Displays 89 (2025) 103070

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

Optimizing fire detection in remote sensing imagery for edge devices: A quantization-enhanced hybrid deep learning model☆

Syed Muhammad Salman Bukhari [a], Nadia Dahmani [b,c], Sujan Gyawali [d], Muhammad Hamza Zafar [e], Filippo Sanfilippo [e,f,*], Kiran Raja [g]

[a] Department of Electrical Engineering, Capital University of Science and Technology, Islamabad, Pakistan
[b] College of Technological Innovation, Zayed University, Abu Dhabi, United Arab Emirates
[c] LARODEC, Institut Superieur de Gestion de Tunis, Tunis, Tunisia
[d] Department of Computer Science, Lamar University, TX, USA
[e] Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway
[f] Department of Software Engineering, Kaunas University of Technology, Kaunas, Lithuania
[g] Norwegian University of Science and Technology (NTNU), Gjøvik, Norway

ARTICLE INFO

Keywords:
Bushfire detection
Quantization
Unmanned aerial vehicles (UAV)
Inception-resNet
Transformer models
Smart city applications

ABSTRACT

Wildfires are increasing in frequency and severity, presenting critical challenges for timely detection and response, particularly in remote or resource-limited environments. This study introduces the Inception-ResNet Transformer with Quantization (IRTQ), a novel hybrid deep learning (DL) framework that integrates multi-scale feature extraction with global attention and advanced quantization. The proposed model is specifically optimized for edge deployment on platforms such as unmanned aerial vehicles (UAVs), offering a unique combination of high accuracy, low latency, and compact memory footprint. The IRTQ model achieves 98.9% accuracy across diverse datasets and shows strong generalization through cross-dataset validation. Quantization significantly reduces the parameter count to 0.09M and memory usage to 0.13 MB, enabling real-time inference in 3 ms. Interpretability is further enhanced through Grad-CAM visualizations, supporting transparent decision-making. While achieving state-of-the-art performance, the model encounters challenges in visually ambiguous fire-like regions. To address these, future work will explore multi-modal inputs and extend the model towards multi-class classification. IRTQ represents a technically grounded, interpretable, and deployable solution for AI-driven wildfire detection and disaster response.

1. Introduction

Bushfires are among the most devastating natural disasters, with their frequency and intensity escalating due to climate change and extreme weather conditions. For instance, the 2019–2020 Australian bushfire season, known as "Black Summer", devastated over 24 million hectares of land, causing 33 direct fatalities and nearly 450 deaths due to smoke inhalation [1]. More recently, the 2025 Palisades and Eaton wildfires in Los Angeles destroyed over 12,000 structures and at least 27 fatalities [2]. These events underscore the critical need for advanced detection and response strategies to mitigate the impacts of such disasters. AI-powered drones have transformed wildfire management by enhancing situational awareness before, during, and after fires [3]. Remote sensing technologies, particularly aerial imagery, are pivotal in identifying affected areas. However, existing fire detection methods face significant challenges, including high false positive rates, limited adaptability to complex environments, and computational inefficiency on edge devices such as drones. For example, methods such as YOLO and Faster R-CNN struggle to generalize across datasets, often misclassifying reflections or bright light as fire [4,5]. Vision Transformers (ViT), while powerful, demand substantial computational resources, making them unsuitable for resource-constrained platforms. Addressing these challenges requires innovative solutions that balance accuracy, efficiency, and adaptability.

This study presents the Inception-ResNet Transformer with Quantization (IRTQ) model, a novel hybrid deep learning framework that integrates Inception-ResNet for multi-scale feature extraction and transformer modules for global contextual understanding. Through advanced quantization, IRTQ achieves high-precision fire detection with

☆ This paper was recommended for publication by Guangtao Zhai.

* Corresponding author at: Department of Engineering Sciences, University of Agder, Grimstad, 4879, Norway.
E-mail addresses: syedsalman.muhammad@gmail.com (S.M.S. Bukhari), nadia.dahmani@zu.ac.ae (N. Dahmani), sgyawali2@lamar.edu (S. Gyawali), muhammad.h.zafar@uia.no (M.H. Zafar), filippo.sanfilippo@uia.no (F. Sanfilippo).

https://doi.org/10.1016/j.displa.2025.103070
Received 23 November 2024; Received in revised form 25 April 2025; Accepted 28 April 2025
Available online 10 May 2025

significantly reduced memory and computation, enabling real-time deployment on resource-constrained platforms such as UAVs and embedded edge devices.

1.1. Motivation and contributions

Accurate and efficient fire detection in real-time, especially on edge devices deployed in disaster-prone or remote environments, remains a critical challenge. Traditional remote sensing and deep learning (DL) techniques often suffer from high false positive rates, limited adaptability, and computational complexity [2]. This research addresses these gaps by introducing a technically grounded and scalable framework with the following key innovations:

- **Hybrid architecture:** Combines Inception-ResNet and transformers to capture both local and global features, improving detection under complex conditions.
- **Quantization-aware efficiency:** Achieves real-time inference (3 ms) and 98.9% accuracy with only 0.09M parameters, ideal for embedded platforms.
- **Cross-dataset generalization:** Validated on geographically diverse data, showcasing adaptability and reliability in real-world environments.

2. Related work

The field of fire detection has seen significant advancements through the application of deep learning techniques. Among these, YOLO (You Only Look Once) algorithms, particularly YOLOv5, have shown success in real-time object detection due to their rapid processing speeds and high accuracy [6]. Variants such as YOLOv5-s and YOLOv5-x strike a balance between computational efficiency and detection accuracy, making them flexible for diverse applications [7]. However, YOLO-based methods often exhibit high false positive rates in complex scenarios, such as distinguishing fire from bright light sources or reflections, limiting their reliability in practical fire detection applications [5]. Faster R-CNN, a two-stage detection framework, enhances precision by refining object proposals [8]. While it works well for general object detection, its high computational demands restrict its use on devices with limited resources, such as drones and mobile edge devices [9]. Although lightweight adaptations for specific datasets, such as synthetic aperture radar (SAR), have emerged, they remain highly specialized and less suitable for real-time fire detection.

The integration of deep learning with drone-based systems has further expanded fire detection capabilities. Titu et al. [10] investigated lightweight deep learning models deployed on drones with edge computing to enable real-time fire detection. This approach reduces latency and computational overhead, proving practical for rapid response scenarios. However, scalability and accuracy under diverse environmental conditions remain key challenges. Transformer-based models, such as the Swin Transformer, have significantly improved high-resolution feature representation and global contextual analysis [11]. However, these models are computationally expensive, making them unsuitable for low-power devices such as drones. Also, DA-Net has shown promise in handling dense object distributions and varying aspect ratios, but its adaptability to diverse fire scenarios remains constrained [5]. Generative adversarial networks (GANs) have also been explored for image translation tasks. For instance, Boroujeni and Razi [12] proposed an improved conditional GAN (IC-GAN) for RGB-to-IR image translation tailored for forest fire monitoring. While effective, the computational complexity of GANs presents challenges for deployment on resource-limited devices.

Recent advancements in quantization techniques intend to address computational inefficiencies in resource-constrained environments. Nagel et al. [13] proposed methods to minimize accuracy degradation during quantization, while Li et al. [14] introduced adaptive quantization for enhanced flexibility. However, these approaches are often not optimized for hybrid architectures combining multi-scale feature extraction and global contextual awareness.

Hybrid architectures integrating Inception and ResNet modules have proved strong capabilities in multi-scale feature extraction and residual learning [15]. Combined with transformers, these architectures address limitations in scale invariance and global feature representation. Despite their potential, few studies have successfully applied such hybrid models for fire detection, leaving a gap that the proposed IRTQ model addresses. The IRTQ model integrates Inception-ResNet with transformers and advanced quantization techniques, providing a lightweight yet robust solution for real-time fire detection on resource-constrained devices. Unlike existing methods, IRTQ shows superior adaptability, achieving only 0.09 million parameters while maintaining high accuracy. Table 1 summarizes the comparison of existing methods and highlights the unique contributions of IRTQ.

