fire

MDPI

Article

Leveraging Large Language Models for Enhanced Classification and Analysis: Fire Incidents Case Study

Eman H. Alkhammash

Department of Computer Science, College of Computers and Information Technology, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia; eman.kms@tu.edu.sa

Abstract: Fire detection and analysis have been a central focus of numerous studies due to their importance in potentially reducing fire's harmful impact. Fire detection and classification using artificial intelligence (AI) methods have drawn significant attention in the literature. These methods often tackle certain aspects of fire, such as classifying fire versus non-fire images or detecting smoke or flames. However, these studies lack emphasis on integrating the capabilities of large language models for fire classification. This study explores the potential of large language models, especially ChatGPT-4, in fire classification tasks. In particular, we utilize ChatGPT-4 for the first time to develop a classification approach for fire incidents. We evaluate this approach using two benchmark datasets: the Forest Fire dataset and the DFAN dataset. The results indicate that ChatGPT has significant potential for timely fire classification, making it a promising tool to complement existing fire detection technologies. Furthermore, it has the capability to provide users with more thorough information about the type of burning objects and risk level. By integrating ChatGPT, detection systems can benefit from the rapid analysis capabilities of ChatGPT to enhance response times and improve accuracy. Additionally, its ability to provide context-rich information can support better decision-making during fire episodes, making the system more effective overall. The study also examines the limitations of using ChatGPT for classification tasks.

Keywords: large language models; ChatGPT; fire classification; fire detection; fire incidents

Academic Editor: W. John Braun

Received: 20 November 2024
Revised: 16 December 2024
Accepted: 23 December 2024
Published: 26 December 2024

Citation: Alkhammash, E.H. Leveraging Large Language Models for Enhanced Classification and Analysis: Fire Incidents Case Study. *Fire* **2025**, *8*, 7. https://doi.org/10.3390/fire8010007

1. Introduction

Large language models (LLMs) are an essential advancement in the fields of natural language processing (NLP) and artificial intelligence. LLMs are trained on large amounts of data and are capable of understanding and producing human-like language. LLMs can interpret and generate text in numerous languages, summarize information, translate languages, compose creative material, answer queries, and even conduct logical reasoning tasks [1–5]. Several sectors have employed large language models in various applications [6–8]. These applications include content creation, customer service, teaching, education, and research [6–9]. LLMs are increasingly utilized in critical sectors such as healthcare. Healthcare is one of the sectors that has utilized models like GPT-3 and GPT-4, and numerous research has investigated the use of LLM for medical diagnosis, medical education, healthcare services, medical question-answering, dialogue summarization, medical education, and decision support [10–14]. These uses demonstrate the versatility of LLMs and across different domains.

It is crucial to note that LLMs have some limitations. There are ethical concerns about biased, harmful, or unethical results produced by these models. These models can produce results based on incorrect information, which may have unexpected consequences when used in real-world applications [15–18].

They often inherit biases from their training data, which can result in biased or harmful outcomes. Additionally, they require substantial computational resources, raising concerns about their environmental impact. LLMs are also known to occasionally generate factually incorrect or fabricated information, a phenomenon known as "hallucination". Data privacy is another important issue regarding the use of LLMs [16].

Ethical concerns are categorized into four main groups: Social Justice and Rights, which includes issues of fairness, social cohesion, and the risks of digital divides [18]. Individual Needs encompasses concerns around autonomy, safety, privacy, and informed consent. Environmental Issues focuses on sustainability and resource consumption, particularly due to the energy demands of AI models [18]. Finally, Cultural Identity addresses the influence of AI on cultural norms, values, and diversity [18].

Fire detection increasingly uses machine learning and deep learning to quickly identify fire and smoke from images and videos. These advanced models are effective for early warning systems and fire detection tools. The accuracy and performance of deep learning make it an essential part of improving fire control. Deep learning models have been widely adopted for fire classification, detection, and segmentation, and their combination due to their ability to recognize complex visual patterns effectively using satellite remote sensing data, fire images, videos, and geospatial and environmental data [19–27]. Different models and approaches are used for different tasks including the classification of images into fire and non-fire categories or the classification of fire and smoke. For instance, some models are designed for early wildfire while others focus on determining the level of risk by evaluating factors like fire intensity and spread potential, estimating severity, and predicting the spread of wildland fires and wildfire. These diverse applications help ensure accurate and timely responses and assist in minimizing risks to people and property. With the increasing need for efficient and accurate fire detection systems, leveraging advanced AI tools such as ChatGPT can offer new approaches to interpreting visual data, providing context to fire events, and assisting with fire classification tasks.

There is a lack of studies that investigate the effective integration of large language models (LLMs) with existing machine learning and deep learning methods for classification and detection tasks. This study aims to fill the gap in research concerning the application of LLMs in fire classification and analysis. By integrating LLMs, we seek to explore innovative approaches that could significantly improve the effectiveness of existing machine learning and deep learning methods in this critical area. This study investigates a new aspect of the potential applications of LLMs in image classification and analysis, an area that has not been explored in existing research. We aim to explore the potential of large language models for fire classification and analysis. Understanding the limitations of LLMs is crucial for enhancing fire detection systems such as biases in training data that can affect the accuracy of detection algorithms. However, this is out of the scope of this study.

The contributions of this study are as follows:

1. This study explores the capability of ChatGPT to analyze characteristics such as color, texture, and context in images to distinguish between different types of fire scenarios in order to offer responsive fire detection solutions and assist decision-makers.
2. The study highlights the potential of ChatGPT to provide more important information related to fire incidents. For example, ChatGPT can describe the objects that have fire and estimate the risk of the fires. These interpretations can assist in understanding fire incidents and enable better decision-making.

3. We also conduct a comparative analysis between ChatGPT, Gemini 1.5 Flash, Microsoft Copilot, and Poe frameworks to evaluate their performance in fire classification and analysis tasks.
4. We outline the lessons learned regarding the capabilities of ChatGPT, and the limitations for fire classification.
5. Although this approach was developed for fire classification, it can easily be adapted to tackle a variety of other classification tasks across various fields and classification problems.

The remainder of this paper is organized as follows: Section 2 covers related work, Section 3 details the methodology, Section 4 presents the results, Section 5 provides comparison between ChatGPT, Gemini 1.5 Flash, Microsoft Copilot, and Poe frameworks. Section 6 provides the discussion, and Section 7 concludes the paper.

2. Related Works

This section highlights several studies that explore various approaches for fire classification utilizing machine learning and deep learning techniques.

Barros-Daza MJ et al. [28] introduces an approach using a feed-forward artificial neural network (ANN) that aims to support real-time decision-making for mining firefighting personnel during underground coal mine fires. A total of 500 simulated fire scenarios that includes different fire sizes are used to train and test the model. ANN achieved high accuracy of 97% and a 96.7% weighted F1-score. The approach is applicable to similar settings like road tunnels.

