Kathford Journal of Engineering and Management,
ISSN: 2661-6106.

https://kathford.edu.np/Journal/

VOL. 4 ISSUE 1 DECEMBER 2024, PP. 73-83

Performance Evaluations of the Deep Learning Models in Reference to Real-Time Fire and Smoke Detections Abilities

Ankit Rajbhandari[1], Devraj Pokhrel[1], Siddartha Gajurel[1], Chaitya Shova Shakya[1,*]

[1]*Department of Computer and Electronics, Communication & Information Engineering, Kathford International College of Engineering and Management (Affiliated to Tribhuvan University), Balkumari, Lalitpur, Nepal*

*Corresponding Author: *chaitya.shakya@kathford.edu.np*

***ABSTRACT*—** This study is primarily focused on developing a real-time fire and smoke detection systems using two predominant deep learning models, viz. VGG16 and RESNET50. Throughout this work, both of these models are operated in parallel, processed the frames from the input video footage, and made all the essential predictions on each frame. In the course of the practical implementations of the models, they are trained and validated well using the Fire-Smoke-Dataset retrieved from the open source code DeepQuestAI. Over the course of 120 epochs, the model's accuracy is found to be improved from 80% to approximately 96% during the training. Their generic systems are noticed to be able enough to classify the fire, smoke, and neutral cases in all the pre-recorded videos and real-time webcam footages. The results presented herewith are basically vowed to demonstrate the efficacy and workability of both of the models, and their working performances (RESNET50 outperforms VGG16, and achieves an accuracy range of 87.67% compared to 82%). In terms of their precision and recall aptitudes, and efficiency & effectivity in identifying the fire and smoke instances, both of the models are found to offer the satisfactory detections mechanisms. We believe that the present in-depth yet critical evaluations and assessments on the workability of the deep learning models can illuminate the potentialities and promising functions of the VGG16 and RESNET50 towards detecting the real-time fire and smoke which in turn are quite indispensable for promoting them in the public safety and security mechanisms installed in the wide range industrial and non-industrial delicate sectors.

KEYWORDS— ***Fire detection, Smoke detection, Transfer learning, VGG16, RESNET50***

1. INTRODUCTION

Accurate fire and smoke detection in public places is becoming more important with the growing need for improved public safety and efficient emergency response systems. In developing countries like Nepal, where resources and infrastructure are often limited, the ability to monitor and respond to fire and smoke incidents in real-time is essential to minimize risks and damage. Traditional fire and smoke detection methods, such as basic sensor-based systems, often lack the accuracy, speed, and scalability required for efficient crisis management. These methods typically rely on temperature, gas, or smoke sensors, which may only activate once a fire is intensified, leading to delayed response times.

Moreover, such systems struggle to detect early-stage fires, especially in large, open areas or outdoor environments with limited sensor coverage. They are also prone to false alarms triggered by environmental factors like dust, humidity, or steam. Due to these limitations, there is a growing demand for automated systems that leverage existing infrastructure, such as public surveillance cameras, to detect fire and smoke in real time. By using computer vision and deep learning models, these systems can detect fire and smoke at earlier stages, reduce false alarms, and enable proactive emergency responses, ultimately enhancing public safety.

This work involves designing and implementing a deep learning system that integrates convolutional neural network (hereafter, CNN) based architectures, specifically VGG16 and RESNET50, for accurate fire and smoke detection. These architectures are known for automatically extracting hierarchical features from images, enabling precise identification of complex patterns, such as flames and smoke textures, even in dynamic and cluttered environments. Their use enhances detection accuracy, reduces false alarms, and enables real-time analysis to trigger immediate response actions for improving public safety and enhancing crisis management capabilities.

2. LITERATURE REVIEW

Traditional fire and smoke detection systems, employing smoke detectors and heat sensors, face significant challenges, particularly in open or large areas. These systems are less effective in detecting fires at their early stages due to their reliance on physical changes like temperature rise or the presence of smoke particles. Such limitations hinder timely detection, which is crucial for preventing the rapid spread of fire. Furthermore, these systems often struggle with false alarms triggered by non-fire-related heat or smoke sources, making them less reliable in dynamic environments. These constraints emphasize the need for advanced, automated detection systems to detect fire and smoke at an earlier stage and in more complex environments.

