IEC
Institute of Electronics
and Computer

Journal of Artificial Intelligence and Systems, 2023, 5, 23-35
https://iecscience.org/journals/AIS
ISSN Online: 2642-2859

Vision Sensor Assisted Fire Detection in IoT Environment using ConvNext

Sana Zahir[a], Arbab Waseem Abbas[b,]*, Rafi Ullah Khan[c], Mohib Ullah[d]

Institute of Computer Sciences and Information Technology, The University of Agriculture Peshawar, Pakistan
Email: [a.] sanakhan131s@gmail.com, [b.]ristocratarbab@aup.edu.pk, [c.]rafiyz@gmail.com, [d.] mrmohibkhan@gmail.com
*Corresponding Author: Arbab Waseem Abbas, Email: ristocratarbab@aup.edu.pk

How to cite this paper: Sana Zahir, Arbab Waseem Abbas, Rafi Ullah Khan, Mohib Ullah (2023). Vision Sensor Assisted Fire Detection in IoT Environment using ConvNext. Journal of Artificial Intelligence and Systems, 5, 23–35. https://doi.org/10.33969/AIS.2023050102.

Received: January 8, 2023
Accepted: March 30, 2023
Published: March 31, 2023

Abstract

To mitigate social, ecological, and financial damage, effective fire detection and control are crucial. Performing real-time fire detection in Internet of Things (IoT) environments, however, presents significant challenges due to limited storage, transmission, and computational resources. Early fire detection and automated response are essential for addressing these challenges. In this paper, we introduce an IoT-supported deep learning model designed for efficient fire detection. The proposed model builds upon the pre-trained weights of the ConvNext convolutional neural network, which excels at detecting minute features and distinguishing between yellow lights and fire patterns. Implemented on an IoT device, the system triggers an alert when a fire is detected, prompting necessary actions. Our method, tested on the forest fire dataset, demonstrated a 4% improvement in accuracy compared to existing deep learning models for fire detection.

Keywords

Artificial Intelligence; Data augmentation; Convolutional Neural Networks; Deep Learning; Fire Detection; IoT

1. Introduction

The Internet of Things (IoT) has enhanced the interconnectivity of smart devices, and their increased processing capabilities coupled with intelligent technologies play a significant role in various applications, such as e-health, self-driving cars, event surveillance, and law enforcement [1]. Numerous abnormal events, including accidents, disasters, medical emergencies, floods, and fires, can occur during periods of disorder. Obtaining early information about these events is crucial in reducing the probability of major failures and managing rare occurrences promptly with minimal damage. Effective management of natural disasters depends on key physical factors, including early detection, prevention, advance warning, and timely notification to the public and relevant authorities.

According to the World Fire Statistics Report 2018, there were approximately 62,000 fatal building fires between 1993 and 2016, with 57 countries reporting fire-related fatalities. South Korea's National Fire Data System (NFDS) reported 24,539 structural fires between September 2020 and September 2021, resulting in 250 deaths, 1,646 injuries, and $705,960 in direct property damage. Additionally, South Korea experienced 78,219 vehicle fires during the same period, causing 461 fatalities, 1,875 injuries, and $357,609 in property damage. Wildfires, in comparison to building and vehicle fires, are the most devastating natural disasters that affect the environment's life cycle. Various factors can trigger wildfires, including high summer temperatures, changing environments, cloud-based lightning, falling rocks, and the friction of dry branches.

These alarming statistics have motivated researchers to develop reliable systems for early fire detection. Several researchers have explored soft computing techniques for preventing the spread of fires using traditional fire alarm systems, conventional fire alert systems (CFAS), and optical sensors [2]. For fire detection in CFAS, various scalar sensors were used, including visual, flame, and smoke sensors that require close proximity to the fire. However, scalar sensor-based systems are limited in providing additional information, such as area coverage, burning intensity, location, and fire size. In this study, we employ a small-scale version of the ConvNext model, suitable for devices with limited resources, offering improved detection accuracy and reduced false alarm rates. The main contributions of this study are as follows:

(i) To address the computational resource constraints of IoT devices, we propose a lightweight deep learning model capable of real-time operation. This model outperforms the well-known lightweight NASNetMobile and MobileNetV1 networks in terms of accuracy while using only 2.01 and 0.94 million fewer learning parameters, respectively.