3. Data analysis and preprocessing

3.1. Dataset 1

To precisely address the issue of forest fire detection, we have carefully selected and organized a well-balanced dataset [18]. The dataset consists of 1900 photos of good quality, each with a size of 250 × 250 pixels. All the images are in RGB format and have three color channels. The photographs were obtained from several search engines by utilizing specific search phrases related to forest fires. The following curation process involved careful cropping and cleaning of images to remove unnecessary elements such as people and fire-extinguishing equipment. This guaranteed that each image showed either the existence or absence of fire, establishing the foundation for a binary classification issue: fire or no fire. The dataset exhibits perfect balance, with an equal distribution of 950 photos for each class. In our experimental setup, we divided the dataset into 80% for training and 20% for testing. This division allowed us to thoroughly evaluate the model's performance in controlled but realistic situations.

3.2. Dataset 2

The DFS dataset [5] is designed primarily to enhance research in fire and smoke detection in diverse real-world situations. DFS is created by systematically gathering and labeling a large number of real-life images (9462 in total). Each image is carefully annotated according to tight rules to reduce the chances of misclassifying items that may resemble fire in terms of color and brightness. In this investigation, we only used images classified as fire from DFS to concentrate on the capabilities of fire detection. The images in the DFS are labeled according to the magnitude of the fire, which is a helpful characteristic to distinguish detection jobs depending on the amount of fire area covered. This dataset is very difficult and accurately represents real-life fire detection situations, providing a strong standard for assessing different object detection techniques. A series of extensive experiments were carried out to create a standard for comparison, and the dataset was evaluated using various training-testing divisions to determine the most effective setup for practical use.

Both datasets were subjected to a variety of preparation processes aimed at optimizing them for use with the proposed model, IRTQ. The procedures include standardizing the size of the images and normalizing the values of the pixel values. Additionally, augmentation methods such as rotations and flips are used to improve the resilience of our model. By engaging in thorough data preparation, we guarantee that our model is trained using high-quality and relevant data. This enables the model to effectively identify forest fires and perform well in various climatic situations, boosting its effectiveness and dependability. The preparation approaches and insights obtained from these datasets are

Table 1
Comparison of existing methods with the proposed IRTQ model.

Methodology	Key features	Limitations	IRTQ advantages
YOLOv5 [6]	Real-time detection, high accuracy	Prone to false positives in complex scenes	Lightweight design reduces false positives
FRCNN [8]	Two-stage detection, high precision	Computationally intensive	Optimized for edge devices
Swin transformer [11]	High-resolution feature representation	High training complexity	Integrates transformers with quantization
DA-Net [5]	Handles dense object distributions	Limited scalability	Generalizable across diverse datasets
FPN [16]	Multi-scale feature extraction	High memory demand	Memory-efficient with superior performance
Hausdorff IoU Integration [17]	Robust accuracy for complex scenarios	Computational overhead	Real-time performance without additional metrics
Proposed IRTQ model	Hybrid architecture, quantization, low parameters	–	Superior accuracy, real-time detection, resource-efficient

almost identical, highlighting the consistency and replicability of our methodology across many datasets. By using a consistent preprocessing method, we can fully utilize our model's efficient and minimalistic architecture, resulting in accurate and fast detection of fires.

3.3. Preprocessing techniques

The effective preprocessing plays a critical role in the success of DL models by ensuring that input data is consistent, representative, and optimized for extracting meaningful features. In this study, we design preprocessing steps to address challenges such as variability in fire characteristics, environmental conditions, and data quality. These techniques are uniformly applied across the datasets provided by Khan et al. [18] and Liu et al. [5].

- **Resizing:** We resize all images to 224 × 224 × 3 pixels to standardize input dimensions and ensure compatibility with the model architecture. This resolution balances computational efficiency with the preservation of essential features, such as complex fire contours, which are critical for accurate detection. Higher resolutions unnecessarily increase computational costs, while lower resolutions risk losing important details. We use bicubic interpolation to preserve image quality, ensuring minimal distortion during resizing.
- **Data Augmentation:** To enhance the diversity of the training data and improve model robustness, we apply the following augmentation techniques:
  - *Rotation:* We rotate images by 90 degrees to simulate diverse fire orientations, reflecting real-world scenarios where UAVs capture images from varying angles. This augmentation increases the model's ability to detect fires regardless of their orientation, which is vital for aerial monitoring.
  - *Flipping:* Horizontal and vertical flipping is performed to double the dataset size and introduce mirrored fire patterns. This ensures the model can identify fires even when their orientations are reversed, addressing scenarios such as reflections or mirrored imagery in UAV feeds.
  - *Cropping (Dataset 1):* For Dataset 1, we crop images to remove irrelevant elements, such as firefighting equipment and unrelated background objects. This step reduces noise and allows the model to focus on fire and non-fire regions, improving classification accuracy.
- **Normalization:** We apply min–max normalization to scale pixel intensity values to the range [0, 1]. This normalization standardizes input data across the datasets, ensuring that the model processes features consistently regardless of variations in lighting conditions or sensor characteristics. By scaling pixel values, we accelerate model convergence during training and minimize numerical instability. The normalization process is expressed as:

$$I'_{ijk} = \frac{I_{ijk} - \min(I)}{\max(I) - \min(I)} \times (\text{maxvalue} - \text{minvalue}) + \text{minvalue}, \quad (1)$$

where I_{ijk} is the original pixel intensity at position (i, j, k), and I'_{ijk} is the normalized value. For this study, minvalue = 0 and maxvalue = 1.

These preprocessing steps address specific challenges presented by the datasets, including variations in fire intensity, orientation, and environmental noise. Resizing ensures uniformity in input dimensions without compromising critical details, while augmentation techniques create a diverse and representative dataset. Normalization further enhances the model's ability to generalize across different environmental conditions by standardizing feature scales. By applying a consistent preprocessing framework, we establish a robust foundation for training and evaluating the proposed model under diverse scenarios.

4. Proposed model

This section presents the IRTQ model for fire detection, which fuses the strengths of Inception, ResNet, and Transformer architectures. Inception modules enable robust multi-scale feature extraction, ResNet blocks mitigate vanishing gradients through residual learning, and transformer modules provide global contextual awareness via self-attention. Together, they form a lightweight yet powerful model capable of high-accuracy classification on resource-constrained platforms. The overall architecture is depicted in Fig. 1.

4.1. Inception-ResNet

The Inception-ResNet backbone combines the advantages of multi-scale feature extraction and residual learning to build a deep and efficient representation of input images. Inception modules capture spatial patterns at varying receptive fields using parallel convolutional branches with different kernel sizes. Each branch b performs a convolution operation on the input $\mathbf{X}$, and the outputs are concatenated to preserve multiple spatial perspectives:

$$\mathbf{F}_{\text{inception}} = \|_{b=1}^{B} \text{Conv}_{k_b, s_b}(\mathbf{X}) \quad (2)$$

Here, Conv_{k_b, s_b} denotes a convolution with kernel size k_b and stride s_b, and B is the number of branches. The use of such diverse filters allows the model to understand visual phenomena like small flames or dispersed smoke at multiple scales.

To address the degradation problem associated with deeper networks, residual connections are applied. The concatenated output of the Inception module is passed through a transformation function $\mathcal{T}$ and added to the original input:

$$\mathbf{F}_{\text{res}} = \mathbf{X} + \mathcal{T}(\mathbf{F}_{\text{inception}}) \quad (3)$$

The transformation $\mathcal{T}$ typically includes batch normalization and non-linear activations. These residual pathways facilitate gradient flow and accelerate convergence during training.

