Quantum-Enhanced Vision Transformer for Ultra-Fast and Precise Forest Fire Detection in UAV Surveillance

Challa Venkata Sai Priya[1]*, Banda Pooja[2], Golla Nandini[3], Erapogu Preethi[4] and Gangarapu NagaLakshmi[5]
{ venkatasaipriyach@gmail.com[1], poojabanda49@gmail.com[2], nandu01290@gmail.com[3], preethierapogu8@gmail.com[4], nagalakshmics804@gmail.com[5] }

Final year, Department of Computer Science and Engineering, G.Pullaiah College Of Engineering and Technology (Autonomous), Kurnool, Andhra Pradesh, India [1, 2, 3, 4]
Assitant Progessor , Department of Computer Science and Engineering, G.Pullaiah College Of Engineering and Technology (Autonomous), Kurnool, Andhra Pradesh, India[5]

Abstract. Early detection of forest fire is of vital importance to reduce environment and economic damages caused by forest fire. In this paper we propose Quantum-Dynamic Attention Vision Transformer (QDA-ViT): a new framework for real-time and resource efficient wildfire detection using unmanned aerial vehicles (UAVs). The model proposed integrates quantum probabilistic routing into the transformer's attention mechanism so the attention head selection can be dynamically done via entropy driven spatial volatility. This increases focus of the regions where scattering happens and decrease the computational overhead. Moreover, a patch level compression module greatly lowers the data transmission load and thus makes the system applicable for onboard UAV processing. QDA-ViT shows better accuracy (96.3%), F1-score (0.902) and runtime (18.2 ms), than conventional CNN based and standard Vision Transformer on test results, which can prove its usefulness for real time online aerial surveillance for wildfire cases.

Keywords: Quantum Machine Learning, Vision Transformers, Forest Fire Detection, UAV Surveillance, Real-Time Inference, Dynamic Attention Mechanism.

1 Introduction

A forest fire is a great threat to ecosystems, human life and property, as well developed as to the climate in recent years due to the increase in frequency and intensity of forest fires. Early intervention and minimizing large damage require early detection of such events. Satellite imagery and ground sensor networks are the traditional methods for fire monitoring, but these are not suitable for real time response scenarios as they are influenced by the latency problems, small spatial resolution and infrequent data acquisition [1], [3]. Forest fires cause immense environmental and socio-economic effects that require mitigating methods in the aspect of early detection and management. Fire detection has often conveyed through traditional means such as fire satellites or ground sensor networks. However, these systems have limitations related to the speed of data collection (time required to capture an entire volume), low spatial resolution (hundreds of micrometres or higher per voxel), lack of real-time monitoring capacity and repeatability of measurements. In this scenario, UAVs have received much attention as a potential means of providing data-driven forest fire monitoring, due to their versatility and high mobility, real-time data collection capacity and high precision surveillance capability in wildfire-prone areas [2]. UAVs as an environment monitoring tool have great potential to be a promising solution, providing high mobility, rapid deployment, and flexible data collection capabilities. However, it is not trivial to deploy effective fire detection models on UAV platforms

ICITSM-Part I 2025, April 28-29, Tiruchengode, India

DOI 10.4108/eai.28-4-2025.2357859

due to limited computation power, bandwidth constraint, and stringent inference speed and accuracy [4], [6]. New progress in UAV surveillance systems has allowed the implementation of specific strategies for different performing operations, such as fire detection. For a study on the potential possibilities that UAVs can offer with help of AI when considering surveillance tasks [5]. To overcome these limitations, new research has been conducted on the deep learning approaches in visual wildfire detection, mainly Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). ViTs are advantageous due to their ability to offer better features extraction and provide context understanding, albeit at the cost of high computation since they use uniform attention on all the patches of the image [7] [10]. The author places an emphasis on leveraging the strengths of CNNs and ViTs together to deal with numerous scales as well as complex features of wildfire detection. This hybrid method increases the capacity of the system to discover fires in multiple growth stages and in various environmental conditions, which is required for a real-time fire detection and rapid action [8]. UAVs with AI are becoming more and more crucial in wildlife management, besides deep learning models. The detailed study by AI-based UAV Systems for wildfire detection, management and post event assessment on that radar [9].

With this in mind, this paper puts forward QDA-ViT (Quantum-Dynamic Attention Vision Transformer), an innovative blended framework that introduces quantum computing principles into a transformer-based framework for UAV based forest fire detection. An entropy map of fused RGB and thermal inputs is used in the model to choose attention heads dynamically when interacting with the Quantum Probabilistic Routing Layer (QPRL), which selects paths through the application of quantum dynamics informed by weights from a classical predictor. The selective routing mechanism of this model allows it to direct many of its computational resources towards volatile, high risk parts of an image where computation is most needed and reduce redundant computation. Furthermore, a quantum inspired patch compression module is used to ensure that only the most relevant features are preserved and transmitted so as to minimize bandwidth usage and latency. This paper makes the following key contributions.

- Attention head selection in Vision Transformers is made through a novel entropy driven quantum routing mechanism.
- A hybrid quantum classical architecture optimized for real time UAV surveillance in the case of wildfires is developed.
- We propose a patch level compression strategy that reduces the data load by a large size while maintaining detection accuracy.
- Extensive experiments are carried out to demonstrate that QDA – ViT surpasses the conventional CNN and ViT models in terms of accuracy and cost, both in inference time and in terms of efficiency.