(ii) Existing datasets for detecting wildfires are homogenous in nature, limiting their generalizability. In contrast, we collected diverse samples from personal repository videos, social media platforms like Facebook and Instagram, news channels, and YouTube videos.

The rest of the article is organized as follows: Section 2 provides a brief review of fire detection literature; Section 3 describes the model's architecture; Section 4 discusses experimental results; and Section 5 concludes the paper.

2. Literature Review

In recent literature, Conventional Fire Alert Systems (CFAS) and vision sensor-based systems have demonstrated their relevance in the field of fire detection. CFAS utilize various environmental sensors, such as smoke detectors, temperature sensors, and cameras, for detecting fires. However, CFAS technologies are ineffective at detecting fires over large distances, particularly in outdoor contexts, and require close proximity to the fire. Moreover, CFAS cannot provide information about the fire's burning conditions and rate. When an alarm is triggered, CFAS systems necessitate human intervention, such as visiting the fire site to confirm its presence. To address these limitations, numerous visual sensor-based fire detection systems have been proposed in the literature.

Vision-based fire detection systems are broadly divided into two categories: Traditional Fire Detection (TFD) and deep learning-based approaches. TFD-based technology relies on pattern recognition and digital image processing. For example, Liu et al. [3] employed temporal, spatial, and spectral analyses to identify fire regions in an image. However, their solution does not consistently account for fire's irregular shape due to the variable form of moving objects. Foggia et al. used motion analysis, color features, and a bag-of-words approach to detect fire [4]. Khan et al. [5] proposed a cost-effective CNN architecture for fire detection, offering reasonable computational complexity and suitability compared to more complex networks like AlexNet. The model is then fine-tuned to improve the accuracy and efficiency of detecting fire in surveillance videos. Furthermore, Khan et al. [6] employed an innovative CNN model for fire detection, localization, and semantic understanding of fire scenarios that is energy-efficient and computationally effective. It utilizes fully connected layers, a small convolutional kernel, and no dense layers. Khan et al. [1] introduced an early fire detection system for managing fire disasters using CCTV cameras and a fine-tuned neural network capable of detecting fire in both indoor and outdoor settings, along with a dynamic channel selection method for reliable data transmission.

Chen et al. [7] presented a decision rule-assisted approach to fire detection, relying on frame-to-frame variation and utilizing the irregular fire attribute for identification. They also applied an RGB and HIS color model to analyze fire's dynamic behavior. Marbach et al. [2] introduced a model for flame detection in tunnels by comparing video frames and their color features. Another algorithm for real-time fire detection in video games was proposed by [2], focusing on the temporal variation of fire intensity and employing a YUV color model in conjunction with motion features to predict fire and non-fire pixels. Han and Lee [8] developed another model for flame detection in tunnels based on video frame comparison and color feature analysis. Researchers in [9] presented a fire detection method for forests using wavelet analysis and FFT-based contours. In [10], an automatic fire detection approach for real-time video footage was investigated, based on temporal changes in fire intensity to distinguish between fire and non-fire patterns.

Current literature shows that various fire detection algorithms have been developed, achieving considerable accuracies. To safeguard lives and property, it is essential to further enhance detection accuracy while reducing false alarm rates. Additionally, these models often require powerful GPUs and TPUs, resulting in high computational costs. In this study, we employ a ConvNext-based deep learning model for fire detection, which boasts excellent detection accuracy and low false alarm rates.

