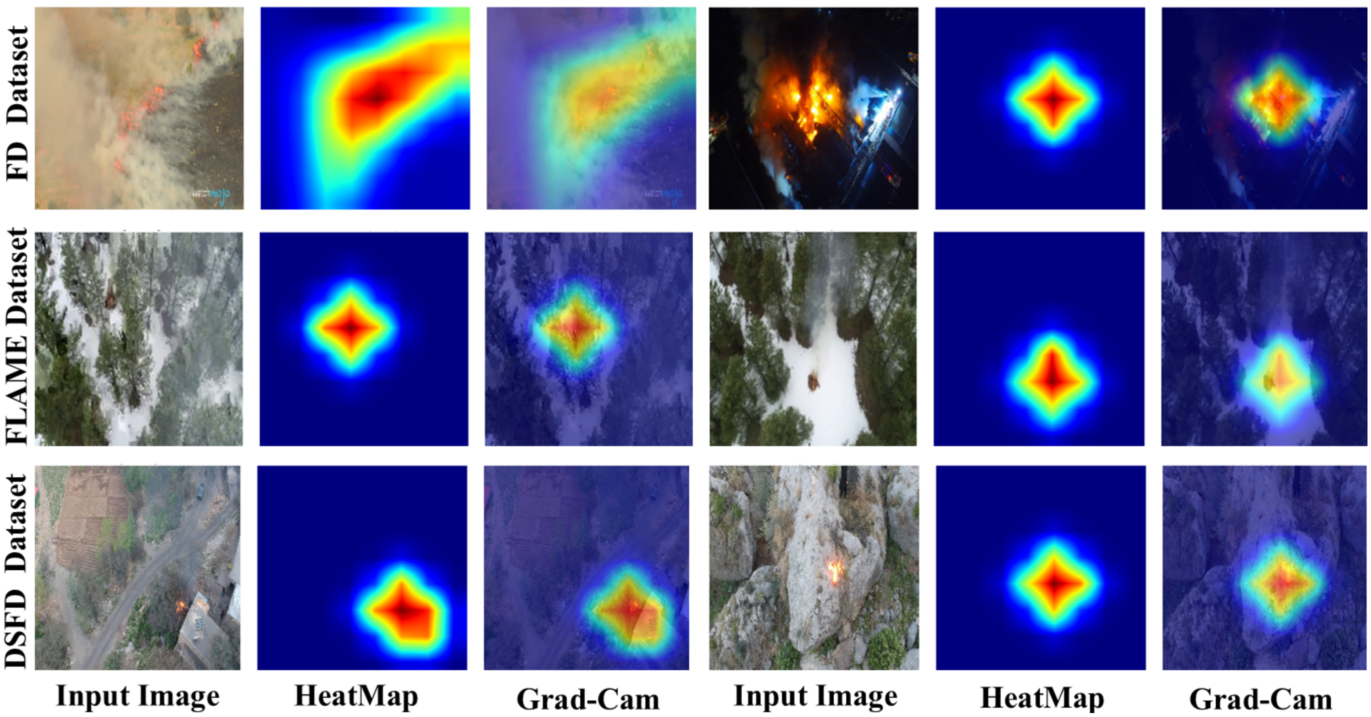

Fig. 3. Visual interpretability results of the proposed HQDL model on FD, FLAME, and ADSF datasets. For each dataset, the figure shows the input aerial image, the corresponding activation heatmap, and the Grad-CAM visualization. The highlighted regions indicate that the model consistently focuses on true fire locations while suppressing background and fire-like distractions, demonstrating robust localization and interpretability across diverse remote sensing scenarios.

TABLE V
COMPARISON OF MODEL SIZE, PARAMETERS, AND INFERENCE SPEED

Reference	Size (MB)	Params (M)	FPS (CPU)	FPS (GPU)	FPS (RPI)
ANetFire [31]	233.0	60.0	17	–	–
ResNetFire [36]	98.0	25.6	2.4	57.3	–
EMNFire [30]	13.0	3.5	2.4	61.2	5.0
GNetFire [31]	43.3	–	4.3	48.2	4.0
DFAN [20]	83.6	23.9	12.90	70.55	0.83
SE-EFFNet [31]	47.7	12.4	–	45.0	6.0
ViT-B-32 [29]	89.8	88.2	20.0	67.0	-
FlareNet [31]	49.1	–	14.0	75.0	-
ADFireNet [37]	38.0	7.2	22.0	72.5	7.2
MAFire-Net [34]	74.4	22.6	14.32	78.31	0.92
EFNet-CSM [15]	38.0	6.4	25.0	82.53	8.0
HQDL (Ours)	**19.3**	**5.03**	**28.42**	**86.48**	**10.49**

