Preprint submitted to the Springer Signal, Image and Video Processing. You can find the published version at:
https://link.springer.com/article/10.1007/s11760-024-03384-x
View-only version: https://rdcu.be/dMeD4

3ENB2: End-to-End EfficientNetB2 Model with Online Data Augmentation for Fire Detection

Ehsanullah Zia[1], Hamed Vahdat-Nejad[1,*], Mohammad Ali Zeraatkar[2], Javad Hassannataj Joloudari[1,3,4], Seyyed Ali Hoseini[1]

[1]Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran
[2]Department of Computer Engineering, Islamic Azad University, Tehran, Iran
[3]Department of Computer Engineering, Islamic Azad University, Babol, Iran
[4]Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran
*Corresponding author: vahdatnejad@birjand.ac.ir

Email: ehsanullah.zia@birjand.ac.ir, vahdatnejad@birjand.ac.ir, zeraatkar@iauk.ac.ir, Javad.hassannataj@birjand.ac.ir, sa.hoseini@birjand.ac.ir

Abstract: Fire is recognized as a destructive disaster in smart environments that causes serious harm to ecosystems and humans. Early and rapid-fire detection in cities and forests can prevent human, economic, and environmental damage. Wireless sensor networks have been used for fire detection, but their deployment is costly and limited to specific locations. On the other hand, cameras are widely deployed in smart cities and interurban areas, as they are cheaper and more pervasive than sensor networks. In this paper, an end-to-end neural network model called EfficientNetB2 (3ENB2) based on transfer learning is proposed for accurate fire detection from images. This model implements an online data augmentation strategy encompassing random rotation and horizontal flip during the data training phase. According to this perspective, the precise count of altered data samples during the training procedure remains unspecified. The results show that the proposed model outperforms the standard ENB2 model with an accuracy rate of 99.04% compared to 98.57%. Additionally, the proposed model provides better localization and representation of fire images.

Keywords: Fire detection, Online data augmentation, Convolutional neural network, Transfer learning, Efficient-NetB2, Grad-CAM.

1. Introduction

Forest fires have the potential to cause environmental disasters and can result in endangering the lives of humans and animals and economic damage. Early detection of fires in forests is crucial for preserving natural resources because once a forest fire spreads to a large area, controlling it becomes extremely difficult and can lead to catastrophic consequences. The studies on fire detection [1] can be broadly categorized into traditional and automated surveillance systems. Traditional surveillance systems mainly consist of thermal sensors, smoke sensors, and lookout towers, which are used for early identification and detection of fires[2]. Deploying these sensors in wide areas is costly, and the aggregation and transmission of their information are usually time-consuming. Moreover, they do not provide precise information about the location of the fire.
Surveillance cameras are expected to fully cover the desired areas in the future. Fire detection using image processing by these cameras can effectively identify fires with minimal cost. To this end, machine learning methods, as a branch of artificial intelligence, are being utilized [3, 4]. These methods can also be used as components in other systems based on new technologies for fire prediction and detection. This trend is towards integrating artificial intelligence with wireless network-based systems to enable automatic fire prediction and detection. Some of the methods used for fire detection include neural networks, regression, decision trees, support vector machines, Bayesian, fuzzy, etc. [5, 6]. The main general problem of this research is how to employ deep learning methods to accurately detect fire by processing surveillance images.

Deep learning methods can process large-scale data and can be applied to various fields such as natural language processing[7], big data analytics [8, 9], object recognition in disaster situations[10], and speech recognition[11], and can be applied to fire detection and prediction in forests [12]. With the emergence of convolutional neural networks (CNNs) and their successful applications in various domains, researchers are using this structure in image analysis. In recent years, the use of CNNs has experienced remarkable growth. CNNs are highly efficient in image classification and machine vision systems [13]. They can be used to improve the accuracy of fire detection and minimize the collateral damage caused by fires. These networks can automatically learn and extract complex features from images effectively [14].

One of the recent effective networks based on CNN in fire detection is called the EfficientNet B0 network [15]. This network has been successfully applied to other tasks such as COVID-19 detection [16], prostate detection [17], analysis of fish feeding behavior[18], and leaf disease classification [19]. The EfficientNet model incorporates a novel scaling method for convolutional networks, which uniformly scales the depth, width, and resolution using a straightforward yet highly effective factor. The specific research question is how the EfficientNet B0 network can be used to accurately detect fire from surveillance pictures.

1.1 Motivation

Today, fires are commonly seen all over the world, and especially forest fires bring great damage to people. More than 10,000 kilometers of forest areas in Europe are burned every year. Therefore, there is a need to use an effective early detection model for fire detection. In this research, a machine-vision-based model for fire detection is proposed and presented[1]

In past research, various transfer learning methods have been used to facilitate fire detection [15, 20, 21]. The main motivation of our proposed model is to use the hybrid structure of an end-to-end neural network under the name of transfer learning based on Efficient Net B2 (3ENB2). Also, this learning model effectively performs feature extraction, feature selection, and classification and has acceptable performance. During the training process, we use online data augmentation techniques in terms of Random Rotation and Horizontal flip. These two techniques can help the model to learn different patterns of fire and smoke in different environmental conditions and thus provide better performance in fire and smoke detection. In addition, the Grad-Cam technique is used to better evaluate the proposed model and identify and locate the important areas of the image. As a result, the detection accuracy rate using the proposed end-to-end model has reached 99.04% compared to the latest results.

1.2 Contribution

In summary, the contributions of this article are as follows:

- The use of online data augmentation techniques to increase the efficiency of the Effi-cientNetB2 network.
- The use of the Grad-cam technique in localizing the detection of input images.
- The elimination of the need for random neuron removal in the convolutional neural network training process.
- The high efficiency of the EfficientNetB2 model in preventing overfitting and increasing the accuracy of fire detection.

The subsequent sections of the article are outlined as follows: Section 2 provides an overview of the related work. Section 3 delves into the explanation of the proposed methodology. The evaluation of the model is discussed in Section 4. Section 5 addresses the conclusion and prospects for future work.

2. Related work

Among various research efforts, machine vision and deep learning-based methods have been used for automatic fire detection. However, the fire detection problem has become a challenging task due to fire similarity to natural objects like sunlight and other existing sources of illumination. In this section, recent works on fire detection from the perspectives of machine learning and deep learning are investigated.

Dogan et al.[22] Proposed two pre-trained ensemble models based on ResNetV1 and ResNetV2 for fire detection. They also utilized four pre-trained ResNet-based networks for feature extraction, and the Support Vector Machine for feature classification. A dataset, including 865 fire images and 875 non-fire images, was used to evaluate their models. As a result, the accuracy achieved for the two models was 98.91% and 99.15%, respectively.

Najab and Khan [23] presented a comprehensive method for extracting features from high-resolution satellite images. This method aims to develop a system that can automatically identify the location of settlements in large-scale images. Additionally, Principal Component Analysis (PCA) was employed to extract features from the images. In their study, Euclidean norm-based classification was used to separate the two classes of settlements and non-settlements using the PCA-extracted features. The images used were collected from various locations by Google Earth. The large images were then cropped into smaller windows with dimensions of 80x80 and 40x40 pixels to match the dimensions of the training images. The results demonstrate that the proposed approach can effectively detect settlements in high-resolution satellite images with an accuracy of 96.43%. In the study by Vikram and Sinha [24], a combined model based on fuzzy neural networks and convolutional neural networks was used for fire identification. In this approach, the forest is divided into multiple regions, each equipped with sensors and cameras. The sensors record the temperature, humidity, drought, and moisture status of the region. The fuzzy neural network model utilizes sensor data for data classification, while the convolutional neural network model utilizes the image data collected by the cameras for fire detection. The combined model achieved an accuracy of 82.29%. Jadon et al [25], used a MobileNet v2 architecture based on convolutional neural networks for fire detection. The utilized dataset consisted of 62690 images and 14 fire videos, along with 16 non-fire videos. They achieved an accuracy of 99%. Ghosh and Kumar [1] proposed a combination of convolutional neural networks and recurrent neural networks for fire detection. They also used fully connected layers for the final classification. In the first stage, preprocessed images are fed into the CNN, which extracts the final feature map of the image. Then, the final features were passed as input to the RNN. The CNN extracts features from different low-level and high-level layers, while the RNN extracts dependent and sequential features. The performance evaluation was carried out on two different public datasets, Mivai and Kaggle. The results of the study showed an accuracy of 99.62% on the Mivai dataset and 99.10% on the Kaggle dataset.