3. Proposed Methodology

This section describes our model and overall framework, including data preparation. The data preparation process readies the data for training and testing. To enhance the number of training examples for better evaluation and generalization of results, data augmentation techniques such as scaling, horizontal flipping, rotation, and contrast enhancement are employed. Augmented datasets are then used to train various ConvNext models. The subsequent sections offer a brief overview of each phase within the proposed framework, with Figure 1 serving as a visual representation.

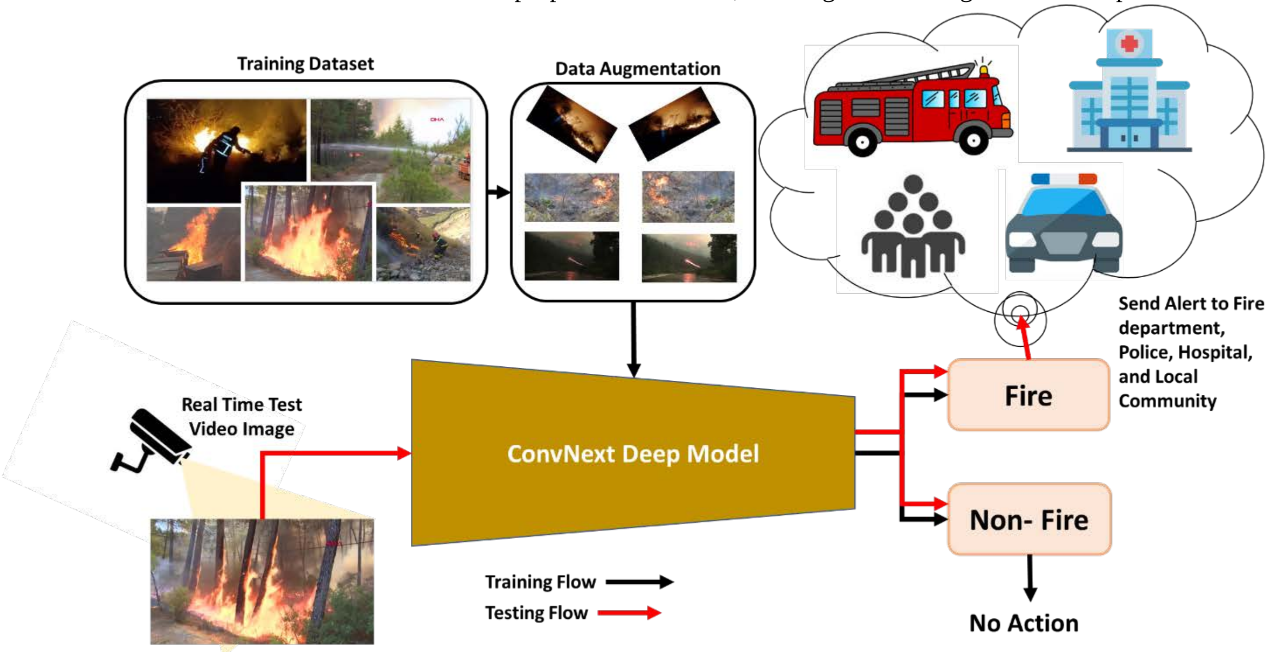

Figure 1: Overall framework of the proposed IoT assisted deep learning-based fire detection.

3.1. ConvNext CNN Model

The ConvNeXT model was placed in a ConvNet for the 2020s [11]. Inspired by the Vision Transformers' architecture, ConvNeXT is a pure convolutional model (ConvNet) that aims to outperform them. The authors employed the AdamW optimizer, trained the model across a larger number of epochs, significantly augmented the data, and utilized regularization techniques. Combining both methods, the performance of ResNet-50 on ImageNet Top1 accuracy improved from 76.1% to 78.8%. The ConvNeXt model does not require specific modules such as shifting window attention or relative position biases and has approximately the same FLOPs, performance, and memory usage as the Swin Transformer. While these results are promising, they are not yet entirely convincing, as ConvNext has only been partially explored to date. However, the scaling behavior of Transformers truly sets them apart.

