MAIKRON

Journal of Artificial Intelligence and Computing Applications

Review article

Convolutional neural networks for identification of forest fires in satellite images: a short narrative review

Randy Santos-Poot[1,*] **and Alejandro Alejandres-Rivera**[1]

[1]Tecnológico Nacional de México / IT de Mérida, Yucatán, México

ABSTRACT

Deforestation, a global phenomenon resulting in massive loss of forest areas, and forest fires, which are increasing in frequency and intensity due to climate change and human activity, present major challenges in managing and reducing these catastrophic events. Forests are essential for biodiversity and, representing about one third of the earth's land surface, require effective protection and conservation strategies as a matter of urgency. The effectiveness demonstrated by the models in detecting forest fires with satellite images is highlighted, allowing a faster response to emergencies. However, some limitations are pointed out, such as satellite capabilities and the need for high quality data to ensure the reliability of CNN model performance. This paper reviews recent advances in this field, highlighting the effectiveness of CNN-based models in identifying fires accurately and in a timely manner.

Keywords: forest fire detection, convolutional neural networks, multispectral satellite images

1. Introduction

Deforestation is a phenomenon that involves the massive loss of forest areas around the world, including forests, which are one of the most important natural resources for humanity. These represent almost a third of the planet's land surface and are home to countless species, especially in tropical areas, where there is usually greater biological diversity [1]. There are 24 areas made up of a large concentration of hotspots threatened by deforestation called "fronts", nine of which are in Latin America [2]. The increasing frequency and intensity of wildfires, exacerbated by climate change and human activity, pose significant challenges in the management and mitigation of these catastrophic events.

Convolutional Neural Networks (CNN) have proven to be promising tools in the field of computer vision and image analysis, particularly in the detection and classification of objects in images [3]. Their ability to learn relevant features automatically makes them ideal for forest fire detection tasks in images. The capacity of these technologies can not only save lives and minimize property damage, but also inform and guide long-term forest management strategies and conservation policies.

Despite significant advances in the application of CNN for wildfire detection in satellite images, significant gaps still remain in the literature. While research has demonstrated the effectiveness of lightweight CNN models in accurately identifying fires [3, 4], there is a lack of consensus on their performance under variable conditions, such as dense vegetation cover or the presence of clouds. Furthermore, the ability of these models to detect fires in incipient stages and their adaptability to different geographic environments also need to be further examined.

The objectives of this review article are multiple.

E-mail address: le20081650@merida.tecnm.mx
https://doi.org/10.5281/zenodo.14165778

Firstly, it seeks to synthesize and critically evaluate the existing literature on the performance and scope of CNN for the detection of forest fires in satellite images, highlighting both the significant advances and the identified limitations. The purpose is to identify specific areas of controversy and gaps in current research to highlight future research opportunities. In line with these objectives, the research questions focus on the effectiveness of CNN models under different environmental conditions, their ability to detect fires in early stages, and their adaptability to various geographic regions.

As this is a short narrative review article, a careful selection of high-quality articles was made, justified by the need to provide readers with a meaningful assessment of the current state of research on CNNs for fire detection. By focusing on the most influential works, a more direct and concise analysis of trends, challenges and opportunities in the field of wildfire detection using CNN can be provided.

In terms of structure, Section 2 will explain the process for selecting articles. Next, Section 3 will critically review the existing literature, highlighting the main findings, limitations and areas of controversy. Subsequently, Section 4 will discuss the implications of the review and possible future research directions in this emerging and vital field. Finally, in Section 5 the conclusions of the work are presented in a brief and concise manner.

2. Methodology

The methodology used in this review is essential to establish the scope, breadth and depth of the research. To ensure a comprehensive and rigorous search of the relevant literature, two search engines were mainly used: Google Scholar and Semantic Scholar. These platforms were selected due to their extensive repository of academic articles in various disciplines. Additionally, these platforms are recognized for their comprehensive indexing, user-friendly search capabilities, and integration with a wide range of academic journals.

The search began with the determination of the following keywords: “multispectral”, “satellite”, “images”, “convolutional”, “neural”, “networks”, “wildfire”, “detection” and “forest”. These keywords were used in combinations to form search terms using Boolean operators (only in the case of Google Scholar). Only those published within the last 5 years were considered, ensuring a balance between seminal works and contemporary ideas and the following search terms were used in the first week of March 2024:

1. “multispectral satellite images” AND “convolutional neural networks” AND “wildfire detection”, which returned 11 results in Google Scholar and 1530 in Semantic Scholar.

2. “multispectral satellite images” AND “convolutional neural networks” AND “forest fire detection” which returned 12 results in Google Scholar and 390 in Semantic Scholar.

In this way, a total result of 1922 articles was obtained in our search.

Inclusion criteria were strict to ensure a high standard of quality and relevance. The search was limited to journals indexed in the Journal Citation Reports (JCR), specifically those classified as Q1 or Q2, which represent higher quality in the field. Also, the scope was limited to works written in English. In addition, works were sought that reached at least 10 citations per year, reflecting their impact and recognition within the academic community. This approach ensured the inclusion of articles of high quality and relevance to the review.

The initial search returned a volume of 1922 articles between the two search terms, which were then filtered using the criteria mentioned above and shown in Figure 1. The approach was to filter these articles considering their relevance, citation count, and their quartile.

Of the 1922 initial articles, 1871 published articles whose title and abstract were not related to the topic were discarded, leaving 51. Then, 31 articles that did not belong to journals were discarded, leaving 20 articles. Of these, all were in English and were within the thematic scope of the research. One that had less than 10 citations per year and 5 that were not indexed in JCR were removed. This process reduced the list to 14 articles that met all inclusion criteria and were considered the most relevant and highest quality for the review. All these articles were published in journals belonging to Q1 or Q2. Table 1 shows the list of the final filtered articles.