Another study by Harkat Houda, et al. [27] uses Support Vector Machine (SVM) with a Radial Basis Function (RBF) kernel to classify fire and non-fire pixels. Datasets used for classification of fire and non-fire are Corsican, FLAME, and Firefront_Gestosa datasets. The information-theoretic feature selection was used to improve performance by reducing redundant features. The proposed model achieved an overall accuracy of 96.21%, sensitivity of 94.42%, specificity of 97.99%, precision of 97.91%, recall of 94.42%, and F-measure and G-mean values of 96.13% and 96.19%, respectively.

Balakrishnan, Vimala, et al. [29] develops a machine learning model to detect fatalities in structure fires. They used a dataset of 11,341 cases that covers the period from 2011 to 2019. Ten machine learning models were tested in the study. The main risk factors found are fires that start in bedrooms, cooking/dining areas, and living areas. The bedding-related fires show the highest fatality rate (20.69%) despite a low incident rate (3.50%). Random Forest (RF) achieved the best accuracy of 86%. The Decision Tree with bagging achieved 84.7% accuracy. Limitations include data quality and grouping of categories in the preprocessing phase.

Vorwerk, Pascal, et al. [30] uses two experiments, the first one uses Linear Discriminant Analysis (LDA) to predict four fire types (smoldering wood, smoldering cable, smoldering cotton, and candle fire) that achieved a classification rate of up to 69% and a Cohen's Kappa of 0.58. The second experiment applied the TrAdaBoost algorithm. Boosting increased from 1% to 30% for specific sensor positions and improved classification rates to 73% and Cohen's Kappa to 0.63. However, excessive boosting led to overfitting on certain sensors which resulted in reducing overall performance.

Wu, Weilin, and Yixiang Chen [31] aims to provide a trustworthy classification model by combining Bayesian Network (BN) and a trustworthy computing approach to classify fire risk in intelligent buildings. BN calculates the risk values of fire-related attributes from seven profiles. The trustworthy computing classifies the risk into five ranks, where higher ranks indicate the greater severity. The study also includes comparisons with three other assessment methods, highlighting the effectiveness of the proposed model.

Bashyal, Shishir et al. [32] proposes a fire-sensing system that aims to provide early fire detection and classification based on the smell of the smoke. The study adopts a sensor array and a neural network for pattern recognition. The neural network is implemented on a general-purpose microcontroller, creating a low-cost, effective fire classifier for real-life applications.

The proposed method of Islam, Al Mohimanul, et al. [33] utilizes a pre-trained EfficientNetB7 and a customized Attention Connected Network (ACNet) for the classification of forest fires. The Bayesian optimization is used for optimizing the model's hyperparameters. The proposed model achieves 97.45% accuracy, 98.20% precision, 97.10% recall, and 97.12% F1-score on the FLAME dataset, and 95.97%, 95.19%, 96.01%, and 95.54%, respectively, on the DeepFire dataset. The model also achieves a TNR of 95.5% and TPR of 99.3% on FLAME, and 94.47% and 96.82% on the DeepFire dataset outperforming many existing approaches. GRAD-CAM is employed to accurately localize fire within feature maps, demonstrating the model's effectiveness in wildfire detection even in low-activity regions.

3. Methodology

This study explores the potential of large language models (LLMs), in particular ChatGPT Plus based on the GPT-4 architecture (https://chatgpt.com/ (accessed on 1 December 2024)), for the classification of fire incidents. We propose a classification-based approach that consists of three main phases as can be seen in Figure 1. In the first phase, the images to be classified are input into the system. Second, prompts are used to guide the LLM (ChatGPT) through the classification process, systematically identifying the type of fire represented in each image. This iterative process continues until all images are classified. Finally, the results are evaluated to determine the accuracy and effectiveness of the proposed approach.

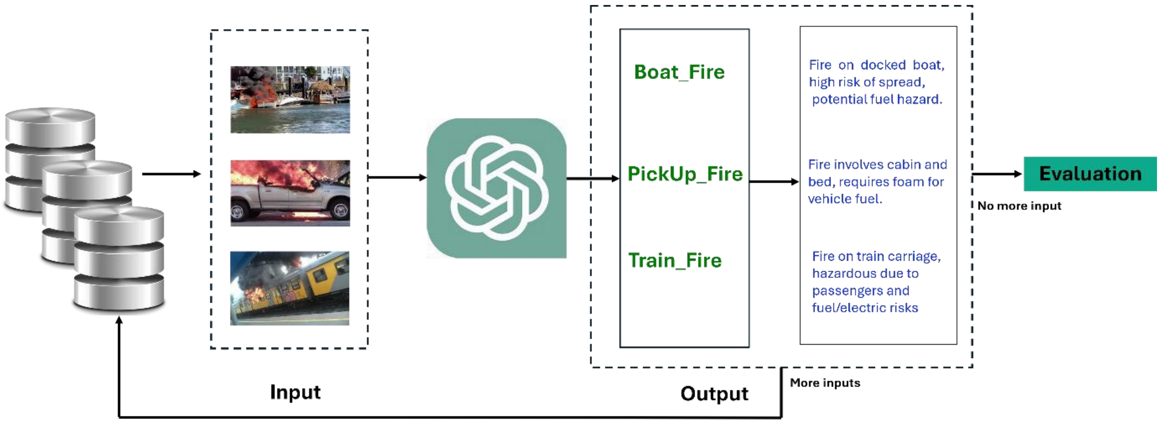

Figure 1. The classification-based approach using ChatGPT.

Our approach aims to leverage the contextual understanding capabilities of large language models (LLMs) to classify different fire types, which can be beneficial in various situations such as fire incidents.

3.1. Large Language Models and ChatGPT

LLMS are built on transformer architectures [34]. The transformer architecture uses a self-attention mechanism to help models grasp context in sentences. ChatGPT is an LLM that has progressed greatly over time, from early versions like GPT-2 (published by OpenAI) to more advanced versions like GPT-4 and ChatGPT. GPT-3 has 175 billion parameters, enabling it to produce highly cohesive and contextually relevant content [35].

The training of ChatGPT is an iterative process, where the model continuously improves as more data are introduced [36,37]. ChatGPT can be fine-tuned for specific applications, enabling it to perform various tasks such as language translation, content generation, and more [38]. To utilize ChatGPT, the user begins by constructing a prompt or question to input into the system. The model then processes this prompt based on its understanding of language patterns and relationships, generating a relevant response. Finally, the response is delivered to the user, who can continue the interaction by asking follow-up questions in an iterative manner. This approach is primarily trained using Reinforcement Learning from Human Feedback (RLHF) [6,36].

3.2. Datasets

3.2.1. Forest Fire Dataset

This dataset [37,39] designed for forest fire detection, contains 3-channel images with a resolution of 250 × 250 pixels. It supports binary classification into "Fire" and "No-Fire" and is balanced with 1900 images (950 per class). A specific "testing" folder holds 20% of the dataset, and all images from this folder were used as inputs for our study (380 images).