A significant challenge in computer vision is determining whether a ConvNet can compete with Swin Transformers on downstream tasks like object detection and semantic segmentation. In this research, we utilized the ConvNext-tiny model for fire detection. We chose the tiny version because it is extremely lightweight and can be easily implemented on IoT devices. Starting with the pre-trained weights from ImageNet, we fine-tuned the model on the forest fire dataset using a low learning rate. The model has effectively learned critical fire patterns, and experimental results demonstrate its robustness in fire detection and ability to ignore non-fire images.

The architecture of ConvNext is inspired by the Vision Transformers' design, but it is a pure convolutional model (ConvNet) that aims to outperform them. ConvNext does not require specific modules, such as shifting window attention or relative position biases, resulting in a more streamlined model with similar FLOPs, performance, and memory usage as the Swin Transformer.

In the ConvNext model, several convolutional layers are stacked together, followed by batch normalization and activation layers. These layers can capture both low-level and high-level features in the input images. The model also incorporates skip connections, which help to maintain the gradient flow during backpropagation and alleviate the vanishing gradient problem. This architecture allows the ConvNext model to learn complex representations of the input data and perform well on various computer vision tasks.

We have addressed the overfitting problem in the ConvNext model by using weight decay and data augmentation techniques. Weight decay adds a regularization term to the loss function, which helps to prevent the model from fitting the noise in the training data. Data augmentation techniques, such as scaling, horizontal flipping, rotation, and contrast enhancement, increase the diversity of the training data, further improving the model's generalization capabilities.

3.2. Data preparation

Preprocessing refers to all modifications applied to the raw data before it is fed into the proposed lightweight Convolutional Neural Network (LW-CNN) architecture. For example, classification performance would be negatively impacted if the CNN architecture is trained on raw data. To enhance classification performance, we expand the input data through data augmentation techniques, creating new images with varied orientations, locations, and sizes. The following sections detail the data augmentation processes employed.

Data augmentation is one of the most widely utilized techniques to generate new images from existing ones, enabling a deep model to handle various variations. This technique increases the diversity of images, making CNN architectures more resilient to challenging situations and improving their classification capabilities. Through data augmentation, the model gains the ability to generalize by learning the same object in the image from multiple perspectives. We used several data augmentation and enhancement techniques before training the model for this purpose.

We implemented data augmentation techniques such as horizontal flipping, contrast adjustment, brightness adjustment, and rotation to increase the variety of training images. Horizontal flipping creates mirror images of the original data, while contrast and brightness adjustments modify the intensity of the image pixels. Rotation, on the other hand, transforms the images by specific angles. These augmentation techniques help the model learn various representations of the input data, thus enhancing its ability to recognize fire patterns and generalize well across diverse scenarios.

3.3. Connectivity with IoT

The IoT is considered the future evolution of the Internet. As technology advances, we are moving towards a society where everything and everyone is connected to the Internet [12]. The primary aim of IoT is to enable autonomous and secure connections and data transfers between real-world devices in various applications. IoT bridges the gap between the physical and digital worlds.

The number of devices connected to the Internet is rapidly growing, including personal computers, laptops, tablets, smartphones, and other portable embedded devices. Most mobile devices are equipped with various sensors and actuators that can perceive and process data, make intelligent decisions, and transmit valuable information collected via the Internet. Utilizing a network of such devices with diverse sensors can lead to a wide array of impressive applications and services that have the potential to positively impact one's life on personal, professional, and financial levels. The Internet of Things (IoT) consists of physical objects, sensor technology, network infrastructure, computing power that can be stored in the cloud, and systems for making decisions and triggering actions. These objects are uniquely identifiable, possess specific attributes, and can be accessed through the Internet. The basic simplified workflow of IoT can be described as follows:

Sensing, identifying, and communicating information about objects. The information is the data detected regarding temperature, direction, motion, vibration, acceleration, humidity, chemical changes in the air, etc., depending on the type of sensor. A variety of sensors can be used to create intelligent services.