Khan and Khan[21]proposed a transfer learning approach based on CNN called MobileNet V2 for the early detection of forest fires in smart city applications. They evaluated their model on a dataset consisting of 1900 images, with 950 fire images and 1950 non-fire images, and achieved a classification accuracy of 98.42%. In another study[6], a fire detection system for classifying fire and non-fire images was proposed. The authors used a dataset called DeepFire, which included 1900 images, with 950 images for each class. Furthermore, for fire classification, they used machine learning algorithms, including K-nearest neighbors, Bayesian network, random forest, support vector machine, logistic regression, and transfer learning based on VGG19. The results showed that the transfer learning method performed better than other methods, with a detection accuracy of 95.70%. Besides, Majid et al [15], presented a CNN-based fire detection approach called EfficientNet B0. The authors combined five different datasets, consisting of a total of 7977 images, with 3988 fire images and 3989 non-fire images. The Grad-Cam method was also applied for localizing fire. Additionally, three deep learning models including VGG16, Google Net, and ResNet 50 were examined, and the highest accuracy was achieved by the EfficientNet B0 model with a rate of 95.40%.

In the study by Xu et al [26], a group learning approach based on convolutional neural networks (CNN) was proposed for fire detection. The Yolov5, Efficient Det, and EfficientNet models were suggested for fire detection. Among these methods, the Yolov5 and EfficientDet models were integrated in parallel for fire detection and showed the highest accuracy in evaluation. However, they have limitations in object detection capability, as the detector may mistakenly identify objects such as sunlight or fire. Therefore, these two methods do not consider the whole image. To address this issue, the authors considered the EfficientNet model, which makes decisions based on the results of the three aforementioned detectors. This effectively improves the detection accuracy and can accurately identify fire-like objects and other objects present in the image. Therefore, the EfficientNet model was trained on a dataset consisting of 10,581 images, including 2,976 fire images and 7,605 non-fire images. Furthermore, the results showed that the EfficientNet model achieved an accuracy of 79.7%.

Saeed et al [27], have proposed a three-stage learning method for fire detection. The first step is a combined neural network model called Multilayer perceptron (MLP) Adaboost. This model utilizes separate sensor data such as smoke, heat, and gas to predict fire detection. In the second stage, they used the Adaboost-local binary pattern (A-LBP) and convolutional neural network for fire detection from images and videos collected by surveillance cameras. The authors evaluated their method on a dataset consisting of 7,303 images, including 5,248 fire and 2,055 non-fire images. The results achieved an accuracy of 99%.

A dual deep-learning system for fire detection has been proposed[28]. The scheme uses complex deep neural network architecture. The first deep learning framework extracts features from the image, including background analysis, color, edge, and smoke. Simultaneously, the second deep learning framework focuses on motion-based features. Ultimately, these features are used to train a deep CNN to extract motion-based features. Both sets of features are used for classification using a Support Vector Machine. The authors evaluated their method on a dataset consisting of 20,000 images and reported a detection accuracy rate of 97.49%. Rodrigues and Riva [29] used machine learning methods based on random forest (RF), boosting regression trees (BRT), and support vector machines (SVM) for fire detection. The results showed that using each of these machine learning algorithms led to improved accuracy in terms of the area under the Curve (AUC) compared to the LR outputs. As a result, the AUC values suggest that Random Forest (RF) and Boosting Regression Trees (BRT) are more suitable methods with AUC values of 0.746 and 0.730, respectively. Additionally, the Support Vector Machine (SVM) model had a higher AUC value compared to Logistic Regression (LR).

Hong et al [30], Provide a comprehensive review of fire detection methods and techniques using transfer learning. Among the topics discussed in their paper are the diversity of training data, and up-to-date techniques to increase the

efficiency of fire detection. The paper is of great importance as a guide for future research in the field of fire detection and prediction and helps to develop more efficient and accurate methods for fire management and prediction.
A summary of related works on fire image detection is shown in Table 1.

Table 1. Related works on fire image detection using deep learning methods

Reference	Base Model	Number of Samples	Performance Metrics				
			Accuracy	**Precision**	**Recall**	**F1-score**	**Loss**
Xu et al. **[26],** 2020	Efficient Net	10581	79.7%	NC	NC	NC	NC
Liand zhao **[31],** 2020	Yolov3	13400	99.63%	NC	NC	NC	NC
Majid et al. [15] 2021	EfficientNet B0	7977	95.45%	91.77%	97.61%	94.76%	NC
Dogan at al. [22]**,** 2022	ReisNetv1	1650	99.25%	99.30 %	NC	99.24%	NC
Khan and Khan [21] , 2022	MobileNetV 2	1900	98.42%	97.42%	99.47%	98.43%	NC
Ghosh and Kumar [1]**,** 2022	CNN	94000	99.62%	NC	NC	NC	NC
Khan at al. [6]**,** 2022	VGG19	1900	95.00%	95.71	94.21	94.96	NC

*ACC, PRE, REC, F1, and NC indicate accuracy, precision, recall, F1-score, not considered, respectively.

Based on the conducted studies [32-34], the initial methods for fire detection have been based on extracting motion, texture, shape, and color features from video and image data. Color features are usually extracted from different color spaces such as RGB, HSV, and YCBCR. LSH is a filter that represents the spatial color information in an image using the values of L, S, and H. H represents the similarity between red, yellow, green, and blue, S represents the intensity of a color, and L represents the balance between white and black in the image. Additionally, the YCBCR filter is used for encoding color information in the image[35]. These methods are essentially based on extracting color models and applying them to the image. One of the disadvantages of these methods is the presence of color phenomena and moving objects with similar colors in video images, which reduces the detection rate and accuracy.

3. Research methodology

In this research, a combination of online data augmentation and EfficientNetB2 is used for fire image detection. The research methodology for fire detection and classification is illustrated in Figure 1. The proposed method consists of three stages: data description, data preprocessing, and classification.