3.2.2. DFAN Dataset

This dataset [40] was collected from YouTube, Facebook, and disaster management agencies. It includes 3804 images spread across different classes: Boat fire (338), Building fire (305), Bus fire (400), Car fire (579), Cargo fire (207), Electric pole fire (300), Forest fire (480), Normal (97), Pick-up fire (257), SUV fire (240), Train fire (300), and Van fire (300). In this study, we randomly selected 10% (380 images) of the dataset as input for our approach.

3.2.3. Real-World Dataset

We captured 30 fire images and collected 191 non-fire images (https://www.kaggle.com/datasets/imankhammash/fireimagesdataset, accessed on 1 December 2024), representing various real-world fire and non-fire scenarios that are not available online. This new dataset is used to validate the performance of the proposed approach in effectively distinguishing between the two classes.

3.2.4. Prompt Engineering

There are three important tasks for using prompts in LLMs [41]. (1) We ensure that prompts are clear and specific, providing enough context for the model to understand the request. (2) Prompt engineering requires an iterative process. We start with an initial prompt, assess the response, and refine it based on the output. (3) We simplify the request as necessary to improve the results. Below are the prompts used in the study:

1. Classify the following images as "fire" or "not fire". Please include the image name along with the classification.
2. Please provide a detailed description of the image.
3. Please classify the following fires into the categories: Electric_Fire, Bus_Fire, Building_Fire, Train_Fire, Non_Fire, Van_Fire, SUV_Fire, PickUp_Fire, Boat_Fire, Forest_Fire, Car_Fire, and Cargo_Fire. Please include the image name along with the classification.
4. Classify the following fires based on: Risk Level, Severity, Confidence Score (accuracy of your classification), Affected Parts, and Estimated Burn Area.

3.2.5. Performance Metrics

We used four important metrics for evaluation of the classification: precision, recall, F1-score, and accuracy. The definitions of these metrics are as follows:

- Precision: This metric represents the ratio of true positives (TP) to the total number of positive predictions (TP + FP). It indicates how many of the predicted positive cases were actually correct:
$$Precision = \frac{TP}{(TP + FP)}$$
- Recall: measures the ability of the model to correctly identify all relevant positive instances. It is the ratio of true positives (TP) to the sum of true positives and false negatives (FN):
$$Recall = \frac{TP}{(TP + FN)}$$
- F1-score: The harmonic mean of precision and recall that provides a balance between the two metrics:
$$F1\ score = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}$$
- Accuracy: The metric measures of how often the model is making correct predictions overall, considering both true positives (TP) and true negatives (TN) in relation to the total number of predictions, which includes false positives (FP) and false negatives (FN) as well:
$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)}$$

4. Results

This section presents the results of the classification-based approach applied to fire datasets. The results for the Forest Fire dataset demonstrate exceptional performance, achieving 100% accuracy, recall, and F1-score, indicating perfect classification across all metrics. For the DFAN dataset, the approach also delivered strong results, with 99% accuracy, recall, and F1-score. These outcomes highlight the effectiveness of the classification approach based on ChatGPT in identifying fire-related images, achieving highly accurate performance on both datasets. Tables 1 and 2 present the evaluation results of the proposed method on two benchmark datasets: Forest Fire and DFAN. The support indicates (Fire, Boat-Fire, etc.) the number of true instances (actual images) of each class in the dataset. The macro average calculates the metric (i.e., precision, recall, F1-score) independently for each class and then takes the average. It treats all classes equally, regardless of their size or frequency. Whereas weighted average calculates the metric for each class and then averages them, it weights each class by its number of instances (support). This means that larger classes contribute more to the final score. Both macro and weighted averages highlight the outstanding ability of the proposed approach to accurately classify and detect both Fire and Non_Fire instances. In these datasets we do not have an imbalanced dataset; macro averages can highlight the performance of minority classes.

Table 1. Performance metrics of the proposed approach on the Forest Fire dataset.

	Precision	Recall	F1-Score	Support
Fire	1.00	0.99	1.00	190
Non_Fire	0.99	1.00	1.00	190
accuracy		1.00		380
macro avg	1.00	1.00	1.00	380
weighted avg	1.00	1.00	1.00	380

Table 2. Performance metrics of the proposed approach on the DFAN dataset.

Fire Type	Precision	Recall	F1-Score	Support
Boat_Fire	1.00	1.00	1.00	42
Building_Fire	1.00	1.00	1.00	28
Bus_Fire	1.00	1.00	1.00	41
Car_Fire	1.00	0.98	0.99	54
Cargo_Fire	1.00	1.00	1.00	24
Electric_Fire	0.98	1.00	0.99	43
Forest_Fire	1.00	1.00	1.00	35
Non_Fire	1.00	1.00	1.00	3
PickUp_Fire	0.96	1.00	0.98	24
SUV_Fire	1.00	1.00	1.00	22
Train_Fire	1.00	1.00	1.00	30
Van_Fire	1.00	0.97	0.99	34
accuracy		0.99		380
Macro avg	0.99	1.00	1.00	380
weighted avg	0.99	0.99	0.99	380

The confusion matrix for the classification-based approach for Forest Fire dataset and DFAN are demonstrated in Figures 2 and 3.

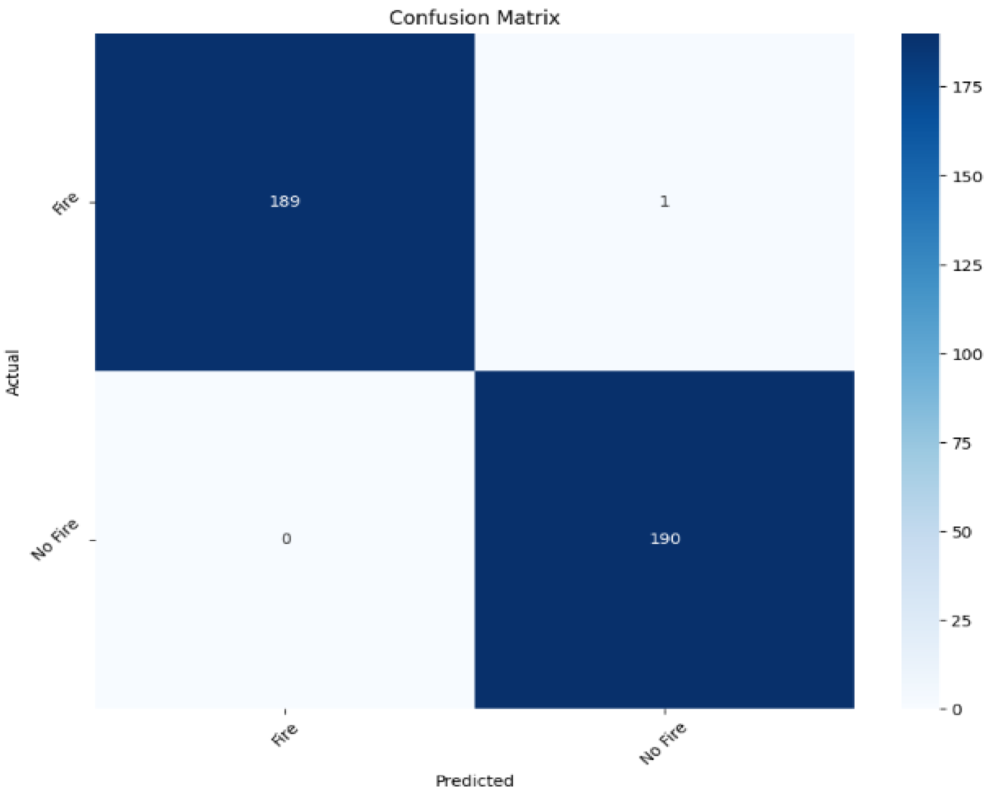

Figure 2. Confusion matrix for the classification-based approach on the Forest Fire dataset.