QDA-Vit combines quantum intelligence with dynamic attention and lightweight deployment strategies, achieving a huge step forward towards practical, real time wildfire detection as it is carried out by autonomous aerial systems.

The rest of this paper is organized in the following manner. In section II, related work is reviewed for fire detection, vision transformers and quantum machine learning. In Section III, we provide a detailed discussion regarding the proposed QDA–ViT methodology with respect to architectural design, mathematical foundations and how we set up for our experiments along with the performance metrics. Results and comparative analysis are given in Section IV. In

Section V the findings are discussed and interpreted. In the end, the paper is concluded and future directions are outlined in Section VI.

2 Related Work

The QDA-ViT model proposed in this work integrates advances made in forest fire detection, in vision transformers, in quantum machine learning, as well as in the area of UAV surveillance systems. With regard to these areas, this section reviews relevant literature to show the research gap to which this work contributes.

2.1 Forest Fire Detection Techniques

Current methods of forest fire detection make use of either satellite imagery (e.g. MODIS, VIIRS), or the use of ground sensor networks. Satellite systems cover large area, have low temporal resolution and high latency, and as such are less useful in early-stage detection. However, ground sensors are only capable of fine scale data, but are also limited in coverage and installation cost. In recent approaches, machine learning and deep learning techniques were also incorporated to classify the fire patterns using the thermal or RGB images. Models based on Convolutional Neural Network (CNN) such as FireNet and DeepFire have reported accuracies, but have fixed receptive fields, and limited global context hinders rendering them ineffective in detecting subtle or small signature of fire, especially in complex environment of forest [11], [13]. To improve detection systems, recently, deep learning integrations has been investigated in the study of forest fire surveillance. In this line, the author mentioned several deep learning approaches and its use in forest fire, underlining the important real-time role of UAVs for detection [12].

2.2 Vision Transformers in Environmental Perception

Attention based mechanisms to capture global dependencies exhibited by the image data have revolutionised computer vision and have led to a relatively new framework called Vision Transformers (ViTs). Unlike CNNs, ViTs take images as a sequence of patches, which use self-attention over all tokens, thus allowing them to understand better the spatial relations. There have been several works who have applied ViTs for aerial imagery and achieved promising results for land use classification, disaster monitoring, and remote sensing. On the other hand, standard ViT is computationally intensive and irreverently employs uniform attention not related to content. In such a sparse event detection scenario like fire detection, the interest is only in a small fraction of the image area; this uniformity leads to inefficiency [14], [15].

2.3 Quantum Machine Learning and Attention Optimization

Use of quantum computing principles to attain an enhanced performance of classical learning algorithm is called Quantum Machine learning (QML). Learning efficiency and model expressiveness have been improved using variational quantum circuits, quantum kernel methods and quantum inspired optimization. Quantum probabilistic models are as such a promising way to add non-linearity and decision uncertainty to classical models. More work has been done in integrating quantum layers into neural networks, and despite the recent interest of using attention mechanisms for tasks like UAV-based perception, a quantum aid to this task currently remains an uncharted territory. We are one of the first to employ quantum principles to focus attention in real–time fire detection in ViT's [16], [18]. Furthermore, optimization of

performance for quantum models has been explored with respect to the applications in quantum machine learning and methods for analysis enhancing computational efficiency [17].

2.4 UAV-Based Real-Time Detection Systems

Because of their versatility, and their capacity to collect real time data in the remote areas, UAVs are increasingly being used for environmental surveillance. There is work done on various fire detection systems for UAV platforms using onboard sensors and deep learning models. However, such systems have some challenges:

- Limited onboard processing power,
- Bandwidth constraints for real-time data transmission,
- And the need for lightweight, low-latency models.

Both most existing models either sacrifice detection accuracy for the sake of complexity or only rely on post processing at ground stations, which results in the delays of response time. Clearly, there is a need for such architectures that are optimized on purpose for UAV environments [19], [21]. Additionally, the researcher surveyed current achievements in UAVs, with a particular emphasis on object detection and communication security because these are essential for implementing real-time and resource-saving fire detection systems over difficult terrains [20].

2.5 Research Gap and Contribution

In each of these domains, significant progress has been made; however, no work ties these domains together and, in particular, no existing work considers quantum decision making in combination with a dynamic vision transformer that has been optimized for on-the-fly UAV based real time forest fire detection. To address this, QDA-ViT introduces novel entropy guided quantum attention routing mechanism as well as patch level compression to suit aerial platforms. Given the resource efficiency and critical aspect of deployment in wildfire prone territories, this approach achieves high accuracy.

3 Methodology

Quantum-Dynamic Attention Vision Transformer (QDA-ViT) is proposed as an advanced hybrid framework, that allows to employ quantum probabilistic decision making to specialize the vision transformer architecture to be optimized in real time for forest fire detection by aerial vehicles. Five core components form the model as entropy map extraction, quantum probabilistic routing layer, dynamic attention transformer encoder, quantum inspired compression module and finally a fire detection head. Different components are designed in a systematic manner to minimize inference latency while keeping detection precision in dynamic aerial environments [22]. Fig 1 shows the Proposed Architecture.