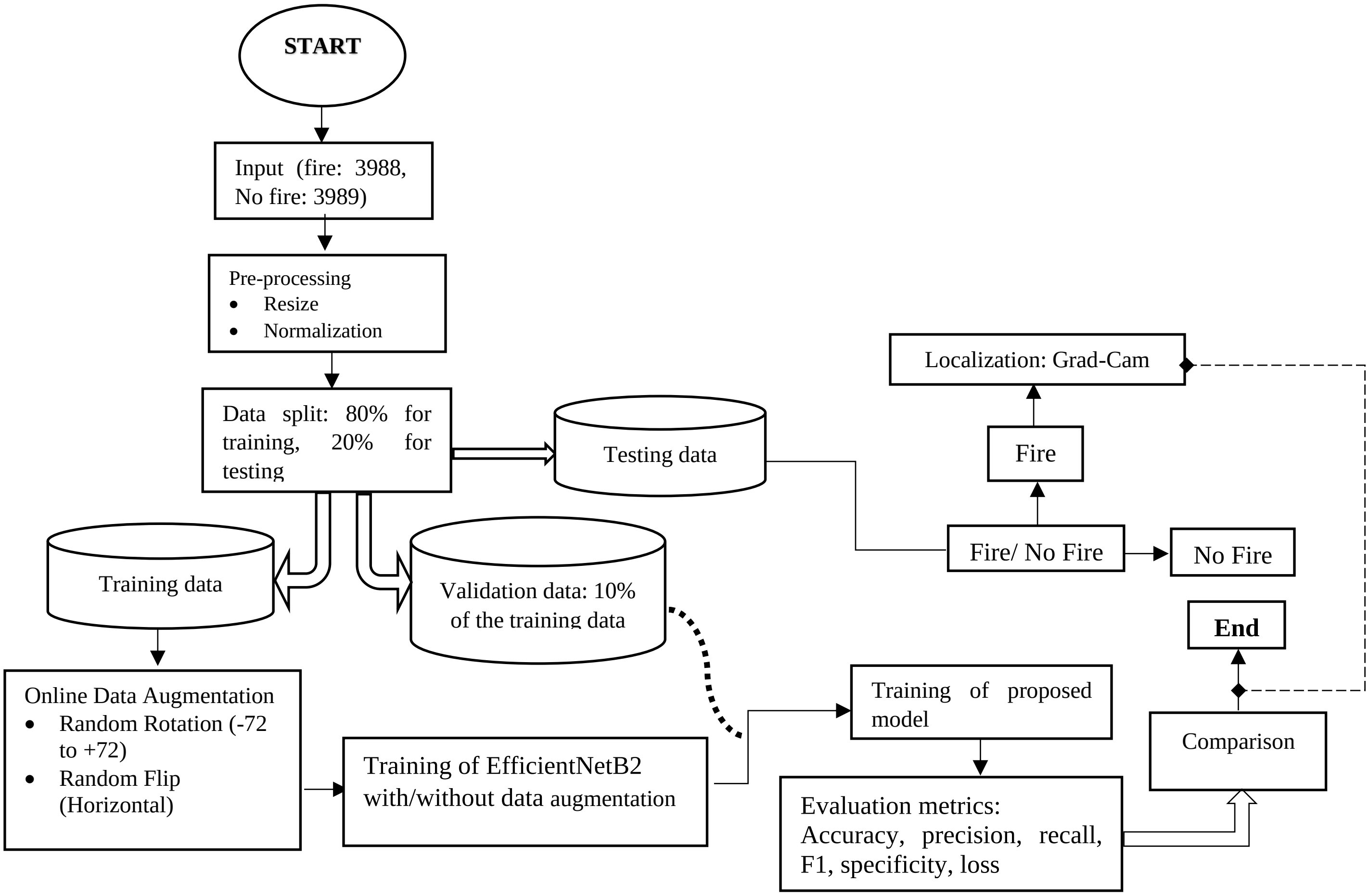


Figure 1. Research methodology

3.1. Data description

According to the conducted investigations regarding the fire image dataset, it was determined that there is a limited availability of samples. In other words, the accessible data samples were insufficient for achieving an efficient model. Therefore, we utilized a mixed dataset consisting of five databases related to fire and non-fire images to create a large generalized dataset and a robust model. This dataset includes images from Dataset-2021[1]; Deep Quest AI-2021[2]; Carlo-2021[3]; Bansal-2021[4]; and Saied-2021[5] [1, 15, 22]. These images encompass a range of environments, including urban streets, hallways, human figures, enclosed spaces, auditoriums, and woodland areas. Additionally, the aggregated dataset comprises diverse colored objects, which makes the fire detection task more challenging. Examples of fire and non-fire images are shown in Figure 2.

Datast-2021, Deep QuestAI-2021, Carlo-2021, and Saied-2021 datasets are more related to the characterization of fire and smoke, while the Bansal2021 dataset is broader and refers to more diverse images that include nature, buildings, and cities. However, this dataset can be generalized for projects related to fire detection [15]. The aggregated dataset includes a wide range of real environments including urban streets, corridors, human figures, closed spaces, halls, and forest areas.

1 https://www.kaggle.com/atulyakumar98/test-dataset
2 https://github.com/DeepQuestAI/Fire-Smoke-Dataset
3 https://www.kaggle.com/datasets/carlo946/fire-and-smokezip
4 https://www.kaggle.com/puneet6060/intel-image-classification
5 https://www.kaggle.com/datasets/phylake1337/fire-dataset/code

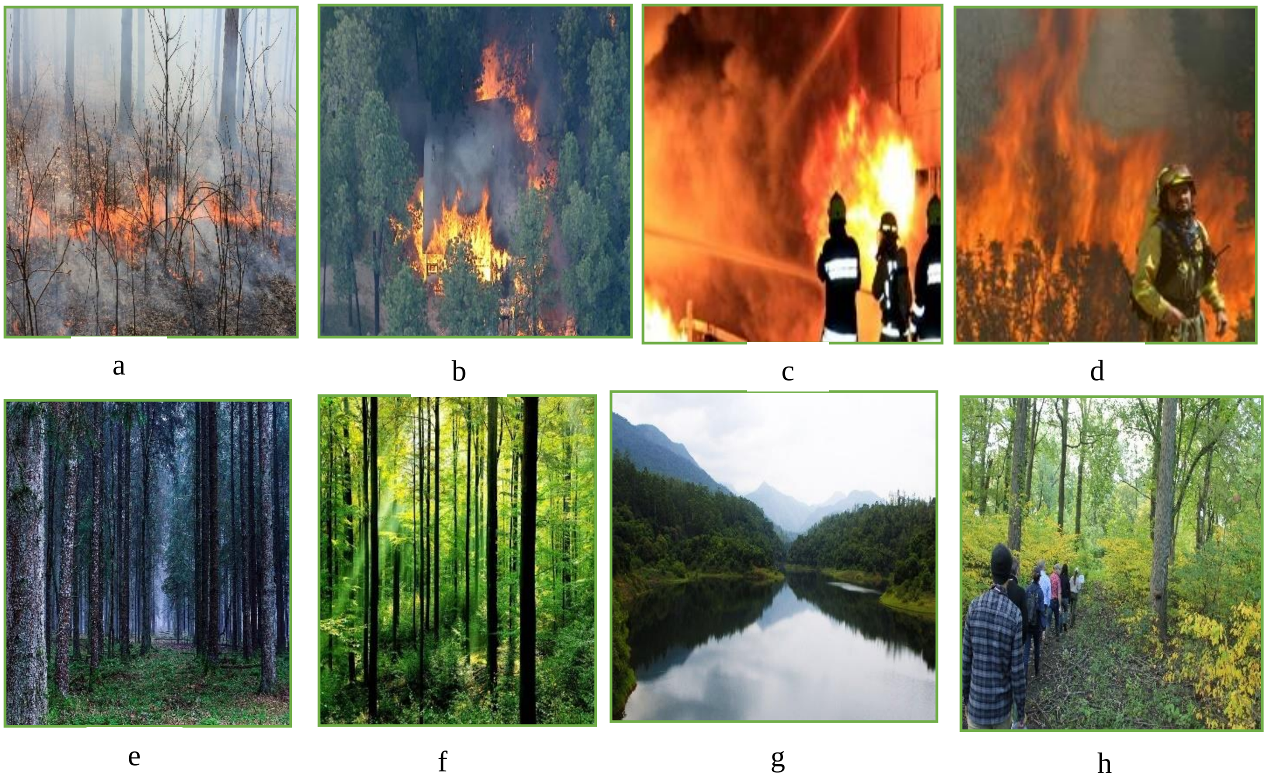


Figure 2. Sample images of fire (a-d) and no-fire (e-h)[15][6]

3.2. Preprocessing

The preprocessing stage is crucial for images that have noise and undesirable lighting situations, as neglecting this stage can have negative effects on image quality. In the preprocessing stage, the image resizing process is performed initially. The available images have different sizes, such as 1920×1080, 640×480, 120×720, etc. In this process, the dimensions of the image are reduced or increased vertically and horizontally. To standardize these sizes, we changed the size of the input images to 150×150. The size of 150×150 is relatively small for an image, which diminishes the computational cost required for image processing in the convolutional neural network (CNN), thus significantly reducing the computational burden on the CNN. Then, the normalization process is applied, which increases the accuracy of the neural network model and reduces the error rate for the network. The normalization method used is the min-max scaling, which transforms the intensity range of the images to [0,1], enhancing the quality of diverse colored images with different light intensities.