ChatGPT not only performs fire classification but also provides additional context and information about the classification results which can significantly aid decision-making. For example, it can identify specific characteristics of the fire, such as affected parts, intensity, potential risks, offering insights that help decision-makers prioritize actions and allocate resources more effectively. This added layer of detail enhances the fire detection system. "The images provided were classified through visual inspection based on key features, including the type of object, contextual cues, and fire characteristics. Each image depicted a clear instance of a fire event, allowing for categorization according to the visible object type and fire context. Contextual cues, combined with visible flames and smoke, helped confirm the classification. Based on these criteria, the images were classified appropriately based on their corresponding fire categories". *ChatGPT (November 2024 version)*. Figure 4 and Table 3 provide examples of the detailed information generated by the classification-based approach using ChatGPT for the Forest Fire dataset.

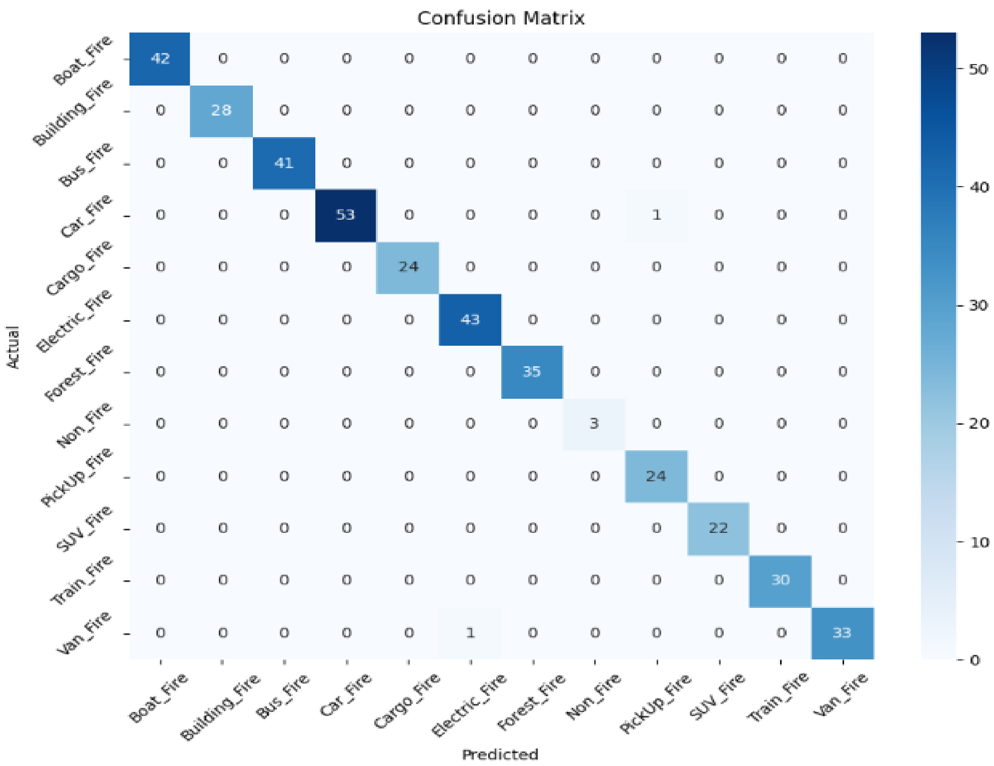

Figure 3. Confusion matrix for the classification-based approach on the DFAN dataset.

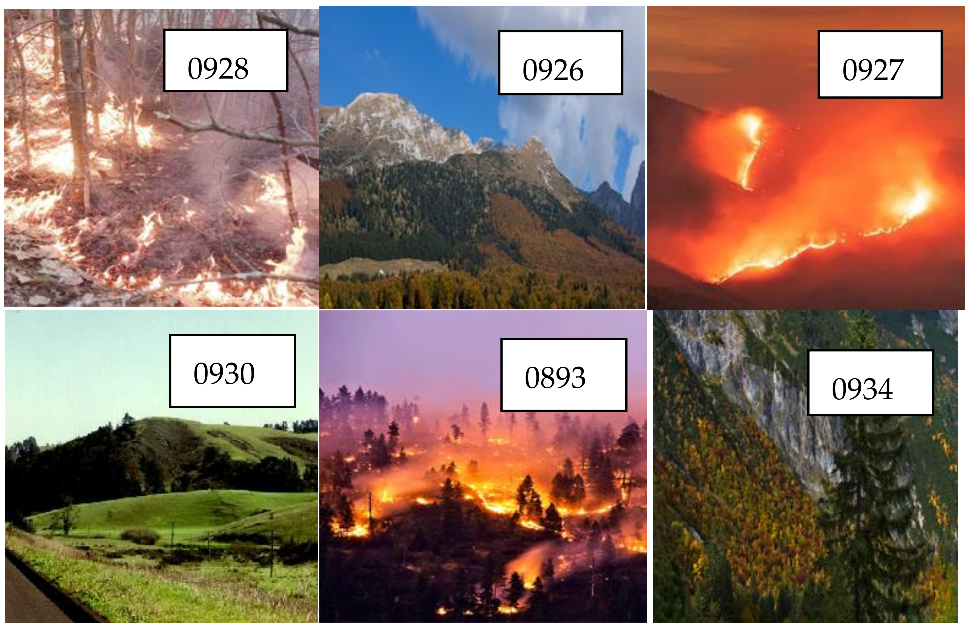

Figure 4. Sample of images from the Forest Fire dataset.

The confidence score added in Table 3 is used for assessing the classification and determining whether additional verification might be needed. The lower the score is, indicates that more checking is needed to verify the results. ChatGPT classified these images based on visual inspection of key fire indicators such as flame intensity, color contrast, smoke and surrounding environment. Any image displaying smoke or flames was classified as "Fire". Images classified as "No Fire" did not have any of these characteristics. For the DFAN dataset as shown in Figure 5 and Table 4, additional features such as object type, color characteristics (e.g., bright colors like red, yellow, or orange indicating flames), scene context (e.g., flames or visible damage to vehicles), and texture and shape (e.g., irregular shapes of flames or smoke) are utilized for better classification.