The obvious limitation of the selection approach used was the rigor of the process, a decision made to maintain a simplified narrative and focus exclusively on the most innovative and highly cited research. While this methodology allows for a concise overview, it may omit certain relevant studies that did not meet strict criteria or that offer alternative or more nuanced perspectives. The potential for such omissions underscores the importance of the research, beyond what this review has encapsulated. Furthermore, an additional limitation was the large number of works in the results, which led to filtering only articles from the beginning, simplifying the process, but potentially omitting relevant studies that did not fit the search criteria. Despite these limitations, the applied methodologies and the selected literature constitute the foundation on which this work is based.

3. Thematic Overview

This analysis identifies several key themes within the detection of forest fires from images, with special attention to satellite data, due to its scope and large-scale monitoring capacity. Three main detection approaches are recognized based on the origin of the images used in the training models: ground, aerial and satellite detection. Although this article focuses on satellite detection, a brief review of ground and airborne approaches is also included to provide a comprehensive perspective.

Within satellite detection, studies are organized

Table 1. The final list of articles used in this review, including information for title, journal, year of publication and citation.

Title	Journal	Year	Citation
Wildfire detection using transfer learning on augmented datasets	Expert systems with applications	2020	[5]
A forest fire smoke detection model combining convolutional neural network and vision transformer	Frontiers in Forests and Global Change	2023	[6]
Forest fire detection in aerial vehicle videos using a deep ensemble neural network model	Aircraft Engineering and Aerospace Technology	2023	[7]
A Small Target Forest Fire Detection Model Based on YOLOv5 Improvement	Forests	2022	[8]
Forest-fire response system using deep-learning-based approaches with CCTV images and weather data	IEEE Access	2022	[9]
Active Fire Detection from Landsat-8 Imagery Using Deep Multiple Kernel Learning	Remote Sensing	2022	[10]
Active Fire Detection Using a Novel Convolutional Neural Network Based on Himawari-8 Satellite Images	Frontiers in Environmental Science	2022	[3]
Comparative Research on Forest Fire Image Segmentation Algorithms Based on Fully Convolutional Neural Networks	Forests	2022	[11]
A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency	GIScience & Remote Sensing	2022	[12]
Super-Resolution Reconstruction of Remote Sensing Data Based on Multiple Satellite Sources for Forest Fire Smoke Segmentation	Remote Sensing	2023	[13]
Autonomous Satellite Wildfire Detection Using Hyperspectral Imagery and Neural Networks: A Case Study on Australian Wildfire	Remote Sensing	2023	[14]
Wildfire Detection Using Convolutional Neural Networks and PRISMA Hyperspectral Imagery A Spatial-Spectral Analysis	Remote Sensing	2024	[15]
Uni-temporal Sentinel-2 imagery for wildfire detection using deep learning semantic segmentation models	Geomatics, Natural Hazards and Risk	2023	[16]
Investigating the Impact of Using IR Bands on Early Fire Smoke Detection from Landsat Imagery with a Lightweight CNN Model	Remote Sensing	2022	[17]

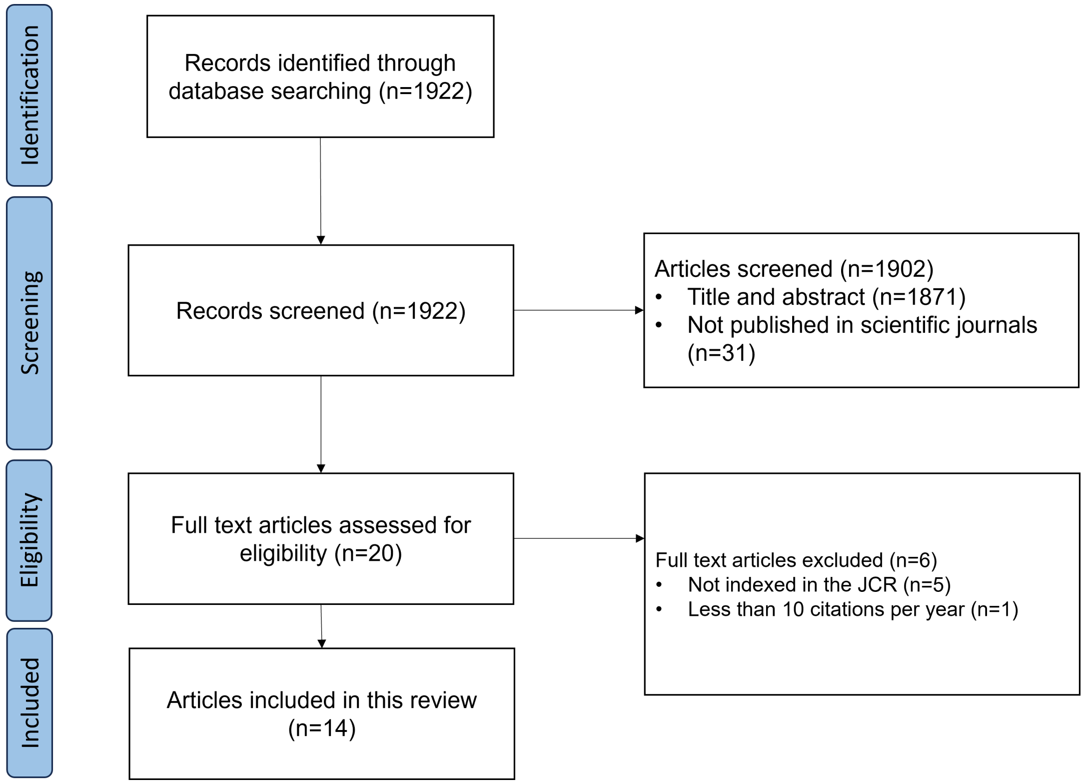

Figure 1. Diagram of the selection process showing the different stages of identification, screening, eligibility and inclusion, filtering from 1922 articles to the final 14 considered in this review.

around three main themes: 1) development of active fire detection models using convolutional neural networks (CNN), which take advantage of satellite data of different spatial and temporal resolutions; 2) comparison and improvement of forest fire image segmentation algorithms, exploring the use of fully convolutional models for accurate fire and affected area detection; and 3) use of geostationary satellite data to minimize latency in fire detection, optimizing early response through high temporal resolution images. Additionally, advanced high-resolution remote sensing image reconstruction and segmentation methods for detecting smoke in wildfires, and the use of hyperspectral imagery to improve the accuracy of fire detection models, are discussed. These topics cover the most recent and sophisticated approaches to spatial and spectral analysis of data, contributing to the accuracy and reliability of detection.