Finally, the online data augmentation method [36], was applied to the training dataset. Online data augmentation is a method in image processing in which data is changed in real-time during model training. The purpose of this work is to increase the variety of training data so that the model can deal with various features of the images and perform better. This method prevents overfitting and increases model performance against new and unknown data [37].

We used the online data augmentation technique from the perspective of random rotation and horizontal flip on the training space. In the online data augmentation method, the exact number of changed data samples under the effect of changes such as random rotation and horizontal flip is not determined during training. These changes are applied randomly while training the data. This continues dynamically during training. When a sample of the dataset is entered into the model, changes are randomly applied to it. For this reason, each new instance given to the model may be

[6] https://www.kaggle.com/datasets/phylake1337/fire-dataset/code

modified in a different way than the previous instances. This data enhancement feature of the online data augmentation makes the model familiar with a large variety of data.
We incorporated online data augmentation techniques in our fire detection model during training. Specifically, we applied horizontal random flipping and random rotation with a maximum angle of 0.2 radians. These augmentations help diversify the training dataset, enhancing the model's ability to generalize to different orientations and variations of fire instances. In general, the logic of these two online augmentation techniques is to see more diversity from the training angle standpoint. This makes the model familiar with different types of light conditions and viewing angles and improves its performance in fire and smoke detection.

3.3. Proposed Architecture (EfficientNet-B2)

The classifier used in this article employs a transfer learning technique called Convolutional neural network-based EfficientNetB2.
Convolutional neural networks (CNNs) are one of the most common methods in the field of computer vision. They are a type of deep learning network designed with multiple convolutional, pooling, and fully connected layers to process and extract important features from data with a network-like structure[38].
Transfer learning is a CNN-based method where pre-trained models are fine-tuned for a specific task. Transfer learning architectures have been developed to prevent overfitting and allow the neural network model to learn from a small subset of data rather than the entire dataset. This means that the created model retains less training data, resulting in an increased ability to generalize to new data[39], Recently, various transfer learning models, including EfficientNetB0, ResNet50, GoogleNet, and VGG16, have been used in fire detection[15].
The EfficientNet generation is based on a novel method for scaling CNN-based models. It uses a highly efficient compound coefficient that differs from other methods and allows for scaling network dimensions such as network depth, network width, and resolution. Increasing each of these features can improve network performance. These three features have a direct relationship with each other, as increasing resolution leads to more features being extracted for training, allowing for a deeper network [40, 41].
EfficientNet-B0 is a base network created using Neural Architecture Search (NAS) with the help of the AutoML MNAS automatic machine learning framework, which searches for a network. In general, EfficientNet is constructed using the Mobile inverted bottleneck convolution (MBConv) block [17, 42].
Therefore, in this paper, an end-to-end neural network model called 3ENB2, based on transfer learning with EfficientNet-B2, is proposed. As mentioned, this model utilizes compound coefficients ∅ to scale the width, depth, and resolution of the network, as described in Equation 1 [16, 19].

$$\begin{aligned} &Depth\ : d = \alpha^{\emptyset} \\ &Width\ : w = \beta^{\emptyset} \\ &Resoultion : r = \Upsilon^{\emptyset} \\ &s.t.\ \alpha \bullet \beta^{2} \bullet \gamma^{2} \approx 2 \\ &\alpha \geq 1, \beta\ "1, \Upsilon\ "1 \end{aligned} \tag{1}$$

The constants α, β, and ϒ can be derived through a small grid search to obtain optimal values. The coefficient ∅ serves as a scaling factor for the model. The computational efficiency of a regular convolution operation is directly proportional to the variables d, w^2, and r^2. By applying a scaling operation with a specific equation, the overall computational load increases, resulting in higher floating point operations per second (FLOPS). To ensure efficient computation, the value of $(\alpha \bullet \beta^{2}, \Upsilon^{2})^{\emptyset}$ is fixed at 2, which effectively sets the total FLOPS to $2^{\emptyset}$.
Simultaneously, a baseline network called Efficient-NetB0 is introduced, where the main network architecture is based on MBConv. Furthermore, the neural architecture is optimized to achieve both accuracy and FLOPS improvement. The resulting network is referred to as Efficient-B0. Subsequently, a compound scaling method is employed to upscale EfficientB0 in two steps. Initially, the value ∅ is set to 1, and a small grid search is conducted based on the aforementioned equation (1). The optimal values for EfficientNet-B0 are found to be $\alpha = 1.2, \beta = 1.1, \Upsilon = 1.15$ while adhering to the given constraint: $\alpha \bullet \beta^{2} \bullet \Upsilon^{\,2} \approx 2$. Secondly, the values of α, β, and ϒ are fixed, and EfficientNet-B0 is scaled up using equation (1) with different ∅ values, resulting in the creation of EfficientNet-B1 to B7.
Considering limited computational resources, EfficientNet-B2, pre-trained on ImageNet, is used as the base model structure for fire detections. The specifications of the end-to-end neural network model based on EfficientNet-B2 for transfer learning, referred to as 3ENB2, are described in Table 2.

Table 2. Specifications of the End-to-End Neural Network Model for Transfer Learning Based on EfficientNet-B2

Stage	Operator	Resolution	Channels	Layer
1	Conv 3×3	260×260	32	1
2	MBCon 1, k3×3	130×130	16	1
3	MBCon6, k 3×3	65×65	24	2
4	MBConv6, k5×5	33×33	48	2
5	MBConv, k3×3	17×17	88	3
6	MBConv6, k5×5	17×17	120	3
7	MBConv6, k5×5	9×9	208	4
8	MBConv6, k3×3	9×9	352	1
9	Conv1×1 & Pooling & FC	9×9	1408	1

One of the challenging tasks in improving the training process of convolutional neural network models is tuning the hyperparameters. The hyperparameter settings for the 3ENB2 are presented in Table 3. We have used the rectified linear unit (ReLU) activation function, which is a modified version of the linear activation function, between all layers except the last layer, where the sigmoid activation function is used. ReLU is a non-linear function that enhances the ability of the neural network model to transform non-linear data. The sigmoid function, on the other hand, is commonly used for predicting independent classes with probability values within the range [0, 1] [33]
.

Table 3. Hyperparameters for fine-tuning the proposed architecture.