Initiating a response; an intelligent device or system processes the provided object information and then determines the automated action that should be executed. The intelligent system/device offers a range of services and has a method for informing the administrator of the system's current state and the outcomes of actions taken.

In our proposed system, we are mindful of privacy concerns related to IoT devices. We have designed the framework to ensure that data collection, transmission, and processing adhere to data privacy regulations and best practices. By incorporating privacy-preserving techniques, our IoT-assisted fire detection system ensures that personal information and sensitive data are protected, while still effectively detecting and responding to fire incidents.

3.4. Alert Generation

Once a fire is detected, generating an alarm is a critical step that enables various emergency services to take prompt action to mitigate the situation. These services include:

- Fire brigade: Alerting the fire brigade allows them to deploy their resources swiftly and efficiently to extinguish the fire before it spreads further, minimizing damage to property and the environment.
- Hospitals and medical teams: Notifying hospitals and medical teams ensures that ambulances are dispatched to the scene of the fire, providing timely assistance to any victims who may require medical attention or evacuation.
- Police department: Informing the police department enables them to

take legal actions, secure the area, and investigate the cause of the fire, helping to prevent future incidents and identify any potential criminal activity associated with the fire.

- Social media alerts: Utilizing social media to inform local residents about the fire's location allows for early evacuation planning and heightens community awareness. This rapid dissemination of information can help to ensure the safety of the neighborhood and prevent further casualties, particularly if the fire is spreading quickly.

Our proposed fire detection system is designed to facilitate efficient communication and collaboration between these various emergency services. By providing real-time alerts and accurate information about the location and extent of the fire, our system enables a coordinated response, ultimately saving lives, property, and minimizing environmental damage..

4. Experimental Evaluation

In this section, we provide a detailed overview of the evaluation measures, statistics, and visual findings. We begin by describing the experimental design and performance measurements, followed by an explanation of the dataset used, and finally, we present the results evaluation. All models, including the proposed model, are trained for 20 epochs with a low learning rate, allowing the network to retain most of the previously learned information. We retrain each network using its default input size, a batch size of 32, and the stochastic gradient descent optimizer with a learning rate set to 1e-4 and a momentum of 0.9. The performance of the proposed method is assessed using four metrics: overall accuracy, precision, recall, and F1-scores, which provide a comprehensive understanding of the model's effectiveness in various aspects of the fire detection task.

$$\text{Accuracy} = \left(\frac{True_{Positives}+True_{Negatives}}{TP+True_{Negatives}+False_{Positives}+False_{Negatives}}\right)$$

$$\text{Precision} = \left(\frac{True_{Positives}}{True_{Positives}+False_{Positives}}\right)$$

$$\text{Recall} = \left(\frac{True_{Positives}}{True_{Positives}+False_{Negatives}}\right)$$

$$\text{f1\_score} = 2 * \left(\frac{precision*recall}{precision + recall}\right)$$

4.1. Dataset

The proposed method is evaluated using a recently created forest fire dataset [13]. This dataset consists of two classes, forest fire and non-fire, with a combined total of 2,000 images, where each class contains 1,000 images. The data has been collected from various sources, including real-world self-recorded videos, Facebook, news channels, and YouTube videos. This dataset poses a significant challenge for evaluation due to the inclusion of numerous non-fire images that closely resemble fire patterns, as illustrated in Figure 3. This diverse collection of images allows for a more robust assessment of the proposed method's performance in detecting fires and distinguishing them from visually similar non-fire scenarios.