Recent advancements in deep learning have enabled the development of more efficient fire and smoke detection models using computer vision techniques. CNNs have become an essential component of video surveillance-based detection systems because of their ability to automatically extract and learn features from raw visual input. Namozov and Cho (2018) proposed a deep learning-based algorithm that could detect fires accurately even with limited data. Their approach utilized a CNN to analyze video frames, demonstrating how deep learning can overcome the constraints of sensor-based systems. This work underscored the importance of CNNs in processing image-based data for early fire detection.

Similarly, Foggia et al. (2015) introduced a real-time fire detection system that utilized expert models based on color, shape, and motion analysis from surveillance videos. By integrating visual data analysis with expert systems, their method improved detection accuracy and significantly reduced false alarms. It demonstrated the practical effectiveness of deep learning models in fire and smoke detection, as the system could operate in real-time and respond more effectively than traditional sensor-based methods.

Kim and Lee (2019) extended these advancements by implementing a video-based fire detection system using the VGG16 architecture. Their system analyzed both temporal and spatial data from video sequences, enhancing the detection of fire and smoke through feature extraction. This approach highlighted the role of CNNs in visual pattern recognition, as they can effectively capture spatial relationships within image frames and temporal variations across video sequences. The system's performance was superior to earlier models, illustrating the advantages of CNNs for real-time fire and smoke detection in video surveillance systems.

Further developments in transfer learning have also contributed to the field. Tammina (2019) and Liang (2020) demonstrated how transfer learning could improve the efficiency of fire and smoke detection by using pre-trained models like VGG16 and RESNET. These models reduced the computational and data requirements typically associated with training deep learning models from scratch. Transfer learning enables the reuse of pre-learned knowledge from large-scale datasets, which is then fine-tuned for fire and smoke detection tasks. This approach significantly reduces the need for large, labeled datasets while accelerating the deployment of detection systems.

CNNs like VGG16 and RESNET are used in video-based fire and smoke detection because of their ability to learn features from raw video data. VGG16 captures low-level features such as edges and textures, and high-level patterns such as object shapes, due to its deep architecture with multiple convolutional layers, distinguishing fire and smoke from other elements in the video. RESNET uses residual connections that allow for deeper networks without the issue of vanishing gradients. This enables RESNET to capture more complex patterns crucial for detecting fire and smoke under various conditions. Both models process visual data locally through convolutional layers, reducing computational complexity and ensuring fast real-time performance.

The ability of these CNNs to generalize across different environments and their efficiency in learning from large datasets, especially through transfer learning, further enhance their accuracy. In essence, the combination of hierarchical feature learning, depth, and computational efficiency makes VGG16 and RESNET highly accurate for fire and smoke detection in video footage.

3. COMPUTATIONAL AND THEORETICAL DETAILS

3.1 Methodology Outlines

The detection of fire and smoke using deep learning involved five different phases, as outlined in Figure 1: data collection, data pre-processing, model selection, model training, and model evaluation.

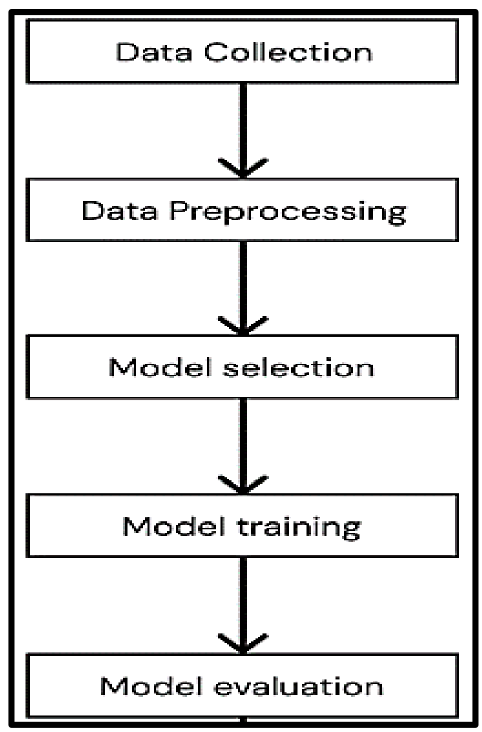

Figure 1. The flow diagram for deep learning models

We used the Fire-Smoke-Dataset from

DeepQuestAI, which consists of 900 images segregated into three classes: fire, smoke, and neutral scenarios, with 90 images reserved for testing purposes. This dataset was chosen for its relevance to fire and smoke detection, as it provides a diverse range of images across different fire and smoke instances. This diversity allows for better generalization, making the models more capable of detecting fire and smoke in real-world settings where such incidents can appear under varying conditions. Additionally, the use of a publicly available dataset allows for benchmarking against other models in the field, ensuring the validity and reliability of the results.