To fuse features across both channel and spatial dimensions, we apply pointwise and spatial convolutions on the residual output:

$$\mathbf{F}_{\text{fusion}} = \text{Conv}_{1\times1}(\mathbf{F}_{\text{res}}) + \text{Conv}_{k_f, s_f}(\mathbf{F}_{\text{res}}) \quad (4)$$

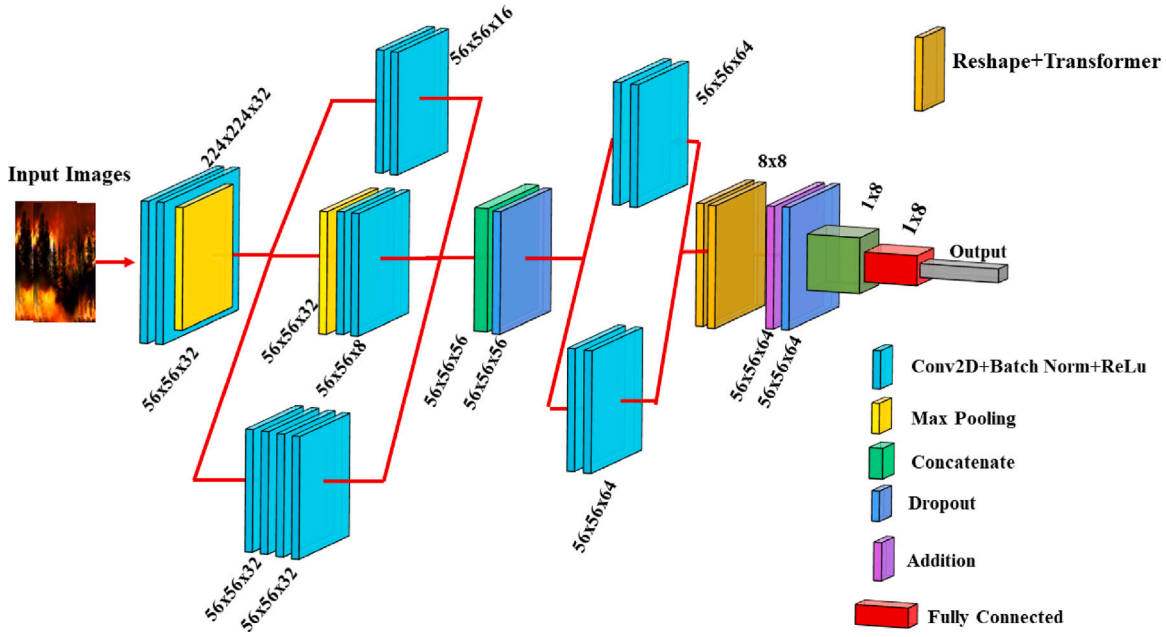

Fig. 1. The architecture of the proposed hybrid deep-learning model for fire detection.

This fusion generates a compact and expressive representation $\mathbf{F}_{\text{fusion}}$, which captures local spatial structures and semantic depth. It serves as the enriched feature map that feeds into the transformer module. Overall, the Inception-ResNet backbone provides a dense multi-resolution representation critical for detecting fires that vary in shape, scale, and intensity.

4.2. Transformer module integration

To enhance the model's understanding of global spatial context, we integrate transformer modules atop the Inception-ResNet backbone. Transformers excel at modeling long-range dependencies, a capability particularly useful in fire detection tasks where contextual cues (e.g., shape continuity or background landscape) are essential to disambiguate fire from non-fire regions [19].

Each transformer block begins with a self-attention mechanism that takes the feature map $\mathbf{F}_{\text{fusion}}$ and projects it into queries $\mathbf{Q}$, keys $\mathbf{K}$, and values $\mathbf{V}$. Attention weights are computed as:

$$\mathbf{A} = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \tag{5}$$

where d_k is the key dimension used for scaling. The attention-weighted output is then given by:

$$\mathbf{O}_{\text{att}} = \mathbf{A}\mathbf{V} \tag{6}$$

This output reflects the global influence of all spatial locations on every feature point. Each transformer block includes a residual connection with the backbone output and a feedforward network with normalization:

$$\mathbf{F}_{\text{transformer}} = \text{FFN}(\text{LayerNorm}(\mathbf{O}_{\text{att}} + \mathbf{F}_{\text{fusion}})) \tag{7}$$

This structure ensures that local features extracted by the convolutional layers are enriched with non-local context, supporting fine-grained scene understanding.

In fire detection, false positives often arise due to bright non-fire regions like sun reflections or vehicle headlights. The transformer's self-attention mitigates this by analyzing patterns across the entire spatial field. For example, transformer modules can distinguish real fire (which exhibits textural irregularities and spreads) from isolated light patches, enhancing robustness in complex scenes. By integrating this global reasoning capability, the model achieves superior classification performance while maintaining lightweight computation.

4.3. Quantization for efficient deep learning deployment

Quantization plays a pivotal role in enabling the deployment of deep learning models on resource-constrained platforms by reducing computational complexity and memory usage. In the IRTQ model, both weights and activations are quantized using 8-bit integer representations, allowing for fast, low-power inference without sacrificing performance.

The quantization process converts floating-point values f to discrete integer values q, using a scale factor s and zero-point z as follows:

$$q = \text{round}\left(\frac{f}{s}\right) + z \tag{8}$$

Here, s maps the real value range to a quantized range (e.g., 0–255), and z anchors the representation of zero. To prevent out-of-bound values, clamping is applied to fit within the data type's range.

In our model, weight quantization is applied statically during model conversion, while activation quantization is performed dynamically at inference time. A representative dataset is used for post-training calibration, during which distribution-aware scale and zero-point parameters are computed to minimize precision loss. This ensures that the quantized model preserves the behavior of its floating-point counterpart.

Quantization enables IRTQ to take advantage of integer-based hardware accelerators, offering significant improvements in inference speed and energy efficiency. In empirical evaluations, the quantized version of IRTQ achieves a memory footprint of only 0.13 MB and an inference latency of 3 ms, with less than a 0.1% drop in accuracy compared to the full-precision model. These improvements are critical for deploying fire detection models on UAVs and embedded edge devices, where power and real-time response are limited.

By tightly integrating quantization into both the training and deployment pipeline, IRTQ provides a highly optimized, resource-aware solution for fire detection without compromising robustness or reliability.

4.4. Mathematical framework of the hybrid model architecture

The IRTQ model can be formally described as a composite function that maps an input image $\mathbf{X}$ to a predicted output $\mathbf{Y}$ through a sequence of learnable transformations. First, the Inception-ResNet backbone, denoted as $f_{\text{Inception-ResNet}}$, extracts rich multi-scale features from the input:

$$\mathbf{F}_{\text{IR}} = f_{\text{Inception-ResNet}}(\mathbf{X}) \tag{9}$$

These features $\mathbf{F}_{\text{IR}}$ capture local patterns such as texture, shape, and edges relevant to identifying fire structures. Next, the transformer module $f_{\text{Transformer}}$ refines these features by incorporating global dependencies:

$$\mathbf{F}_{\text{T}} = f_{\text{Transformer}}(\mathbf{F}_{\text{IR}}) \tag{10}$$

To create a unified representation that retains both local and global characteristics, we fuse the two feature maps:

$$\mathbf{F} = \mathbf{F}_{\text{IR}} \oplus \mathbf{F}_{\text{T}} \tag{11}$$

Here, $\oplus$ represents an element-wise addition operation, selected for its simplicity and compatibility with quantized inference. The fused representation $\mathbf{F}$ is then passed through fully connected layers to yield the final prediction:

$$\mathbf{Y} = f_{\text{FC}}(\mathbf{F}) \tag{12}$$

This end-to-end formulation reflects a hierarchical processing strategy: the Inception-ResNet component excels at detecting detailed spatial structures, while the transformer layers provide a holistic view of the scene. Such synergy is crucial for distinguishing fire from visually similar non-fire phenomena — e.g., sunlight reflections or artificial light sources — which often confound traditional classifiers.