Hyperparameter	Hyperparameter space
Parameter Model	EfficientNetB2
Input image Size	150× 150
Activation function	Sigmoid, Rectified linear unit (ReLu)
Regularization used	L1,L2
Dropout Rate	[0.2, 0.4, 0.6]
Batch Size	100
Optimizer	Adam
Learning Rate	0.001
Loss function	Binary cross-entropy
Epochs	30
Test proportion	20%

4. Evaluation

This section discusses the experimental setup, evaluation metrics, and evaluation results of the proposed model. All experimental setups were performed on a computer with the following specifications:

- 16GB Of GPU -GPU: Nvidia Tesla P100
- 13GB Of RAM
- CPU: Intel(R) Xeon(R) CPU @ 2.00GHz
- The proposed model was implemented in TensorFlow, and Adam was used as the optimizer.

After building the proposed system, classification time is important to be validated. Classification or prediction time refers to the state when the trained model is run on previously unseen data to predict the result. This step includes running the model on the unseen data and producing the final result based on the specified criteria. This period is provided below. The results are calculated as the average of 100 runs on the Nvidia P100 GPU.

- Average Inference Time per Batch of 64 samples: 0.08986 seconds
- Average Inference Time per sample in a batch of 64 samples: 0.00140 seconds
- Average Inference Time per sample: 0.04767 seconds

According to this information, the average inference time for each sample is negligible. It is very fast for this application and shows that the model can be suitable for even real-time applications. Indeed, a processing time of less than 0.1 seconds is considered suitable for many real-time applications.

4.1. Evaluation metrics

Evaluation metrics such as accuracy, precision, recall, specificity, f1 score, AUC, and loss are used for evaluation. The model's performance in terms of these metrics is determined by a confusion matrix. This matrix includes true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). The criteria are defined as follows [15, 22, 43].

(2) $$Accuracy = \frac{TN+TP}{\mathrm{TN+FP+TP+FN}}$$

(3) $$Precision = \frac{TP}{\mathrm{FP+TP}}$$

(4) $$Recall = \frac{TP}{\mathrm{TP+FN}}$$

(5) $$F1-score = 2\frac{\mathrm{precision.Recall}}{precision + Recall}$$

(6) $$Specificity = \frac{TN}{\mathrm{TN+FP}}$$

(7) $$Loss = -\frac{1}{n}\sum_{i=1}^{n}[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$$

4.2. Results

The compiled dataset consists of 7977 images of fire and non-fire. Out of this number, 3988 images are fire, and 3989 images are non-fire [15]. The non-fire images contain images of nature, such as forests, trees, grass, rivers, people, misty forests, and lakes. The details of the image dataset are described in Table 4. In addition, the data related to two classes of fire and non-fire are divided into 80% for the training process and 20% for the testing process. Also, 10% of the training data is allocated for the validation process.

The 3ENB2 was trained multiple times on the fire image dataset based on the available parameter settings to determine the best model. The comparison results of the proposed model and previous models are shown in Table 5**.** Based on these results, increasing the accuracy of the model by using online data addition shows that the model can improve its performance by receiving new data. Also, the proposed model shows the best performance in response to the variability and changes in data and conditions compared to the model without using online data augmentation. Indeed, by increasing the variety of training data, the model can learn deeper patterns and improve its performance. Similarly, online data augmentation can prevent overfitting and increase model stability because the model experiences improvements that make it more stable against changes in the input data.

Table 4. Details of the Fire Image Dataset.

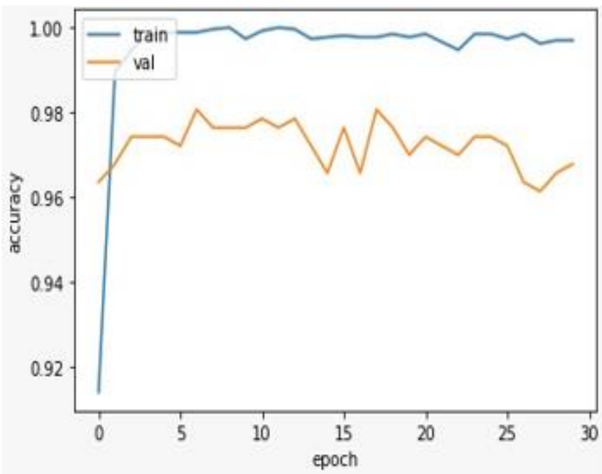
Dataset	Fire	No Fire	Total
Train	3417	3417	6834
Test	571	572	1143
Total	3988	3989	7977

Table 5. Comparison results of the proposed model with the previous model

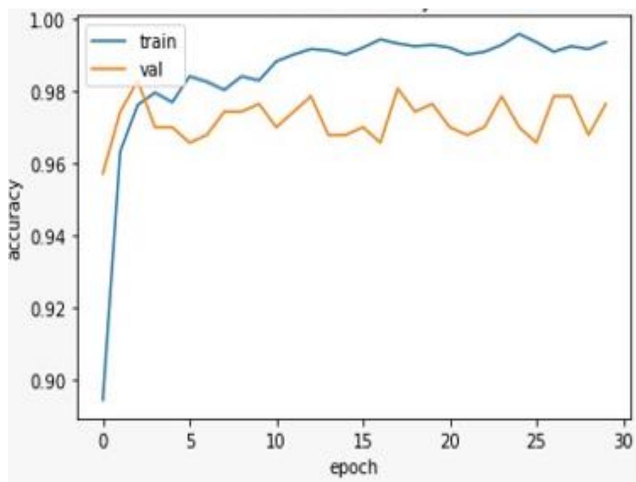
Models	Accuracy	Precision	Recall	F1-score	Specificity	AUC	Loss
Efficient Net B0 [15],	95.40	91.77	97.61	94.76	-	-	-

EfficientNetB2	98.32	98.27	98.51	98.39	98.12	-	0.0751
without Augmentation	98.57	98.61	98.67	98.63	98.46	-	0.0643
EfficientNetB2	98.71	99.00	98.51	98.71	98.91		0.0454
With online Augmentation	**99.04**	**99.51**	**98.86**	**99.18**	**99.23**	**97.67**	**0.0436**

Additionally, the accuracy and loss metrics of the end-to-end neural network model based on transfer learning using EfficientNetB2 (3ENB2) are shown in Figures 4 and 5. In Figure 3, the accuracy rate is considered in the (80-20) split ratio. The accuracy of our training data starts at around 80% and increases as the number of training epochs progresses, reaching an accuracy of 99.04%. The blue and orange lines represent the training and validation processes over 30 epochs, respectively. It can be observed that the best results were obtained using the 3ENB2 model based on EfficientNetB2. In Figure 4, it is observed that as the loss on the training data approaches zero, there is less error in recognizing the input images. The
validation data should follow the training curve. If the gap between these two curves increases, the network tends to overfit.



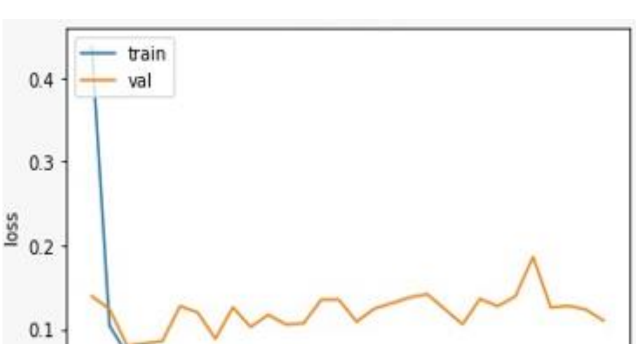
(a)

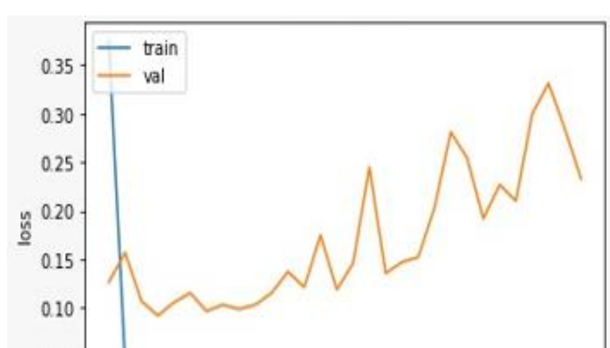


(b)

Figure 3. Accuracy curves, (A) DA-ENB2 Model and (B) ENB2 Model.