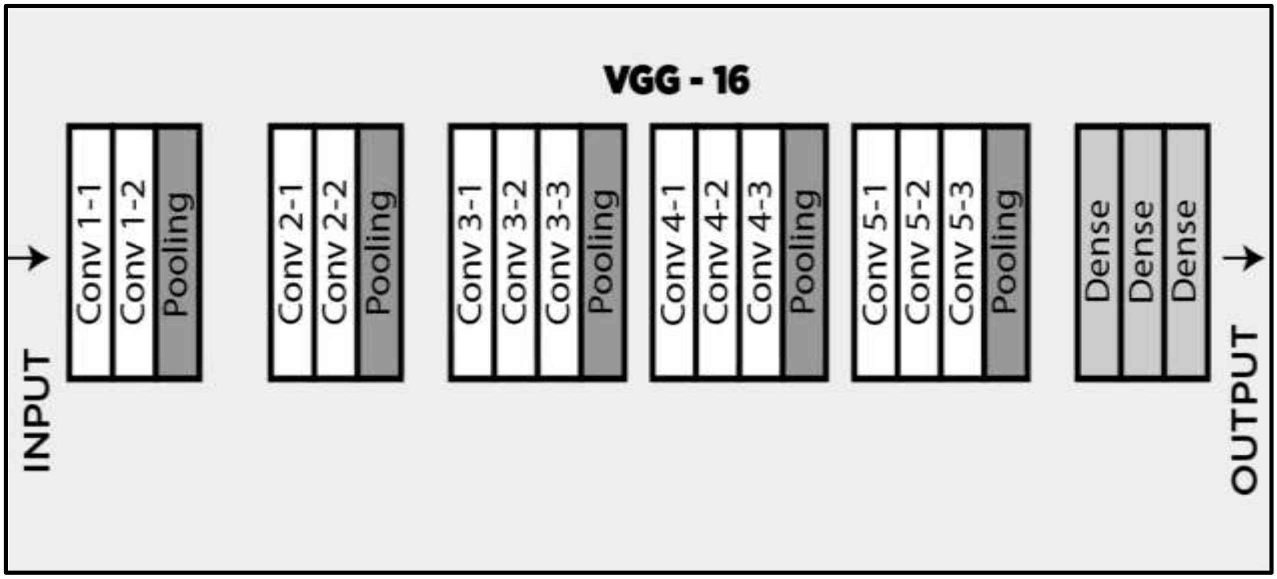

Figure 2. VGG16 architecture map§ (§*www.geeksforgeeks.org*)

The images were pre-processed to prepare them for training deep learning models. Each input frame was resized to a fixed size of 64x64 pixels. Subsequently, the images were transformed into PyTorch tensors. During this pre-processing phase, pixel values were normalized to the range [0, 1], standardizing the input data. This normalization ensures consistency in the scale of input features, which enhances the model's performance. Both models were trained on a pre-processed dataset for 120 epochs. During training, the models progressively improved their accuracy by learning patterns corresponding to fire, smoke, and neutral classes. The cross-entropy loss function was used to minimize classification errors, and optimization was performed using Stochastic Gradient Descent (SGD) with momentum. This approach improved both convergence speed and accuracy.

After training, the models were validated by evaluating the models on a reserved set of 90 test images that had not been used during the training phase. Evaluation metrics such as accuracy, precision, recall, and F1-score were employed to assess the models' effectiveness in classifying fire, smoke, and neutral images. Furthermore, the models were tested on real-time webcam footage to determine how well

Kathford Journal of Engineering and Management,
ISSN: 2661-6106.

https://kathford.edu.np/Journal/
VOL. 4 ISSUE 1 DECEMBER 2024, PP. 73-83

they functioned under dynamic conditions.

3.2 Model Architectures

We used two CNN models, VGG16 and RESNET50, to detect fire and smoke in real-time. Both architectures are well-known in image classification due to their ability to extract relevant features from raw image data efficiently.