By integrating quantization directly into this hybrid architecture, IRTQ preserves accuracy while operating efficiently on low-power platforms. The entire model contains only 0.09 million parameters, demonstrating that the proposed architecture balances complexity, interpretability, and deployability—an essential combination for real-time wildfire monitoring systems.

4.5. Hyperparameters and model configuration

The configuration of hyperparameters plays a pivotal role in shaping the IRTQ model's performance, particularly its ability to generalize across diverse datasets while maintaining computational efficiency. Table 2 summarizes the final hyperparameter settings, which were selected through empirical tuning based on cross-validation results and deployment constraints.

A dropout rate of 0.5 was used to prevent overfitting by randomly deactivating neurons during training, enhancing the model's generalization ability. With only 0.09 million trainable parameters, the model retains a compact structure suitable for real-time applications on edge devices, without compromising performance. The learning rate was initialized at 0.01 to enable substantial weight updates in the early stages of training. A scheduled decay — reducing the rate by a factor of 0.25 every 15 epochs — supports convergence by allowing finer adjustments as the model approaches optimality. Training was conducted using an 80-10-10 train-validation-test split, ensuring fair evaluation and sufficient exposure to unseen data.

The transformer module is configured with a single block to preserve efficiency, given the lightweight design objectives. An embedding dimension of 8, two attention heads, and a feed-forward size of 32 strike a balance between model expressiveness and computational cost. A transformer dropout rate of 0.1 was used to prevent overfitting in attention layers. For the convolutional backbone, the initial convolutional layer uses 32 filters to extract fundamental spatial features. The Inception module utilizes a multi-branch structure with filter sizes [16, 32, 32, 8] to facilitate multi-scale feature learning. Residual modules with 64 filters support deeper integration of hierarchical features. Both the transformer sequence length and feature dimension were set to 8 to maintain alignment with the compact feature map size, optimizing throughput during inference.

Table 2
Hyperparameter settings for the hybrid model.

Hyperparameter	Value
Dropout Rate	0.5
Trainable Parameters	0.09M
Initial Learning Rate	0.01
Learning Rate Schedule	Decrease by a factor of 0.25 every 15 epochs
Training Epochs	Adaptive with a base of 15 epochs
Train-Validation-Test Split	80%–10%–10%
Number of Transformer Blocks	1
Transformer Embedding Dimension	8
Transformer Number of Heads	2
Transformer Feed-Forward Dimension	32
Transformer Dropout Rate	0.1
Initial Convolutional Layer Filters	32
Inception Module Filters	[16, 32, 32, 8]
Residual Module Filters	64
Transformer Sequence Length	8
Transformer Feature Dimension	8

This hyperparameter configuration was deliberately selected to support the IRTQ model's goal of high predictive accuracy under strict computational constraints. The integration of tuned architectural settings with quantization-aware design ensures reliable performance across real-world fire detection scenarios, including deployment on UAVs and embedded systems.

5. Results

This section provides a comprehensive empirical assessment and performance analysis of the IRTQ model, specifically designed for detecting fires in remote sensing data. Our assessment system goes beyond traditional measures and includes accuracy, precision, recall, and the F1 score. Additionally, we enhance our evaluation with innovative approaches like gradient-weighted class activation mapping (Grad-CAM) to provide meaningful visualizations of image gradients. These extensive metrics provide a thorough perspective on the model's ability to recognize and its effectiveness in applying knowledge to various datasets. The efficacy and adaptability of our approach are evaluated using a unique two-phase experimental configuration. At first, the model is trained separately using a main dataset to accurately capture the complex patterns and properties that are necessary for fire detection. The same model is assessed on a separate dataset without reutilizing the acquired weights from the first dataset. This strategy involves retraining the model using the second dataset and then assessing its performance on the first dataset. This cross-validation technique not only assesses the model's resilience and flexibility but also demonstrates its capability for practical use in early bushfire detection situations. By doing rigorous testing and analysis on these datasets, our goal is to provide a thorough showcase of the model's exceptional performance and significant contributions to the crucial field of fire detection.

5.1. Evaluation metrics

We evaluate our model using the following metrics, which are crucial for assessing its performance in fire detection tasks:

- **Accuracy**: Measures the proportion of correct predictions — both true positives and true negatives — among the total number of cases examined.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{13}$$

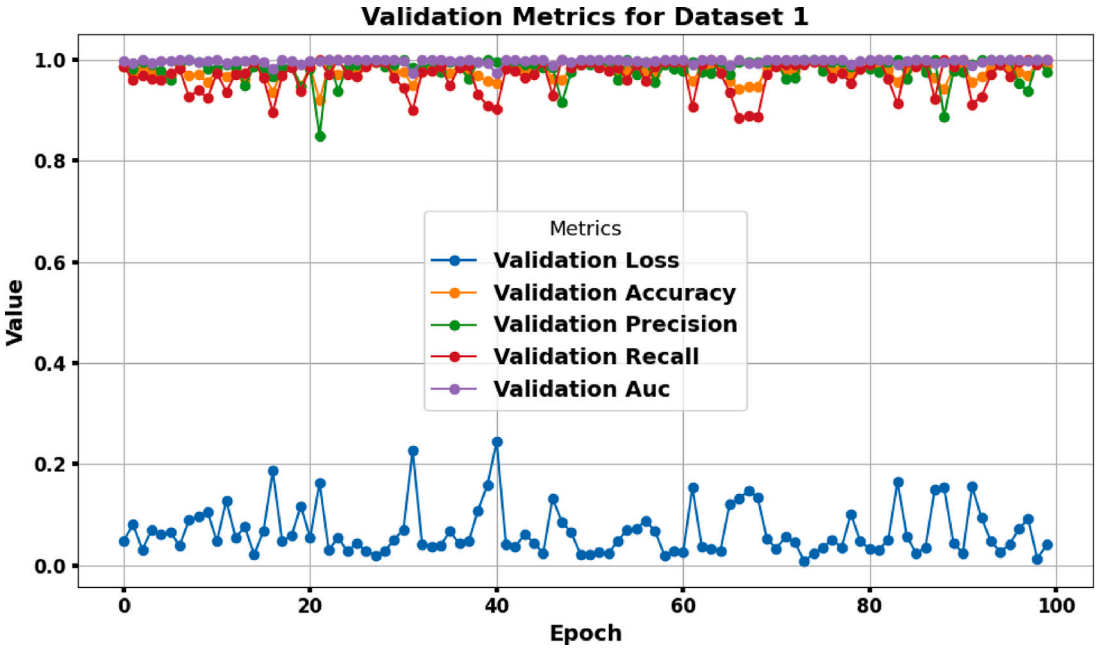

Fig. 2. Validation metrics over 100 epochs for Dataset 2. The graph illustrates validation loss, accuracy, precision, recall, and AUC, with loss displaying more variability compared to the relatively stable other metrics.

- **Precision**: Indicates the accuracy of positive predictions.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{14}$$

- **Recall** (Sensitivity): Reflects the model's ability to identify all relevant instances.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{15}$$

- **F1 Score**: Combines precision and recall into a single metric by taking their harmonic mean.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

- **ROC Curve**: Visualizes the trade-off between the true positive rate (sensitivity) and the false positive rate, essential for choosing an optimal threshold.

$$\text{TPR} = \text{Recall}, \quad \text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{17}$$

- **Grad-CAM (Gradient-weighted Class Activation Mapping)**: Uses gradients to show the target concept in the final convolutional layer, highlighting the important parts of the input image.

$$\text{Grad-CAM} = \text{ReLU}\left(\sum_k \alpha_k^c \cdot \text{A}^k\right) \tag{18}$$

These metrics will help validate the effectiveness of our model in real-world bushfire detection scenarios, demonstrating its potential as a robust tool in the domain of remote sensing for fire detection.