MobileViT as the backbone. It includes quantum-enhanced versions of convolutional layers and dense layers to achieve both high accuracy and efficient computation. Our model was evaluated on three standard fire-detection datasets, including FD, FLAME, and ADSF. Our hybrid model is shown to be more accurate than the current CNN, attention, and transformer-based models. It also demonstrates fast inference times, making it suitable for real-time deployment on resource-constrained platforms. Also, it was shown that quantum feature extraction outperforms classical approaches in terms of time required to process high-dimensional data. Visualizations of the model's decisions during the evaluation phase show that it can focus on the fire-related region and ignore background regions. The robustness of the model across different types of fire scenarios, combined with its computational efficiency, demonstrates that combining deep learning with quantum computing offers value for overcoming some of the challenges associated with detecting fire in complex remote sensing environments.

However, a few limitations should be addressed in future work. While the model demonstrates excellent performance for image-based fire detection, its applicability to video-based fire detection remains a challenge. Real-time Video analysis typically involves processing continuous frames, which may introduce temporal dependencies and require further optimizations to handle large amounts of data in real-time without degrading performance. Extending this model to handle Video data effectively while maintaining computational efficiency and accuracy would be a key area for future research.

Future research could expand the capabilities of fire detection systems by integrating vision-language models, enabling the combination of both visual and textual data for more comprehensive scene understanding. This approach would facilitate the detection of fire intensity levels and classification of fire nature, providing critical information for emergency response teams. Additionally, segmentation techniques can be explored to precisely delineate fire boundaries and improve spatial understanding of fire regions. The integration of detection and segmentation models would enable the system to not only detect fires but also accurately map their spread, enhancing response efforts. Moreover, leveraging multi-modal data from different sensor types would enhance the model's robustness in varying environmental conditions and fire scenarios, ensuring better performance in real-time applications.

REFERENCES

[1] B. Jankovic, S. Jangirova, W. Ullah, L. U. Khan, and M. Guizani, "Disaster detection on the fly: Optimized transformers for uavs," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025.

[2] W. Ullah, L. U. Khan, M. Guizani, C.-D. Wang, and D. Wu, "Graph-based temporal attention network for anomaly recognition in internet of things video surveillance," *IEEE Internet of Things Journal*, 2025.

[3] G. Wang, H. Li, S. Ye, H. Zhao, H. Ding, and S. Xie, "Rfwnet: A multiscale remote sensing forest wildfire detection network with digital twinning, adaptive spatial aggregation, and dynamic sparse features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–23, 2024.

[4] K. Avazov, M. Mukhiddinov, F. Makhmudov, and Y. I. Cho, "Fire detection method in smart city environments using a deep-learning-based approach," *Electronics*, vol. 11, no. 1, p. 73, 2021.

[5] I. association of fire and rescue services, "102 million hectares has burned globally in 2025 - compare with the record years of 2012 and 2015," 2025, access date 12-08-2025 https://ctif.org/news/102-million-hectares-has-burned-globally-2025-compare-record-years-2012-and-2015.

[6] O. W. I. Data, "Annual area burnt by wildfires," 2025, access date 12-08-2025 https://ourworldindata.org/grapher/annual-area-burnt-by-wildfires.

[7] W. Ullah, F. Outay, L. U. Khan, and M. Guizani, "Split federated learning for real-time aerial video event recognition in uav-based geospatial monitoring," *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[8] H. Yar, A. S. Imran, Z. A. Khan, M. Sajjad, and Z. Kastrati, "Towards smart home automation using iot-enabled edge-computing paradigm," *Sensors*, vol. 21, no. 14, p. 4932, 2021.

[9] S. Jangirova, B. Jankovic, W. Ullah, L. U. Khan, and M. Guizani, "Real-time aerial fire detection on resource-constrained devices using knowledge distillation," *International Journal of Applied Earth Observation and Geoinformation*, vol. 142, p. 104665, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1569843225003127

[10] A. Secilmis, N. Aksu, F. A. Dael, I. Shayea, and A. A. El-Saleh, "Machine learning-based fire detection: A comprehensive review and evaluation of classification models," *JOIV: International Journal on Informatics Visualization*, vol. 7, no. 3-2, pp. 1982–1988, 2023.

[11] F. Huot, R. L. Hu, N. Goyal, T. Sankar, M. Ihme, and Y.-F. Chen, "Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[12] P. Foggia, A. Saggese, and M. Vento, "Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion," *IEEE TRANSACTIONS on circuits and systems for video technology*, vol. 25, no. 9, pp. 1545–1556, 2015.

