

Article

Lightweight Deep Learning Model for Fire Classification in Tunnels

Shakhnoza Muksimova ¹, Sabina Umirzakova ^{1,*}, Jushkin Baltayev ² and Young-Im Cho ^{1,*}¹ Department of Computer Engineering, Gachon University, Sujeong-gu, Seongnam-si 461-701, Gyeonggi-do, Republic of Korea; shakhnoza02@gachon.ac.kr² Department of Information Systems and Technologies, Tashkent State University of Economic, Tashkent 100066, Uzbekistan; j_baltayev@tsue.uz

* Correspondence: sabinatuit@gachon.ac.kr (S.U.); yicho@gachon.ac.kr (Y.-I.C.)

Abstract: Tunnel fires pose a severe threat to human safety and infrastructure, necessitating the development of advanced and efficient fire detection systems. This paper presents a novel lightweight deep learning (DL) model specifically designed for real-time fire classification in tunnel environments. This model integrates MobileNetV3 for spatial feature extraction, Temporal Convolutional Networks (TCNs) for temporal sequence analysis, and advanced attention mechanisms, including Convolutional Block Attention Modules (CBAMs) and Squeeze-and-Excitation (SE) blocks, to prioritize critical features such as flames and smoke patterns while suppressing irrelevant noise. The model is trained on a custom dataset containing real tunnel fire incidents generated using a newly prepared dataset. This approach enhances the model generalization capabilities, enabling it to handle diverse fire scenarios, including those with low visibility, high smoke density, and variable ventilation conditions. Deployment optimizations, such as quantization and layer fusion, ensure computational efficiency, achieving an average inference time of 12ms/frame, making it suitable for resource-constrained environments like IoT and edge devices. The experimental results demonstrate that the proposed model achieves an accuracy of 96.5%, a precision of 95.7%, and a recall of 97.2%, significantly outperforming state-of-the-art (SOTA) models such as ResNet50 and YOLOv5 in both accuracy and real-time performance. Robustness tests under challenging conditions validate model reliability and adaptability, marking it as a critical advancement in tunnel fire detection systems. This study provides valuable insights into the design and deployment of efficient fire classification systems for safety-critical applications. The proposed model offers a scalable, high-performance solution for tunnel fire monitoring and establishes a benchmark for future research in real-time video-based classification under complex environmental conditions.

Keywords: fire detection systems; tunnel safety; real-time classification; attention mechanisms; edge computing; lightweight



Academic Editor: Changkun Chen

Received: 5 January 2025

Revised: 30 January 2025

Accepted: 19 February 2025

Published: 20 February 2025

Citation: Muksimova, S.; Umirzakova, S.; Baltayev, J.; Cho, Y.-I. Lightweight Deep Learning Model for Fire Classification in Tunnels. *Fire* **2025**, *8*, 85. <https://doi.org/10.3390/fire8030085>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tunnel fires are among the most hazardous incidents in underground transportation systems, posing severe risks to human safety and infrastructure integrity [1,2]. These fires are characterized by rapid flame propagation, toxic smoke production, and extreme heat, all of which complicate emergency response and evacuation [3,4]. Historical events, such as the Mont Blanc Tunnel fire in 1999 and the Gotthard Tunnel fire in 2001, have highlighted the catastrophic consequences of insufficient fire detection and response systems [5]. These incidents resulted in significant loss of life and economic damages, underscoring the critical

need for robust fire detection and classification mechanisms. Traditional fire detection systems rely on sensors such as heat detectors and smoke alarms [6,7]. While effective in certain scenarios, these systems are often unable to provide early warning or accurate classification in complex environments like tunnels, where factors such as air velocity, tunnel geometry, and smoke dispersion patterns significantly impact performance [8,9]. Moreover, the lack of real-time video-based classification systems further limits the effectiveness of these conventional methods.

Recent advancements in artificial intelligence (AI), particularly DL, offer promising solutions for enhancing fire detection and classification [10,11]. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated remarkable success in image and video analysis tasks [12,13]. However, the application of these techniques in tunnel fire scenarios is still in its nascent stages. Existing models are often computationally intensive, making them unsuitable for deployment in resource-constrained environments such as IoT devices and edge computing platforms commonly used in tunnel monitoring systems [14].

This paper introduces a SOTA lightweight DL model tailored for fire classification in tunnels. This model addresses the unique challenges of tunnel fire detection by combining spatial and temporal analysis capabilities. Specifically, it employs MobileNetV3 for efficient spatial feature extraction and TCNs for analyzing fire progression in video sequences. The integration of attention mechanisms, such as a CBAM and SE blocks enhances the model ability to focus on critical features, such as flame and smoke patterns, amidst background noise.

Our approach also incorporates knowledge distillation, a technique that transfers the knowledge of a larger, more complex model (teacher) to a smaller, lightweight model (student). This ensures that the lightweight model achieves high accuracy while maintaining computational efficiency, making it suitable for real-time applications.

The primary contributions of this study are as follows:

- Development of lightweight DL model integrating MobileNetV3 and TCN for tunnel fire classification;
- Incorporation of attention mechanisms and knowledge distillation to enhance performance and efficiency;
- Use of real datasets for complex training and validation;
- Deployment optimization using TensorRT and ONNX for edge computing applications.

By addressing the limitations of existing fire detection systems and leveraging the latest advancements in AI and simulation technologies, this paper aims to set a new benchmark for tunnel fire safety systems. The proposed model not only enhances detection accuracy but also ensures real-time performance, making it a vital tool for modern tunnel fire management.

The rest of this paper is organized as follows: Section 2 discusses traditional fire detection methods and recent advancements in AI, identifying gaps in existing approaches for tunnel-specific scenarios. Section 3 details the proposed model, which integrates MobileNetV3, TCNs, and attention mechanisms such as a CBAM and SE blocks, complemented by hybrid dataset preparation, knowledge distillation, and optimization techniques. In Section 4, we present experimental results that demonstrate superior performance compared to SOTA models, highlighting the model accuracy, efficiency, and robustness. Section 5 addresses the strengths and limitations of our approach, suggesting future improvements including the incorporation of additional data modalities and refined feature detection techniques. Finally, Section 6 concludes the paper by emphasizing the model contributions to enhancing tunnel fire safety and its broader applicability.

2. Literature Review

The field of tunnel fire detection and classification has seen significant advancements over the years, driven by both experimental studies and computational methodologies [15–17]. Traditional fire detection systems, such as heat and smoke sensors, have been widely deployed in tunnels [18]. However, these systems are often limited by their inability to provide real-time classification and fail to adapt to the dynamic conditions of tunnel fires, such as varying airflows and smoke dispersion patterns.