The designed predictions were made using a method that extracts information from the test dataset of the end-to-end neural network model based on transfer learning using EfficientNetB2 (3ENB2). Furthermore, the Grad-CAM method is one of the most famous and popular techniques developed for interpreting and representing important regions of images to evaluate the created neural network. This method is specifically designed for networks used in image classification tasks. The Grad-CAM method identifies influential regions in the network's output by examining the backpropagation of gradients. By applying these gradients to the weights of the convolutional layers of the network and finally combining them with the feature map output from the last layer, a heat map is obtained. This heat map indicates which parts of the image had the most influence on the network's decision-making [44]. Figure 5 displays examples of images analyzed using Grad-CAM. The feature maps on the images indicate which parts of the image were highly influential in predicting the desired class by the neural network. Additionally, the highlighted regions on the images pinpoint specific areas where the model accurately detected fire. However, according to the DA-ENB2 model, Figure 6 reveals instances where the model produces incorrect results when using Grad-CAM on the dataset. The highlighted areas on the images may indicate points where the model mistakenly detected fire, leading to false positives. These occurrences could be due to unintended changes in the images or similarities between specific image regions that misguide the model. In summary, Grad-CAM serves as an interpretive tool to enhance understanding of the performance of neural network models. It can assist researchers and developers in gaining a deeper insight into the intricacies of neural networks, ultimately contributing to the improvement of results.





(a)

Figure 4. Loss curves, (A) DA-ENB2 Model, and (B)ENB2 Model.

Figure 6. GRAD-CAM visualization of the test images, employing the DA-3ENB2 model

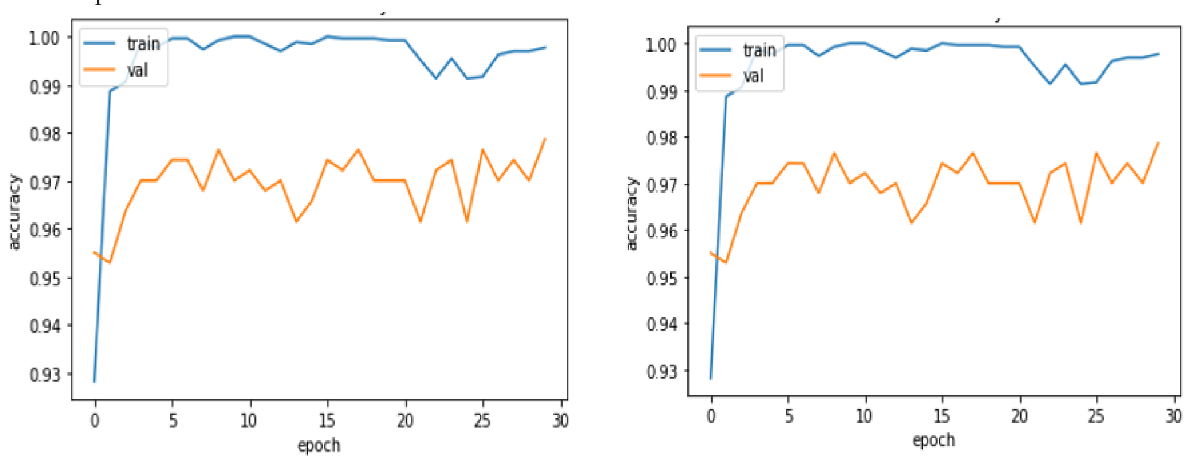
4.3. Effect of noise

To investigate the effect of noise on the performance of the proposed model, we introduced Gaussian noise with a standard deviation of 10 to the images in the input scale of 255 RGB. While noise was added in the preprocessing step for the validation set, it was dynamically applied to the training set during each training session, generating new noise patterns to help better generalize during model training [45]. The results with online data augmentation and without data augmentation are shown in Table 6.

Table 6. Comparing the influence of noise with online data augmentation and without data augmentation

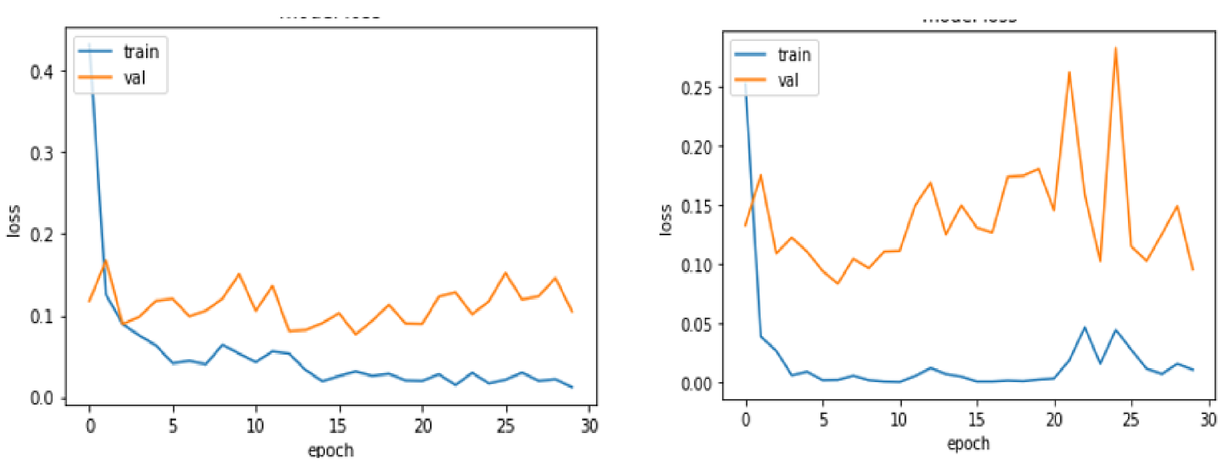
Models	Accuracy	Precision	Recall	F1-score	Specificity	AUC
EfficientNetB2 without Augmentation	98.7%	98.26%	98.01%	98.14%	98.12%	98.07%
EfficientNetB2 With online Augmentation	98.45%	98.75%	98.26%	98.14%	98.66%	98.46%

The accuracy and loss criteria of the proposed model are specified in Figures 7 and 8. In Figure 7, the accuracy division ratio of the training data starts from approximately 80% and increases with the progress of the number of training courses until it reaches an accuracy of 98.45%. Overall, the accuracy of the model improves significantly with the addition of Gaussian noise. This shows that the model is robust against noise. In addition, it can be seen in Figure 8 that as the loss of training data gets closer to zero, the error in recognizing the input images decreases. Blue lines represent the model losses during the training process and orange lines represent the values of losses during the validation process.



(a) (b)

Figure 7. Accuracy for (a) DA-ENB2 model and (b) ENB2 model.



(a) (b)

Figure 8. Loss for (a) DA-ENB2 model and (b) ENB2 model

4.4. Ethical considerations

Deploying fire detection systems in smart environments offers extensive benefits for safety and public health. However, ethical concerns and societal implications should also be considered. These concerns include preserving privacy, avoiding bias in data collection, and understanding the social implications of this technology. Therefore, it is essential to maintain ethical considerations in research to address potential privacy concerns, data collection bias, and social implications.