The VGG16 was chosen for its simplicity, deep structure, and proven effectiveness in learning complex image features. The architecture, developed by Simonyan and Zisserman (2014) as shown in Figure 2. It consists of 13 convolutional layers organized into blocks. Each block is followed by max-pooling layers to reduce spatial dimensions and computational complexity. After these layers, three fully connected layers are used for classification. The final output layer generates class probabilities through a softmax function.

Key features of VGG16 include:

- Convolutional Layers extract features such as edges, textures, and object parts from the input image.
- Activation Function ReLU is used for allowing the model to learn non-linear patterns.
- The final output layer generates a probability distribution over the possible class labels (fire, smoke, and neutral).

On the other hand, RESNET50, a deep residual network, uses residual learning to train very deep networks. Research by He et al. (2016) and other related studies has highlighted its ability to learn deep, hierarchical features from images, making it an ideal choice for applications such as fire and smoke detection.

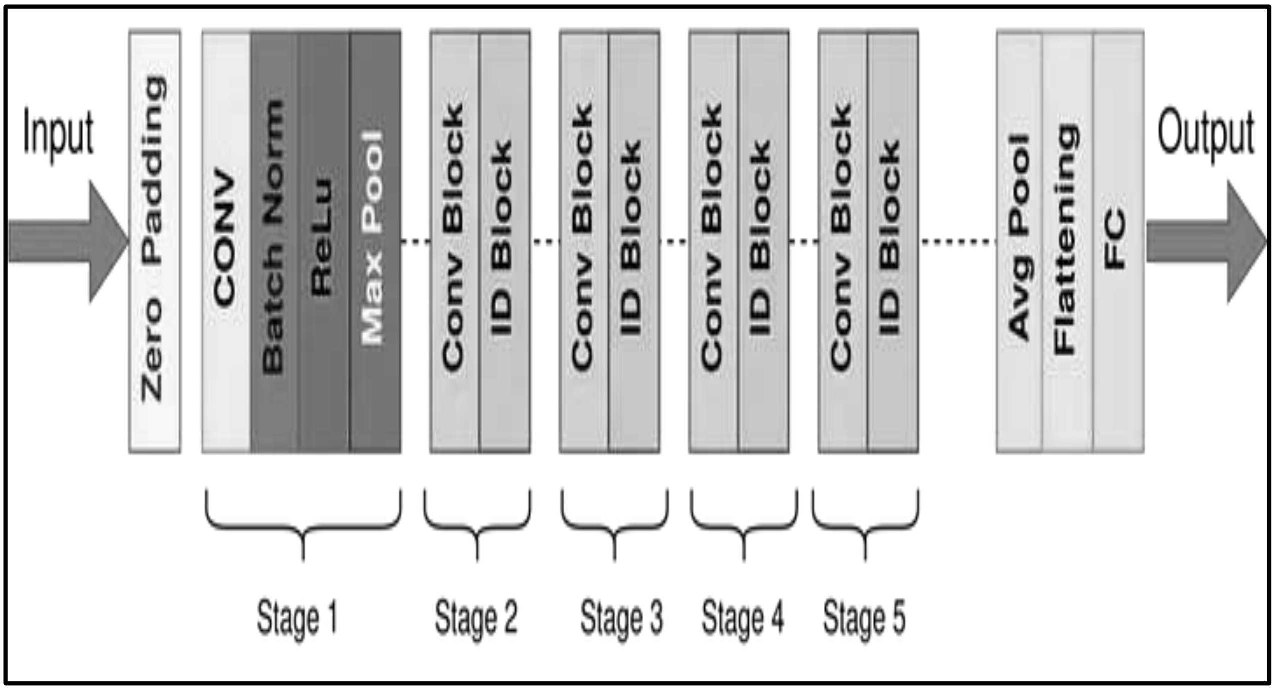

Figure 3. RESNET50 architecture map[ξ] *([ξ]www.medium.com)*

The RESNET50 architecture consists of 50 layers with residual blocks as the primary components as shown in Figure 3. These blocks contain skip connections that bypass one or more layers to add the input to the output, mitigating the vanishing gradient problem. This design enhances learning efficiency and enables the model to capture complex patterns from input data.