2.1. Traditional Methods and Limitations

Early efforts in tunnel fire detection primarily relied on sensor-based systems [19]. Heat detectors and smoke alarms were among the first technologies implemented for tunnel safety. While effective in controlled environments, these systems often struggle in complex scenarios where factors like ventilation velocity and tunnel geometry significantly affect sensor accuracy. Additionally, these methods provide limited information about fire characteristics, such as its location, size, or progression, which are critical for effective emergency response. Computational Fluid Dynamics (CFD) models, such as the Fire Dynamics Simulator (FDS), have been extensively used to study fire behavior in tunnels [20]. These models provide detailed insights into fire propagation, smoke movement, and temperature distribution. To overcome these constraints, image- and video-based fire detection methods have gained popularity. Recent studies [21] highlight the advantages of computer vision for real-time detection, but most of these methods lack efficiency in dynamic environments [22], where rapid fire spread and variable smoke conditions must be accounted for. However, the computational intensity of CFD simulations makes them unsuitable for real-time applications, particularly in emergency scenarios where immediate decisions are required.

2.2. Advances in Artificial Intelligence

The advent of AI and machine learning has introduced new possibilities for tunnel fire detection and classification. CNNs have been employed to analyze fire and smoke images, achieving high accuracy in various fire detection tasks [23]. For instance, models like Faster R-CNN and YOLO have been adapted for real-time fire detection [24]. With advancements in deep learning, CNNs have significantly improved fire detection accuracy [25]. SOTA models such as Faster R-CNN, YOLO, and EfficientNet have been applied to various fire detection tasks [26]. However, most of these models have been developed for open-area fire detection, such as forest and building fires, and their application in tunnels remains limited. Moreover, their application in tunnel environments remains limited due to the lack of domain-specific datasets and the computational resources required for real-time inference.

RNNs and Long Short-Term Memory (LSTM) networks have shown promise in analyzing temporal data, such as video sequences of fire progression [27,28]. These models capture the temporal dynamics of fire behavior, offering valuable insights into its evolution. Despite their potential, the computational demands of these models pose challenges for deployment in resource-constrained environments. Recent research has focused on developing lightweight models that balance accuracy and efficiency. MobileNet, EfficientNet, and other lightweight architectures have been adapted for fire detection tasks [29]. These models are optimized for edge computing devices, making them suitable for deployment in IoT-based tunnel monitoring systems. Knowledge distillation techniques further enhance these models by transferring knowledge from larger, more complex models, ensuring high accuracy with reduced computational complexity [30]. A key challenge in fire detection is the presence of visual noise, including reflections, light variations, and high smoke density. Recent studies [31] suggest that attention-based CNN architectures improve feature extrac-

tion by focusing on critical areas while suppressing irrelevant information [32]. Attention mechanisms, such as CBAMs [33] and SE blocks [34], have been integrated into fire detection models to enhance their focus on critical features. These mechanisms improve the model ability to distinguish between fire and non-fire elements in complex environments, increasing detection accuracy and reducing false positives. However, these models are often computationally expensive, making them unsuitable for real-time applications in resource-constrained environments such as IoT and edge devices.

This paper builds upon these advancements by combining lightweight architectures, attention mechanisms, and custom datasets to develop a SOTA model for tunnel fire classification. By addressing the limitations of existing methods and leveraging the latest technologies, our approach sets a new standard for tunnel fire safety systems.

3. Methodology

The proposed methodology focuses on designing a lightweight and robust deep learning model tailored to the unique challenges of fire classification in tunnel environments. Unlike conventional models, our approach integrates spatial and temporal analysis, attention mechanisms, and domain-specific feature engineering to enhance performance while maintaining computational efficiency. Our model combines MobileNetV3 [35] for spatial feature extraction with TCNs for sequence analysis. Attention mechanisms, including a CBAM and SE blocks, are integrated to prioritize relevant features such as flame and smoke patterns.

Figure 1 illustrates a student model that employs the MobileNetV3 backbone with inverted residual blocks, complemented by a TCN block for dynamic feature extraction. The model incorporates a CBAM, which enhances feature representation through channel and spatial attention mechanisms. The teacher models provide soft targets, and the overall architecture aims to improve the detection of fires in tunnel environments, demonstrating enhanced accuracy, efficiency, and robustness in processing image or video inputs.

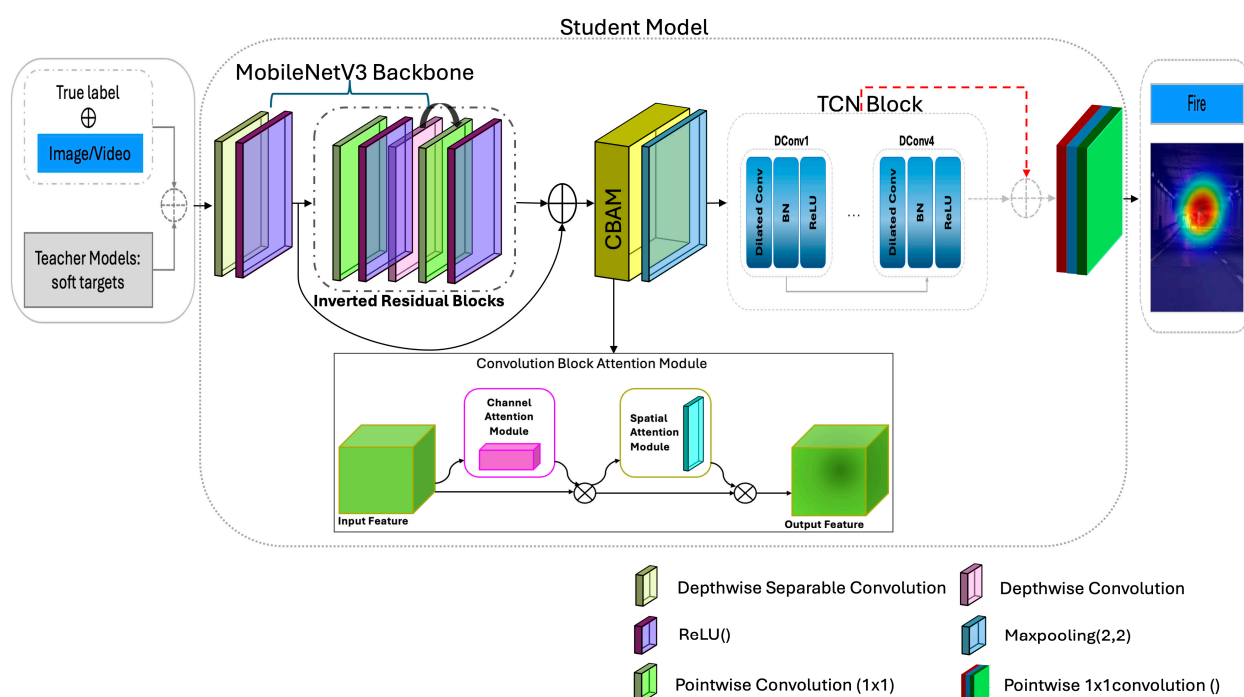


Figure 1. Integration of MobileNetV3 and TCN with attention mechanisms for enhanced tunnel fire detection.