5.2. Comparative analysis of Scenario 1

In this section, we undertake a comprehensive evaluation of our proposed model, the IRTQ, under various testing conditions to elucidate its performance and robustness. The IRTQ model was first trained on Dataset 2 and then rigorously evaluated on Dataset 1 to assess its generalization capabilities across different datasets. This cross-dataset evaluation strategy is crucial for understanding the model's adaptability and robustness when applied to new, unseen data environments. The performance of the IRTQ model is assessed both with and without quantization to investigate the impact of computational optimizations on its efficiency and effectiveness. The results of this testing offer a thorough understanding of the IRTQ model's practical utility and demonstrate its robustness under real-world operational constraints.

5.2.1. Training and validation on Dataset 2

This subsection explores the training and validation phases of the model on the dataset, particularly focusing on how the model's performance metrics evolved over epochs. As depicted in Fig. 2, the model demonstrates strong and stable performance in terms of accuracy, precision, recall, and AUC, with most metrics consistently maintaining above 0.8 after initial fluctuations. Notably, the validation loss shows significant variability, reflecting potential challenges in the model's convergence during the initial epochs yet stabilizes towards the later epochs.

5.2.2. Testing without quantization on Dataset 1

The proposed IRTQ model was tested on Dataset 1 without quantization after being trained on Dataset 2 to see how well it could work with other datasets. The classification report, illustrated in Fig. 3, shows high precision, recall, and F1-scores for both 'Non-Fire' and 'Fire' classes. Notably, the model achieves a precision above 0.97 and a recall above 0.96 across both categories, with F1-scores similarly high. These metrics suggest the robust ability of the IRTQ model to generalize from Dataset 2 to Dataset 1 while maintaining high accuracy levels.

The confusion matrix in Fig. 4 shows that the model did a great job, with a lot of true positives (753 out of 760) and true negatives (758 out of 760), which means that it was very accurate overall and had very few false positives and false negatives. These results underscore the IRTQ model's reliable predictive power and its potential utility in real-world scenarios, where accurate differentiation between 'Fire' and 'Non-Fire' is crucial.

A comparative analysis of various models evaluated on Dataset 1 without quantization. The models compared include the proposed model, a CNN, and a Vision Transformer (ViT). The comparison focuses on key performance metrics such as precision, recall, and F1-score, highlighting the differences in their ability to classify fire and non-fire incidents accurately. The purpose of this analysis is to illustrate the efficacy of the proposed model relative to other well-established

Fig. 3. Classification report for the proposed IRTQ model evaluated on Dataset 1 showing precision, recall, and F1-score for 'Non-Fire' and 'Fire' classes.

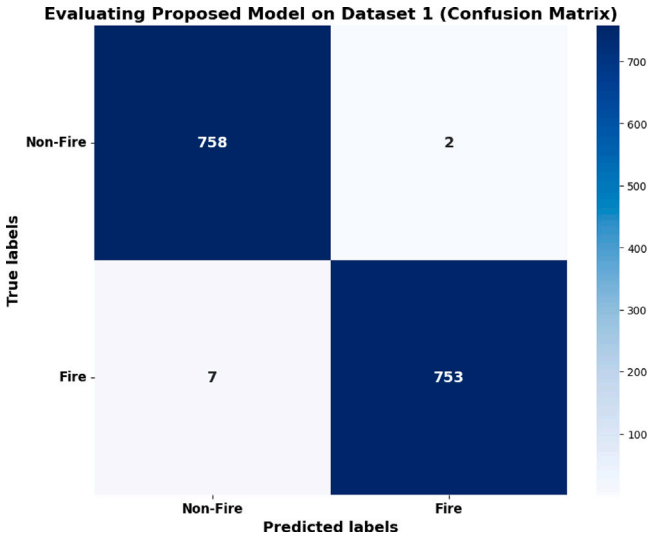

Fig. 4. Confusion matrix for the proposed IRTQ model evaluated on Dataset 1, detailing true positive, false positive, true negative, and false negative counts.

Table 3
Comparison of model performance on Dataset 1 without quantization.

Metric	Model	Non-Fire	Fire
Precision	Proposed model	0.99	1.00
	CNN	0.93	0.92
	ViT	0.96	0.95
Recall	Proposed model	1.00	0.99
	CNN	0.92	0.91
	ViT	0.96	0.96
F1-Score	Proposed model	0.99	0.99
	CNN	0.92	0.92
	ViT	0.95	0.93

architectures in handling the specific challenges presented by Dataset 1 (see Table 3).

Table 4
Comparison of different models with quantization on Dataset 1.

Metric	Model	Non-Fire	Fire
Precision	Proposed model	0.99	1.00
	CNN	0.95	0.95
	ViT	0.97	0.94
Recall	Proposed model	1.00	0.99
	CNN	0.91	0.92
	ViT	0.96	0.96
F1-Score	Proposed model	0.99	0.99
	CNN	0.95	0.95
	ViT	0.95	0.96

The table demonstrates the proposed model's superior precision and recall across both the fire and non-fire classes, affirming its robustness and efficiency compared to CNN and ViT. These metrics are crucial for ensuring that the model not only detects fires accurately but also minimizes false alarms, which are particularly critical in real-world fire detection scenarios.

5.2.3. Testing with quantization on Dataset 1

Quantization typically reduces the model size significantly, which can influence the precision, recall, and F1 score. This subsection details the performance of the quantized proposed model on Dataset 1, offering insights into the impact of quantization on model efficiency and effectiveness. The comparison between the non-quantized and quantized versions of the model highlights subtle variations in metric performance, demonstrating the model's robustness despite the size reduction. The following table compares the performance metrics before and after quantization, illustrating that while there is a slight decrease in precision for the Fire class, the model maintains high overall effectiveness (see Table 4):

This table shows that the quantized model achieves nearly identical scores compared to the non-quantized model, affirming the effectiveness of quantization in maintaining high performance while reducing computational requirements. This is particularly advantageous for deployment in resource-constrained environments.

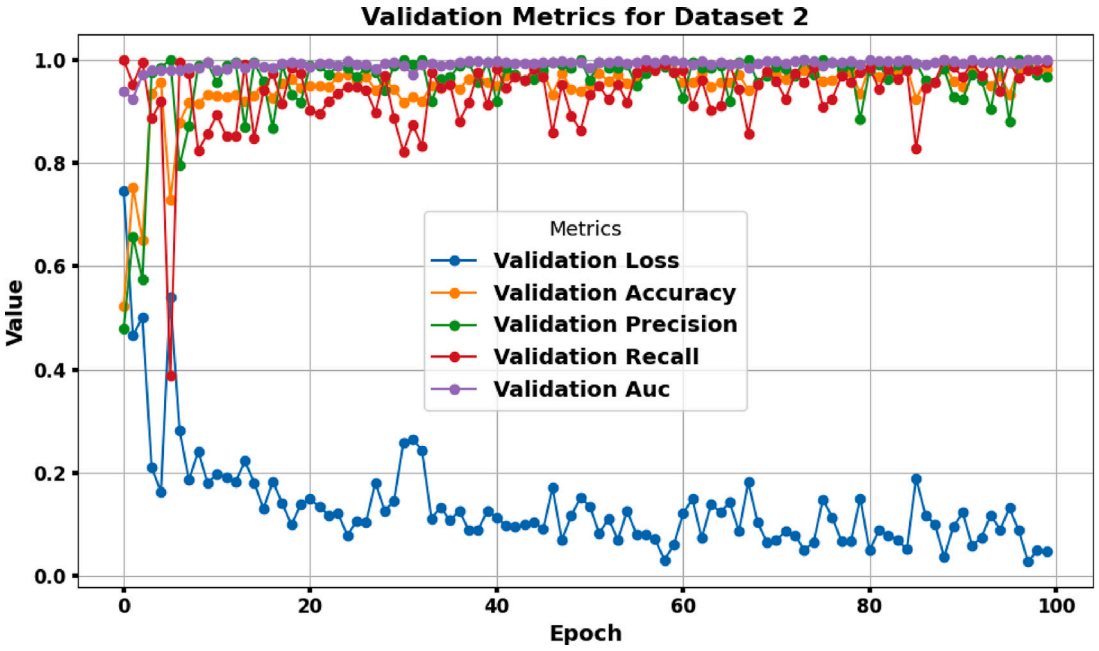

Fig. 5. Validation metrics for Dataset 2 over 100 epochs, showing trends in validation loss, accuracy, precision, recall, and AUC.