Key Features of RESNET50 include

- Residual blocks maintain the gradient flow during training, making it easier to train deeper networks.
- Bottleneck design reduces computational costs by using 1x1 convolutions, improving efficiency.
- The final layer generates the class probabilities via a softmax function.

4. RESULTS AND DISCUSSIONS

The evaluation of the fire and smoke detection models using VGG16 and RESNET50, provided significant insights into their effectiveness in identifying fire hazards. After prolonged training, both the VGG16 and RESNET50 models demonstrated steady improvement. Figures 4 and 5 show VGG16 and RESNET50's performance at various stages of training, highlighting reduction in training loss and validation loss over multiple epochs respectively. The training and validation loss curves for both VGG16 and RESNET50 demonstrate consistent improvement across 20 epochs, as illustrated in Figure 4 & 5, reflecting their learning efficiency. For VGG16, the training loss decreases from approximately 1.2 to 0.35, while the validation loss drops from 1.3 to 0.44, indicating slight overfitting and the need for improved generalization strategies. In contrast, RESNET50 achieved a training loss reduction from 1.4 to 0.37 and a validation loss drop from 1.45 to 0.52, with a smaller gap between the losses, indicating better

Figure 4. Training and Validation loss in VGG16

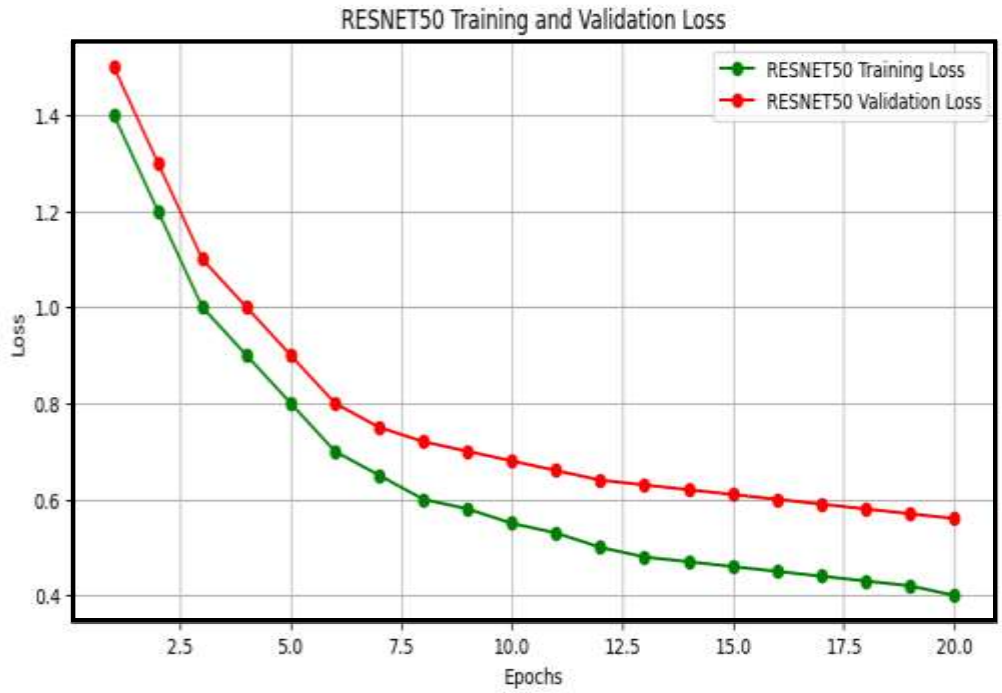

Figure 5. Training and Validation loss in RESNET50

Figure 6. Testing on pre-recorded instance

generalization. RESNET50 also has a smoother convergence pattern, which

Kathford Journal of Engineering and Management,
ISSN: 2661-6106.