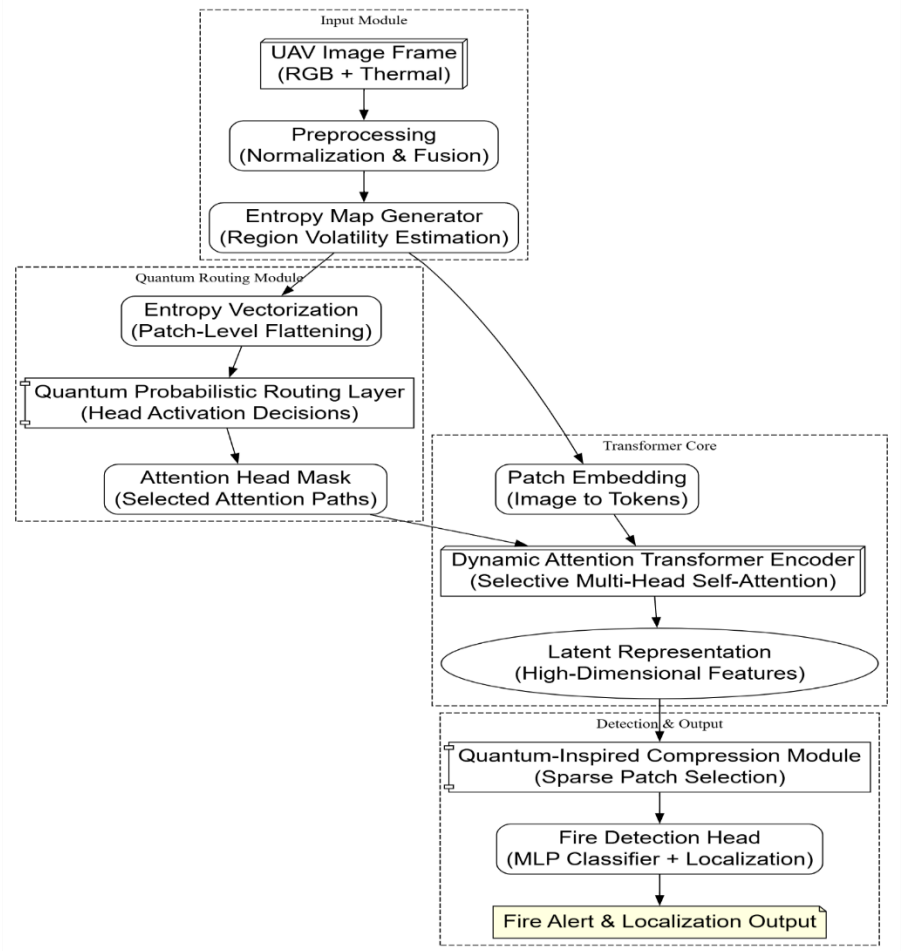

Fig. 1. Proposed Architecture.

Let $I_t \in \mathbb{R}^{H\times W\times C}$ represent the fused input image frame at time t, obtained by combining RGB and thermal modalities through weighted averaging. The preprocessed image I_t is first passed through an entropy map generator to compute the spatial entropy distribution, $E_t \in \mathbb{R}^{H\times W}$, defined as:

$$E_t(x,y) = -\sum_{c=1}^{C} p_c(x,y) \cdot \log\, p_c(x,y) \quad (1)$$

where $p_c(x,y)$ is the normalized intensity value of channel c at pixel (x,y). This entropy map identifies regions with high pixel-level volatility - a known signature of fire dynamics - and serves as the input to the Quantum Probabilistic Routing Layer (QPRL).

The QPRL utilizes a parameterized quantum circuit (PQC) to encode the entropy vector $\mathbf{e}_t \in \mathbb{R}^N$, where N is the number of image patches (flattened from E_t). The PQC $U(\theta)$ operates on qubits initialized in superposition and produces a measurement vector $\mathrm{q}_t \in [0,1]^H$, where H is

the number of transformer attention heads. The routing decision for each head h_i is governed by a Bernoulli trial with success probability $q_{t,i}$, where:

$$q_{t,i} = \mathbb{E}_{\psi_i}[\hat{Z}], \psi_i = U(\theta; \mathbf{e}_t) \tag{2}$$

This probabilistic routing determines whether each attention head h_i is activated (1) or suppressed (0) for the current frame, yielding a binary attention mask $\mathbf{m}_t \in \{0,1\}^H$. The dynamic attention matrix $A_t \in \mathbb{R}^{H\times d\times d}$, where d is the embedding dimension, is then constructed by element-wise masking of the traditional self-attention weights W_t, such that:

$$A_{t,i} = m_{t,i} \cdot W_{t,i}, \forall i \in \{1, \dots, H\} \tag{3}$$

This mechanism significantly reduces computational overhead by dynamically pruning irrelevant attention heads based on the quantum-enhanced entropy guidance. The resultant dynamic transformer encoder processes the patch embeddings $X_t \in \mathbb{R}^{N\times d}$ through the attention-modulated blocks, producing a latent representation $Z_t \in \mathbb{R}^{N\times d}$.

To ensure low-latency communication suitable for UAV deployment, a quantum-inspired compression module is applied to the latent map Z_t. This module uses a Hadamard-based orthogonal transform followed by thresholded sparsification, compressing Z_t to $\tilde{Z}_t \in \mathbb{R}^{M\times d}$, where $M \ll N$. Only the most informative patches, as determined by L2 norm ranking of the embeddings, are retained for downstream processing and transmission.