3.1 Types of images for forest fire detection

Forest fire detection using images is divided into three forms depending on the source of the data: satellite, aerial and ground detection. Aerial images, taken from manned or unmanned aircraft (UAV), offer high spatial resolution, allowing specific areas to be observed in great detail. Its flexibility to fly over specific areas is an advantage, although its coverage is more limited compared to satellite images. Terrestrial imaging, on the other hand, allows for constant monitoring of the monitored area, as the cameras can operate without interruptions, although they are restricted to the places where they have been physically installed, thus providing a more limited range in relation to the others. types of images.

The satellite form of detection allows large areas of land to be covered, although, unlike terrestrial detection, it does not have fixed cameras to continuously monitor the same site due to the orbit of the satellites, which leaves an interval between images of the same place depending on the temporal resolution of the satellite. Satellites commonly used for creating datasets include Landsat-8, Himawari-8, PRISMA, VIIRS, and Sentinel-2. High spatial resolution satellites, such as Landsat-8 and Sentinel-2, allow smaller-scale fire detection, while high temporal resolution satellites, such as Himawari-8, facilitate early detection thanks to their orbits with reduced revisit time [10]. The satellites generate multispectral images (Landsat-8, Sentinel-2, VIIRS, Himawari-8) and also hyperspectral images (PRISMA). The difference between these types lies in the number of bands: multispectral images contain a small number of bands that cover wide ranges of the electromagnetic spectrum, while hyperspectral images, such as those from PRISMA, can contain up to 230 finer bands [15, 3, 10].

3.2 Moments of detection in the fire cycle

It is possible to define different approaches for fire detection, divided into three categories according to the time of the fire in which the detection is carried out: early, active and post-fire detection. Early detection is generally related to smoke detection, as satellite sensors often identify smoke before fire, since smoke spreads more quickly and fire detection is based on infrared bands. These bands allow the temperature of the Earth's surface to be captured, although the flames are required to have a minimum size, determined by the spatial resolution of the sensor [6]. Post-fire detection, on the other hand, focuses on identifying damaged areas after the fire.

3.3 Detection models based on Deep Learning

Deep learning (DL) is generally used as an improvement over threshold-based methods, as CNNs allow for more accurate pattern detection. Because wildfire detection using CNN is a recent approach, there are few large-scale datasets specifically designed for training fire detection models in the literature. Therefore, it is common for studies where models are proposed to generate their own datasets by labeling the phenomenon to be detected, such as smoke, fire, and burned areas.

These datasets are mostly designed for segmentation models, where detection is done at the pixel level, that is, each pixel of an image is labeled as a specific class, either in multiclasses or binary classes. There are also models where detection is done at the scene level [17, 6, 3, 10], so the label is applied to the entire image. Pixel-level detection allows you to observe the shape of the fire and its edges with greater precision. The input characteristics of the CNNs include reflectance values of the bands of each satellite, which vary depending on the radiometric resolution. Although the characteristics used in each model depend on the approach, in general spectral characteristics (band data) are used, and in some cases temporal characteristics (multitemporal band information) applicable to early detection. These datasets come from diverse regions, from small areas within a country [17, 16, 3] to global data sets with images of fires on multiple continents [10].

Regarding the architectures used, a wide variety of technologies and modules are observed, ranging from simple CNN architectures to hybrid models that combine CNN with Vision Transformer (ViT). Techniques such as multiscale convolution, residual edges, depth-separable convolution, inverted residual blocks, and spatial and channel attention modules have been used. Specialized architectures such as U-Net, ResNet, LinkNet, DeeplabV3, and Inception, among others, have also been identified. Models with lower parametric complexity, known as lightweight models, allow inferences to be carried out with lower latency, being suitable for early detection approaches [17, 6].

CNNs have limitations in their generalization capacity related with smoke [6], which has motivated the development of hybrid models such as SR-Net [6]. This model combines CNN and ViT to take advantage of both approaches and improve smoke detection. The training is carried out with true color images from the Himawari-8 satellite, which has a very high temporal resolution, ideal for early detection of fires. This model has been shown to be computationally more efficient and stable compared to other reference models, such as AlexNet, MobileNet, GoogLeNet and ResNet50, improving both adaptability and stability, evaluated using activation techniques such as Gradient-weighted Class Activation Mapping.

3.4 Optimization and improvement of detection models

The use of the visible spectrum bands (Red, Green, Blue) may not be sufficient to differentiate smoke from other aerosols in the atmosphere, which can lead to false alarms; Therefore, the use of infrared (IR) bands to improve the accuracy of smoke detection is common in the literature. It has been shown that the combined information from the IR bands can significantly increase the detection accuracy. An example is the lightweight VI_SD model, which performs scene-level classification [17], using a simplified but efficient structure, with spatial and channel attention modules that highlight relevant features. In addition, it incorporates residual modules to improve model learning. This model was compared with other state-of-the-art models, such as SmokeNet, SAFA, and Inception-ResNetV2, showing competitive performance with fewer parameters. When analyzing the contribution of the IR bands, it was found that the inclusion of the NIR band improved the model performance, while other band combinations yielded variations in the model accuracy.

[3] also uses images from the Himawari-8 satellite to demonstrate the potential of multiscale convolution and residual structures in fire detection efficiency. A CNN model called FireCNN is proposed, designed for real-time detection of forest fires. Classification is performed at the pixel level using a fully connected layer, similar to SimpleCNN, which processes and fuses features from multiple scale modules to calculate the probability that a pixel belongs to a fire, using Softmax. This approach optimizes the accuracy of wildfire detection by analyzing features at multiple scales, evidencing the effectiveness of FireCNN in accurate fire detection.