https://kathford.edu.np/Journal/
VOL. 4 ISSUE 1 DECEMBER 2024, PP. 73-83

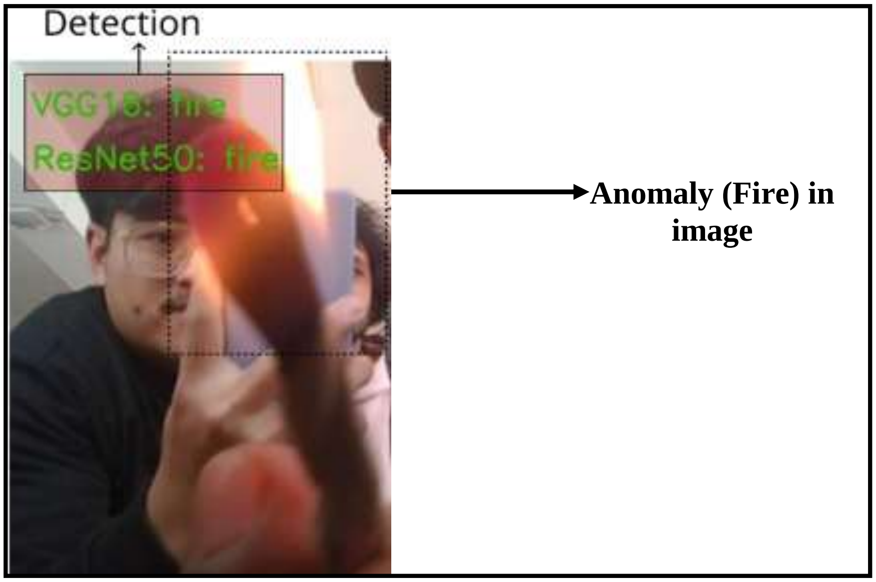

Figure 7. Testing on real-time instance

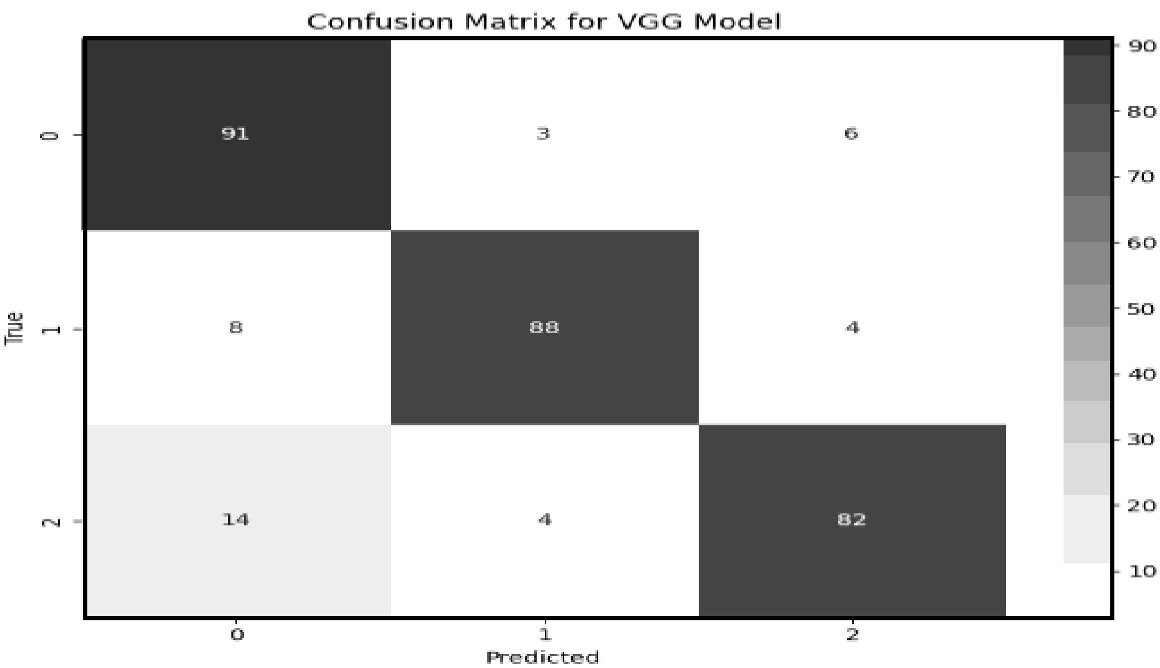

Figure 8. Confusion matrix for VGG16

demonstrate its robust feature extraction capabilities. Both architectures effectively processed the pre-recorded videos across diverse settings, such as varying lighting conditions and crowded scenes etc., and detected fire and smoke accurately as illustrated in Figure 6. The models were also able to correctly identify fire and smoke events during real-time testing on the video footage taken by laptop webcam as shown in Figure 7.