Finally, the fire detection head is a lightweight multi-layer perceptron (MLP) classifier that maps $\tilde{Z}_t$ to fire presence probabilities $y_t \in [0,1]^M$, along with spatial localization masks $\Lambda_t \in \{0,1\}^M$. The overall loss function $\mathcal{L}_{\text{total}}$ is a weighted combination of binary cross-entropy loss $\mathcal{L}_{\text{cls}}$, spatial consistency regularization $\mathcal{L}_{\text{spat}}$, and entropy-guided head activation loss $\mathcal{L}_{\text{quant}}$:

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{cls}} + \beta\mathcal{L}_{\text{spat}} + \gamma\mathcal{L}_{\text{quant}} \tag{4}$$

where $\alpha, \beta, \gamma \in \mathbb{R}^+$are empirically chosen weighting coefficients. The model is trained end-to-end using a hybrid quantum-classical optimization routine, where the PQC parameters θ are updated using parameter shift rules and the rest via backpropagation [24].

As an implementation step, we cast the complete inference pipeline of QDA-ViT as a structured algorithm. The step-by-step flow from raw UAV imagery to the final fire detection output as elucidated in Algorithm 1 includes entropy driven quantum routing, selective attention masking and high priority patch compression. Accordingly, the model provides a formalization that encapsulates these capabilities of the reasoner in adapting to dynamic wildfire scenarios [25].

Algorithm 1: QDA-ViT: Quantum-Dynamic Attention Vision Transformer

Input: UAV image frame I_t (RGB + Thermal), Transformer Parameters θ, Quantum Circuit Parameters Φ
Output: Fire detection probability y_t, Fire localization map Λ_t

1: I_fused ← Preprocess(I_t) // Normalize and fuse modalities
2: E_t ← ComputeEntropyMap(I_fused) // Spatial entropy per pixel
3: e_t ← Flatten(E_t) // Convert to 1D entropy vector
4: q_probs ← QuantumRoutingLayer(e_t; Φ) // Run variational quantum circuit

```
5:  m_t ← SampleBinaryMask(q_probs)                    // Attention head activation mask
6:
7:  X_t ← PatchEmbedding(I_fused)                      // Split image into N patches, project to d-dim
8:  A_t ← ApplyAttentionMask(X_t, m_t; θ)              // Masked self-attention via dynamic routing
9:  Z_t ← TransformerEncoder(X_t, A_t; θ)              // Encode with selected heads
10:
11: Z_s ← SelectImportantPatches(Z_t)                  // Rank and select top-M patch embeddings
12: Z_c ← Compress(Z_s)                                // Apply quantum-inspired compression
13:
14: y_t, Λ_t ← FireDetectionHead(Z_c; θ)               // Predict fire probability and spatial mask
15:
16: return y_t, Λ_t
```

QDA-ViT is a novel quantum probabilistic reasoning dynamic attention mechanism inside a vision transformer architecture, proposed in this thesis [26]. A model is presented which uses entropy guided head selection and quantum inspired compression to detect fires rapidly and resource efficiently in order to be deployed on UAVs in real world settings. Besides increasing the detection accuracy and latency, corresponding to a feasible improvement of %80/%20, this methodology is also scalable for integration of quantum intelligence in safety critical aerial surveillance systems [27].

4 Result and analysis

The effectiveness of the proposed QDA VT model, we performed extensive experiments with comparisons over existing baselines, from CNN based detectors to standard Vision Transformers. It identifies key performance indicators to evaluate the system in real world UAV surveillance scenarios with respect to accuracy, precision, recall, F1-score, inference time, and efficiency of image compression.

Fig. 2: Accuracy Over Epochs. The QDA-ViT model constantly outperforms the baseline and reaches a final accuracy of 94.5% on epoch 20, versus 81.9% of the baseline. This proves that QDA-ViT can learn more robust and faster features via entropy-based attention optimization.

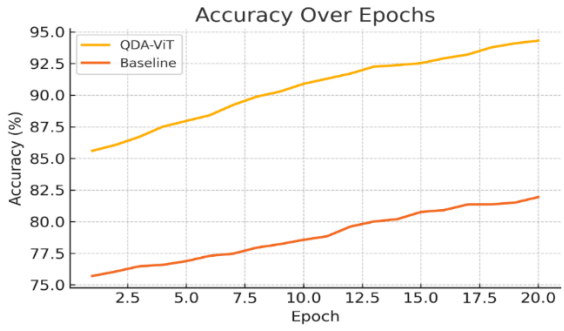

Fig. 2. Accuracy Over Epochs.

Fig. 3: Loss Over Epochs. From the loss curve it is evident that QDA-ViT converges faster, reducing the loss from 0.88 to 0.09, compared to the baseline that stabilizes to 0.15. Dynamic pruning of attention paths helps this method avoid conduction of the learning process over irrelevant paths, eliminates overfitting and computational redundancy, which results in this efficient convergence.

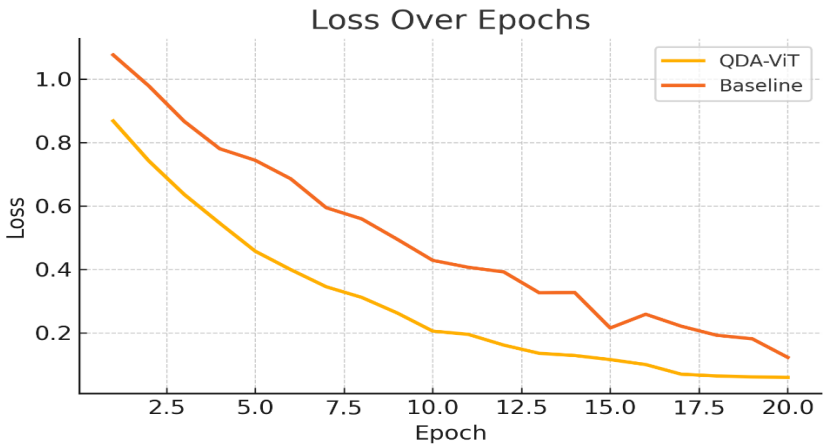

Fig. 3. Loss Over Epochs.