[13] H. Xu, G. Zhang, R. Chu, J. Zhang, Z. Yang, X. Wu, and H. Xiao, "Detecting forest fire omission error based on data fusion at subpixel scale," *International Journal of Applied Earth Observation and Geoinformation*, vol. 128, p. 103737, 2024.

[14] H. Singh, L.-M. Ang, and S. K. Srivastava, "Benchmarking artificial neural networks and u-net convolutional architectures for wildfire susceptibility prediction: Innovations in geospatial intelligence," *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[15] H. Yar, F. U. M. Ullah, Z. A. Khan, M. J. Kim, and S. W. Baik, "Efnet-csm: Efficientnet with a modified attention mechanism for effective fire detection," *Knowledge-Based Systems*, p. 114353, 2025.

[16] W. Bo, J. Liu, X. Fan, T. Tjahjadi, Q. Ye, and L. Fu, "Basnet: Burned area segmentation network for real-time detection of damage maps in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[17] Q. Li, H. Li, Y. Huang, and H. Ji, "Pdenet: A high-precision forest fire detection network enabled by dual-stream hierarchical and elastic response," *IEEE Transactions on Geoscience and Remote Sensing*, 2025.

[18] Z. Daoud, A. Ben Hamida, and C. Ben Amar, "Fireclassnet: a deep convolutional neural network approach for pjf fire images classification," *Neural Computing and Applications*, vol. 35, no. 26, pp. 19 069–19 085, 2023.

[19] S. Li, Q. Yan, and P. Liu, "An efficient fire detection method based on multiscale feature extraction, implicit deep supervision and channel attention mechanism," *IEEE Transactions on Image Processing*, vol. 29, pp. 8467–8475, 2020.

[20] H. Yar, T. Hussain, M. Agarwal, Z. A. Khan, S. K. Gupta, and S. W. Baik, "Optimized dual fire attention network and medium-scale fire classification benchmark," *IEEE Transactions on Image Processing*, vol. 31, pp. 6331–6343, 2022.

[21] T. Khan, Z. A. Khan, and C. Choi, "Enhancing real-time fire detection: An effective multi-attention network and a fire benchmark," *Neural Computing and Applications*, vol. 37, no. 18, pp. 11 693–11 707, 2025.

[22] H. Yar, Z. A. Khan, I. Rida, W. Ullah, M. J. Kim, and S. W. Baik, "An efficient deep learning architecture for effective fire detection in smart surveillance," *Image and Vision Computing*, vol. 145, p. 104989, 2024.

[23] S. U. Khan, S. Danish, E. Iqbal, D. Gupta, Y. Zweiri, and Y. Abdulrahman, "Firemamba: An efficient visual representation learning framework for fire detection," *Alexandria Engineering Journal*, vol. 132, pp. 133–144, 2025.

[24] N. Dilshad, S. U. Khan, N. S. Alghamdi, T. Taleb, and J. Song, "Toward efficient fire detection in iot environment: A modified attention network and large-scale data set," *IEEE Internet of Things Journal*, vol. 11, no. 8, pp. 13 467–13 481, 2023.

[25] S. Chen, W. Li, Y. Cao, and X. Lu, "Combining the convolution and transformer for classification of smoke-like scenes in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2022.

[26] H. Yar, Z. A. Khan, T. Hussain, and S. W. Baik, "A modified vision transformer architecture with scratch learning capabilities for effective fire detection," *Expert Systems with Applications*, vol. 252, p. 123935, 2024.

[27] M. V. Cruz, N. Kumar, A. Subramanian, S. C. Sethuraman *et al.*, "Wildfire detection using vision transformer (vit)," in *2025 6th International Conference on Artificial Intelligence, Robotics and Control (AIRC)*. IEEE, 2025, pp. 402–406.

[28] M. Shahid, H.-C. Wang, Y.-Y. Chen, and K.-L. Hua, "Hybrid cnn-vit architecture to exploit spatio-temporal feature for fire recognition trained through transfer learning," *Multimedia Tools and Applications*, vol. 84, no. 8, pp. 4703–4732, 2025.

[29] M. Shahid and K.-l. Hua, "Fire detection using transformer network," in *Proceedings of the 2021 international conference on multimedia retrieval*, 2021, pp. 627–630.

[30] A. Shamsoshoara, F. Afghah, A. Razi, L. Zheng, P. Z. Fulé, and E. Blasch, "Aerial imagery pile burn detection using deep learning: The flame dataset," *Computer Networks*, vol. 193, p. 108001, 2021.