Figure 8 and Figure 9 illustrated the confusion matrices for VGG16 and RESNET50, respectively. The confusion matrices for the VGG16 and ResNet50 models, shown in Figure 8 and Figure 9, respectively, provide an insightful comparison of their classification performance across three classes: fire, neutral, and smoke. VGG16 demonstrated strong performance in classifying fire instances, correctly predicting 91 out of 100 fire cases, though it misclassified 3 as neutral and 6 as smoke. For the neutral class, VGG16 accurately identified 88 instances, while misclassifying 8 as fire and 4 as smoke. In the smoke class, it correctly classified 82 instances but made 14 misclassifications as fire and 4 as neutral. In contrast, ResNet50 correctly classified 84 fire instances, with 1 misclassified as neutral and 15 as smoke. For neutral instances, ResNet50 predicted 81 correctly, while misclassifying 6 as fire and 13 as smoke. The model excelled in the smoke class, correctly identifying 94 instances, with only 4 misclassifications as fire and 2 as neutral. While both models demonstrated effective classification capabilities, their respective strengths and misclassification patterns—VGG16's accuracy in detecting fire and ResNet50's proficiency in identifying smoke—highlight key differences in their predictive behaviors, offering opportunities for further refinement tailored to specific detection needs.

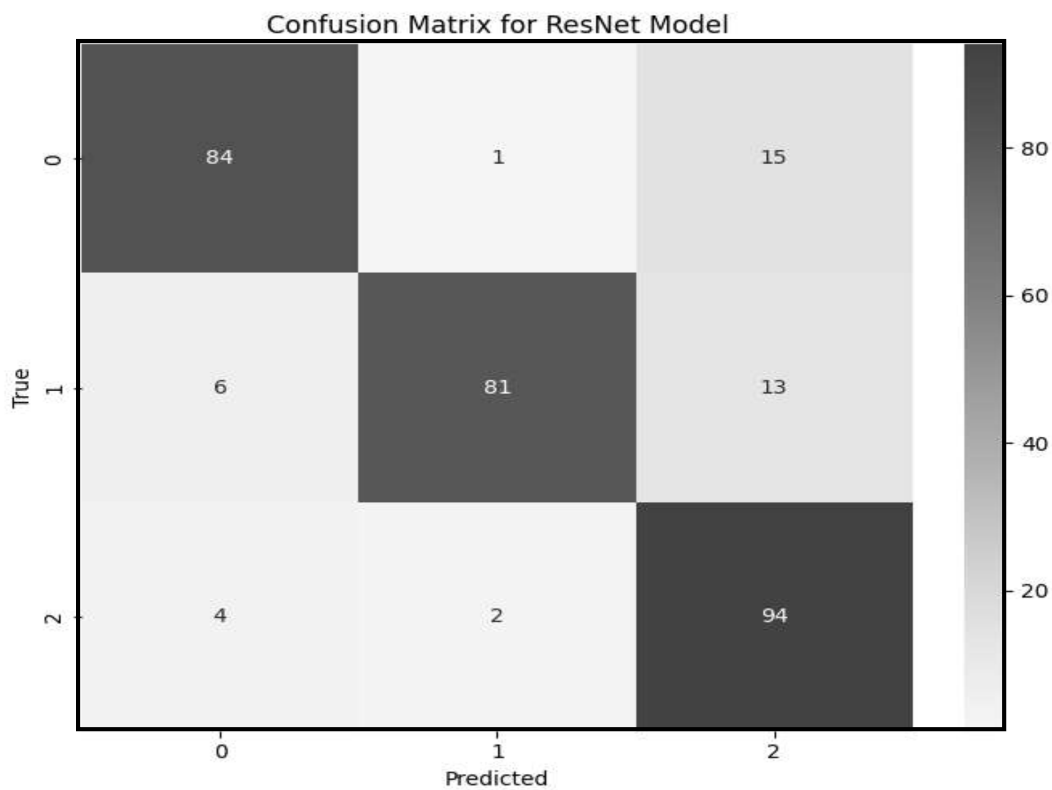

Figure 9. Confusion matrix for RESNET50

Performance metrics obtained during the

training and testing phases offered a quantitative measure of their ability to detect fire and smoke accurately while minimizing false alarms. The metrics used—validation loss, accuracy, precision, recall, and F1 score—are key indicators in classification tasks.

Validation loss reflects the model's generalization ability, with lower values indicating better performance.