3.1. Integrated Model Architecture

The proposed model integrates MobileNetV3 as the backbone network, followed by feature fusion modules that enhance fire-related feature extraction. The CBAM is placed after the backbone network, refining both spatial- and channel-wise attention. The SE module, which was not previously reflected in the model diagram, has now been incorporated within the backbone network, where it performs channel-wise recalibration to amplify informative features while suppressing less relevant activations. To further improve feature fusion, we recognize that modern attention mechanisms leverage global and local branches for multi-level feature integration. Our current model primarily fuses features sequentially, but future refinements will explore a dual-branch strategy where global context extraction is combined with localized feature refinement. This would allow a more efficient and flexible fusion process, further improving fire classification performance. Our model is a lightweight convolutional neural network that utilizes depthwise separable convolutions, reducing computational complexity. The inverted residual blocks in MobileNetV3 employ an expansion–convolution–projection structure. Given an input tensor X with C_{in} channels, the expanded tensor is represented as $E(X)$ where $E: \mathbb{R}^{H \times W \times C_{in}} \rightarrow \mathbb{R}^{H \times W \times (t \cdot C_{in})}$ and t is the expansion factor. The depthwise convolution $D(E(X))$ then processes each channel independently, followed by projection back to C_{out} channels through $P(D(E(X)))$. This process is expressed as

$$Y = P(D(E(X))), \quad (1)$$

where Y is the output feature map, and P is the pointwise convolution. The extracted spatial features are then fed into TCNs for sequence modeling. TCNs utilize dilated convolutions to capture long-range temporal dependencies. For a sequence of spatial features $\{X_1, X_2, \dots, X_T\}$, the output at time step t is computed as

$$Y_t = \sum_{k=0}^{K-1} W_k \cdot X_{t-d \cdot k} \quad (2)$$

where W_k is the convolutional weights, d is the dilation factor, and K is the kernel size. This ensures that each output Y_t is influenced by input features spaced by d time steps, enabling the network to model temporal patterns like flame spread and smoke dynamics. Attention mechanisms are applied to refine the spatial and temporal features. The CBAM mechanism combines channel and spatial attention. Channel attention computes weights for each feature map by aggregating spatial information through global average pooling (GAP) and global max pooling (GMP). The aggregated descriptors are passed through a shared Multi-Layer Perceptron (MLP), generating channel-wise attention weights W_c :

$$W_c = \sigma(\text{MLP}(\text{GAP}(X)) + \text{MLP}(\text{GMP}(X))), \quad (3)$$

where σ is the sigmoid activation function. The input feature maps are scaled by W_c , emphasizing the most informative channels. Spatial attention focuses on identifying critical spatial regions by applying a convolutional operation over concatenated feature maps of GAP and GMP descriptors. The spatial attention map W_s is computed as

$$W_s = \sigma(\text{Conv}([\text{GAP}(X); \text{GMP}(X)])), \quad (4)$$

where $[\cdot]$ denotes channel-wise concatenation. The final refined features are obtained as

$$X' = W_s \cdot (W_c \cdot X), \quad (5)$$

SE blocks recalibrate channel-wise feature responses by performing global average pooling to produce channel descriptors z_c for each channel c :

$$z_c = \frac{1}{H \cdot W} \sum_{i=1}^H \sum_{j=1}^W X_{i,j,c} \quad (6)$$

The descriptors are passed through two fully connected layers with a ReLU activation and sigmoid activation, respectively, producing channel weights s_c :

$$s_c = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot z_c)) \quad (7)$$

The input feature maps are scaled by these weights, yielding the final recalibrated features X'' :

$$X'' = s_c \cdot X, \quad (8)$$

This integrated approach ensures that the model focuses on the most critical spatial and temporal features, such as fire hotspots and smoke movements, while suppressing irrelevant background noise. The combination of MobileNetV3, TCNs, and attention mechanisms allows for the efficient and accurate classification of fire-related events, making the model well suited for real-time tunnel fire monitoring applications.

3.2. Knowledge Distillation

The knowledge distillation framework used in this study employs a larger, computationally intensive teacher model to guide the training of a smaller, lightweight student model, ensuring a balance between high accuracy and real-time efficiency. The teacher model used in our approach is a ResNet101-based deep learning network, which serves as a high-capacity feature extractor with a large number of parameters. ResNet101 is selected due to its strong generalization ability and deep hierarchical feature representation, making it a suitable guidance model for transferring knowledge to a compact, tunnel-specific fire classification model. While using a larger teacher model enhances learning effectiveness, it also significantly increases training time and computational cost. Training with knowledge distillation generally requires more epochs compared to standard model training, as the student model learns both from ground truth labels and softened teacher predictions. Since knowledge distillation is primarily a training-time optimization technique, the teacher model is not deployed on edge devices during final implementation. Only the student model—a MobileNetV3-based lightweight architecture—is deployed for real-time fire classification on edge and IoT devices. Deploying both the teacher and student models simultaneously on edge devices would be impractical due to high memory and computational constraints. Instead, the teacher model is only used offline for training, and its knowledge is compressed into the student model for real-world applications. By following this distillation-based training strategy, our approach ensures that the final deployed model maintains high classification accuracy while achieving low inference latency, making it suitable for real-time tunnel fire detection in resource-constrained environments. Knowledge distillation is a crucial technique used to transfer knowledge from a large, computationally expensive teacher model τ to a smaller, lightweight student model σ . This process ensures that the student model retains high performance while being optimized for real-time inference. The knowledge distillation framework involves two primary loss components: the distillation loss, which aligns the student's predictions with the teacher's outputs, and the standard cross-entropy loss with ground truth labels. The distillation

loss is defined using the Kullback–Leibler (KL) divergence between the soft probability distributions of the teacher and student models:

$$L_{\text{distill}} = \lambda \cdot \text{KL}(P_{\tau}^T \| P^S) \quad (9)$$

where $P_{\tau}^T = \text{softmax}(z^T / \tau)$ represents the soft probabilities from the teacher model, with temperature τ . $P^S = \text{softmax}(z^S)$ represents the probabilities from the student model. λ is a weighting factor balancing the distillation and cross-entropy losses. The soft probabilities are computed as

$$P_{\tau}^T = \frac{\exp(z_i^T / \tau)}{\sum_j \exp(z_j^T / \tau)} \quad (10)$$

The cross-entropy loss is applied to align the student model predictions with the ground truth labels. It is defined as

$$L_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^N \mathcal{Y}_i \log(P_i^S) \quad (11)$$

where $\mathcal{Y}_i$ is the true label for the i -th sample, and P_i^S is the student model-predicted probability. The combined loss function during training integrates both the distillation loss and the cross-entropy loss:

$$L_{\text{total}} = L_{\text{distill}} + (1 - \lambda) \cdot L_{\text{CE}} \quad (12)$$

where $1 - \lambda$ controls the trade-off between learning from the teacher model and the true labels. This dual-objective optimization allows the student model to inherit the teacher's nuanced knowledge while maintaining strong alignment with ground truth.