Multiscale convolution has been integrated into other architectures, such as Multiscale-Net, proposed in [10]. Multiscale-Net also uses different dilation rates and is trained with Landsat-8 images of various continents, allowing it to cover a variety of geographic scenarios and environmental conditions, ensuring the generalization and robustness of the model. Architectures like U-Net and FCN are also used in image segmentation; The encoder part of these architectures extracts features and generates lower resolution maps, while the decoder part produces segmentation masks. These masks integrate information from both intermediate and initial layers, combining local and global details to avoid the loss of important information and improve segmentation accuracy.

The inclusion of temporal information in model training has been shown to be relevant to improve early fire detection. The study in [12] explores latency reduction in wildfire detection using time series data from the Himawari-8 satellite, compared to MODIS and VIIRS. Three types of input features were evaluated: spectral, temporal and spatial, in regions of South Korea, North Korea and China. After classification, post-processing was applied to eliminate false alarms, using thresholds based on MODIS land cover ratios, helping to optimize the overall accuracy of the model.

Although Landsat-8 does not have high temporal resolution, its images are also useful in early detection, as the study by [13] shows. This work focuses on improving smoke detection in images from the VIIRS sensor, which has limited spatial resolution compared to sun-synchronous orbit satellites such as Landsat-8. To address this limitation, a CNN designed for super-resolution reconstruction of VIIRS images was used, in order to obtain high temporal and spatial resolution images. The CNN performs the super-resolution reconstruction, while the Smoke-Net network is responsible for segmenting the smoke in the enhanced images. Matched Landsat-8 and VIIRS images were used and groundtruth images were manually labeled. The VIIRS RGB images were adapted to the Landsat-8 RGB domain using a CycleGAN, allowing their reconstruction in super-resolution. The enhanced VIIRS images were then used for smoke segmentation with Smoke-Net, showing performance close to that of the original Landsat-8 images.

Detection latency plays a crucial role in early fire detection, and may depend on satellite orbit, model inference time, and image acquisition. To address this challenge, [14] proposes to obtain only sensing data from satellites instead of all sensor data. In this context, they developed and trained an optimized CNN for deployment on reliable autonomous satellites (TASO), adjusting its complexity to meet onboard processing requirements. A one-dimensional CNN (1-DCNN) proved to be effective for wildfire classification and, after evaluating various hardware options, the model was implemented on data processing devices on the satellite itself. This approach promises to significantly improve response capacity in natural disaster management.

3.5 Limitations and challenges in forest fire detection

Regarding the PRISMA satellite, there is a lack of suitable datasets for training segmentation models, which led [15] to manually create a dataset with the objective of evaluating the generalization capacity of four models. The models, which shared a basic structure of three hidden layers and an output layer with softmax for classification into seven classes, varied in complexity depending on their inputs. Although all models achieved good results in terms of accuracy, those that incorporated spatial information (neighboring pixels) in the input layer showed greater robustness and a lower false alarm rate.

Of the articles used in the review, the only recent study that has used data from the Sentinel-2 satellite is that of [16], in which several CNN models for semantic segmentation are compared. 14 models based on five encoder-decoder architectures were evaluated, including variations with encoders such as MobileNet and versions of ResNet. Among these, the U-Net model with ResNet50 as encoder showed outstanding performance and was selected for the creation of a pre-trained model using the Keras library. This model was evaluated with images of forest fires from several countries in climatic conditions similar to those of Turkey between 2021 and 2022, facing the challenge of adapting its performance to diverse conditions and the lack of specific datasets.

The accurate detection of forest fires is crucial for the prevention and mitigation of this natural disaster, the lack of datasets is a problem that not only affects satellite detection, this does not allow the training of sufficiently robust models. Although the images used for this type of detection are different from satellite images, which generally cover a large amount of space, they are designed to detect fires on a smaller scale and, therefore, are better in that regard than satellite detection but having limitations in the territory covered. In general, this field is relatively more advanced, such as the complexity of the architectures used, but they are more focused on active detection.

Two recent studies address this challenge from complementary perspectives. [5] highlights the scarcity of large-scale databases with real fire images, proposing an approach based on transfer learning combined with data augmentation techniques. This study used databases from Portugal and Corsica to retrain the Inception-v3 model, previously trained on ImageNet, in order to improve its performance in image-level fire detection using binary classification. Limitations were identified, including classification errors due to specific patterns and problems of reduced spatial scales, thus recommending a multiclass formulation and the consideration of consecutive frames to improve reliability. The second study [8], also uses transfer learning to overcome the limitation of a small dataset, focusing on improving the YOLOv5 model for real-time detection of forest fires. Tweaks include replacing the SPPF module with SPPFP to improve global information retention and adding a CBAM attention module, among other structural changes, optimizing multi-scale feature fusion. Although the system faces challenges such as false positives, proposed future optimizations could improve its real-time performance and applicability, especially for applications in drones or helicopter-mounted cameras.

A promising approach to improve fire detection accuracy is the use of deep neural network ensemble models, as proposed by [7]. In this study, an ensemble model was developed that combines four CNN models (Faster R-CNN, RetinaNet, Yolov2 and Yolov3) trained on the Corsica fire dataset, using data augmentation techniques to optimize performance. The model fuses the outputs of all four networks to improve fire detection accuracy, leveraging the strengths of each individual model. Although this approach notably improves accuracy, it was observed that each combination variant is more effective under certain specific conditions and that high computational complexity represents a major barrier to its real-time implementation on resource-limited platforms. However, the results suggest that the improvement in precision justifies the use of an ensemble model, especially in critical situations where early detection is key to preventing fire expansion.

Recently, the literature has begun to analyze in detail some cutting-edge architectures in fire detection, since these play a crucial role in extracting features from images, directly impacting the accuracy and efficiency of the model. Studies have shown that the choice of backbone can significantly influence model performance, as evidenced by research by [11] and [9]. In [11], a comparison was performed between four semantic segmentation networks to evaluate which is more effective in distinguishing between flame and forest background pixels in images captured by UAV. To improve generalization, data augmentation, random noise, and variations of environmental conditions were applied. The models evaluated included architectures such as FCN, U-Net, PSPNet and DeepLabV3+, tested with two backbones: VGG16 and ResNet50. Although similar performances were observed, the choice of ResNet50 as the backbone proved to be more effective overall, despite its higher computational complexity, which represents a challenge in terms of efficiency and applicability in resource-limited environments.