Table 3. Detailed classification information for selected images from the Forest Fire dataset using ChatGPT-4 (November 2024).

Image Name	Classification	Additional Information	Risk Level	Severity	Confidence Score	Affected Parts	Estimated Burn Area	Vegetation Type
0928.jpg	Fire	Moderate fire spread, visible flames with some smoke, affecting limited areas.	Medium	Medium	85%	Lower region	800 m²	Shrubland
0926.jpg	No Fire	Natural landscape, no signs of fire or smoke. Colors are mostly greens and browns.	Low	Low	95%	None	N/A	Grassland
0927.jpg	Fire	Active fire visible with flames and smoke. Bright colors (orange, red, yellow) suggest medium to high severity.	High	High	92%	Central and lower areas	1500 m²	Forest
0930.jpg	No Fire	Grassy or open landscape, no flames or smoke visible. Calm environment.	Low	Low	97%	None	N/A	Grassland
0893.jpg	Fire	Significant fire spread visible, flames and smoke affecting large areas. Bright colors and dense smoke indicate severe fire.	High	High	90%	Entire region	2500 m²	Mixed Forest
0934.jpg	No Fire	Landscape with changing leaf colors, no visible fire or smoke. Colors due to seasonal foliage.	Low	Low	96%	None	N/A	Forest

Figure 5. Sample of images from DFAN dataset.

Table 4. Detailed classification information for selected images from the DFAN dataset using ChatGPT-4 (November 2024).

Image Name	Fire Type	Details
063551.jpg	Building Fire	Fire on a high-rise building's balcony, with significant black smoke and visible flames.
113133.jpg	SUV Fire	Dark-colored SUV with fire coming from underneath, near the rear section.
19916.jpg	Electric Fire	Electric pole with flames, suggesting a fire caused by electrical malfunction.
27111.jpg	Toy Car Fire	Small toy blue pickup truck model set on fire, with flames coming from the roof area.
28539.jpg	PickUp Fire	Pickup truck engulfed in flames, mostly from the cabin and rear areas.
14453.jpg	Car Fire	Race car (National Guard #88) with flames near the rear right section, likely caused by mechanical issues.

5. Comparison Between ChatGPT, Gemini 1.5 Flash, Microsoft Copilot, and Poe Frameworks

In this section, we present a comparison between ChatGPT, Gemini 1.5 Flash, Microsoft Copilot, and the Poe framework, utilizing a new dataset containing 30 fire images and 191 non-fire images. All frameworks were able to classify these images with 100% accuracy. Interestingly, images with strong sunlight were not misclassified. Additionally, we observed that these frameworks provide very detailed insights about the images. For example, when inquiring about an image featuring a sparkle, the response from ChatGPT is as follows: "The image features a sparkler, which emits bright flames and sparks. While visually similar to fire, a sparkler is not considered a conventional fire class in most datasets". *ChatGPT (December 2024 version)*. Figure 6 displays sample images from the new dataset, while Table 5 presents the responses generated by ChatGPT (https://chatgpt.com/ (accessed on 1 December 2024)), Gemini 1.5 flash (https://gemini.google.com/ (accessed on 1 December 2024)), Microsoft Copilot (https://copilot.microsoft.com/ (accessed on 1 December 2024)), and Poe (https://poe.com/ (accessed on 1 December 2024)). Responses of these frameworks highlight their ability to deliver clear and detailed descriptions of the image, covering both its main features and contextual background.

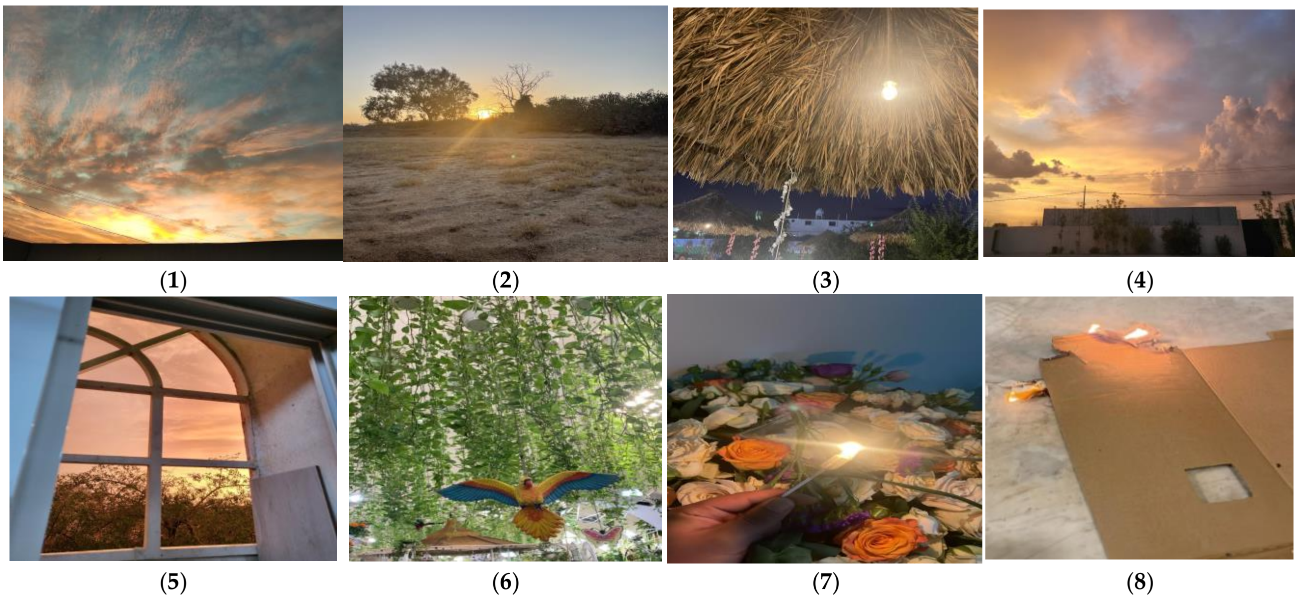

Figure 6. Samples of the images of new dataset (**1**) Sky-focused sunset, (**2**) ground-focused sunset, (**3**)Thatched roof with lightbulb, (**4**) Sunset sky and clouds, (**5**) Sunset through a window, (**6**) Indoor hanging plants and bird decoration, (**7**) Hand lighting a candle near flowers, and (**8**) Cardboard catching fire.

Table 5. Detailed classification information for selected images from the new dataset using ChatGPT-4, Gemini 1.5 flash, Microsoft Copilot, and Poe (December 2024).