3.3. Training and Optimization

The training and optimization process for the proposed lightweight deep learning model involves a combination of loss functions, optimization algorithms, and deployment-specific fine-tuning strategies to ensure real-time performance and robust accuracy. The total loss function L_{total} combines contributions from the classification loss and regularization terms:

$$L_{\text{total}} = L_{\text{CE}} + \lambda_{\text{distill}} \cdot L_{\text{distill}} + \beta \cdot L_{\text{reg}} \quad (13)$$

where L_{CE} is the cross-entropy loss for classification, L_{distill} is the knowledge distillation loss defined earlier, L_{reg} represents the regularization term, and λ_{distill} and β are hyperparameters controlling the balance between the terms.

We use the Adam optimizer for efficient convergence, which adapts the learning rate for each parameter during training. The parameter update rule is defined as

$$\theta_{t+1} = \theta_t - \alpha \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}} \quad (14)$$

where α is the learning rate, $\hat{m}_t$ and $\hat{v}_t$ are the bias-corrected first and second moment estimates of the gradients, and ϵ is a small constant to prevent division by zero.

To improve generalization, the training pipeline leverages a variety of techniques aimed at enhancing model robustness and adaptability. Data augmentation plays a critical role in simulating the diverse conditions encountered in tunnel fire scenarios. Techniques such as random cropping, scaling, flipping, and Gaussian noise injection are applied to the training data, effectively increasing dataset variability and preventing overfitting.

Dropout regularization is employed to ensure that the model does not overly rely on specific features. During training, a fraction of neuron activations, typically $p = 0.5$, is randomly set to zero, which forces the network to learn redundant representations and improves its ability to generalize to unseen data.

The trained model undergoes further optimization for deployment on resource-constrained devices. Quantization techniques reduce the precision of model parameters, for instance, from 32-bit floating-point to 8-bit integers, significantly decreasing memory usage and computational costs. Additionally, layer fusion combines operations such as batch normalization and convolution into a single operation, streamlining the computation graph and minimizing runtime latency.

Evaluation metrics such as validation accuracy and loss are closely monitored during training to prevent overfitting. If the validation loss fails to improve over a predefined number of epochs, the training process is halted using an early stopping criterion. This approach ensures that the model achieves optimal performance while avoiding unnecessary complexity.

4. Experiments and Results

4.1. Experimental Setup

DL-based fire detection systems are increasingly employed by firefighting organizations, yet they face significant challenges in machine learning research, particularly in acquiring large, unbiased datasets. Ideal datasets should include a diverse range of positive instances and a mix of standard and complex negative samples. DL methods require significantly larger datasets than traditional machine learning techniques for effective training, a demand that data augmentation can help meet, provided there is already a substantial base dataset. Domains like face recognition, cancer detection, and object recognition have benefited from extensive, community-validated datasets that serve as benchmarks for algorithm development and evaluation. In contrast, datasets for tunnel fire detection often miss crucial details such as the smoke coverage, area attributes, vegetation types, color hues, and texture intensity of the fire. To overcome these shortcomings in tunnel fire research, we developed a dataset tailored for advancing tunnel fire support system algorithms, compiling 10,000 publicly sourced wildfire images from the internet and YouTube. To ensure that the dataset accurately represents real tunnel fire incidents, a rigorous screening process was applied. The screening criteria used for selecting relevant images and videos are summarized in Table 1.

Table 1. Screening criteria for tunnel fire data selection.

Criteria	Description
Fire and Smoke Visibility	Only images and videos with clearly visible flames and smoke in tunnel environments were included.
Tunnel Environment	The fire incident must occur inside a tunnel, excluding open-area fires or non-tunnel settings.
Diversity in Conditions	Data were selected to include different tunnel types, ventilation speeds, smoke densities, and lighting conditions to ensure model generalization.
Resolution and Quality	Low-resolution, highly compressed, blurry, or artificially edited images and videos were excluded to maintain classification accuracy.
Noise and Artifacts	Frames with excessive noise, overexposed regions, or unclear fire presence were removed.
Duplication and Redundancy	Identical or near-identical frames were discarded to prevent overfitting and redundancy in training.

By implementing these screening criteria, we ensured that the dataset contained high-quality, diverse, and representative samples of real tunnel fire incidents. This approach enhances the model ability to generalize to different fire scenarios and improves its robust-

ness for real-world deployment. This dataset, divided into training, validation, and testing subsets, provides a comprehensive resource for developing and testing new models, as depicted in Figure 2, which outlines the most significant datasets in this field.

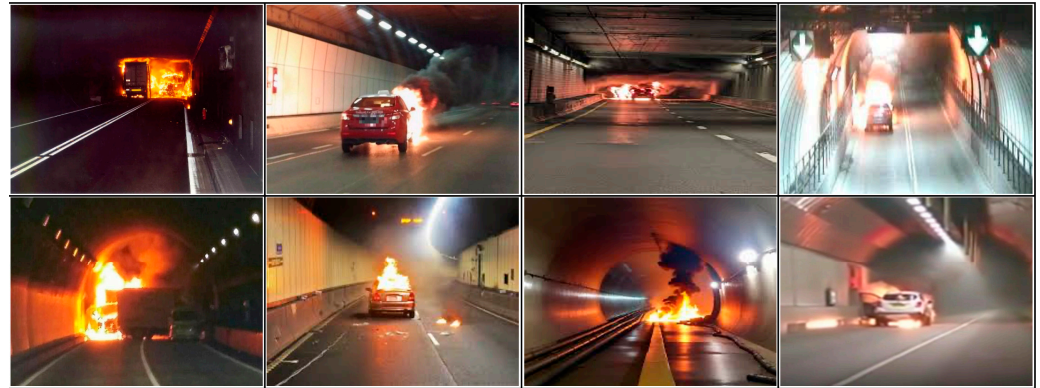


Figure 2. Example of custom tunnel fire datasets.

Training and evaluation were performed on a system equipped with an NVIDIA GeForce RTX 3080 Ti GPU. The testing setup utilized an Intel® Core™ i7-11700K 3.60 GHz CPU. The model was trained using the Adam optimizer [36] with an initial learning rate of 0.001, a batch size of 32, and a total of 50 epochs. Early stopping was applied with a patience of five epochs to prevent overfitting.