5.3. Comparative analysis of Scenario 2

This section provides an in-depth analysis of how the proposed model, which was trained on Dataset 1 and validated on Dataset 2, performs under different testing conditions—both with and without quantization. This evaluation is crucial to understanding the model's adaptability and effectiveness when applied in varying operational environments.

5.3.1. Training and validation on Dataset 1

The training and validation of the dataset were crucial in fine-tuning the model's parameters and assessing its ability to generalize effectively. Throughout the epochs of training, the validation metrics, as depicted in Fig. 5, demonstrated significant improvements and stabilization. Initially, the validation loss exhibited a sharp decrease, stabilizing at a low level as the epochs progressed, which indicates effective learning without overfitting. Validation accuracy, precision, recall, and the AUC consistently maintained high values, predominantly above 0.8, reflecting the robust generalization capabilities of the model. Notably, after initial fluctuations, all performance metrics converged to stable high values, indicating the model's resilience to overfitting and its strong predictive performance. This consistent high performance across metrics suggests that the model was trained with careful regularization and parameter tuning, which allowed it to keep its high accuracy and other important metrics. This proves that the model can effectively make predictions in complex real-world situations.

5.3.2. Testing without quantization on dataset 2

Following the training phase, the IRTQ model was tested on Dataset 2 without any quantization to evaluate its raw performance capabilities. This test aimed to assess how effectively the model could apply its learned capabilities to new, unseen data within the same dataset, and the results were exemplary. As shown in Fig. 6, the model achieved impressive precision, recall, and F1 scores across both non-fire and fire classes. The precision for the non-fire class was notably high and nearly perfect, while the fire class also showed strong precision and recall, demonstrating the model's accurate detection capabilities. The F1-score, a harmonic mean of precision and recall, was similarly high for both classes, indicating balanced performance. The confusion matrix, illustrated in Fig. 7, further supports the model's effectiveness.

Table 5
Comparison of model performance on Dataset 2 without quantization.

Metric	Model	Non-Fire	Fire
Precision	Proposed model	1.00	0.96
	CNN	0.92	0.90
	ViT	0.96	0.94
Recall	Proposed model	0.96	1.00
	CNN	0.91	0.91
	ViT	0.93	0.91
F1-Score	Proposed model	0.98	0.98
	CNN	0.91	0.90
	ViT	0.95	0.92

It correctly identified 2411 Non-Fire cases with only 92 false positives and almost perfectly categorized the Fire cases with only one false negative. Such results underscore the model's high reliability and robustness in classifying fire events accurately. These metrics highlight the model's stability and reliability in predicting the correct classes without computational optimizations. This performance is critical for real-world applications where accurate and timely fire detection is essential for effective response and mitigation efforts.

A comparative analysis of various models evaluated on Dataset 2 without quantization. The analysis includes the proposed model, a CNN, and a ViT. The comparison highlights the precision, recall, and F1-score of each model to evaluate their effectiveness in accurately classifying fire and non-fire events within Dataset 2. The goal is to showcase the relative performance of the proposed model against other standard architectures in terms of accuracy and reliability (see Table 5).

The table delineates the performance of each model, illustrating that the proposed model consistently achieves higher precision and recall rates compared to CNN and ViT. This superior performance underlines the proposed model's enhanced ability to identify true fire and non-fire scenarios accurately, thereby reducing the risk of false alarms and missed detections. Such capabilities make it exceptionally suited for practical applications where reliability and precision are paramount.

5.3.3. Testing with quantization on dataset 2

Quantization is a technique aimed at reducing the computational demands of models, making them more suitable for deployment in

Fig. 6. Classification report for the IRTQ model on Dataset 2 showing precision, recall, and F1-score for Non-Fire and Fire classes.

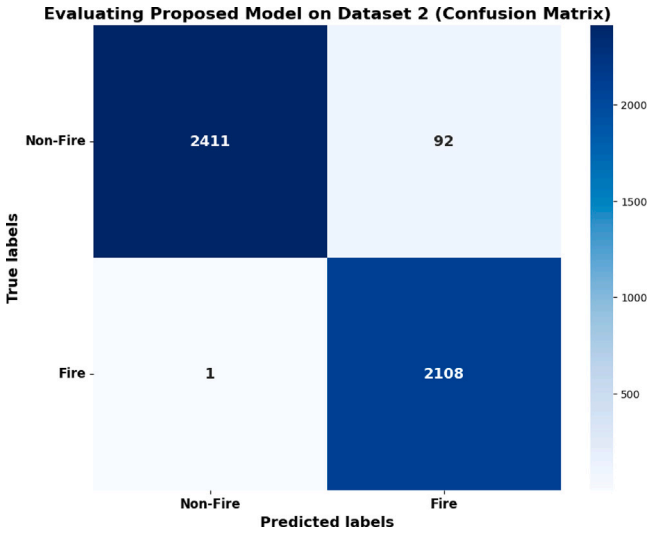

Fig. 7. Confusion matrix for the IRTQ model tested on Dataset 2, detailing the true and predicted classifications for the non-fire and Fire categories.

environments with limited computational resources. This subsection assesses the effect of quantization on the performance of the IRTQ model when tested on Dataset 2, with a focus on precision, recall, F1-score, and overall accuracy. Quantization did not have a significant impact on the model's predictive performance. The classification report indicates slight changes in precision and recall values before and after quantization, but the overall effectiveness of the model remains consistent. Specifically, the precision for non-fire cases remained perfect at 1.00 before and after quantization, though the recall slightly decreased from 0.97 to 0.96. For fire cases, the recall improved to a perfect score from 0.97 to 1.00, demonstrating the model's enhanced ability to identify all fire cases correctly, albeit at a slightly reduced precision from 1.00 to 0.96 (see Table 6).

The data shows that quantization effectively maintains high-performance levels with small changes in precision and recall. This supports the model's usefulness in environments with limited resources

Table 6
Comparison of different models with quantization on Dataset 2.

Metric	Model	Non-Fire	Fire
Precision	Proposed model	1.00	0.96
	CNN	0.93	0.91
	ViT	0.97	0.97
Recall	Proposed model	0.96	1.00
	CNN	0.92	0.90
	ViT	0.96	0.96
F1-Score	Proposed model	0.98	0.98
	CNN	0.95	0.93
	ViT	0.97	0.95

without significantly lowering its accuracy or dependability. These findings highlight the model's robustness and adaptability, affirming its potential for real-world applications where computational efficiency is as crucial as predictive accuracy.