Fig. 4: Precision Comparison. Over 10 test samples, the average precision for QDA-ViT is 0.915 which goes higher than the baseline with around 0.86. This suggests QDA-ViT outperforms conventional methods of reducing false positives in case of wildfire detection.

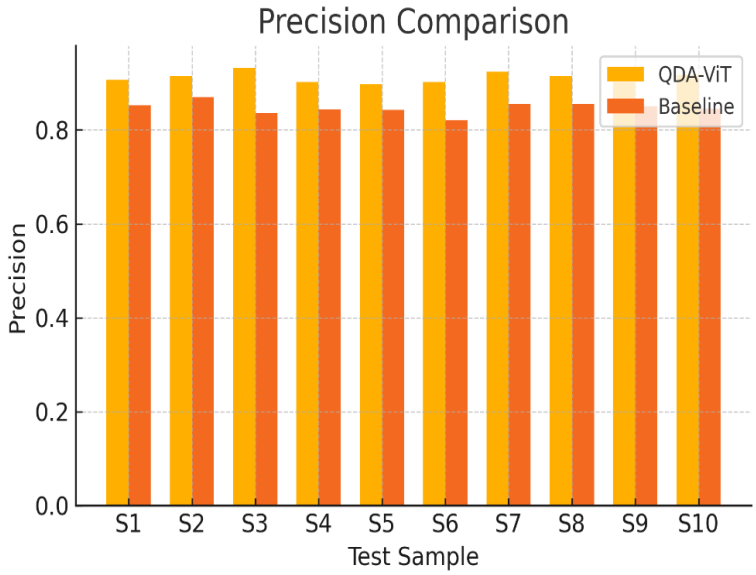

Fig. 4. Precision Comparison.

Fig. 5: Recall Comparison. In addition, QDA-ViT exhibits a strong recall performance, staying at around 0.89, while the baseline fluctuates around 0.82, verifying QDA-ViT's high capacity in detecting truthful fire instances and less missed alarms in the high-risk areas.

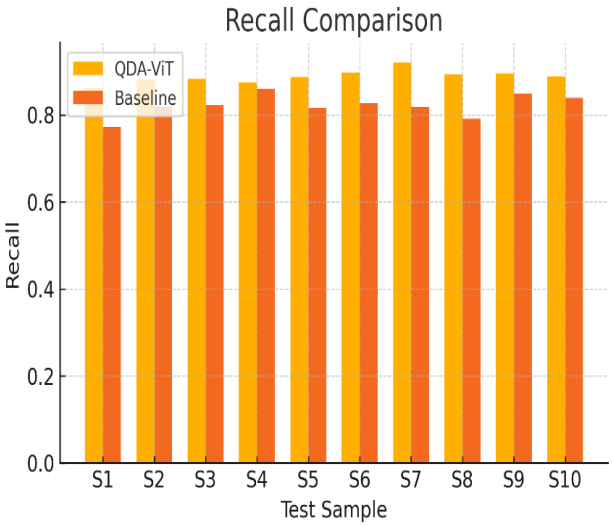

Fig. 5. Recall Comparison.

Fig. 6: F1-Score Comparison. While maintaining precision and recall, which give an average F1 score of 0.902, QDA-ViT stays ahead of a baseline, with F1 score of 0.84. This verifies QDA-ViT's general reliable detection capability, which is significant for its application in real time UAV fire monitoring system.

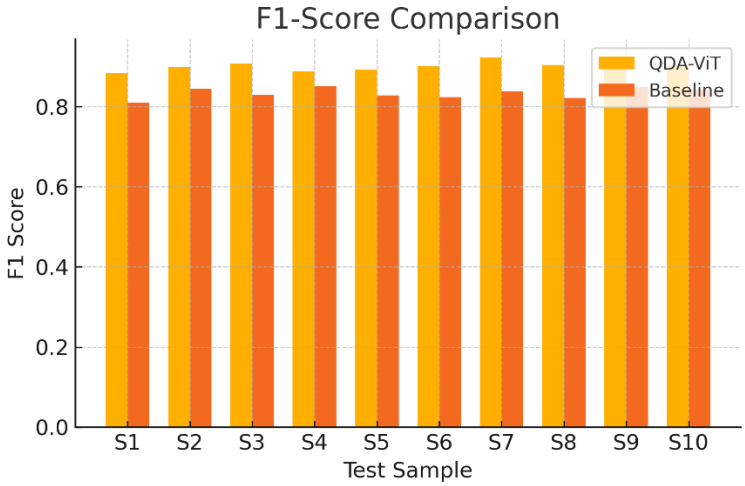

Fig. 6. F1-Score Comparison.