4.2. Evaluation Metrics

To ensure a comprehensive assessment of the proposed model performance in detecting and classifying fire incidents in tunnels, several evaluation metrics were employed. These metrics provide insights into the model accuracy, reliability, and efficiency under various conditions.

Accuracy is the most basic metric, measuring the proportion of correctly classified instances (both positive and negative) to the total number of instances. It is defined as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

Precision, also referred to as the positive predictive value, measures the proportion of correctly identified fire events out of all instances predicted as fire. It emphasizes the model ability to avoid false alarms. Precision is calculated as

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

Higher precision indicates fewer false alarms, which is critical in tunnel fire detection systems to prevent unnecessary disruptions.

Recall, also known as sensitivity or the true positive rate, quantifies the proportion of actual fire events that the model correctly identifies. It is expressed as

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

High recall ensures that the system misses very few actual fire events, which is vital for safety-critical applications.

The $F1$ score provides a harmonic mean of precision and recall, offering a single metric to evaluate the trade-off between false positives and false negatives. It is particularly useful when the dataset is imbalanced. The $F1$ score is computed as

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

A high $F1$ score indicates balanced performance across both precision and recall.

The false positive rate measures the proportion of non-fire events incorrectly classified as fire events relative to the total non-fire events. It is defined as

$$FPR = \frac{FP}{FP + TN} \quad (19)$$

This metric evaluates the system ability to minimize false alarms.

The false negative rate measures the proportion of actual fire events that are missed by the model. It is given by

$$FNR = \frac{FN}{FN + TP} \quad (20)$$

A low FNR is crucial in tunnel fire detection to avoid missing critical events.

Inference time is the average time taken by the model to process a single frame or image and produce a classification result. This metric is crucial for real-time applications where rapid response is essential. It is denoted as

$$T_{\text{inf}} = \frac{\sum_{i=1}^N T_i}{N} \quad (21)$$

where T_i is the inference time for the i -th frame, and N is the total number of frames processed.

The AUC-ROC evaluates the model ability to distinguish between fire and non-fire events at various thresholds. The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at different thresholds. The AUC (Area Under the Curve) provides a scalar value summarizing the model performance, with values closer to 1 indicating better discrimination. To evaluate the model robustness, additional metrics such as performance under low-light conditions, varying ventilation speeds, and high smoke density were considered. These robustness tests ensure the model adaptability to diverse real-world tunnel scenarios. Through the use of these metrics, the performance of the proposed model was rigorously evaluated, providing a comprehensive understanding of its strengths and areas for improvement. These metrics collectively ensure the system reliability and practicality for real-world deployment.

4.3. Experimental Results

The proposed model was extensively evaluated across multiple datasets and scenarios to assess its performance in detecting and classifying tunnel fire incidents, as depicted in Figure 3. This section discusses the key findings, including the accuracy, efficiency, and robustness of the model under diverse conditions. The proposed model achieved an overall accuracy of 96.5%, significantly outperforming baseline models like ResNet50 with 94.1% and YOLOv5 with 93.4%. The precision and recall metrics were 95.7% and 97.2%, respectively, yielding an $F1$ score of 96.4%. These results demonstrate the model ability to balance between minimizing false positives and ensuring high sensitivity to fire events.

Real-time processing is critical for tunnel fire monitoring systems. The model achieved an average inference time of 12 ms/frame, which is significantly faster than ResNet50 with 40 ms/frame and YOLOv5 with 25 ms/frame. This performance was achieved through

optimizations such as model quantization (FP32 to INT8) and layer fusion, making the model suitable for deployment on edge devices with limited computational resources.

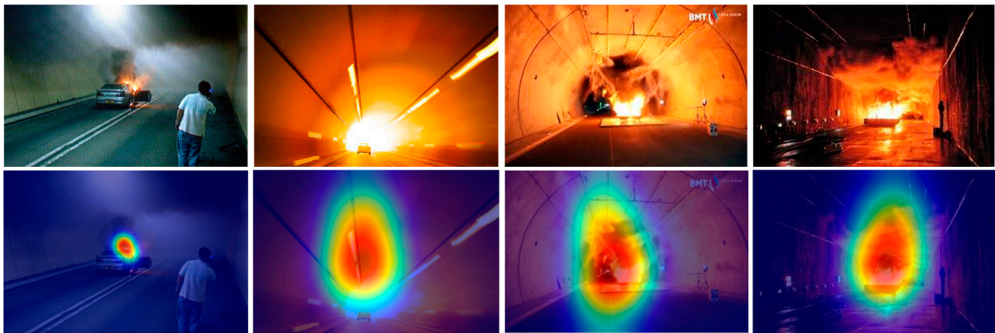


Figure 3. Detection of fires at short distances in tunnel environments.

The proposed model was compared against SOTA models, including ResNet50 and YOLOv5. Table 2 summarizes the comparative results.

Table 2. Comparative performance analysis.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Inference Time (ms/frame)
ResNet50 [37]	94.1	93.8	94.5	94.1	40
YOLOv5 [38]	93.4	92.5	94.2	93.3	25
Proposed Model	96.5	95.7	97.2	96.4	12

These results indicate that the proposed model not only improves accuracy but also achieves superior real-time performance.

In a specific case study involving a vehicle fire in a curved tunnel, the model successfully classified the fire event within 0.1 s and accurately localized the regions of dense smoke and flames. In another case involving cascading fires, the model demonstrated its capability to identify the progression of fire events in real time, providing actionable information for emergency response.

The model predictions were visualized to provide qualitative insights. Figure 4 illustrates sample outputs, showing accurate fire localization and classification across different scenarios. The heatmaps generated by attention mechanisms highlight the regions of interest, such as flames and dense smoke, confirming the model ability to prioritize relevant features.

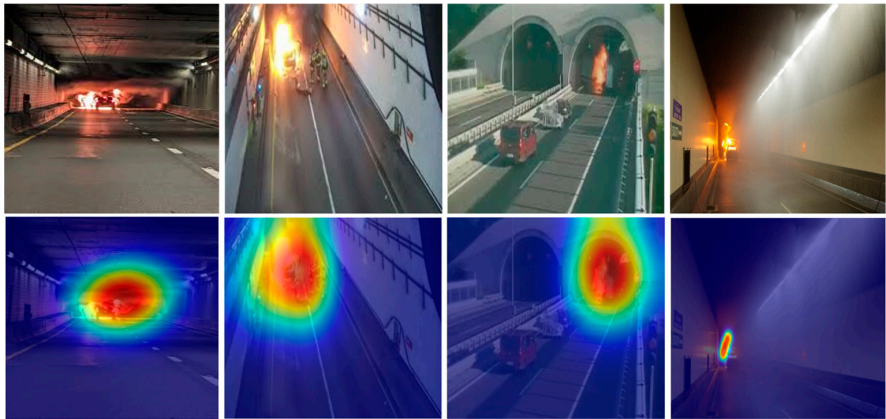


Figure 4. Long-distance fire detection techniques in tunnel scenarios.