4.2. Comparison with State-of-the-Art

The proposed model is compared with other deep learning methods, as shown in Table 1. Our method achieved the highest fire detection precision of 0.98, representing a 4% improvement over the previous high of 0.94 achieved by the VGG16 CNN model, and a high recall of 1.0. For Non-Fire classification, our model attained a precision of 1.0 and a recall of 0.98, surpassing the previously high-performing LW-CNN and VGG16 models. With the highest F1-score of 0.99 for both fire and non-fire classes, the proposed method demonstrates its robustness in detecting fires with a low false alarm rate on the forest fire dataset.

Table 1: Comparison with other fire detection deep learning models on Forest-fire dataset. The values of all models given in the table are collected from article [13].

Methods	**Fire Precision**	**Fire Recall**	**Fire F1-Score**	**Non-Fire Precision**	**Non-Fire Recall**	**Non-Fire F1-Score**
AlexNet	0.83	0.93	0.86	0.91	0.79	0.8
ResNet50	0.82	0.96	0.87	0.95	0.77	0.86
NASNetMobile	0.90	0.95	0.92	0.94	0.88	0.91
MobileNetV1	0.87	0.99	0.93	0.98	0.86	0.92
VGG16	0.94	0.93	0.93	0.92	0.94	0.93
LW-CNN	0.91	0.98	0.95	0.98	0.91	0.94
Proposed Method	**0.98**	**1.0**	**0.99**	**1.0**	**0.98**	**0.99**

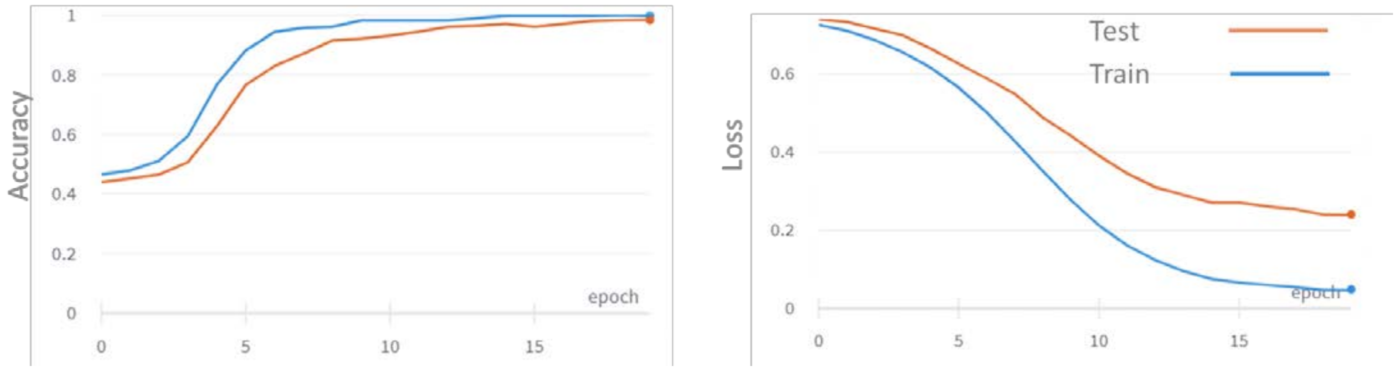

Figure 2: The learning graphs, including train, test accuracies, and losses of the ConvNext-tiny model