Images	Fire	Description	Framework
(1)	No	A scenic sky with clouds during sunset, no visible fire.	ChatGPT
	No	The image shows a beautiful sunset with a colorful sky. The sky is filled with different shades of pink, orange, and yellow, creating a stunning display of natural colors.	Gemini 1.5 flash
	No	The image shows a beautiful sunset with a colorful sky. The sky is filled with different shades of pink, orange, and yellow, creating a stunning display of natural colors.	Copilot
	No	The image shows a beautiful sky at sunset, with colorful clouds and a gradient of warm hues.	Poe
(2)	No	A clear sky with trees and a bright sun at sunset, no visible fire.	ChatGPT
	No	The image shows a beautiful sunset. The sun is setting behind some trees, casting a warm glow on the surrounding landscape.	Gemini 1.5 flash
	No	The image captures a tranquil sunset scene with a golden sky gradient. There are a few trees, some dry and sparse grass patches, and a sandy foreground. The sun creates a beautiful lens flare effect.	Copilot
	No	The image appears to capture a serene sunset scene, featuring silhouettes of trees and vegetation against a glowing horizon.	Poe
(3)	No	A light bulb glowing under a thatched roof at night, no visible fire.	ChatGPT
	No	The image shows a thatched roof with a light bulb hanging from it. The background is dark, suggesting it is nighttime.	Gemini 1.5 flash
	No	The image shows an outdoor setting at night with several thatched-roof structures, likely huts or gazebos, illuminated by hanging light bulbs.	Copilot
	No	This image features a cozy outdoor setting, likely at night, with a thatched roof and a hanging light bulb. The warm glow from the bulb creates an inviting atmosphere, contrasting with the darker surroundings. The texture of the thatch adds a rustic charm to the scene.	Poe
(4)	No	Beautiful sunset with clouds and clear sky, no visible fire.	ChatGPT
	No	The image shows a beautiful sunset with a colorful sky. The sky is filled with different shades of pink, orange, and yellow, creating a stunning display of natural colors.	Gemini 1.5 flash
	No	The image you uploaded shows a beautiful sunset with a mix of orange, yellow, and purple hues in the sky. There are large, fluffy clouds and some smaller clouds scattered throughout the sky.	Copilot
	No	This image captures a stunning sunset with vibrant colors and dramatic cloud formations. The warm hues of orange, pink, and purple contrast beautifully with the sky, creating a serene and picturesque scene.	Poe
(5)	No	A warm-colored sky seen through a window, resembling a sunset. No visible fire.	ChatGPT
	No	The image shows a beautiful sunset viewed through a window. The window has a unique arched design and is divided into several panes.	Gemini 1.5 flash
	No	The image shows a view through an open, grid-like window. Outside, there is a warm, orange sky with a tree silhouetted against it. It is likely a sunset or sunrise, creating a beautiful and peaceful scene	Copilot
	No	This image shows a window with a beautiful view of a sunset, featuring soft orange and pink hues in the sky. The window frame adds a structural element to the composition, while the greenery outside provides a natural contrast to the warm colors of the sky.	Poe

Table 5. *Cont.*

Images	Fire	Description	Framework
(6)	No	An indoor display with plants and decorative birds, no visible fire.	
	No	The image shows a colorful parrot decoration hanging from a ceiling covered in lush green plants. It looks like an indoor garden or a themed display.	Gemini 1.5 flash
	No	The image shows an indoor garden with a lush, green canopy of hanging plants. There is a colorful parrot decoration with outstretched wings hanging from the ceiling, surrounded by various plants and flowers. Several people are observing the exhibit, creating a vibrant and lively scene	Copilot
	No	This image features a vibrant indoor setting filled with hanging greenery and decorative elements, including a colorful bird. The lush plants create a lively atmosphere, while the bright colors of the bird add a playful touch to the scene.	Poe
(7)	Yes	A glowing light source resembling fire or a small flame-like light, held near a bouquet of colorful flowers.	ChatGPT
	No	The image shows a lit match held in a person's hand. The match is surrounded by a bed of colorful flowers.	Gemini 1.5 flash
	No	The image shows a hand holding a lit matchstick above a bouquet of flowers, including roses in shades of orange, white, and purple. The matchstick is glowing brightly, illuminating the surrounding flowers. There is no fire apart from the lit matchstick.	Copilot
	Yes	This image captures a hand holding a matchstick or lighter, illuminating a beautiful arrangement of flowers. The flowers, in various colors, create a vibrant and romantic backdrop, while the light adds warmth and a sense of intimacy to the scene.	Poe
(8)	Yes	Cardboard ignited with visible flames on the surface, clearly showing fire.	ChatGPT
	Yes	The image shows a cardboard box on fire. The flames are visible on one side of the box.	Gemini 1.5 flash
	Yes	A small fire is burning on the edge of a piece of cardboard on a tiled floor. Remember to stay safe and keep flammable materials away from potential ignition sources!	Copilot
	Yes	The image shows flames emerging from a cardboard surface, indicating that the cardboard is burning	Poe

Observations on the Comparison of LLM Frameworks: ChatGPT, Microsoft Copilot, Gemini, and Poe

This section provides a detailed examination of the key differences and similarities among the frameworks of ChatGPT, Microsoft Copilot, Gemini, and Poe.

- All frameworks ChatGPT, Microsoft Copilot, Gemini, and Poe deliver highly accurate results in image classification tasks, effectively identifying and categorizing images based on content.
- Microsoft Copilot and Gemini restrict users to uploading one image at a time, which can slow down analysis, whereas ChatGPT allows up to 10 images to be uploaded simultaneously for greater convenience. Poe supports multiple uploads but analyzes individual images more effectively; if too many images are uploaded, it notifies users with a message stating, "Message or attachment too large. Please shorten the message or upload a smaller attachment, or consider using a different bot that supports larger messages".
- LLMs employ varying strategies for data privacy. For instance, Microsoft Copilot automatically hides faces in images to address privacy concerns, a feature not present in some other frameworks.
- Some LLMs prioritize user safety by providing immediate advisories when images are classified as depicting fire. For instance, Microsoft Copilot instruct users of the following:

  "*Remember to stay safe and keep flammable materials away from potential ignition sources!*"

"It's a reminder of how easily flammable materials can catch fire and the importance of fire safety. It's always good to be cautious with items like cardboard and to ensure they're kept away from any potential ignition sources. If you need any tips on fire safety or have other questions, feel free to ask!"

- Some models emphasize user confidentiality and implement features to mask sensitive information. Microsoft Copilot also treats fire images as sensitive content, displaying messages like, "The content here is sensitive; can you try a different file?" and "I'm afraid I can't discuss that topic; sorry about that," to maintain privacy and safety concerns.

6. Discussion

This section provides a brief overview of the benefits and drawbacks of using a ChatGPT classification-based approach for fire incidents to support fire detection, highlighting its ease of use and speed, while also noting its limitations.