A modified Faster R-CNN architecture is proposed in [9], using DetNAS, a variant of Neural Architecture Search (NAS) for object detection networks. NAS is an automated method that searches the space of possible architectures to find the most suitable one for specific tasks, without relying exclusively on manual design. This study used DetNAS to identify an optimal backbone in the Faster R-CNN architecture, obtaining a lightweight model suitable for real-time detection. The resulting architecture was compared with others, such as ResNet, VoVNet and FBNetV3, the latter also based on NAS. The application of NAS for architectural optimization represents progress, although the model continues to face real-time implementation challenges in environments with limited computational resources.

As a conclusion to this section, the topics presented show a comprehensive overview of current techniques and approaches in forest fire detection using images. As part of the review of the different topics, the different types of images used (terrestrial, aerial and satellite), temporal detection approaches, as well as advances and improvements in the accuracy of the models through optimized architectures and the use of multispectral and hyperspectral bands. Although these techniques have allowed notable progress, challenges persist, particularly regarding the availability and quality of datasets, computational limitations, and the generalization capacity of models in different contexts and environmental conditions. These discussion points will be discussed further in the next section.

4. Discussion

In this study, a comprehensive literature review has been carried out to evaluate the performance and scope of CNNs in detecting forest fires using satellite images. In addition, we sought to identify the most effective model among those reviewed. The importance of this review lies in the growing interest in improving the accuracy and speed of fire detection, considering that prevention and early response are essential to mitigate the environmental and economic damage caused by forest fires. Although satellites offer significant advantages in coverage and cost, their spatial and temporal resolution can limit early detection, especially for smaller-scale fires. Satellite images, in particular, face the restriction of discontinuous coverage, where detection depends on the temporal resolution of the satellite and the sensitivity of its spectral bands. For example, detecting low-intensity fires may require satellites with high radiometric resolution, something that is not always available in current satellite orbits. This study also examines how DL techniques, such as CNNs, are beginning to overcome some of these limitations, proposing improvements in accuracy and responsiveness through advanced techniques such as spectral band combining and the use of hybrid architectures to address specific challenges of smaller-scale fires and in areas of high geographic diversity.

The results of the review indicate that CNNs have shown strong performance in fire detection at both the pixel and scene levels, as shown in Table 2. Although the models achieve good accuracy scores, these largely depend on the difficulty of the classification task and the characteristics of the dataset. The review reveals a clear limitation in terms of the lack of large-scale datasets, which prevents a robust evaluation of the generalization ability of the models. For example, in [15]'s study, the small size of the dataset, limited to images from high-biodiversity places like Australia, Sicily, and Oregon, resulted in a 20% drop between the test set metrics and the training set. This finding highlights the need for larger datasets that include greater geographic and seasonal variability for adequate evaluation and robustness of the models. Likewise, although no reduction in performance was observed in [14], the test set was composed of data obtained from a single image, which represents a considerable limitation, since it does not reflect all the variability of the patterns of forest fires in different environments and climates. These examples highlight the importance of expanding and diversifying datasets to improve the reliability of models in practice.

Among the high temporal resolution satellite datasets, those developed from Himawari-8 images stand out, with studies such as [6, 3, 12, 14] and [15] taking advantage of their high temporal frequency for the early detection. These sets offer the advantage of

Table 2. Comparison of technologies used in CNN models for forest fire detection.

Model	Metrics	Technology	Focus	Level
MultiScale-Net [10]	F1: 91%, IoU: 84.54%	U-Net, multiscale	Active, Fire	Pixel
SR-NET [6]	Recall: 97%, F1: 95%	CNN, ViT, multi-head attention	Early, Smoke	Scene
FireCNN [3]	Precision: 0.998, Recall: 0.999	Multiscale, Fully Connected	Active, Fire	Pixel
SN [12]	F1: 0.74	Classic CNN, Fully Connected	Early, Fire	Pixel
Smoke-Net [13]	IoU: 0.742	Residuals, attention, layer skip	Active, Smoke	Pixel
1-DCNN [14]	Accuracy: 97.83%	Classic CNN, Fully Connected	Damage/Active, Fire/Smoke	Pixel
3DCNN [15]	Average Accuracy: 0.68	3D Input, Classic CNN	Damage/Active, Fire/Smoke	Pixel
SN [16]	IoU: 97.98%	U-Net, ResNet50	Damage, Burned Area	Pixel
VIB_SD [17]	Accuracy: 93.57%	Spatial attention, channel, residuals	Early, Smoke	Scene

continuous coverage over time, although they present limitations in their spatial resolution that restrict their applicability in the detection of smaller-scale fires. In [6], 4,000 images of fires in China and Australia were used to classify between clouds and smoke, while the dataset in [3] focused on specific provinces in China, using the full disk product Himawari-8 L1. Preprocessing in [12], on the other hand, incorporated multiple thermal infrared bands (bands 7, 12, 13, 14, and 15) to improve accuracy by calculating the Brightness Temperature (BT). These variations in datasets reveal the need to choose resolution and features based on the specific geographic and temporal conditions of each fire scenario.

Higher spatial resolution datasets come from satellites such as Landsat-8 and Sentinel-2, and are especially useful for active detection and post-fire analysis. However, some studies have also used them in smoke-based early detection, such as in [13]. The latter is one of the largest sets, both in number of images and geographical coverage, although its temporal resolution limits its usefulness in real-time detection. In [17], an Australian-specific dataset was developed, based on Landsat multispectral imagery and processed by NBART, ideal for segmentation of burned areas. On the other hand, [16] focused on images of Türkiye with Sentinel-2, focused on evaluating post-fire effects through 21,690 images from 13 bands. These sets allow for detailed segmentation, but being geographically limited to regions such as Australia and Turkey, their global applicability is lower compared to global datasets such as [10]. The latter used Landsat-8 to build a global dataset, although its labels were generated algorithmically and not manually, which could affect the accuracy of the segmentation.