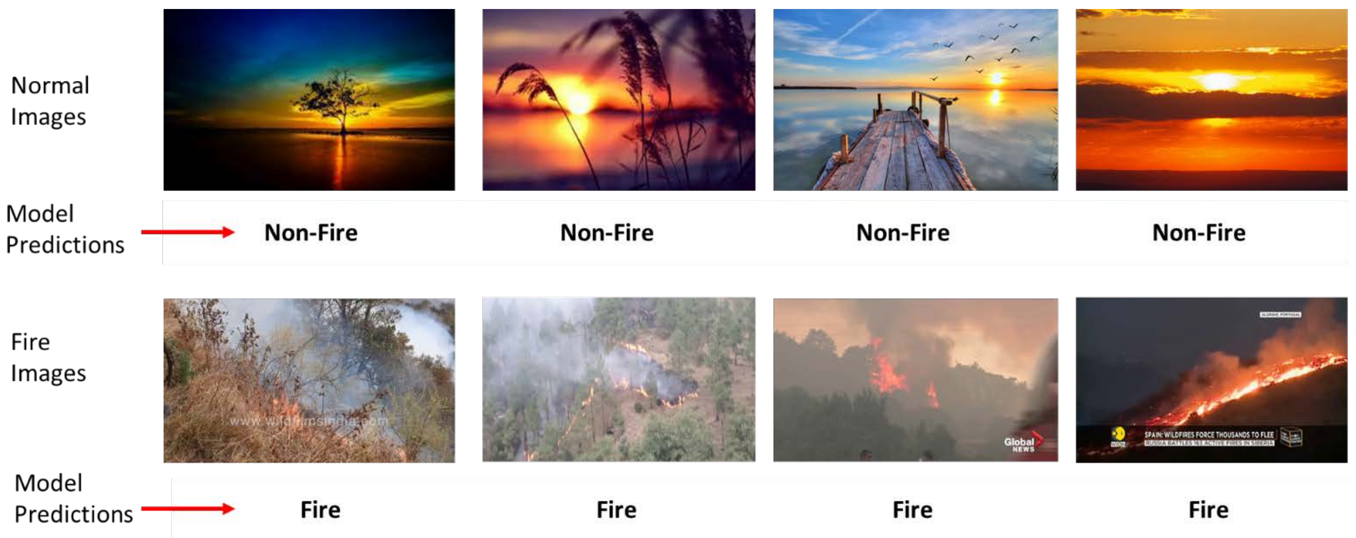

Figure 3: Visual findings of the proposed model achieved for challenging images in the dataset for normal and fire predictions.

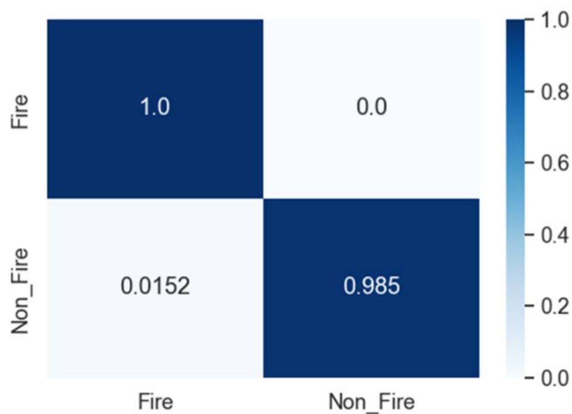

Figure 4: Confusion matrix of the proposed model for forest-fire dataset.

Figure 2 shows the learning graphs for the proposed model. The training and test actuaries are increasing to highest point with any sparks in the graph. Same is for training and testing losses, which indicates that the model is not overfitting. We also visualized some of the visual results in Figure 3. It can be seen from Figure 4 top row that images are very similar to fire, however, our model is so robust to detect those patterns as non-fire class. On the other hand, in second row we can see that fire is very tiny portion of the image, but it is correctly detected by the proposed model. Figure 4 shows the confusion matrix of the proposed model over test data using forest fire dataset. The proposed model obtained 99% and 98% accuracy for the fire and non-fire classes, respectively. This illustrates the effectiveness of the suggested model over forest fire dataset. The confusion matrix provides a summary of the classification performance of the proposed model on the forest fire dataset. In this case, the confusion matrix shows the following results:

- Fire detected as fire: 1.0
- Non-fire detected as fire: 0.0152
- Non-fire detected as non-fire: 0.985

These results indicate that the model has excellent performance in detecting fires, with a true positive rate of 1.0. This means that all fire instances in the test dataset were accurately identified as fires by the model. However, the model also shows a small false alarm rate, as it incorrectly classifies 1.52% (0.0152) of non-fire instances as fires. While this rate is low, it still represents a minor limitation of the model. The false alarm rate is an important metric to consider, as false alarms can lead to unnecessary emergency responses, wasting resources, and potentially causing panic.