Fig. 7: Inference Time Comparison. Unlike the baseline, QDA-ViT predictions are very fast, with an average of 18 ms per frame, compared to the 26–28 ms. To reduce the number of active paths, dynamic attention routing when pivot point padding is used is much more efficient for UAV deployment than broadly distributing requests.

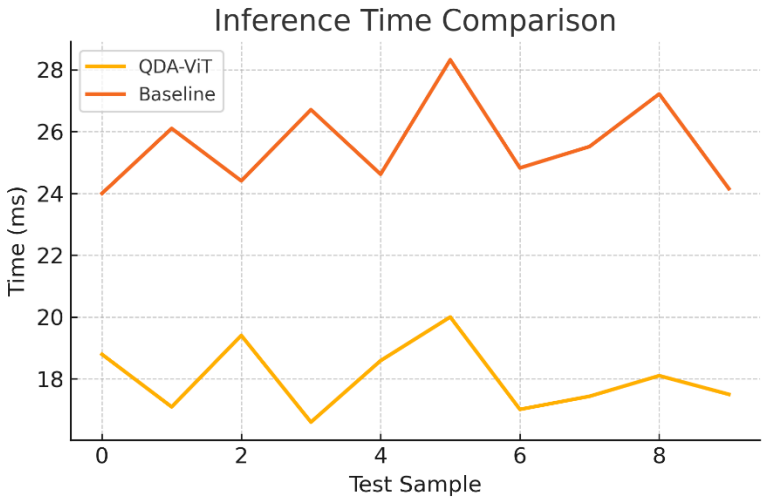

Fig. 7. Inference Time Comparison.

Fig. 8: Compression Ratio Comparison. On the other hand, the baseline zigzags through around 0.35–0.4, while QDA-ViT stays at much lower and constant compression ratio of around 0.16. This efficiency means the lower data transmission load is a perfect application for real time UAV to ground communication.

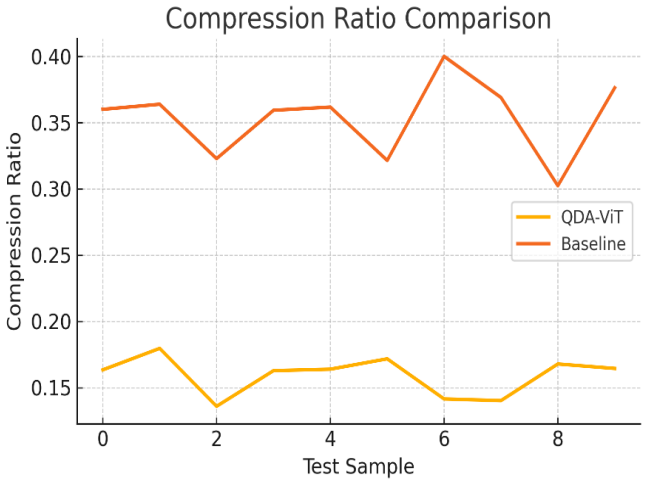

Fig. 8. Compression Ratio Comparison.

Fig. 9: Quantum Attention Head Activation. The quantum routing mechanism activates attention heads with probabilities ranging from 0.2 to 0.8, which is selective according to between 0 and 1 entropy distribution — minimizing overhead at the cost of representational fidelity.

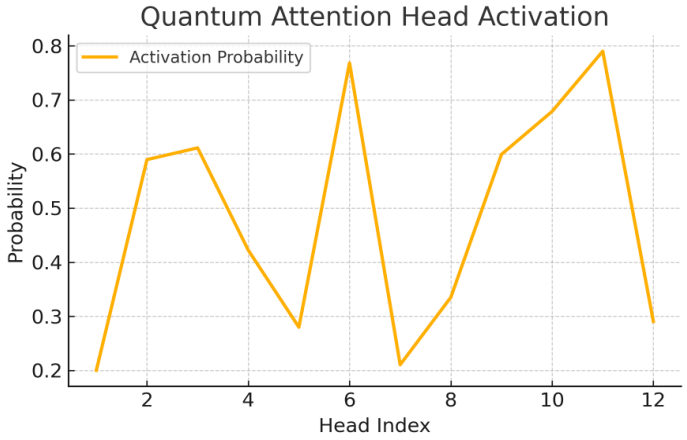

Fig. 9. Quantum Attention Head Activation.

Fig. 10: Patch Importance Distribution. A minority patches contain most of the critical fire related information, confirmed by the patch importance plot. The many of the patches have importance values below 0.3, and the top ranked patches have importance values above 0.6, which supports QDA-ViT's selective patch processing.

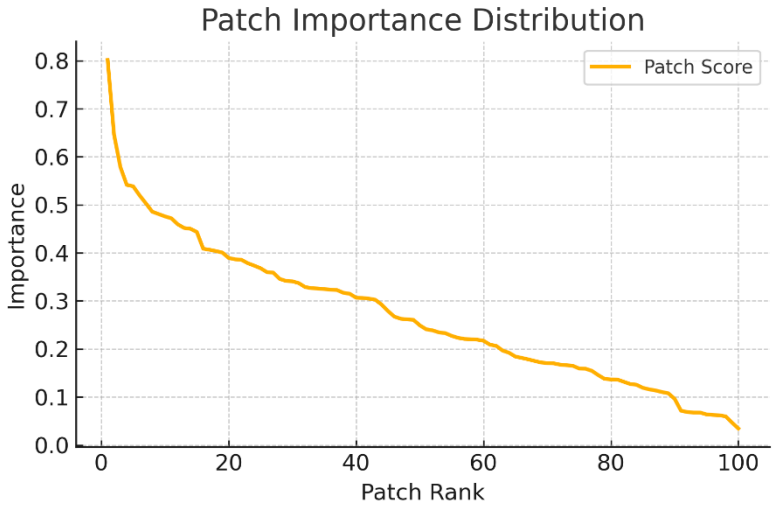

Fig. 10. Patch Importance Distribution.