Both sets, being limited to specific regions such as Australia and Turkey, offer less global applicability compared to other datasets such as the one presented in [10], which takes advantage of the global coverage of Landsat-8 to develop a set of data spanning all continents. This set employs the SWIR2, SWIR1 and Blue bands, highlighting the sensitivity of SWIR2 to radiation emitted by fire, and employs the Active Fire Index to highlight fire areas while reducing smoke interference. However, this dataset was labeled using algorithms, in contrast to the manual labeling used in the other studies, which could affect the accuracy of pixel-level segmentation.

One of the common challenges in pixel-level datasets is class imbalance due to the size of fire-affected areas compared to the background. This makes it necessary to use specific metrics such as F1 score and IoU (Intersection over Union) to properly evaluate the performance of models on imbalanced datasets and precise segmentation.

In terms of architecture, most of the models reviewed use traditional CNN structures, and those that do not use them still achieve similar performance metrics, as seen in Table 3. This suggests that current architectures may be sufficient for the available datasets, although not necessarily optimized for complex conditions. Among the models analyzed, FireCNN and Multiscale-Net stand out. FireCNN, designed for real-time detection, achieved near 100% accuracy. However, this performance was evaluated in a very limited study area, centered on southern China, and was not validated in multiple contexts, raising questions about its generalizability. In contrast, Multiscale-Net, which used a large and diverse dataset (including images from different continents), achieved an IoU metric of 84.54%, demonstrating a high capacity to adapt to variability in fire patterns and environmental conditions. This architecture makes use of multi-scale dilation and convolution rates, allowing it to detect fires of different scales and improve its accuracy compared to standard CNN networks.

However, the effectiveness of these models remains limited by the quality and diversity of the available

Table 3. Datasets used in CNN models for forest fire detection.

Model	Period	Private	Region	Satellite	Size	Bands	Label
SR-Net [6]	2015-2022	Yes	CN, AU	Himawari-8	4,000	RGB	Manual, Smoke
Multiscale-Net [10]	2020	No	Global	Landsat-8	150,000	SWIR1, SWIR2, Blue	Algorithms, Active Fire
FireCNN [3]	2020	Yes	CN	Himawari-8	5,469	1-16	Manual, Active Fire
SN [12]	2015-2019	Yes	KR, CN	Himawari-8	2,157	7, 12, 13, 14, 15	Himawari Wild Fire L2
Smoke-Net [13]	2016-2020	Yes	SA, AS, SIB, AU, NA	Landsat-8	54,270	RGB, SWIR2, TIRS1	Manual, Smoke
1-DCNN [14]	2016-2020	No	AU	PRISMA	259 px	RGB, SWIR2, TIRS1	Manual, Fire/Burned Area
3DCNN [15]	2019-2021	No	AU, US, IT	PRISMA	593 px	1-230	Manual, Fire/Burned Area
SN [16]	2019-2021	Yes	TR	Sentinel-2	21,690	1-12	Manual, Fire
VIB_SD [17]	2010-2020	Yes	AU	Landsat 5.8	1.836	RGB, NIR, SWIR1, SWIR2	Manual, Smoke

datasets. In particular, it is noted that the lack of large-scale datasets and the difficulty in obtaining satellite images with high spatial and temporal resolution limit the evaluation and improve the generalization of these models. Studies using additional spectral bands, such as infrared (IR), have shown improvements in fire detection accuracy. In the case of VI_SD, for example, the addition of the NIR band increased accuracy by 6%, while other combinations of IR bands decreased performance, highlighting the need to carefully select the most suitable spectral bands. However, the problem of spectral similarity between smoke and other aerosols, such as clouds or dust, has not yet been completely resolved, and continues to cause false alarms, although to a lesser extent with the use of combined bands.

The findings of this review present significant implications for research and practice. The effectiveness of CNNs in detecting wildfires suggests that these techniques can be integrated into real-time satellite monitoring systems for early warnings. CNN-based models can also improve accuracy in post-fire damage assessment, which is critical for recovery planning and prevention of future fires. However, to maximize the benefits of these technologies, it is essential to address limitations such as the spatial and temporal resolution of satellites and explore onboard processing solutions, which can reduce detection latency. The implementation of models in hardware accelerators, as demonstrated in [14], is a step forward in this direction. Additionally, multi-class datasets represent a considerable challenge, as they require precise and detailed labeling. To resolve the lack of these sets, it is necessary to explore more efficient labeling methods, such as the large-scale dataset generation algorithm used in [10].

To maximize the benefits of these models, it is essential to address current limitations related to the spatial and temporal resolution of satellites, and explore onboard processing to reduce detection latency. Models such as [14]'s 1-DCNN, implemented on hardware accelerators such as Intel Movidius NCS-2 and Nvidia Jetson Nano, have shown that on-board data processing is a viable possibility and could significantly reduce processing times. answer. As for datasets, multiclass datasets have proven to be considerably more complex to classify than binary datasets, due to the greater precision required in labeling, as seen in Table 3. Datasets such as [15], which uses PRISMA images from multiple geographic regions, have revealed up to a 20% drop in accuracy between training and test sets, underscoring the importance of having datasets larger and more diverse.

Furthermore, the useful life of satellites and their orbits directly limit the amount and coverage of data available for fire detection, a fundamental aspect for the development of more robust and generalizable models.

Compared to previous studies, this review highlights significant progress in the accuracy and robustness of CNN models applied to wildfire detection. While previous research focused on threshold-based methods or algorithms with low generalizability, the studies reviewed here demonstrate that combining multiple spectral bands—including the use of infrared bands and hyperspectral imaging experiments such as PRISMA—together With the use of various advanced CNN features and architectures, it enables more accurate and faster fire detection. These approaches have overcome several of the shortcomings of traditional methods, such as the difficulty in setting appropriate thresholds under changing conditions and the high rate of false alarms that resulted from these approaches. However, significant challenges remain, particularly in terms of data quality and the generalizability of models to different geographic contexts. Studies such as [15] have revealed that models that integrate spectral data and spatial information from neighboring pixels can improve accuracy, although only partially, over varied geographic regions.