[31] R. Ghali, M. A. Akhloufi, and W. S. Mseddi, "Deep learning and transformer approaches for uav-based wildfire detection and segmentation," *Sensors*, vol. 22, no. 5, p. 1977, 2022.

[32] B. Hopkins, L. ONeill, M. Marinaccio, E. Rowell, R. Parsons, S. Flanary, I. Nazim, C. Seielstad, and F. Afghah, "Flame 3 dataset: Unleashing the power of radiometric thermal uav imagery for wildfire management," *arXiv preprint arXiv:2412.02831*, 2024.

[33] S. Treneska and B. R. Stojkoska, "Wildfire detection from uav collected images using transfer learning," in *Proceedings of the 18th International Conference on Informatics and Information Technologies, Skopje, North Macedonia*, 2021, pp. 6–7.

[34] Y.-P. Liang, C.-M. Chang, and C.-C. Chung, "Implementation of lightweight convolutional neural networks with an early exit mechanism utilizing 40 nm cmos process for fire detection in unmanned aerial vehicles," *Sensors*, vol. 24, no. 7, p. 2265, 2024.

[35] H. Zhao, Y. Liu, Y. Han, X. Li, Y. Guo, and J. Jin, "A zero-shot high-performance fire detection framework based on large language models," *Neurocomputing*, p. 131403, 2025.

[36] Z. Yan, L. Wang, K. Qin, F. Zhou, J. Ouyang, T. Wang, X. Hou, and L. Bu, "Unsupervised domain adaptation for forest fire recognition using transferable knowledge from public datasets," *Forests*, vol. 14, no. 1, p. 52, 2022.

[37] L. Zhang, M. Wang, Y. Fu, and Y. Ding, "A forest fire recognition method using uav images based on transfer learning," *Forests*, vol. 13, no. 7, p. 975, 2022.

[38] J. Sharma, O.-C. Granmo, M. Goodwin, and J. T. Fidje, "Deep convolutional neural networks for fire detection in images," in *International conference on engineering applications of neural networks*. Springer, 2017, pp. 183–193.

[39] K. Muhammad, J. Ahmad, and S. W. Baik, "Early fire detection using convolutional neural networks during surveillance for effective disaster management," *Neurocomputing*, vol. 288, pp. 30–42, 2018.

[40] 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 Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 7, pp. 1419–1434, 2018.

[41] K. Muhammad, S. Khan, M. Elhoseny, S. H. Ahmed, and S. W. Baik, "Efficient fire detection for uncertain surveillance environment," *IEEE*

Transactions on Industrial Informatics, vol. 15, no. 5, pp. 3113–3122, 2019.

[42] B. Yousaf, A. F. Mirza, M. Irfan, M. Mansoor, and Z. Yang, "Flarenet: A feature fusion based method for fire detection under diverse conditions," 2024.

[43] J. Feng and Y. Sun, "Multiscale network based on feature fusion for fire disaster detection in complex scenes," *Expert Systems with Applications*, vol. 240, p. 122494, 2024.

[44] G.-s. Chen, T. Chandarasupsang, X.-d. Luo, A. Tananchana, and L. Mu, "Iefdnet: Detection of fire using partial convolution inception," in *2024 4th International Conference on Big Data Engineering and Education (BDEE)*. IEEE, 2024, pp. 32–37.

[45] K. Chen, W. Wen, F. Feng, X. Xu, and Y. Liang, "Fire detection based on flame enhancement for weak fires," in *Chinese Conference on Image and Graphics Technologies*. Springer, 2024, pp. 105–117.

[46] A. U. Haider, S. Khan, M. J. Ahmed, and T. A. Khan, "Strip pooling coordinate attention with directional learning for intelligent fire recognition in smart cities," *ICCK Transactions on Sensing, Communication, and Control*, vol. 2, no. 4, pp. 263–275, 2025.

[47] H. Yar, W. Ullah, Z. A. Khan, and S. W. Baik, "An effective attention-based cnn model for fire detection in adverse weather conditions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 206, pp. 335–346, 2023.

[48] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, "Convolutional neural networks based fire detection in surveillance videos," *Ieee Access*, vol. 6, pp. 18 174–18 183, 2018.

[49] Z. Deng, S. Hu, S. Yin, Y. Wang, A. Basu, and I. Cheng, "Multi-step implicit adams predictor-corrector network for fire detection," *IET Image Processing*, vol. 16, no. 9, pp. 2338–2350, 2022.