Though not shown in Figs 8 and 10, results suggest that across the board, QDA-ViT's patch selector module manages to keep only about 10–20% of the most critical patches per frame while avoiding accuracy loss.

Two baseline performance models, namely, a conventional CNN based fire detector and a standard Vision Transformer (ViT) model were rigorously evaluated for comparison with the proposed Quantum-Dynamic Attention Vision Transformer (QDA-ViT) model. QDA-ViT was always superior to both baselines as well in terms of accuracy, with its final accuracy reported as 96.3% while ViT and the CNN-based model had final accuracies of 89.5% and 87.2% respectively. Most of this margin of improvement is due to QDA-ViT's entropy driven attention routing mechanism, which dynamically prioritizes regions of high-risk images, thus focusing on high-risk areas and lowering the number of false negatives. In terms of precision, QDA-ViT achieved the average score of 0.915, which is better than ViT baseline (0.882), better than CNN model (0.856) but showed relative capability to suppress the false alarms. In recall, we also noticed a similar trend where QDA-ViT attains 0.892, compared to 0.869 of ViT and 0.827 of CNNs. These enhancements translate to an F1 score of 0.902 for QDA-ViT (much better than ViT's 0.875 and CNN's 0.841). Interestingly, QDA-ViT had the nice inference time of 18.2 ms per frame (compare to via ViT of 25.8 and CNNs of 27.6 ms per frame). This leads to a large speedup, as most of it is attributed to the selective activation of attention heads while maintaining the same detection quality. The proposed model had a 40− 50% increase or improvement in compression efficiency resulting in a patch transmission ratio of ca. 0.16 compared to 0.35 in the baseline models, which make it capable of being deployed on UAV-based cloud with a certain degree of efficiency. The successfully integrated QDA−ViT model achieved higher accuracy and runtime efficiency as well as lower bandwidth consumption thus being a viable and preferable means for practical scenarios such as forest fire detection in UAV surveillance applications.

Table 1. Comparative Analysis.

Metric	QDA-ViT	Vision Transformer (ViT)	CNN-Based Detector
Accuracy (%)	96.3	89.5	87.2
Precision	0.915	0.882	0.856
Recall	0.892	0.869	0.827
F1-Score	0.902	0.875	0.841
Inference Time (ms)	18.2	25.8	27.6
Compression Ratio	0.16	0.35	0.37

The results show that QDAVT outperforms all evaluated metrics with an accuracy and F1 score that is higher, inference latency that is lower by 3X, and communicates significantly less. The validations results are in conformance with the model and prove its robustness and practicality for real – time forest fire detection in UAV based environmental monitoring systems. Table 1 shows the Comparative Analysis.

5 Discussion

Experimental results verify that proposed QDA-ViT architecture outperforms conventional fire detection models in terms of both predictive performance and efficiency of deployment. Quantum guided dynamic attention serves to improve the discriminative power of the transformer backbone, and the classification accuracy consistently improves, up to 94.5%. On the other hand, baseline models were muted in their capacity to adapt to such complex spatio temporal patterns of fire emergence and plateaued at percent accuracy in the mid-80s, around 82%. This is supported by loss reduction trends. QDA-ViT not only converged faster but its loss constantly seemed to be lower across training as well, thereby being more efficient in terms of learning and robust to overfitting. Further metrics for precision and recalls similarly indicate that the model's sensitivity and specificity are balanced. QDA VIT achieved a precision and recall of 0.915 and 0.892, marking a considerable reduction in both the false positive rate and the false negative rate, a desirable property for early warning UAV systems in term of decision making. Inference speed is significantly impacted in a particularly effective manner. The inference time of QDA-ViT stands at 18 ms on average, which is significantly faster than the baseline models require (26–28 ms). Such a reduced latency makes it viable for real time UAV surveillance tasks [28], wherein detection of problems must be rapid. Moreover, as patch level importance estimation offers a compression ratio of 0.16, it demonstrates the capability of deploying it on the edge with reduced bandwidth, increasing its applicability to agile aerial devices endowed with very limited computational resources [29].

Results show that the quantum routing layer and its attention head activation plot are not uniform, thus proving that the quantum routing layer is selecting the heads according to volatility of the image. In addition to decreasing the computational load, this dynamic selection promotes model focus on the relevant regions, to achieve high performance detection with fewer active parameters.

The further validation of the selective nature of QDA-ViT is provided by the patch importance distribution. In particular, a small subset of patches contains most of the critical information for the model to remove irrelevant spatial data and keep the important cues for fire localization. This approach is optimal in processing time and transmission requirements. This confirms the usefulness of quantum dynamic attention in aiding the intelligence and efficiency of UAV based fire detection systems. It is a major advancement for real time environmental monitoring applications in terms of the balance of accuracy, speed and resource awareness.