One of the main concerns is the collection and use of data related to people's movements and activities in smart environments. Fire detection systems can collect large amounts of sensitive data, including images, and location data. This data can be misused to identify and track people, create behavioral profiles, and even predict their future behavior. Using this data without the consent or knowledge of people can violate their privacy and lead to abuse. Strong privacy mechanisms must be adopted to protect the data collected by fire detection systems [46]. These mechanisms should include transparency about how data is collected, used, and stored, as well as strong controls for accessing and sharing data. Also, another concern is data collection bias, where fire detection systems may accidentally collect data that overrepresents some areas over others. This bias can lead to false and discriminatory results. For example, if a fire detection system disproportionately focuses on certain areas, it may increase the rate of false alarms for people who live or work in those areas, while missing the fires in other areas. This can lead to unnecessary discomfort and anxiety, as well as the reduction of trust in the system.

To reduce bias in data collection, fire detection systems must be designed and implemented in such a way that they collect primary data from the entire population. This may include using random or purposive sampling methods to collect data as well as analyzing the data to identify and correct any biases. Therefore, the deployment of fire detection systems in smart environments can have wider consequences. For example, these systems can be used to monitor people's behavior and impose restrictions on their movements. This can lead to a loss of privacy and can also lead to an under-surveillance society where people feel less freedom [47, 48].

4.5. Comparison

We used various methods for comparison to assess the convolutional neural network's performance. The model parameters have been comprehensively illustrated in Table 7, indicating a notably larger number of trainable parameters for our proposed solution in contrast to other existing approaches addressing this issue.

Although the VGG network possesses greater trainable parameters than our approach, it fails to effectively identify fire occurrences in images. A comprehensive presentation of the comparative results concerning accuracy can be found in Table 8.

Table 7. Parameter details of the proposed model comparison with other models.

Base Model	Total	Trainable	Non-Trainable
FileNet Model[49]	649,182	649,182	0
VGG[50]	134,342,526	134,342,526	0
Mobile Net[51]	2,28,646	1,249,534	34,112
EfficinetNet B0 [15]	4,0964,378	4,051,802	44,576
EfficinetNet B2(proposed)	25,854,274	18,088,705	7,768,569

Table 8. The comparison between the EfficientNetB2 and related research on a similar dataset.

References Model	**Performance**						
	Accuracy	**Precision**	**Recall**	**F- Score**	**Specificity**	**AUC**	**Loss**
Gideon et al[49]	87.31%	89.59%	84.41%	88.92	N/R	N/R	N/R
Namozov and Manchuk[50]	75.15%	67.06	98.77	79.88	N/R	N/R	N/R
Dua et al[51]	87.60%	80.80%	98.77%	88.88%	N/R	N/R	N/R
Majid et.al 2021[15]	95.40%	91.77%	97.61%	94.76%	N/R	N/R	N/R
Proposed models (EfficientNetB2 with online data augmentation)	99.04%	99.15%	98.86%	99.18%	99.23%	97.67%	0,0454 %

*N/R: Not Reported

Based on Table 8, the performance of our proposed model using EfficientNetB2 with online data augmentation method achieved an accuracy of 99.04% and a recall of over 98.86%, which is higher than other studies.

5. Conclusion

Forest fires cause an imbalance in the ecosystem and pose a threat to the lives of humans, animals, and the environment. Early detection of fires is essential to addressing this problem. In this article, an end-to-end neural network model that uses transfer learning based on EfficientNetB2 (3ENB2) has been proposed for fire detection. Specifically, we implemented the transfer learning methodology through the process of refining the model derived from the renowned group of EfficientNet models. We also used online data augmentation techniques to modify images in terms of random rotation and horizontal flip and employed the Grad-cam method for better localization of the proposed model on images.

Although the detection accuracy of the proposed method is high, it is not absolute. As a result, there are rare cases that the model can not accurately predict. This limitation can be due to the insufficient variety of features in the training data, which does not match the features in the test image. In other words, the test images may have features that the model has not encountered during training, so it cannot recognize them correctly. The second limitation is that we need more data sets in real-world scenarios that are not currently available. The third limitation could be the higher computational cost of our model compared to traditional machine learning models. In future work, the proposed model is intended to be augmented and used for practical applications in fire identification and detection in real environments. Furthermore, other neural network architectures, such as the Generative Adversarial Network (GAN) algorithm, can be combined with the Genetic Algorithm (GA) as an evolutionary algorithm for fire detection.

Statements and Declarations

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical approval

Not applicable

Funding

Not applicable

Authorship statement

Ehsanullah Zia: Conceptualization, Software, Original draft preparation

Hamed Vahdat-Nejad: -Supervision- Methodology- Reviewing and Editing

Mohammad-Ali Zeraatkar: Software - Methodology

Javad Hassannataj Joloudari: Methodology- Reviewing and Editing

Seyyed-Ali Hoseini: Supervision

Code & data

https://github.com/mmdalix/fire-detection

References

[1] R. Ghosh and A. Kumar, "A hybrid deep learning model by combining convolutional neural network and recurrent neural network to detect forest fire," *Multimedia Tools and Applications,* vol. 81, no. 27, pp. 38643-38660, 2022.

[2] P. Dasari, G. K. J. Reddy, and A. Gudipalli, "Forest fire detection using wireless sensor networks," *International Journal on Smart Sensing and Intelligent Systems,* vol. 13, no. 1, pp. 1-8, 2020.

[3] F. Bu and M. S. Gharajeh, "Intelligent and vision-based fire detection systems: A survey," *Image and vision computing,* vol. 91, p. 103803, 2019.

[4] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR).[Internet],* vol. 9, no. 1, pp. 381-386, 2020.

[5] S. G. Kong, D. Jin, S. Li, and H. Kim, "Fast fire flame detection in surveillance video using logistic regression and temporal smoothing," *Fire Safety Journal,* vol. 79, pp. 37-43, 2016.

[6] A. Khan, B. Hassan, S. Khan, R. Ahmed, and A. Abuassba, "DeepFire: A novel dataset and deep transfer learning benchmark for forest fire detection," *Mobile Information Systems*, vol. 2022, 2022.

[7] D. W. Otter, J. R. Medina, and J. K. Kalita, "A survey of the usages of deep learning for natural language processing," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 2, pp. 604-624, 2020.

[8] S. Sohangir, D. Wang, A. Pomeranets, and T. M. Khoshgoftaar, "Big Data: Deep Learning for financial sentiment analysis," *Journal of Big Data*, vol. 5, no. 1, pp. 1-25, 2018.

[9] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146-157, 2018.

[10] I. Khan, Z. Guo, K. Lim, J. Kim, and Y.-W. Kwon, "Assessment of indoor risk through deep learning-based object recognition in disaster situations," *Multimedia Tools and Applications*, pp. 1-22, 2023.

[11] W. Algihab, N. Alawwad, A. Aldawish, and S. AlHumoud, "Arabic speech recognition with deep learning: A review," in *Social Computing and Social Media. Design, Human Behavior and Analytics: 11th International Conference, SCSM 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part I 21*, 2019: Springer, pp. 15-31.

[12] Z. Jiao *et al.*, "A deep learning based forest fire detection approach using UAV and YOLOv3," in *2019 1st International conference on industrial artificial intelligence (IAI)*, 2019: IEEE, pp. 1-5.

[13] S. Aslan, U. Güdükbay, B. U. Töreyin, and A. E. Çetin, "Deep convolutional generative adversarial networks for flame detection in video," in *International Conference on Computational Collective Intelligence*, 2020: Springer, pp. 807-815.

[14] A. Gómez-Ríos, S. Tabik, J. Luengo, A. Shihavuddin, B. Krawczyk, and F. Herrera, "Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation," *Expert Systems with Applications*, vol. 118, pp. 315-328, 2019.