5.4. Gradient-based class activation mappings

In Fig. 8, we present Grad-CAM visualizations derived from randomly selected images across both datasets. These visual aids are instrumental in showing how the quantized models, using weights from their respective datasets, maintain their focus on critical regions within the images — the actual fire zones — while disregarding irrelevant noise or distractors. The heatmaps provide empirical evidence of the model's capacity to accurately highlight fire locations, ensuring reliable fire detection across diverse scenarios. This precision is crucial for the dependability of fire detection systems, particularly in time-sensitive applications such as wildfire monitoring. The visualizations reveal the models' ability to discern fire's shape, color, and texture amidst complex visual patterns in remote sensing data. Additionally, Grad-CAM visualizations are not only valuable for model interpretation but also play a significant role in model refinement. For example, heat maps indicating sporadic attention to non-fire regions may highlight the need for additional training, improved dataset diversity, or the inclusion of advanced regularization techniques. These insights can be leveraged to fine-tune model parameters and optimize performance further. However, despite their utility, Grad-CAM visualizations have

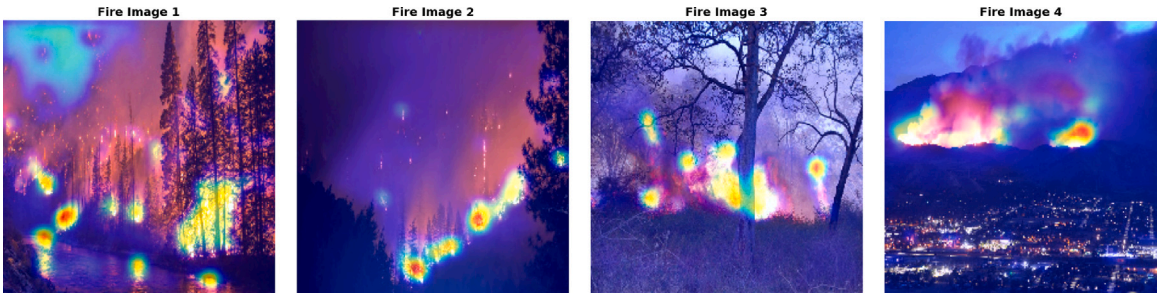

Fig. 8. Grad-CAM visualizations highlighting the model's focal points in fire detection across images from both datasets.

limitations, particularly when applied to ambiguous fire-like regions. For instance, areas with sunlight reflections, smoke patterns, or orange-hued objects may exhibit similar visual characteristics to actual fire zones, potentially leading to misleading activations in the heatmaps. This limitation highlights a potential vulnerability in interpretability, as false positives could arise in visually complex environments. Such occurrences underscore the need for caution when relying solely on Grad-CAM visualizations for model evaluation. To address these limitations, future work could explore the integration of Grad-CAM with complementary interpretability techniques, such as Local Interpretable Model-agnostic Explanations (LIME) or SHapley Additive exPlanations (SHAP), to validate and cross-check activation regions. Additionally, domain-specific preprocessing steps, such as spectral filtering or temporal analysis, could help differentiate fire from fire-like features. By combining these approaches, the reliability of interpretability frameworks can be significantly enhanced, particularly for challenging and ambiguous scenarios. The subplots in Fig. 8 illustrate the models' targeted attention on fire regions under diverse conditions, showcasing their adaptability and effectiveness in real-world fire detection tasks. While these results validate the robustness of the models, the discussion of limitations provides a roadmap for future advancements in both model development and interpretability methods.

5.5. Comparative analysis of deep learning models

In this subsection, we present a comparative analysis of several DL models, including the proposed IRTQ model, to assess their effectiveness in real-time fire detection tasks. The key performance metrics, accuracy, precision, recall, parameter size, memory usage, and inference time, were evaluated across diverse datasets to highlight each model's strengths and limitations in edge computing environments. Table 7 summarizes the results. The IRTQ model consistently outperforms traditional and recent architectures by achieving 98.9% across accuracy, precision, and recall. Its compact architecture (only 0.09M parameters), minimal memory footprint (0.13 MB), and extremely low inference time (3 ms) show its suitability for deployment on resource-constrained devices such as UAVs. These performance gains are primarily attributed to the hybrid architecture and advanced quantization techniques that preserve accuracy while reducing computational cost. To strengthen the comparative rigor, two recent models have been included. YOLOGX [20], a YOLOv8-based model optimized for forest fire detection, delivers competitive recall and achieves high inference speed (8.7 ms) with only 6.2M parameters. Meanwhile, FireNet [21], a hybrid CNN-Transformer architecture, shows strong classification accuracy (98.43%) with an ultra-lightweight design (0.318M parameters), highlighting the feasibility of transformer-based models for edge scenarios. IRTQ exceeds current fire detection models by providing a better combination of accuracy, speed, and ease of implementation. This makes IRTQ a top choice for fire detection due to its strong technical features and real-world usefulness.

6. Discussion

This section provides a critical analysis of the findings presented in the results, highlighting key insights, limitations, and potential opportunities for future research. The goal is to place the study's contributions in context while acknowledging areas for improvement and expansion. The proposed IRTQ model shows significant advancements in bushfire detection through a hybrid architecture combining Inception-ResNet and transformers, enhanced with advanced quantization techniques. The key findings are summarized as follows:

- The model achieves high accuracy (98.9%) across diverse datasets, outperforming benchmark methods such as CNN and ViT.
- Quantization reduces computational demands, enabling real-time deployment on resource-constrained devices, with minimal impact on performance.
- Grad-CAM visualizations validate the model's ability to focus on critical fire regions, ensuring interpretability.

These findings highlight the model's potential for practical applications in time-critical scenarios such as wildfire monitoring and management.

6.1. Limitations

While the proposed IRTQ model offers substantial improvements, certain limitations need to be acknowledged:

- **Ambiguity in Fire-Like Regions**: The model occasionally misclassifies fire-such as regions caused by reflections, smoke, or orange-hued objects, leading to false positives in complex scenes.
- **Dataset Dependency**: The model's performance is highly dependent on the quality and diversity of training datasets. Limited representation of specific environmental conditions may impact generalizability.
- **Trade-offs in Quantization**: Although quantization significantly enhances computational efficiency, slight reductions in precision were observed in some scenarios, particularly for fire classification in edge cases.
- **Limited Multi-Class Capability**: The current implementation focuses solely on binary classification (fire vs. non-fire). Expanding to multi-class classification could improve usability in broader disaster management contexts.

Addressing these limitations is crucial for further enhancing the model's reliability and applicability.

6.2. Future work

Building on the limitations and findings of this study, several directions for future research are proposed:

Table 7
Performance comparison of various fire detection models.

Model	Image set size	Memory (MB)	Accuracy (%)	Precision (%)	Recall (%)	TP	FP	Parameters (M)	Inference time (ms)
CNN [22]	224	1.7	92.04	90.0	93.0	–	9.34	–	43.68
CNN-LSTM [23]	–	2.0	91.5	92.0	91.0	–	–	–	20.0
Faster R-CNN [24]	–	2.9	93.36	95.0	92.0	–	5.61	–	46.3
YOLOv5-s [25]	320	3.1	95.0	95.5	94.3	–	–	–	–
YOLOv5-s [26]	640	3.5	94.6	90.1	90.56	–	–	7.26	15.7
YOLOGX [20]	–	–	–	77.20	75.70	–	–	6.2	8.7
FireNet [21]	–	–	98.43	–	–	–	–	0.318	–
Proposed IRTQ model	**462**	**0.13**	**98.9**	**98.9**	**98.9**	**98.9**	**0.09**	**0.09**	**3.0**

- **Enhancing Interpretability**: Integrate complementary interpretability techniques, such as LIME or SHAP, to improve the reliability of Grad-CAM visualizations, particularly in ambiguous fire-like regions.
- **Integrating Multi-Modal Data**: Explore the use of multi-modal inputs, such as combining RGB and thermal imagery, to improve detection robustness across diverse conditions.
- **Expanding Datasets**: Collect and curate more diverse datasets, including images from different geographic and environmental contexts, to enhance model generalizability and adaptability.
- **Advanced Optimization Techniques**: Investigate alternative optimization strategies, such as pruning and mixed precision training, to further reduce computational demands while maintaining performance.
- **Multi-Class Classification**: Extend the model to support multi-class classification, enabling differentiation between fire, smoke, and other related phenomena.
- **Real-Time Validation**: Test the model in real-world operational scenarios using UAVs to evaluate its performance and reliability under dynamic conditions.