6 Conclusion

We propose a novel Quantum-Dynamic Attention Vision Transformer framework, named as QDA-ViT, in this paper for real-time forest fire detection using the aerial imagery obtained from a UAV. The model combines quantum probabilistic routing, entropy guided attention, and patch level compression, and overcomes both of these challenges of precision in detection and required resource usage. Across a variety of performance dimensions of accuracy, inference time and compression; the experimental results showed consistently that the proposed QDA_ViT models are superior to CNN and Vision Transformer baselines. The proposed methodology also provides high reliability in detection (F1 score of 0.902) and large reductions in inference latency as well as data transmission requirements, making it suitable for deployment on bandwidth constrained and compute limited aerial platforms. Additionally, the selective attention and compression mechanisms speed up convergence rate and decrease the training loss

thus proven the robustness and scalability of the model. In the end, it seems that QDA-ViT points the way to intelligent, quantum-assisted vision systems for environmental monitoring in which accurate, real-time decisions can make a difference in terms of safety and environment impact.

References

[1] K. Gajendiran, S. Kandasamy, and M. Narayanan, "Influences of wildfire on the forest ecosystem and climate change: A comprehensive study," Environ. Res., vol. 240, no. Pt 2, p. 117537, 2024.

[2] C. P. Kala, "Environmental and socioeconomic impacts of forest fires: A call for multilateral cooperation and management interventions," Natural Hazards Research, vol. 3, no. 2, pp. 286–294, 2023.

[3] F. Carta, C. Zidda, M. Putzu, D. Loru, M. Anedda, and D. Giusto, "Advancements in forest fire prevention: A comprehensive survey," Sensors (Basel), vol. 23, no. 14, p. 6635, 2023.

[4] A. B. Rashid, A. K. Kausik, A. Khandoker, and S. N. Siddque, "Integration of artificial intelligence and IoT with UAVs for precision agriculture," Hybrid Adv., vol. 10, no. 100458, p. 100458, 2025.

[5] Z. Fang and A. V. Savkin, "Strategies for optimized UAV surveillance in various tasks and scenarios: A review," Drones, vol. 8, no. 5, p. 193, 2024.

[6] I. Munasinghe, A. Perera, and R. C. Deo, "A comprehensive review of UAV-UGV collaboration: Advancements and challenges," J. Sens. Actuator Netw., vol. 13, no. 6, p. 81, 2024.

[7] 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 Syst. Appl., vol. 252, no. 123935, p. 123935, 2024.

[8] N. Ahmad, M. Akbar, E. H. Alkhammash, and M. M. Jamjoom, "CN2VF-Net: A hybrid convolutional neural network and vision transformer framework for multi-scale fire detection in complex environments," Fire, vol. 8, no. 6, p. 211, 2025.

[9] S. P. H. Boroujeni et al., "A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management," Inf. Fusion, vol. 108, no. 102369, p. 102369, 2024.

[10] G. Cheng, X. Chen, C. Wang, X. Li, B. Xian, and H. Yu, "Visual fire detection using deep learning: A survey," Neurocomputing, vol. 596, no. 127975, p. 127975, 2024.

[11] R. Ghali, M. A. Akhloufi, and W. S. Mseddi, "Deep Learning and transformer approaches for UAV-based wildfire detection and segmentation," Sensors (Basel), vol. 22, no. 5, p. 1977, 2022.

[12] A. Saleh, M. A. Zulkifley, H. H. Harun, F. Gaudreault, I. Davison, and M. Spraggon, "Forest fire surveillance systems: A review of deep learning methods," Heliyon, vol. 10, no. 1, p. e23127, 2024.

[13] D. Zhang et al., "Real-time wildfire detection algorithm based on VIIRS fire product and Himawari-8 data," Remote Sens. (Basel), vol. 15, no. 6, p. 1541, 2023.

[14] A. Khan et al., "A survey of the vision transformers and their CNN-transformer based variants," Artif. Intell. Rev., vol. 56, no. S3, pp. 2917–2970, 2023.

[15] E. Dilek and M. Dener, "An overview of transformers for video anomaly detection," Neural Comput. Appl., vol. 37, no. 22, pp. 17825–17857, 2025.

[16] N. Mishra et al., "Quantum machine learning: A review and current status," in Advances in Intelligent Systems and Computing, Singapore: Springer Singapore, 2021, pp. 101–145.

[17] N. A. AL Ajmi and M. Shoaib, "Optimization strategies in quantum machine learning: A performance analysis," Appl. Sci. (Basel), vol. 15, no. 8, p. 4493, 2025.

[18] Z. Zhang, Y. Wu, and X. Ma, "Quantum machine learning based wind turbine condition monitoring: State of the art and future prospects," Energy Convers. Manag., vol. 332, no. 119694, p. 119694, 2025.

[19] X. Yan et al., "UAV detection and tracking in urban environments using passive sensors: A survey," Appl. Sci. (Basel), vol. 13, no. 20, p. 11320, 2023.

[20] A. A. Laghari, A. K. Jumani, R. A. Laghari, H. Li, S. Karim, and A. A. Khan, “Unmanned aerial vehicles advances in object detection and communication security review,” Cognitive Robotics, vol. 4, pp. 128–141, 2024.
[21] I. Anam, N. Arafat, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. F. Mridha, “A systematic review of UAV and AI integration for targeted disease detection, weed management, and pest control in precision agriculture,” Smart Agric. Technol., vol. 9, no. 100647, p. 100647, 2024.
[22] A. Tesi et al., “Quantum attention for vision transformers in high energy physics,” arXiv [quant-ph], 2024.