Accuracy shows the proportion of correct predictions, though it may not capture model performance in imbalanced datasets.

Precision measures the correctness of positive predictions (fire or smoke), while **recall** indicates how well the model identifies all true instances of fire and smoke.

F1 score balances precision and recall, providing a comprehensive measure. These metrics are crucial for fire and smoke detection as they assess both the accuracy of predictions and the model's ability to minimize false negatives, which is vital in safety-critical applications. The correlation between precision, recall, and F1 score reflects the trade-off between detecting fire/smoke and avoiding misclassifications. Table 1 compares the performance metrics of VGG16 and ResNet50 for fire and smoke detection. The comparison shows that ResNet50 outperforms VGG16 across all metrics except validation loss. ResNet50 achieved a higher accuracy (87.67% vs. 82%), precision (88.44% vs. 82.55%), recall (87.67% vs. 82%), and F1 score (87.76% vs. 81.99%), indicating its superior overall performance in fire and smoke detection. ResNet50's consistently higher accuracy and other metrics suggest it is more effective at distinguishing between fire, smoke, and neutral classes in the dataset. The results indicate that both models show potential for fire and smoke hazard detection, but RESNET50's advanced architecture led to better accuracy and fewer detection errors.

Table 1. Performance metrics comparison for VGG16 and ResNet50

Metric	VGG16	RESNET50
Validation Loss	55.37%	67.28%
Accuracy	82.00%	87.67%
Precision	82.55%	88.44%
Recall	82.00%	87.67%
F1 Score	81.99%	87.76%

5. CONCLUSION

This study designed and evaluated the deep learning-based systems for real-time fire and smoke detection using VGG16 and RESNET50 architectures. The results demonstrated that both models were effective, with RESNET50 outperforming VGG16, achieving an accuracy of 87.67% compared to 82%. The model also exhibited high precision and recall, effectively identifying fire and smoke events. These findings emphasize the potential of deep learning in real-time fire and smoke detection scenarios to enhance safety and security in public and industrial settings. Among VGG16 and RESNET50, RESNET50 exhibited better performance during both training and real-time testing, making it a strong candidate for future deployment in fire and smoke alarm/safety systems.

6. LIMITATION AND RECOMMENDATION

A key limitation of the fire and smoke detection system was the lack of diverse datasets, which restricted the model's ability to generalize across various fire and smoke scenarios. Further testing and validation using diverse datasets in more complex, real-world environments is needed to confirm their effectiveness for deployment in safety systems. Future improvements may involve integrating additional data types, such as thermal imaging and environmental sensor readings, to strengthen the system. Additionally, connecting the system to IoT platforms could improve safety measures with automated alerts and proactive fire prevention.

7. ACKNOWLEDGMENTS

The authors would like to thank Department of Computer and Electronics, Communication & Information Engineering, Kathford International College of Engineering and Management for the timely initiation of this work. The invaluable assistance, continuous support, and prompt guidance we received from the specific faculties are commendable. The concerned editor and reviewer are very much acknowledged for the constructive comments and critical evaluations of the manuscript.

REFERENCES

1. Namozov, A., and Cho, Y. I. (2018). An Efficient Deep Learning Algorithm for Fire and Smoke Detection with Limited Data. *Advances in Electrical and Computer Engineering*, 18(4), 121–128.

2. Foggia, P., Saggese, A., and Vento, M. (2015). 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*, 25(9), 1545–1556.

3. Kim, B., and Lee, J. (2019). A Video-Based fire detection using deep learning models. *Applied Sciences*, 9(14), 2862.

4. Yusof, N. F. M., Roshdi, F. R. M., Saharuddin, S., and Noor, S. N. A. M. (2021). Fire safety management in Malaysian higher educational institutions. *Malaysian Journal of Real Estate/ International Journal of Real Estate Studies*, 15(S1), 70–81.

5. Tammina, S. (2019). Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images. *International Journal of Scientific and Research Publications.*

6. Liang, J. (2020). Image classification based on RESNET. *Journal of Physics: Conference Series.* 1634. 012110.

7. Shees, A., Ansari, M. S., Varshney, A., Asghar, M. N., and Kanwal, N. (2023). FireNet-v2: Improved Lightweight Fire Detection Model for Real-Time IoT Applications. *Procedia Computer Science, 218, 2233–2242.*