Regarding model architectures, although in terrestrial and aerial detection it has been mentioned that it is advisable to try models specifically designed for remote sensing applications, it could also be beneficial to explore architectures developed in other areas, such as medical image processing or computer vision in general, which could provide advanced techniques in terms of precision and sensitivity. Additionally, techniques such as neural architecture search (NAS), ensemble models, or transfer learning could optimize model performance more efficiently than manual optimization. In studies such as [9], where NAS was used to optimize a Faster R-CNN architecture in the detection of different types of smoke, an improvement in accuracy and greater adaptability to different conditions was observed, indicating that this approach could be relevant to improve robustness in forest fire detection.

Despite the progress made, this review also reveals several limitations in using CNN for wildfire detection. One of the main limitations is the need for large labeled datasets to train these models, which is expensive and laborious to build. In several studies, such as [15], the lack of sufficiently diverse data led to a significant drop in model performance, revealing a 20% decrease in accuracy between the test and training set when data was insufficient. Furthermore, the generalizability of the models remains a challenge, especially when applied in different geographic regions or under varying environmental conditions. For example, models developed with data from a specific region, such as those from [3] in southern China, may show decreased performance when applied in scenarios with different vegetation or climatic conditions. Another important limitation is the latency in detection, which may be related to the limited spatial resolution of some satellites and the time required to process and transmit the images to Earth, reducing the effectiveness in early detection of fires in real time.

Additionally, the complexity of remote sensing imagery, which typically covers large areas with lower pixel density compared to smaller-scale imagery, and complex backgrounds with different vegetation types or geographic structures, pose additional challenges for accurate fire classification. Although ground and aerial detection faces the same problem of lack of datasets, the creation of these is usually relatively simpler compared to satellite remote sensing, which requires experts who can interpret the complex spectral signatures captured by the satellite sensors. This advantage in data collection for ground and airborne methods could allow for more rapid progress in these areas, suggesting that it would be beneficial to closely monitor these developments and adapt, where possible, the advances made to satellite detection to improve its performance in practice.

Based on the identified limitations, it is recommended that future research focus on the development of more robust models that can generalize better in different contexts and environmental conditions. This is especially relevant considering previous studies, such as [15] and [14], which highlighted how variability in geographic conditions affects the performance of models trained in specific regions. Additionally, it would be beneficial to explore hybrid approaches that combine CNN with other machine learning or image processing techniques, such as transformer-based learning, to improve accuracy and reduce false alarms in different environments. Future research should also focus on the use of hyperspectral imaging, which could offer significant advantages in the detection of active fires and in the differentiation of complex spectral signatures, since hyperspectral images contain up to 230 bands, as in PRISMA [15] , providing critical spectral details.

Likewise, it is crucial to encourage the creation and access to larger and more diverse datasets, as well as the development of technologies that enable data processing on board satellites, thereby reducing latency and improving real-time response capacity. To solve the problem of missing datasets, it is necessary to explore faster and more accurate labeling methods, such as the one used in [10], which used an algorithm to create a large-scale dataset automatically. This method is promising for multiclass datasets, especially if applied to hyperspectral images, as suggested by [12], achieving performance improvements by incorporating a combination of IR and RGB spectral bands.

Furthermore, domain adaptation could be a viable solution to address the lack of specific datasets, allowing the transfer of knowledge from one domain (e.g., images from one satellite) to another (images from another satellite), thus improving the ability of the model to generalize in different scenarios. Creating datasets with similar characteristics, from satellites such as Landsat-8 and Sentinel-2, along with domain adaptation, could result in more robust and accurate models for fire detection. It would be extremely beneficial for future research to make their datasets public, since, as seen in Table 3, most of these are private. The availability of these ensembles would not only allow for more complete evaluations and comparisons between models, but would also help mitigate the problem of missing data for wildfire detection. Once this obstacle is overcome, research could focus on comparing and optimizing specific models for active detection or remote sensing, seeking the optimal architecture for each type of detection.

The review conducted demonstrates that CNNs have considerable potential in wildfire detection using satellite images, although they still face significant challenges, especially in terms of generalization and data availability. Recent advances, such as the combination of multiple spectral bands (including infrared and hyperspectral bands) and the development of advanced architectures, such as FireCNN and Multiscale-Net, have improved detection accuracy and capability. However, maximizing the impact of these technologies in practical applications requires overcoming current limitations, such as the need for large labeled datasets and detection latency. Future studies should focus on the development of robust and diversified datasets, the use of hybrid approaches that integrate various DL techniques, and the implementation of processing on board satellites. These strategies can improve the accuracy and speed of detection, allowing a more effective and timely response to forest fires and their mitigation. Ultimately, these efforts will solidify the role of artificial intelligence in environmental conservation.

However, to maximize the impact of these technologies in practical applications, it is crucial to address current limitations, such as reliance on large labeled datasets and latency in detection. Future research should focus on creating more robust datasets, developing hybrid approaches that integrate various DL techniques, and deploying models with onboard satellite processing. These strategies can not only improve the accuracy and speed of fire detection, but also enable a more effective and timely response in wildfire management and mitigation.

Early detection of wildfires using satellite imagery and CNN offers an invaluable tool for protecting ecosystems and communities. Although technology has advanced significantly, the real challenge lies in achieving models that are robust, accurate and scalable in real environments. Addressing current limitations and promoting access to open datasets will not only strengthen emergency response capacity, but will also enable the full potential of artificial intelligence in environmental conservation and protection.

5. Conclusion

This work has reviewed the state of the art in forest fire detection using CNN using satellite images. Although CNNs have proven to be powerful tools for fire classification at the pixel and scene level, their effectiveness largely depends on data quality and availability. The integration of multiple spectral bands, including hyperspectral bands, has improved detection accuracy; however, significant challenges remain related to generalizability and latency in real-time detection.

The findings of this review underline the relevance of CNNs not only in academic research, but also in practical applications, such as real-time monitoring, with direct impact on wildfire response capacity. Improvements in precision and speed of detection offer the possibility of transforming prevention and response strategies, promoting the development of policies that promote the use of advanced technologies in the management of natural disasters.