1. The ChatGPT-based approach is a quick and accessible solution for fire classification, capable of classifying a variety of images with high accuracy and speed, but it requires very precise prompts to ensure accurate results.
2. ChatGPT not only classifies fire images but also provides important contextual details that can aid in decision-making. These additional insights, such as identifying fire severity, affected areas, or potential risks, can support more informed responses and planning in emergency situations.
3. Analyzing a large number of images at once with ChatGPT can limit its ability to focus on each image effectively. This can result in a loss of detail and accuracy. Processing fewer images at a time ensures more reliable and precise results. In this experiment, analyzing three images yields an accuracy of about 99%, whereas analyzing 10 images at once may reduces the accuracy to below 95%.
4. ChatGPT does not allow analyzing too many images in a single session due to its limitations. As a result, there is a restriction on how many images can be processed at a time, requiring a few hours to wait before the next set of images can be analyzed.
5. Despite the high accuracy of classifications, mistakes can still occur, and ChatGPT may sometimes misclassify images that were previously classified correctly, leading to different results each time. Results can vary according to prompt and time. They can change, but overall, the results are good. To minimize such errors, it is important to provide clear and accurate prompts and to limit the number of images analyzed at once for the best performance.
6. Classification based ChatGPT can also be used to verify and validate dataset contents. For instance, the classification based ChatGPT was successful in discovering two images in the DFAN dataset that were mistakenly classified as Car_Fire and Electric_Fire.

1361452

221353

Image Name	ChatGPT based classification approach	DFAN classification	Information
1361452.jpg	Non_Fire	Electric_Fire	Utility pole with transformer; no visible signs of fire.
221353.jpg	PickUp_Fire	Car_Fire	Visible fire and smoke around the pickup truck.

7. LLMs and ChatGPT can complement existing technologies for fire detection and classification, adding an extra layer that provides detailed classification information to support informed decision-making.
8. LLMs can provide direct translation of responses into multiple languages, which is a crucial feature for fast communication in critical situations, such as fire incidents. This capability ensures that vital information can be understood by individuals, regardless of their language, facilitating quicker decision-making and response during emergencies.
9. Comparisons between ChatGPT, Gemini 1.5 Flash, Microsoft Copilot, and Poe frameworks reveal that all frameworks achieve a high level of accuracy in image classification tasks but they differ in how they handle user uploads and whether or not to treat fire images as sensitive content.

7. Conclusions

This study highlights the promising role of ChatGPT-4 for enhancing fire classification capabilities. Most fire detection systems have focused on using AI to identify specific fire-related features, such as flames, smoke, or differentiating between fire and non-fire scenarios. However, this approach lacked the integration of advanced language models that can offer deeper contextual understanding. By leveraging ChatGPT-4, this study provides a classification method that not only identifies fire incidents but also adds valuable contextual insights.

The evaluation of the method using two benchmark datasets—the Forest Fire dataset and the DFAN dataset—demonstrated the significant potential of ChatGPT-4 for timely and accurate fire classification. The integration of ChatGPT-4 into fire detection systems can complement existing technologies by providing rapid analysis that leads to quicker identification and categorization of fire incidents. Moreover, its capability to understand and communicate detailed information, such as the type of burning materials and associated risk levels, makes it a good tool for improving early warning mechanisms. While ChatGPT-4 shows promising results for fire classification, it has some limitations. It cannot process too many images simultaneously, requiring a waiting period between sessions, which can hinder efficiency in time-sensitive scenarios. Additionally, despite high accuracy, misclassifications can occur, and the model may inconsistently classify the same images over different sessions. Results can also vary based on the prompts provided and the time of analysis, indicating a need for consistent input to achieve reliable outcomes. The fire classification approach introduced in this study can be effectively adapted for a variety of other classification tasks. This makes it a good tool for addressing diverse classification challenges, ensuring better outcomes in various contexts and applications. For future work, we plan to integrate LLMs with other computer vision tasks, such as image segmentation and object detection, to improve the identification of fire-related features and enhance the classification of elements like smoke and flames. Furthermore, a key area of future research will involve studying the risks and limitations of LLMs in classification and detection tasks.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in Kaggle at https://www.kaggle.com/datasets/imankhammash/fireimagesdataset (accessed on 1 December 2024).

Acknowledgments: The researcher would like to acknowledge the Deanship of Graduate Studies and Scientific Research, Taif University, Saudi Arabia, for funding this work. She also acknowledges the use of ChatGPT and Grammarly for improving the manuscript's syntax and grammar, thereby enhancing its readability.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Mohamadi, S.; Mujtaba, G.; Le, N.; Doretto, G.; Adjeroh, D.A. ChatGPT in the age of generative AI and large language models: A concise survey. *arXiv* **2023**, arXiv:2307.04251.
2. Liu, Y.; Han, T.; Ma, S.; Zhang, J.; Yang, Y.; Tian, J.; He, H.; Li, A.; He, M.; Liu, Z.; et al. Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology* **2023**, *1*, 100017. [CrossRef]
3. Zhu, Y.; Yuan, H.; Wang, S.; Liu, J.; Liu, W.; Deng, C.; Chen, H.; Liu, Z.; Dou, Z.; Wen, J.R. Large language models for information retrieval: A survey. *arXiv* **2023**, arXiv:2308.07107.
4. Hadi, M.U.; Al Tashi, Q.; Shah, A.; Qureshi, R.; Muneer, A.; Irfan, M.; Zafar, A.; Shaikh, M.B.; Akhtar, N.; Wu, J.; et al. Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects. *Authorea* **2024**. [CrossRef]
5. Kenton, J.D.M.W.C.; Toutanova, L.K. Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), Minneapolis, MN, USA, 2–7 June 2019.
6. Zhou, J.; Ke, P.; Qiu, X.; Huang, M.; Zhang, J. ChatGPT: Potential, prospects, and limitations. *Front. Inf. Technol. Electron. Eng.* **2023**. [CrossRef]
7. Motlagh, N.Y.; Khajavi, M.; Sharifi, A.; Ahmadi, M. The impact of artificial intelligence on the evolution of digital education: A comparative study of openAI text generation tools including ChatGPT, Bing Chat, Bard, and Ernie. *arXiv* **2023**, arXiv:2309.02029.
8. Biswas, S. Role of ChatGPT in Computer Programming. *Mesopotamian J. Comput. Sci.* **2013**, *2013*, 9–15.
9. Adeshola, I.; Adepoju, A.P. The opportunities and challenges of ChatGPT in education. *Interact. Learn. Environ.* **2023**, 1–14. [CrossRef]
10. Li, J.; Dada, A.; Puladi, B.; Kleesiek, J.; Egger, J. ChatGPT in healthcare: A taxonomy and systematic review. *Comput. Methods Programs Biomed.* **2024**, *245*, 108013. [CrossRef] [PubMed]
11. Tustumi, F.; Andreollo, N.A.; Aguilar-Nascimento, J.E.D. Future of the language models in healthcare: The role of chatGPT. *J. Med. Syst.* **2023**, *36*, e1727.
12. Javaid, M.; Haleem, A.; Singh, R.P. ChatGPT for healthcare services: An emerging stage for an innovative perspective. *BenchCouncil Trans. Benchmarks Stand. Eval.* **2023**, *3*, 100105. [CrossRef]