Future research will explore YOLOv7-YOLOv12 for a more complete comparison. While excluded here due to computational demands, evaluating their performance with quantization or pruning could reveal their potential for edge-based fire detection, further refining our approach and advancing fire detection technology.

The discussion highlights the strengths of the IRTQ model while critically evaluating its limitations and identifying opportunities for future research. By addressing these challenges, the proposed framework can evolve into a more robust and versatile solution for wildfire management and beyond.

7. Conclusion

This study presents a significant advancement in bushfire detection through the innovative Inception-ResNet Transformer with Quantization (IRTQ) model, which integrates Inception-ResNet and transformer architectures with advanced quantization techniques. By addressing the limitations of existing methods, the proposed model delivers precise, efficient, and reliable fire detection personalized for deployment on resource-constrained platforms such as UAVs and edge devices. Extensive cross-dataset evaluations show the model's robustness, achieving exceptional accuracy, precision, and recall across diverse scenarios, making it well-suited for real-world applications. The integration of quantization substantially enhances computational efficiency while maintaining high performance, enabling real-time deployment in time-critical situations. Grad-CAM visualizations further validate the interpretability of the model by highlighting critical regions relevant to fire detection, ensuring transparency. This interpretability also provides actionable insights for refining the model's focus in challenging scenarios, such as ambiguous fire-like regions. While this research establishes a strong foundation, several opportunities for future advancements remain. Addressing limitations, such as the occasional misclassification of fire-like regions and dependency on dataset quality, is critical for enhancing generalizability. Future work should explore multi-modal data integration, including RGB and thermal imagery, to strengthen robustness under diverse conditions. The IRTQ model represents a transformative approach to wildfire management by balancing high performance and computational efficiency. Its adaptability, coupled with hard validation, highlights its potential for broader disaster management applications. The insights gained from this research cover the way for future developments in resource-efficient, interpretable, and reliable AI-driven fire detection systems.

CRediT authorship contribution statement

Syed Muhammad Salman Bukhari: Conceptualization, Methodology, Supervision, Resources, Project administration. **Nadia Dahmani:** Validation, Data curation, Software. **Sujan Gyawali:** Conceptualization, Formal analysis, Investigation. **Muhammad Hamza Zafar:** Visualization, Data curation. **Filippo Sanfilippo:** Supervision, Project administration, Formal analysis, Funding acquisition. **Kiran Raja:** Investigation, Software.

Declaration of competing interest

All authors claim that there is not any conflict of interest regarding the above submission. The work of this submission has not been published previously. It is not under consideration for publication elsewhere. Its publication is approved by all authors and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

Data availability

The data used in this work is publically available.

References

[1] G. Cook, A. Dowdy, J. Knauer, M. Meyer, P. Canadell, P. Briggs, Australia's Black Summer of fire was not normal, CSIRO (2021).
[2] T. Magazine, How the Los Angeles fires compare to historic wildfires, Time (2025).
[3] S.P.H. Boroujeni, A. Razi, S. Khoshdel, F. Afghah, J.L. Coen, L. O'Neill, K.G. Vamvoudakis, A comprehensive survey of research towards AI-enabled unmanned aerial systems in pre-, active-, and post-wildfire management, Inf. Fusion 102369 (2024).
[4] N. Su, Z. Huang, Y. Yan, C. Zhao, S. Zhou, Detect larger at once: Large-area remote-sensing image arbitrary-oriented ship detection, IEEE Geosci. Remote. Sens. Lett. 19 (2022) 1–5.
[5] Z. Liu, et al., Deep learning based method for fire detection, 2023.
[6] J.M. Topple, J.A. Fawcett, MiNet: Efficient deep learning automatic target recognition for small autonomous vehicles, IEEE Geosci. Remote. Sens. Lett. 18 (6) (2020) 1014–1018.
[7] M.L. Mekhalfi, C. Nicolò, Y. Bazi, M.M. Al Rahhal, N.A. Alsharif, E. Al Maghayreh, Contrasting YOLOv5, transformer, and EfficientDet detectors for crop circle detection in desert, IEEE Geosci. Remote. Sens. Lett. 19 (2021) 1–5.
[8] Z. Zhao, P. Tang, L. Zhao, Z. Zhang, Few-shot object detection of remote sensing images via two-stage fine-tuning, IEEE Geosci. Remote. Sens. Lett. 19 (2021) 1–5.
[9] Y. Li, S. Zhang, W.-Q. Wang, A lightweight faster R-CNN for ship detection in SAR images, IEEE Geosci. Remote. Sens. Lett. 19 (2020) 1–5.

[10] M.F.S. Titu, M.A. Pavel, G.K.O. Michael, H. Babar, U. Aman, R. Khan, Real-time fire detection: Integrating lightweight deep learning models on drones with edge computing, Drones 8 (9) (2024) 483.
[11] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, IEEE Trans. Pattern Anal. Mach. Intell. 44 (10) (2022) 7839–7853.
[12] S.P.H. Boroujeni, A. Razi, Ic-gan: An improved conditional generative adversarial network for rgb-to-ir image translation with applications to forest fire monitoring, Expert Syst. Appl. 238 (2024) 121962.
[13] M. Nagel, M. Fournarakis, R.A. Amjad, A white paper on neural network quantization, 2023, arXiv preprint arXiv:2301.05678.
[14] Y. Li, W. Zhang, H. Zhou, Adaptive quantization for efficient deployment of neural networks on edge devices, IEEE Trans. Neural Netw. Learn. Syst. 35 (1) (2024) 123–134.
[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
[16] X. Dong, R. Fu, Y. Gao, Y. Qin, Y. Ye, B. Li, Remote sensing object detection based on receptive field expansion block, IEEE Geosci. Remote. Sens. Lett. 19 (2021) 1–5.
[17] L. Wang, X. Mu, C. Ma, J. Zhang, Hausdorff iou and context maximum selection nms: Improving object detection in remote sensing images with a novel metric and postprocessing module, IEEE Geosci. Remote. Sens. Lett. 19 (2021) 1–5.
[18] A. Khan, B. Hassan, Dataset for forest fire detection, Mendeley Data 1 (2020) 2020.
[19] Y. Liu, G. Sun, Y. Qiu, L. Zhang, A. Chhatkuli, L. Van Gool, Transformer in convolutional neural networks, 2021, arXiv preprint arXiv:2106.03180. 3.
[20] X. Li, W. Zhang, R. Chen, H. Gao, YOLOGX: An improved forest fire detection algorithm based on YOLOv8, Front. Environ. Sci. 12 (2024) 1486212.
[21] Y. He, A. Sahma, X. He, R. Wu, R. Zhang, FireNet: A lightweight and efficient multi-scenario fire object detector, Remote. Sens. 16 (21) (2024) 4112.
[22] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, S.W. Baik, Efficient deep CNN-based fire detection and localization in video surveillance applications, IEEE Trans. Syst. Man Cybern.: Syst. 49 (7) (2018) 1419–1434.
[23] M.D. Nguyen, H.N. Vu, D.C. Pham, B. Choi, S. Ro, Multistage real-time fire detection using convolutional neural networks and long short-term memory networks, IEEE Access 9 (2021) 146667–146679.
[24] C. Chaoxia, W. Shang, F. Zhang, Information-guided flame detection based on faster R-CNN, IEEE Access 8 (2020) 58923–58932.
[25] K. Guo, C. He, M. Yang, S. Wang, A pavement distresses identification method optimized for YOLOv5s, Sci. Rep. 12 (1) (2022) 3542.
[26] S. Liu, J. Feng, Q. Zhang, B. Peng, A real-time smoke and fire warning detection method based on an improved YOLOv5 model, in: 2022 5th International Conference on Pattern Recognition and Artificial Intelligence, PRAI, IEEE, 2022, pp. 728–734.