The field of remote sensing fire detection is still emerging and faces important limitations, such as the need for large-scale labeled datasets and the difficulty of generalizing models in diverse geographic contexts. These limitations reflect the need to improve both the quality and availability of data and explore new techniques, such as the use of NAS and Ensemble Models in the optimization of architectures such as U-Net, testing their performance on satellite images of wildfires.

Future research should focus on the development of more robust models that can adapt to diverse contexts, thus improving generalization and reducing false alarms. To achieve this, the creation of larger and more accessible datasets could be explored using automatic labeling algorithms that speed up the manual process. Also, the development of hybrid approaches and the implementation of models onboard satellites are promising areas with the potential to significantly reduce latency and improve real-time detection of fires.

This review offers a comprehensive and critical perspective on the use of satellite data and artificial intelligence techniques for wildfire detection. Current advances in hybrid models, advanced architectures, and the integration of multiple data sources reflect significant progress in the field, although some fundamental issues, such as the lack of labeled data, remain unresolved. Addressing these limitations will be essential to achieving the full potential of fire detection technologies. Ultimately, continued research and commitment to the development of innovative solutions in this area will contribute to more effective protection of ecosystems and a faster response in wildfire management.

References

[1] M. N. Suratman, N. H. A. Hamid, M. D. M. Sabri, M. Kusin, and S. A. K. Yamani, *Changes in Tree Species Distribution Along Altitudinal Gradients of Montane Forests in Malaysia*. Springer International Publishing, 2015, pp. 491–522. [Online]. Available: https://link.springer.com/10.1007/978-3-319-12859-7_19

[2] "Deforestation fronts: Drivers and responses in a changing world." [Online]. Available: https://wwfint.awsassets.panda.org/downloads/deforestation_fronts___drivers_and_responses_in_a_changing_world___full_report_1.pdf

[3] Z. Hong, Z. Tang, H. Pan, Y. Zhang, Z. Zheng, R. Zhou, Z. Ma, Y. Zhang, Y. Han, J. Wang, and S. Yang, "Active fire detection using a novel convolutional neural network based on himawari-8 satellite images," *Frontiers in Environmental Science*, vol. 10, 3 2022.

[4] V. E. Sathishkumar, J. Cho, M. Subramanian, and O. S. Naren, "Forest fire and smoke detection using deep learning-based learning without forgetting," *Fire Ecology*, vol. 19, 12 2023.

[5] M. J. Sousa, A. Moutinho, and M. Almeida, "Wildfire detection using transfer learning on augmented datasets," *Expert Systems with Applications*, vol. 142, p. 112975, 3 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0957417419306931

[6] Y. Zheng, G. Zhang, S. Tan, Z. Yang, D. Wen, and H. Xiao, "A forest fire smoke detection model combining convolutional neural network and vision transformer," *Frontiers in Forests and Global Change*, vol. 6, 4 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/ffgc.2023.1136969/full

[7] N. S. Basturk, "Forest fire detection in aerial vehicle videos using a deep ensemble neural network model," *Aircraft Engineering and Aerospace Technology*, vol. 95, pp. 1257–1267, 7 2023. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/AEAT-01-2022-0004/full/html

[8] Z. Xue, H. Lin, and F. Wang, "A small target forest fire detection model based on yolov5 improvement," *Forests*, vol. 13, p. 1332, 8 2022. [Online]. Available: https://www.mdpi.com/1999-4907/13/8/1332

[9] D. Q. Tran, M. Park, Y. Jeon, J. Bak, and S. Park, "Forest-fire response system using deep-learning-based approaches with cctv images and weather data," *IEEE Access*, vol. 10, pp. 66 061–66 071, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9801825/

[10] A. Rostami, R. Shah-Hosseini, S. Asgari, A. Zarei, M. Aghdami-Nia, and S. Homayouni, "Active fire detection from landsat-8 imagery using deep multiple kernel learning," *Remote Sensing*, vol. 14, p. 992, 2 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/4/992

[11] Z. Wang, T. Peng, and Z. Lu, "Comparative research on forest fire image segmentation algorithms based on fully convolutional neural networks," *Forests*, vol. 13, p. 1133, 7 2022. [Online]. Available: https://www.mdpi.com/1999-4907/13/7/1133

[12] Y. Kang, E. Jang, J. Im, and C. Kwon, "A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency," *GIScience Remote Sensing*, vol. 59, pp. 2019–2035, 12 2022. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/15481603.2022.2143872

[13] H. Liang, C. Zheng, X. Liu, Y. Tian, J. Zhang, and W. Cui, "Super-resolution reconstruction of remote sensing data based on multiple satellite sources for forest fire smoke segmentation," *Remote Sensing*, vol. 15, p. 4180, 8 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/17/4180

[14] K. Thangavel, D. Spiller, R. Sabatini, S. Amici, S. T. Sasidharan, H. Fayek, and P. Marzocca, "Autonomous satellite wildfire detection using hyperspectral imagery and neural networks: A case study on australian wildfire," *Remote Sensing*, vol. 15, p. 720, 1 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/3/720

[15] D. Spiller, A. Carbone, S. Amici, K. Thangavel, R. Sabatini, and G. Laneve, "Wildfire detection using convolutional neural networks and prisma hyperspectral imagery: A spatial-spectral analysis," *Remote Sensing*, vol. 15, p. 4855, 10 2023. [Online]. Available: https://www.mdpi.com/2072-4292/15/19/4855

[16] A. M. Al-Dabbagh and M. Ilyas, "Uni-temporal sentinel-2 imagery for wildfire detection using deep learning semantic segmentation models," *Geomatics, Natural Hazards and Risk*, vol. 14, 12 2023. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/19475705.2023.2196370

[17] L. Zhao, J. Liu, S. Peters, J. Li, S. Oliver, and N. Mueller, "Investigating the impact of using ir bands on early fire smoke detection from landsat imagery with a lightweight cnn model," *Remote Sensing*, vol. 14, p. 3047, 6 2022. [Online]. Available: https://www.mdpi.com/2072-4292/14/13/3047