On the other hand, the model has a high true negative rate of 0.985, meaning that 98.5% of non-fire instances were accurately identified as non-fires. This high true negative rate, combined with the low false alarm rate, demonstrates the overall robustness and effectiveness of the proposed model in detecting fires and minimizing false alarms in the forest fire dataset.

4. Conclusion

In this study, we presented a deep learning model with IoT support for effective fire detection. The pre-trained weights of the ConvNext model are utilized to identify minute details and clearly distinguish between lights and fire patterns. The model is fine-tuned for the forest fire dataset. We employed an IoT system that functions when a fire is detected. The IoT device is mounted with a camera, and when it detects fire, it generates an alarm and sends it to concerned departments so that they can take the required steps to stop its spread. By comparing our approach to the current state-of-the-art model, we were able to increase accuracy by 4% using the forest fire dataset.

One limitation of our method is its struggle to detect long-range fires due to the unavailability of data representing such scenarios in the dataset. In the future, we will work on the efficiency of the deep models to run them in real-time on small IoT devices. Additionally, we will investigate more scenarios and datasets, including those containing long-range fire examples, to address this limitation and further improve the performance of our fire detection system..

Conflicts of Interest

The authors declare no Conflict of Interest.

References

[1] K. Muhammad, R. Hamza, J. Ahmad, J. Lloret, H. Wang, and S. W. Baik, "Secure surveillance framework for IoT systems using probabilistic image encryption," *IEEE Transactions on Industrial Informatics,* vol. 14, no. 8, pp. 3679-3689, 2018.

[2] G. Marbach, M. Loepfe, and T. Brupbacher, "An image processing technique for fire detection in video images," *Fire safety journal,* vol. 41, no. 4, pp. 285-289, 2006.

[3] C.-B. Liu and N. Ahuja, "Vision based fire detection," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, 2004, vol. 4: IEEE, pp. 134-137.

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

[5] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, and S. W. Baik, "Efficient deep CNN-based fire detection and localization in video surveillance applications," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 49, no. 7, pp. 1419-1434, 2018.

[6] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, "Convolutional neural networks based fire detection in surveillance videos," *IEEE Access,* vol. 6, pp. 18174-18183, 2018.

[7] T.-H. Chen, P.-H. Wu, and Y.-C. Chiou, "An early fire-detection method based on image processing," in *2004 International Conference on Image Processing, 2004. ICIP'04.*, 2004, vol. 3: IEEE, pp. 1707-1710.

[8] R. D. Lascio, A. Greco, A. Saggese, and M. Vento, "Improving fire detection reliability by a combination of videoanalytics," in *International Conference Image Analysis and Recognition*, 2014: Springer, pp. 477-484.

[9] Z. Zhang, J. Zhao, D. Zhang, C. Qu, Y. Ke, and B. Cai, "Contour based forest fire detection using FFT and wavelet," in *2008 International conference on computer science and software engineering,* 2008, vol. 1: IEEE, pp. 760-763.

[10] T. Celik and H. Demirel, "Fire detection in video sequences using a generic color model," *Fire safety journal,* vol. 44, no. 2, pp. 147-158, 2009.

[11] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 11976-11986.

[12] N. Dilshad, A. Ullah, J. Kim, and J. Seo, "LocateUAV: Unmanned Aerial Vehicle Location Estimation via Contextual Analysis in an IoT Environment," *IEEE Internet of Things Journal*, 2022.

[13] H. Yar, T. Hussain, Z. A. Khan, D. Koundal, M. Y. Lee, and S. W. Baik, "Vision sensor-based real-time fire detection in resource-constrained IoT environments," *Computational intelligence and neuroscience,* vol. 2021, 2021.