[15] S. Majid, F. Alenezi, S. Masood, M. Ahmad, E. S. Gündüz, and K. Polat, "Attention based CNN model for fire detection and localization in real-world images," *Expert Systems with Applications*, vol. 189, p. 116114, 2022.

[16] M. M. A. Rahhal, Y. Bazi, R. M. Jomaa, M. Zuair, and F. Melgani, "Contrasting EfficientNet, ViT, and gMLP for COVID-19 Detection in Ultrasound Imagery," *Journal of Personalized Medicine*, vol. 12, no. 10, p. 1707, 2022.

[17] K. Ikromjanov *et al.*, "Region Segmentation of Whole-Slide Images for Analyzing Histological Differentiation of Prostate Adenocarcinoma Using Ensemble EfficientNetB2 U-Net with Transfer Learning Mechanism," *Cancers*, vol. 15, no. 3, p. 762, 2023.

[18] L. Yang *et al.*, "A dual attention network based on efficientNet-B2 for short-term fish school feeding behavior analysis in aquaculture," *Computers and Electronics in Agriculture*, vol. 187, p. 106316, 2021.

[19] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model," *Ecological Informatics*, vol. 61, p. 101182, 2021.

[20] A. Khan, B. Hassan, S. Khan, R. Ahmed, and A. Abuassba, "DeepFire: A novel dataset and deep transfer learning benchmark for forest fire detection," *Mobile Information Systems*, vol. 2022, pp. 1-14, 2022.

[21] S. Khan and A. Khan, "Ffirenet: Deep learning based forest fire classification and detection in smart cities," *Symmetry*, vol. 14, no. 10, p. 2155, 2022.

[22] S. Dogan *et al.*, "Automated accurate fire detection system using ensemble pretrained residual network," *Expert Systems with Applications*, vol. 203, p. 117407, 2022.

[23] A. Najab, I. Khan, M. Arshad, and F. Ahmad, "Classification of settlements in satellite images using holistic feature extraction," in *2010 12th International Conference on Computer Modelling and Simulation*, 2010: IEEE, pp. 267-271.

[24] R. Vikram and D. Sinha, "A multimodal framework for Forest fire detection and monitoring," *Multimedia Tools and Applications*, vol. 82, no. 7, pp. 9819-9842, 2023.

[25] A. Jadon, A. Varshney, and M. S. Ansari, "Low-complexity high-performance deep learning model for real-time low-cost embedded fire detection systems," *Procedia Computer Science*, vol. 171, pp. 418-426, 2020.

[26] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, no. 2, p. 217, 2021.

[27] F. Saeed, A. Paul, P. Karthigaikumar, and A. Nayyar, "Convolutional neural network based early fire detection," *Multimedia Tools and Applications*, vol. 79, pp. 9083-9099, 2020.

[28] A. S. Pundir and B. Raman, "Dual deep learning model for image based smoke detection," *Fire technology*, vol. 55, no. 6, pp. 2419-2442, 2019.

[29] M. Rodrigues and J. De la Riva, "An insight into machine-learning algorithms to model human-caused wildfire occurrence," *Environmental Modelling & Software*, vol. 57, pp. 192-201, 2014.

[30] Z. Hong, E. Hamdan, Y. Zhao, T. Ye, H. Pan, and A. E. Cetin, "Wildfire detection via transfer learning: a survey," *Signal, Image and Video Processing*, vol. 18, no. 1, pp. 207-214, 2024.

[31] P. Li and W. Zhao, "Image fire detection algorithms based on convolutional neural networks," *Case Studies in Thermal Engineering,* vol. 19, p. 100625, 2020.
[32] Y.-H. Kim, A. Kim, and H.-Y. Jeong, "RGB color model based the fire detection algorithm in video sequences on wireless sensor network," *International Journal of Distributed Sensor Networks,* vol. 10, no. 4, p. 923609, 2014.
[33] T. Celik, H. Demirel, H. Ozkaramanli, and M. Uyguroglu, "Fire detection using statistical color model in video sequences," *Journal of Visual Communication and Image Representation,* vol. 18, no. 2, pp. 176-185, 2007.
[34] P. Gomes, P. Santana, and J. Barata, "A vision-based approach to fire detection," *International Journal of Advanced Robotic Systems,* vol. 11, no. 9, p. 149, 2014.
[35] P. Messias, M. J. Sousa, and A. Moutinho, "Color-based superpixel semantic segmentation for fire data annotation," in *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2021: IEEE, pp. 1-7.
[36] Z. Tang, Y. Gao, L. Karlinsky, P. Sattigeri, R. Feris, and D. Metaxas, "OnlineAugment: Online data augmentation with less domain knowledge," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*, 2020: Springer, pp. 313-329.
[37] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data,* vol. 6, no. 1, pp. 1-48, 2019.
[38] Y. LeCun, K. Kavukcuoglu, and C. Farabet, "Convolutional networks and applications in vision," in *Proceedings of 2010 IEEE international symposium on circuits and systems*, 2010: IEEE, pp. 253-256.
[39] K. You, Y. Liu, J. Wang, and M. Long, "Logme: Practical assessment of pre-trained models for transfer learning," in *International Conference on Machine Learning*, 2021: PMLR, pp. 12133-12143.
[40] H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, "Classification of remote sensing images using EfficientNet-B3 CNN model with attention," *IEEE access,* vol. 9, pp. 14078-14094, 2021.
[41] Z. A. Khan, T. Hussain, F. U. M. Ullah, S. K. Gupta, M. Y. Lee, and S. W. Baik, "Randomly initialized CNN with densely connected stacked autoencoder for efficient fire detection," *Engineering Applications of Artificial Intelligence,* vol. 116, p. 105403, 2022.
[42] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, 2019: PMLR, pp. 6105-6114.
[43] L. Huang, G. Liu, Y. Wang, H. Yuan, and T. Chen, "Fire detection in video surveillances using convolutional neural networks and wavelet transform," *Engineering Applications of Artificial Intelligence,* vol. 110, p. 104737, 2022.
[44] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?," *arXiv preprint arXiv:1611.07450,* 2016.
[45] M. Mafi, H. Martin, M. Cabrerizo, J. Andrian, A. Barreto, and M. Adjouadi, "A comprehensive survey on impulse and Gaussian denoising filters for digital images," *Signal Processing,* vol. 157, pp. 236-260, 2019.
[46] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science,* vol. 347, no. 6221, pp. 509-514, 2015.
[47] L. Floridi *et al.*, "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations," *Minds and machines,* vol. 28, pp. 689-707, 2018.
[48] P. R. Wolpe, K. S. Rommelfanger, Drafting, and R. D. o. t. B. W. Groups, "Ethical principles for the use of human cellular biotechnologies," *Nature Biotechnology,* vol. 35, no. 11, pp. 1050-1058, 2017.
[49] A. Jadon, M. Omama, A. Varshney, M. S. Ansari, and R. Sharma, "FireNet: a specialized lightweight fire & smoke detection model for real-time IoT applications," *arXiv preprint arXiv:1905.11922,* 2019.
[50] A. Namozov and Y. Im Cho, "An efficient deep learning algorithm for fire and smoke detection with limited data," *Advances in Electrical and Computer Engineering,* vol. 18, no. 4, pp. 121-128, 2018.
[51] M. Dua, M. Kumar, G. S. Charan, and P. S. Ravi, "An improved approach for fire detection using deep learning models," in *2020 International Conference on Industry 4.0 Technology (I4Tech)*, 2020: IEEE, pp. 171-175.