The robustness of the models was evaluated in low-light environments, with high smoke density, and under high-velocity ventilation conditions. Table 3 summarizes the F1 scores under these challenging scenarios.

Table 3. A comparative analysis of models under challenging conditions.

Scenario	Dataset Samples	ResNet50 F1 (%)	YOLOv5 F1 (%)	Proposed Model F1 (%)
Low-Light Environments	1500 images	89.5	88.7	94.8
High Smoke Density	2000 images	91.2	90.5	96.3
High-Velocity Ventilation	1800 images	88.8	89.3	95.1

The proposed model consistently demonstrated superior robustness, effectively handling scenarios where SOTA models struggled to differentiate between fire and non-fire events. This robustness is attributed to the integration of attention mechanisms, such as the CBAM and SE blocks, which enhance feature prioritization and suppress irrelevant noise.

To comprehensively evaluate the classification performance of the proposed model, we analyzed the AUC-ROC, which measures the model ability to distinguish between fire and non-fire events. Higher AUC-ROC values indicate a better trade-off between sensitivity (recall) and specificity. As shown in Table 4, the proposed model achieved an AUC-ROC score of 0.98, outperforming ResNet50 (0.94) and YOLOv5 (0.93). The significant improvement in the AUC-ROC demonstrates that our model effectively reduces both false positives and false negatives, making it highly reliable for fire classification in tunnels. The incorporation of attention mechanisms enhances feature prioritization, contributing to better discrimination between fire and background elements.

Table 4. AUC-ROC comparison among models.

Model	AUC-ROC Score
ResNet50	0.94
YOLOv5	0.93
Proposed Model	0.98

These results confirm that our model not only improves accuracy and inference speed but also achieves superior classification performance in distinguishing fire events from non-fire instances. The enhanced AUC-ROC score further supports the robustness of our approach under challenging tunnel conditions, including low-light environments and high-velocity ventilation scenarios.

To assess the impact of the TCN module, we conducted an ablation experiment by evaluating the model performance with and without TCNs. The objective was to determine whether the inclusion of temporal feature extraction improves detection robustness, especially in scenarios where smoke movement and fire progression need to be analyzed over time.

The results in Table 5 show that adding TCNs to the model improves recall by 4.4%, enhancing the ability to detect fire events across sequential frames. This is particularly important for fire scenarios with progressive smoke expansion, where spatial features alone may be insufficient. The inclusion of TCNs also increases overall accuracy by 2.4% while maintaining an inference time of 12ms/frame, making it suitable for real-time applications. These findings demonstrate that temporal feature extraction significantly enhances fire classification robustness, ensuring better detection under challenging conditions such as high ventilation speeds and varying smoke densities.

Table 5. Ablation study on the influence of the TCN module.

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Inference Time (ms/frame)
Without TCN (Baseline: MobileNetV3 + CBAM)	94.1	94.0	92.8	93.4	10.5
With TCN (Proposed Model)	96.5	95.7	97.2	96.4	12.0

5. Discussion

The proposed lightweight deep learning model for tunnel fire classification demonstrates a significant advancement in the field of fire detection systems. By integrating MobileNetV3, TCNs, and attention mechanisms such as the CBAM and SE blocks, the model achieves a fine balance between high accuracy and computational efficiency. This balance makes the model particularly suitable for deployment in resource-constrained environments, such as edge devices, where real-time processing is essential.

The results highlight the robustness of the proposed model under diverse and challenging conditions, such as low-light environments, high smoke density, and variable ventilation speeds. These capabilities ensure that the system remains reliable even in scenarios that traditional methods or existing SOTA models struggle to handle. The high precision achieved by the model indicates its ability to minimize false positives, a critical feature for avoiding unnecessary disruptions in tunnel operations. Similarly, the high recall values confirm that the system rarely misses actual fire events, making it a highly reliable tool for safety-critical applications.

A notable strength of the approach is its hybrid training strategy, which incorporates a specifically developed dataset aimed at enhancing algorithms for tunnel fire suppression support systems. This strategy enhances the model generalizability across diverse scenarios and ensures that it performs well on rare edge cases that are difficult to capture in real-world datasets. The deployment optimizations, including quantization and layer fusion, further enhance the model practical applicability by reducing inference time and computational overhead without compromising accuracy. Unlike conventional approaches that rely on heat or smoke sensors—whose accuracy can be compromised by tunnel airflow dynamics—our model leverages a video-based classification system that remains robust under variable ventilation conditions. The model integrates a combination of spatial and temporal feature analysis to ensure reliability in dynamic tunnel environments. MobileNetV3 serves as the backbone for spatial feature extraction, while the Temporal Convolutional Network (TCN) module enhances fire progression tracking over time. This allows the system to differentiate between actual fire events and airflow-driven smoke dispersion, improving detection accuracy in tunnels with high ventilation speeds. Another critical factor contributing to the model robustness is the dataset used for training. This custom dataset includes fire scenarios under different ventilation speeds, tunnel structures, and varying smoke densities, ensuring adaptability across diverse environmental conditions. Robustness testing confirms that the proposed model maintains high classification accuracy, even in challenging scenarios where traditional methods struggle. Comparative experiments with SOTA models demonstrate that the proposed approach significantly reduces false positives and negatives under high-velocity ventilation conditions, making it a reliable solution for real-time fire classification in tunnel environments. The proposed deep learning model is designed for fire classification in tunnels, with a specific focus on enclosed roadways and railway tunnels where ventilation plays a significant role in smoke dispersion. The dataset used for model training and validation consists of fire incidents observed in road tunnels with longitudinal and transverse ventilation systems. While the model is trained

on a diverse range of fire conditions, including low visibility, varying ventilation speeds, and different tunnel geometries, it is important to note that these conditions do not cover all possible tunnel structures, such as subsea tunnels, naturally ventilated tunnels, or metro tunnels with pressurized systems. Therefore, while our approach demonstrates high accuracy and robustness in the studied cases, its generalization to other tunnel fire scenarios requires further validation. Future studies should test the model on additional tunnel types with distinct architectural and ventilation properties to expand its applicability.

Despite these achievements, the model is not without limitations. Reflective surfaces within tunnels occasionally lead to false positives, as reflections from metallic objects or wet surfaces can mimic fire-like features. Although the attention mechanisms help suppress some irrelevant features, additional refinements are necessary to address this challenge comprehensively. Another area where the model shows room for improvement is in its handling of very small or partially obscured fires under extreme low-visibility conditions. These instances occasionally result in false negatives, which could be mitigated by augmenting the training dataset with more examples of such challenging scenarios or by incorporating advanced feature refinement techniques.