13. Sallam, M. ChatGPT Utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare* **2023**, *11*, 887. [CrossRef]
14. Bender, E.M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S. On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual, 3–10 March 2021; pp. 610–623.
15. Wang, C.; Liu, S.; Yang, H.; Guo, J.; Wu, Y.; Liu, J. Ethical considerations of using ChatGPT in health care. *J. Med. Internet Res.* **2023**, *25*, 48009. [CrossRef] [PubMed]
16. Parray, A.A.; Inam, Z.M.; Ramonfaur, D.; Haider, S.S.; Mistry, S.K.; Pandya, A.K. ChatGPT and global public health: Applications, challenges, ethical considerations and mitigation strategies. *Glob. Transit.* **2023**, *5*, 50–54. [CrossRef]
17. Hua, S.; Jin, S.; Jiang, S. The limitations and ethical considerations of chatgpt. *Data Intell.* **2024**, *6*, 201–239. [CrossRef]
18. Stahl, B.C.; Eke, D. The ethics of ChatGPT–Exploring the ethical issues of an emerging technology. *Int. J. Inf. Manag.* **2024**, *74*, 102700. [CrossRef]
19. Seydi, S.T.; Saeidi, V.; Kalantar, B.; Ueda, N.; Halin, A.A. Fire-Net: A Deep Learning Framework for Active Forest Fire Detection. *J. Sens.* **2022**, *2022*, 8044390. [CrossRef]
20. Zhang, Q.; Xu, J.; Xu, L.; Guo, H. Deep convolutional neural networks for forest fire detection. In Proceedings of the 2016 International Forum on Management, Education and Information Technology Application, Guangzhou, China, 30–31 January 2016; pp. 568–575.
21. Avazov, K.; Jamil, M.K.; Muminov, B.; Abdusalomov, A.B.; Cho, Y.I. Fire detection and notification method in ship areas using deep learning and computer vision approaches. *Sensors* **2023**, *23*, 7078. [CrossRef]
22. Sathishkumar, V.E.; Cho, J.; Subramanian, M.; Naren, O.S. Forest fire and smoke detection using deep learning-based learning without forgetting. *Fire Ecol.* **2023**, *19*, 9. [CrossRef]
23. Ghali, R.; Akhloufi, M.A. Deep learning approaches for wildland fires using satellite remote sensing data: Detection, mapping, and prediction. *Fire* **2023**, *6*, 192. [CrossRef]
24. Saleh, A.; Zulkifley, M.A.; Harun, H.H.; Gaudreault, F.; Davison, I.; Spraggon, M. Forest fire surveillance systems: A review of deep learning methods. *Heliyon* **2024**, *10*, e23127. [CrossRef]
25. Idroes, G.M.; Maulana, A.; Suhendra, R.; Lala, A.; Karma, T.; Kusumo, F.; Hewindati, Y.T.; Noviandy, T.R. TeutongNet: A fine-tuned deep learning model for improved forest fire detection. *Leuser J. Environ. Stud.* **2023**, *1*, 1–8. [CrossRef]
26. Saha, S.; Bera, B.; Shit, P.K.; Bhattacharjee, S.; Sengupta, N. Prediction of forest fire susceptibility applying machine and deep learning algorithms for conservation priorities of forest resources. *Remote Sens. Appl.* **2023**, *29*, 100917. [CrossRef]
27. Harkat, H.; Nascimento, J.M.; Bernardino, A.; Ahmed, H.F.T. Fire images classification based on a handcraft approach. *Expert Syst. Appl.* **2023**, *212*, 118594. [CrossRef]
28. Barros-Daza, M.J.; Luxbacher, K.D.; Lattimer, B.Y.; Hodges, J.L. Real time mine fire classification to support firefighter decision making. *Fire Technol.* **2022**, *58*, 1545–1578. [CrossRef]
29. Balakrishnan, V.; Mohammed Hashim, A.N.; Lee, V.C.; Lee, V.H.; Lee, Y.Q. Fatal structure fire classification from building fire data using machine learning. *Int. J. Intell. Comput. Cybern.* **2024**, *17*, 236–252. [CrossRef]
30. Vorwerk, P.; Kelleter, J.; Müller, S.; Krause, U. Classification in Early Fire Detection Using Multi-Sensor Nodes—A Transfer Learning Approach. *Sensors* **2024**, *24*, 1428. [CrossRef] [PubMed]
31. Wu, W.; Chen, Y. A Trustworthy Classification Model for Intelligent Building Fire Risk. *IEEE Access* **2022**, *10*, 10371–10383. [CrossRef]
32. Bashyal, S.; Venayagamoorthy, G.K.; Paudel, B. Embedded neural network for fire classification using an array of gas sensors. In Proceedings of the 2008 IEEE Sensors Applications Symposium, Atlanta, GE, USA, 12–14 February 2008; IEEE: Atlanta, GA, USA, 2008; pp. 146–148.
33. Islam, A.M.; Masud, F.B.; Ahmed, M.R.; Jafar, A.I.; Ullah, J.R.; Islam, S.; Shatabda, S.; Islam, A.M. An attention-guided deep-learning-based network with Bayesian optimization for forest fire classification and localization. *Forests* **2023**, *14*, 2080. [CrossRef]
34. Vaswani, A. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**. Available online: https://user.phil.hhu.de/~cwurm/wp-content/uploads/2020/01/7181-attention-is-all-you-need.pdf (accessed on 1 December 2024).
35. Brown, T. Language Models are Few-Shot Learners. *arXiv* **2020**, arXiv:2005.14165.
36. Mok, A.; Zinkula, J. ChatGPT may be coming for our jobs. Here are the 10 roles that AI is most likely to replace. *Bus. Insid.* **2023**, *2*.
37. Khan, A.; Hassan, B.; Khan, S.; Ahmed, R.; Abuassba, A. DeepFire: A Novel Dataset and Deep Transfer Learning Benchmark for Forest Fire Detection. *Mob. Inf. Syst.* **2022**, *2022*, 5358359. [CrossRef]
38. Kalla, D.; Smith, N.; Samaah, F.; Kuraku, S. Study and analysis of chat GPT and its impact on different fields of study. *Int. J. Innov. Sci. Res. Technol.* **2023**, *8*. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4402499 (accessed on 1 December 2024).
39. Forest Fire Dataset. 2022. Available online: https://www.kaggle.com/datasets/alik05/forest-fire-dataset (accessed on 1 October 2024).

40. Yar, H.; Hussain, T.; Agarwal, M.; Khan, Z.A.; Gupta, S.K.; Baik, S.W. Optimized dual fire attention network and medium-scale fire classification benchmark. *IEEE Trans. Image Process.* **2022**, *31*, 6331–6343. [CrossRef] [PubMed]
41. Prompt Engineering. 2024. Available online: https://help.openai.com/en/articles/10032626-prompt-engineering-best-practices-for-chatgpt#h_6eb43fed58 (accessed on 1 November 2024).