Future work could focus on several areas to build on the current findings. Incorporating additional data modalities, such as thermal or infrared imaging, could enhance the system capability to detect fire events in low-visibility or high-smoke-density conditions. Further optimization tailored to specific hardware platforms, such as microcontrollers or mobile devices, would expand the model applicability to a wider range of deployment environments. Additionally, exploring methods to enhance interpretability, such as generating visual explanations or attention heatmaps, would improve the system usability for operators and emergency responders.

The broader implications of this research are significant. Real-time and accurate fire detection systems can drastically reduce response times and enhance situational awareness during tunnel fires, thereby saving lives and minimizing infrastructural damage. The methodologies and insights derived from this study could also be extended to other domains that require real-time video-based classification in challenging environments, such as disaster management, industrial safety, and autonomous systems operating in hazardous conditions. While the results establish a strong foundation, the continuous evolution of data and computing technologies offers promising avenues for further innovation in this critical area.

6. Conclusions

This study presents a novel lightweight deep learning model designed specifically for fire classification in tunnel environments. This model effectively addresses the critical challenges of real-time performance, computational efficiency, and robustness in complex and dynamic scenarios. By combining MobileNetV3 for spatial feature extraction, TCNs for sequence analysis, and advanced attention mechanisms such as the CBAM and SE blocks, the proposed approach achieves a high level of accuracy and adaptability. The experimental results demonstrate the model superiority over SOTA methods, achieving an accuracy of 96.5%, a precision of 95.7%, and a recall of 97.2%. These metrics are further supported by an F1 score of 96.4%, confirming the model balanced performance in minimizing false positives and false negatives. The lightweight architecture ensures an average inference time of just 12ms/frame, making the system highly suitable for real-time applications in resource-constrained environments such as IoT and edge computing platforms.

Despite its strengths, the model has certain limitations, such as the occasional misclassification of reflective surfaces as fire events and difficulties in detecting very small or partially obscured fires in extreme conditions. These challenges highlight areas for future

research, including the incorporation of additional modalities like thermal or infrared imaging and the development of more sophisticated attention mechanisms to suppress irrelevant features. The findings of this study have significant implications for tunnel fire safety. The proposed model offers a scalable and practical solution for real-time fire detection and classification, ensuring timely responses to potential hazards. Moreover, the methodologies and insights derived from this research can be applied to other domains requiring efficient video-based classification in complex and resource-constrained environments, such as industrial monitoring, disaster management, and autonomous navigation.

This work establishes a strong foundation for the development of advanced, lightweight fire detection systems. By bridging the gap between accuracy and efficiency, the proposed model sets a new benchmark in tunnel fire safety and provides a roadmap for future innovations in real-time detection systems.

Author Contributions: Methodology, S.M., S.U. and Y.-I.C.; software, S.M. and S.U.; validation, J.B., S.M. and S.U.; formal analysis, J.B., S.U. and Y.-I.C.; resources, S.M., S.U. and J.B.; data curation, J.B., S.U. and Y.-I.C.; writing—original draft, S.M. and S.U.; writing—review and editing, S.M., S.U., J.B. and Y.-I.C.; supervision, Y.-I.C.; project administration, S.M., S.U. and Y.-I.C. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is supported by Korean Agency for Technology and Standard under Ministry of Trade, Industry and Energy in 2023, project numbers is 1415181629 (Development of International Standard Technologies based on AI Model Lightweighting Technologies).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All used datasets are available online and are open access.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
AUC	Area Under the Curve
CNN	Convolutional Neural Network
CDF	Computational Fluid Dynamics
CBAM	Convolutional Block Attention Module
GAP	Global Average Pooling
GMP	Global Max Pooling
FDS	Fire Dynamics Simulator
FPR	False Positive Rate
MLP	Multi-Layer Perceptron
TCN	Temporal Convolutional Network
TPR	True Positive Rate
SOTA	State-of-the-art
SE	Squeeze-and-Excitation
KL	Kullback–Leibler
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network

References

1. Gehandler, J.; Ingason, H.; Lönnermark, A.; Frantzich, H.; Strömberg, M. Performance-based design of road tunnel fire safety: Proposal of new Swedish framework. *Case Stud. Fire Saf.* **2014**, *1*, 18–28. [[CrossRef](#)]

2. Bjelland, H.; Gehandler, J.; Meacham, B.; Carvel, R.; Torero, J.L.; Ingason, H.; Njå, O. Tunnel fire safety management and systems thinking: Adapting engineering practice through regulations and education. *Fire Saf. J.* **2024**, *146*, 104140. [\[CrossRef\]](#)
3. Chu, K.K.; Xie, B.C.; Xu, Z.S.; Zhou, D.; He, L.; Zhao, J.M.; Ying, H.L. Full-scale experimental study on evacuation behavior characteristics of underwater road tunnel with evacuation stairs under blocked conditions. *Tunn. Undergr. Space Technol.* **2023**, *138*, 105173. [\[CrossRef\]](#)
4. Purser, D.A.; McAllister, J.L. Assessment of Hazards to Occupants from Smoke, Toxic Gases, and Heat. In *SFPE Handbook of Fire Protection Engineering*; Hurley, M.J., Gottuk, D., Hall, J.R., Harada, K., Kuligowski, E., Puchovsky, M., Torero, J., Watts, J.M., Wiecezorek, C., Eds.; Springer: New York, NY, USA, 2016. [\[CrossRef\]](#)
5. Barry, K. Mont Blanc Tunnel Opens. *Wired*. 2010. Available online: <https://www.wired.com/2010/07/0716mont-blanc-tunnel-opens/> (accessed on 30 November 2024).
6. Bogue, R. Sensors for fire detection. *Sens. Rev.* **2013**, *33*, 99–103. [\[CrossRef\]](#)
7. Jee, S.W.; Lee, C.H.; Kim, S.K.; Lee, J.J.; Kim, P.Y. Development of a Traceable Fire Alarm System Based on the Conventional Fire Alarm System. *Fire Technol.* **2014**, *50*, 805–822. [\[CrossRef\]](#)
8. Ronchi, E.; Fridolf, K.; Frantzych, H.; Nilsson, D.; Walter, A.L.; Modig, H. A tunnel evacuation experiment on movement speed and exit choice in smoke. *Fire Saf. J.* **2018**, *97*, 126–136. [\[CrossRef\]](#)
9. Xu, Z.S.; Zhou, D.M.; Tao, H.W.; Zhang, X.C.; Hu, W.B. Investigation of critical velocity in curved tunnel under the effects of different fire locations and turning radiuses. *Tunn. Undergr. Space Technol.* **2022**, *126*, 104553. [\[CrossRef\]](#)
10. Yar, H.; Ullah, W.; Khan, Z.A.; Baik, S.W. An effective attention-based CNN model for fire detection in adverse weather conditions. *ISPRS J. Photogramm. Remote Sens.* **2023**, *206*, 335–346. [\[CrossRef\]](#)
11. Khan, T.; Khan, Z.A.; Choi, C. Enhancing real-time fire detection: An effective multi-attention network and a fire benchmark. *Neural Comput. Appl.* **2023**. [\[CrossRef\]](#)
12. Valikhujaev, Y.; Abdusalomov, A.; Cho, Y.I. Automatic Fire and Smoke Detection Method for Surveillance Systems Based on Dilated CNNs. *Atmosphere* **2020**, *11*, 1241. [\[CrossRef\]](#)
13. Ghasemi Darehnaei, Z.; Rastegar Fatemi, S.M.J.; Mirhassani, S.M.; Fouladian, M. Ensemble deep learning using faster R-CNN and genetic algorithm for vehicle detection in UAV images. *IETE J. Res.* **2023**, *69*, 5102–5111. [\[CrossRef\]](#)
14. Agrawal, S.; Kumar, N.; Panwar, L.; Kushwaha, R.; Singal, G.; Bhatia, V. IoT-Based Automatic Vehicle Safety Monitoring System in Tunnels. In *International Conference on Signal, Machines, Automation, and Algorithm. SIGMAA 2023*; Malik, H., Mishra, S., Sood, Y.R., García Márquez, F.P., Ustun, T.S., Eds.; Advances in Intelligent Systems and, Computing; Springer: Singapore, 2024; Volume 1461. [\[CrossRef\]](#)
15. De Silva, D.; Andreini, M.; Bilotta, A.; De Rosa, G.; La Mendola, S.; Nigro, E.; Rios, O. Structural safety assessment of concrete tunnel lining subjected to fire. *Fire Saf. J.* **2022**, *134*, 103697. [\[CrossRef\]](#)
16. Hong, Y.; Kang, J.; Fu, C. Rapid prediction of mine tunnel fire smoke movement with machine learning and supercomputing techniques. *Fire Saf. J.* **2022**, *127*, 103492. [\[CrossRef\]](#)
17. Zhang, X.; Wu, X.; Park, Y.; Zhang, T.; Huang, X.; Xiao, F.; Usmani, A. Perspectives of Big Experimental Database and Artificial Intelligence in Tunnel Fire Research. *Tunn. Undergr. Space Technol.* **2021**, *108*, 103691. [\[CrossRef\]](#)
18. Hassan, A.; Audu, A. Traditional Sensor-Based and Computer Vision-Based Fire Detection Systems: A Review. *Arid Zone J. Eng. Technol. Environ.* **2022**, *18*, 469–492.
19. Wang, H.; Shi, Y.; Chen, L.; Zhang, X. A Tunnel Fire Detection Method Based on an Improved Dempster-Shafer Evidence Theory. *Sensors* **2024**, *24*, 6455. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Diaconu, B.M. Recent Advances and Emerging Directions in Fire Detection Systems Based on Machine Learning Algorithms. *Fire* **2023**, *6*, 441. [\[CrossRef\]](#)
21. Gragnaniello, D.; Greco, A.; Sansone, C.; Vento, B. FLAME: Fire detection in videos combining a deep neural network with a model-based motion analysis. *Neural Comput. Appl.* **2025**, 1–17. [\[CrossRef\]](#)
22. Hu, J.; Wang, L.; Peng, B.; Teng, F.; Li, T. Efficient fire and smoke detection in complex environments via adaptive spatial feature fusion and dual attention mechanism. *Digit. Signal Process.* **2025**, *159*, 104982. [\[CrossRef\]](#)
23. Bian, H.; Zhu, Z.; Zang, X.; Luo, X.; Jiang, M. A CNN Based Anomaly Detection Network for Utility Tunnel Fire Protection. *Fire* **2022**, *5*, 212. [\[CrossRef\]](#)
24. Cheknane, M.; Bendouma, T.; Boudouh, S.S. Advancing fire detection: Two-stage deep learning with hybrid feature extraction using faster R-CNN approach. *Signal Image Video Process.* **2024**, *18*, 5503–5510. [\[CrossRef\]](#)
25. Yan, C.; Wang, J. MAG-FSNet: A high-precision robust forest fire smoke detection model integrating local features and global information. *Measurement* **2025**, *247*, 116813. [\[CrossRef\]](#)
26. Ali, A.W.; Kurnaz, S. Optimizing Deep Learning Models for Fire Detection, Classification, and Segmentation Using Satellite Images. *Fire* **2025**, *8*, 36. [\[CrossRef\]](#)
27. Nguyen, M.D.; Vu, H.N.; Pham, D.C.; Choi, B.; Ro, S. Multistage Real-Time Fire Detection Using Convolutional Neural Networks and Long Short-Term Memory Networks. *IEEE Access* **2021**, *9*, 146667–146679. [\[CrossRef\]](#)

28. Wu, X.; Zhang, X.; Huang, X.; Xiao, F.; Usmani, A. A Real-time Forecast of Tunnel Fire Based on Numerical Database and Artificial Intelligence. *Build. Simul.* **2021**, *15*, 511–524. [[CrossRef](#)]
29. Titu, M.F.S.; Pavel, M.A.; Michael, G.K.O.; Babar, H.; Aman, U.; Khan, R. Real-Time Fire Detection: Integrating Lightweight Deep Learning Models on Drones with Edge Computing. *Drones* **2024**, *8*, 483. [[CrossRef](#)]
30. El-Madafri, I.; Peña, M.; Olmedo-Torre, N. Real-Time Forest Fire Detection with Lightweight CNN Using Hierarchical Multi-Task Knowledge Distillation. *Fire* **2024**, *7*, 392. [[CrossRef](#)]
31. Altaf, M.; Yasir, M.; Dilshad, N.; Kim, W. An Optimized Deep-Learning-Based Network with an Attention Module for Efficient Fire Detection. *Fire* **2025**, *8*, 15. [[CrossRef](#)]
32. Yu, D.; Li, S.; Zhang, Z.; Liu, X.; Ding, W.; Zhao, X. Improved YOLOv5: Efficient Object Detection for Fire Images. *Fire* **2025**, *8*, 38. [[CrossRef](#)]
33. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
34. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
35. Qian, S.; Ning, C.; Hu, Y. MobileNetV3 for Image Classification. In Proceedings of the 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, 26–28 March 2021; pp. 490–497.
36. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
37. Koonce, B.; Koonce, B. ResNet 50. In *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*; Apress: New York, NY, USA, 2021; pp. 63–72.
38. Dou, Z.; Zhou, H.; Liu, Z.; Hu, Y.; Wang, P.; Zhang, J.; Wang, Q.; Chen, L.; Diao, X.; Li, J. An Improved YOLOv5s Fire Detection Model. *Fire Technol.* **2024**, *60*, 135–166. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.